Zborník vybraných prác študentskej vedeckej a odbornej činnosti

Študentská vedecká a odborná činnosť
8. apríl 2025

Organizátor:

Slovenská technická univerzita v Bratislave
Fakulta elektrotechniky a informatiky

Editori:

Peter Benko
Stanislav Sojak

Vydala Slovenská technická univerzita v Bratislave vo Vydavateľstve SPEKTRUM STU.
Náklad 50 ks CD nosičov

ISBN 978-80-227-5491-0

Obsah

Vyhodnotenie prehliadky prác ŠVOČ **8**

1 Aplikovaná informatika **17**

1.1 Implementácia šifrovacieho stroja Condenser PBJ
Jan Mikulec, Eugen Antal 18

1.2 Deciding Primality of Permutation Automata
Alexandra Reviľáková, Andrej Madzochci 24

1.3 Vplyv viackrokového dotrénovania na úspešnosť a vysvetliteľnosť klasifikácie obrazov z optickej koherenčnej tomografie
Samuel Gibala, Veronika Kurilová 31

1.4 Detekcia štandardnej kvantizačnej matice JPEG obrázka
Martin Mydliar, Matúš Jókay 37

1.5 Mikroslužbový systém na automatizované spracovanie, testovanie a vyhodnotenie používateľského kódu
Filip Švec, Martin Pajpach 42

1.6 Interaktívna vizualizácia 3D modelov
Alex Mihalik, Erik Kučera 49

1.7 Detekcia phishingových e-mailov
Simona Zlatohlávková, Miloš Oravec 55

1.8 Webová aplikácia na hľadanie spolubývajúcich
Teak Min Guillaume Oh, Oľga Stašová 60

2 Mechatronika I **65**

2.1 Tvorba modulárneho grafického rozhrania pre zmiešanú realitu
Tomáš Sedláček, Erik Kučera 66

2.2 Rozpoznávanie gest pre úlohy ovládania mechatronického systému
Miloš Ilovský, Oto Haffner 71

2.3 Vzdelávacia aplikácia pre počítačové videnie
Dominik Kysel, Oto Haffner 77

2.4 Využitie výpočtovej inteligencie pre úlohy kontroly kvality
Eunika Farkašová, Oto Haffner 83

2.5 EKV - elektronická kontrola vstupu
Andrej Marián Drugda, Felix Jozef Juraj Vítek, Katarína Žáková 90

3 Mechatronika II **99**

3.1 Referenčný systém merania teploty pre teplotné vláknové segmentové interferometrické senzory
Andrea Kuňáková, Richard Balogh 100

3.2 Guľôčka na ramene
Daniel Žula, Richard Schwarcz 105

3.3 Aplikácia hlbokých neurónových sietí pre analýzu kožných lézií
Sabína Ovčiariková, Erik Kučera 111

3.4 AI chatbot a RPA - inteligentná automatizácia
Andrea Sigetiová, Peter Drahoš 117

3.5 BabyGuard: Pokojný spánok dieťaťa vďaka moderným technológiám
Matej Bašista, Lukáš Beňo 122

4 Elektronika a fotonika, Elektrotechnika, Elektroenergetika **127**

4.1 Projektovanie VN sietí a zariadení v BIM a analýza siete prostredníctvom integrácie ETAP s Revitom
Patrik Hakala, Peter Janiga 128

4.2 Analýza výsledkov cyklického výkonového testu MOS tranzistora podporená 3D elektrotepelnými simuláciami
Patrik Suhaj, Aleš Chvála 132

4.3 Návrh nízkopríkonového rekonfigurovateľného FIR filtra v HDL jazyku
Richard Bagín, Lukáš Nagy 137

4.4 Návrh μLaborátoria satelitu pre meranie viskozity mazív
Martin Brázda 143

4.5 Simulácie pre optimalizáciu batériového systému v elektrickom autobuse MHD
Boris Bajla, Miroslav Mikolášek 147

4.6 Zlepšenie elektrochemických vlastností elektród na báze MoS_2 prostredníctvom chemickej aktivácie
Marcel Husár, Peter Ondrejka 153

4.7 Memristor Technology for Neural Network Hardware Synapses
Natalia Kamila Calvo, Boris Hudec 157

4.8 Napájací zdroj s ultra nízkym šumom a ultra nízkym únikovým prúdom
Viktor Štovčík, Daniel Valúch 162

5 Jadrové a fyzikálne inžinierstvo **167**

5.1 Možnosť implementácie DT neutrónového generátora do novobudovaného laboratória
Matej Kleštinec, Branislav Vrban 168

5.2 Simulácia radiačnej situácie v oblasti jadrovej elektrárne Mochovce v prípade ťažkej havárie
Brigita Kovácsová, Jana Šimeg Veterníková 173

5.3 Predlžovanie životnosti jadrových elektrární a využitie pozitrónovej anihilačnej spektroskopie pri testovaní starnutia materiálov
Marek Kováč, Jana Šimeg Veterníková 179

5.4 Registrácia 14 MeV neutrónov s Si Timepix detektorom
Matej Balušík, Andrea Šagátová 183

5.5 Nelineárna rezonancia magneticky budeného kyvadla
Roman Michelko, Peter Bokes 189

5.6 Supravodivosť vs. Prvák FEI STU & vice versa
Ján Virág, Pavol Valko 195

5.7 Optimalizácia procesu prípravy vrstiev pomocou rotačného nanášania
Mykhailo Olkhovyk, Patrik Novák 200

6 Robotika a kybernetika **206**

6.1 Riadenie mobilných robotov
Filip Kráľ, František Duchoň 207

6.2 Generation of AI Images in Ophthalmology
Dávid Szépvologyi, Michal Kováč 211

6.3 Simulation Analysis of a Sensor Utilizing Phototransistors for Measuring the Gravitational Constant in a Microsatellite
Viktor Hrúz, Ľuboš Chovanec 217

6.4 Polohovanie dotykovej ultrazvukovej sondy priemyselným kolaboratívnym robotom
Juraj Kováč, Ľuboš Chovanec 223

6.5 Variable Structure Control for a Motion System: Synthesis and Performance Assessment
Radovan Jakubčík, Marián Tárník 229

6.6 Model Reference Adaptive Control: Design and Implementation for a Laboratory Dynamic System
Adrián Jakubčík, Marián Tárník 235

6.7 Comparative Analysis of Neural Network Architectures for Sentiment Classification in Slovak Textual Reviews
Matej Medvedzky, Jarmila Pavlovičová 241

6.8 Riadenie mobilného robota
Yehor Kiriienko, František Duchoň 245

7 Informačné a komunikačné technológie **248**

7.1 Data science application in IKT
Dmitrii Savin 249

7.2 Creation of interactable environment for Virtual Reality using Unreal Engine 5
Yelyzaveta Marchenko 252

7.3 Simplistic simulation model for impairments in DNA strings in selected system
Pavlo Melnyk, Peter Farkaš 256

7.4 Translation codes constructions supported by computer
Mykola Melnyk, Peter Farkaš 261

7.5 Analýza nákladov ochrany prevádzky v pasívnych optických sieťach WDM-PON
Filip Fuňák, Rastislav Róka 265

7.6 Optimizing signal transmission through cables using machine learning
Valerii Cherniatiev 271

7.7 Photogrammetry
Anastasiia Bondarenko, Andrii Smetaniuk 275

7.8 Virtuálne laboratórium pre optické vláknové médium
Kateryna Vynnyk, Rastislav Róka 278

7.9 Návrh a realizácia modelu letiska M.R. Štefánika
Matúš Dúbravka 284

Členovia programového a recenzného výboru

Ing. Šimon Berta
Ing. Peter Benko, PhD.
Ing. Ján Cigánek, PhD.
RNDr. Juraj Chlpík, PhD.
Ing. Marián Choma
doc. Mgr. Daniela Chudá, PhD.
Ing. Michal Dobiš, PhD.
doc. Ing. Gabriel Farkas, PhD.
doc. Mgr. Miroslava Farkas Smitková, PhD.
Ing. Michal Hlavatý, PhD.
doc. Mária Hypiusová, PhD.
doc. Ing. Jaroslav Kováč, PhD.
prof. Ing. Alena Kozáková, PhD.
Ing. Branislav Lojko, PhD.
Ing. Michal Miloslav Uličný
RNDr. Martin Nehéz, PhD.
Ing. Vladislav Novák
Ing. Patrik Novák, PhD.
Ing. Kristína Okienková
prof. Ing. Márius Pavlovič, PhD.
doc. Ing. Jana Paulusová, PhD.
Ing. Richard Schwarz
Ing. Juraj Slačka, PhD.
Ing. Michal Šovčík, PhD.
Ing. Michal Štibraný

Vyhodnotenie prehliadky prác ŠVOČ

Úvod

Na Fakulte elektrotechniky a informatiky STU v Bratislave sa dňa 08. apríla 2025 konala prehliadka prác študentskej vedeckej a odbornej činnosti (ŠVOČ). Táto študentská vedecká konferencia dáva príležitosť študentom bakalárskeho a inžinierskeho štúdia prezentovať výsledky, ktoré vypracovali samostatne, alebo pod vedením pedagogických a výskumných pracovníkov ústavov FEI STU, prípadne významných odborníkov z praxe.

V roku 2025 bolo vyhlásených osem sekcií zodpovedajúcich študijným programom akreditovaným na FEI STU. Počet prác prihlásených do jednotlivých sekcií je uvedený v nasledovnej tabuľke.

	Sekcia	Ústav	Počet prác	Počet študentov
1	Aplikovaná informatika	ÚIM	8	8
2	Elektrotechnika	ÚE	1	1
3	Elektroenergetika	ÚEAE	1	1
4	Elektronika a fotonika	ÚEF	6	6
5	Informačné a komunikačné technológie	ÚMIKT	9	10
6	Jadrové a fyzikálne inžinierstvo	ÚJFI	7	7
7	Mechatronika	ÚAMT	10	11
8	Robotika a kybernetika	ÚRK	8	8
			50	52

Prihlásené práce boli prezentované v šiestich sekciách nasledovne:

1. Aplikovaná informatika: 8 príspevkov
2. Mechatronika I: 5 príspevkov,
3. Mechatronika II: 5 príspevkov
4. Elektronika a fotonika, Elektrotechnika, Elektroenergetika: 8 príspevkov
5. Informačné a komunikačné technológie: 9 príspevkov
6. Jadrové a fyzikálne inžinierstvo: 7 príspevkov
7. Robotika a kybernetika: 8 príspevkov

Príspevky v jednotlivých sekciách hodnotili najmenej trojčlenné komisie.

V každej sekcii boli najlepšie práce ocenené Cenou dekana FEI STU v Bratislave a ďalšie vynikajúce práce získali Diplom dekana FEI STU v Bratislave. Obe ocenenia sú spojené s jednorazovým mimoriadnym štipendiom.

Československá sekcia IEEE venovala do súťaže polročné členstvo v IEEE pre jednu prácu z každej technickej sekcie. Ceny odovzdala členka predstavenstva československej sekcie IEEE Ing. Elena Cocherová, PhD.

Spoločnosť HUMUSOFT s.r.o. ocenila Cenou HUMUSOFTu najlepšie príspevky vypracované s využitím systému MATLAB a COMSOL Multiphysics. Autori víťazných prác získali voucher na online školenie podľa vlastného výberu a ďalšie vecné dary. Ceny odovzdal zástupca spoločnosti HUMUSOFT na Slovensku Ing. Martin Foltin, PhD.

Zväz slovenských vedeckotechnických spoločností (ZSVTS) venoval do súťaže Cenu ZSVTS, ktorá bola určená pre najlepšiu prácu študentskej vedeckej konferencie, pričom nositeľ ceny dostal možnosť stať sa Ambasádorom ZSVTS, ktorého úlohou bude sprostredkovávať informácie o svojich vedeckých aktivitách aj o ZSVTS nielen na univerzite, ale aj vo svojom okolí. Práce nominované na toto ocenenie navrhli komisie v jednotlivých sekciách. Najlepšej práci odovzdal ocenenie v mene ZSVTS riaditeľ Zväzu slovenských vedeckotechnických spoločností Ing. Ivan Janáč.

V dvoch sekciách uskutočnených v študijnom odbore elektrotechnika bola najlepšej práci z tohto odboru udelená Cena onsemi. Cenu odovzdal Ing. Branislav Lojko, PhD.

Ocenené práce a ďalšie práce odporučené komisiami v sekciách boli zaradené do zborníka, ktorý bol vydaný v elektronickej forme.

Zoznam ocenených prác je uvedený v nasledujúcich tabuľkách.

Prehľad ocenení

Cena dekana FEI STU v Bratislave

Meno	Názov práce	Vedúci práce	Sekcia
Bc. Richard Bagín	Návrh nízkopríkonového rekonfigurovateľného FIR filtra v HDL jazyku	Ing. Lukáš Nagy, PhD.	Elektronika a fotonika, Elektrotechnika, Elektroenergetika
Bc. Matej Bašista	BabyGuard: Pokojný spánok dieťaťa	Ing. Lukáš Beňo, PhD.	Mechatronika II.
Bc. Andrej Marián Drugda, Bc. Felix Jozef Juraj Vítek	EKV - elektronická kontrola vstupu	prof. Ing. Katarína Žáková, PhD.	Mechatronika I.
Bc. Juraj Kováč	Polohovanie dotykovej ultrazvukovej sondy priemyselným kolaboratívnym robotom	Ing. Ľuboš Chovanec, PhD.	Robotika a kybernetika
Mykola Melnyk	Translation codes constructions supported by computer	prof. Ing. Peter Farkaš, DrSc.	Informačné a komunikačné technológie
Bc. Roman Michelko	Nelineárna rezonancia magneticky budeného kyvadla	doc. Ing. Peter Bokes, PhD.	Jadrové a fyzikálne inžinierstvo
Bc. Ján Mikulec	Implementácia šifrovacieho stroja Condenser PBJ	Ing. Eugen Antal, PhD.	Aplikovaná informatika

Diplom dekana FEI STU v Bratislave

Meno	Názov práce	Vedúci práce	Sekcia
Bc. Matej Balušík	Registrácia 14 MeV neutrónov s Si Timepix detektorom	doc. Ing. Andrea Šagátová, PhD.	Jadrové a fyzikálne inžinierstvo
Anastasiia Bondarenko, Andrii Smetaniuk	Photogrammetry	-	Informačné a komunikačné technológie
Bc. Martin Brázda	Návrh μLaborátoria satelitu pre meranie viskozity mazív	-	Elektronika a fotonika, Elektrotechnika, Elektroenergetika
Bc. Eunika Farkašová	Využitie výpočtovej inteligencie pre úlohy kontroly kvality	doc. Ing. Oto Haffner, PhD.	Mechatronika I.
Bc. Samuel Gibala	Vplyv viackrokového dotrénovania na úspešnosť a vysvetliteľnosť klasifikácie obrazov z optickej koherenčnej tomografie	MUDr. Veronika Kurilová, PhD.	Aplikovaná informatika
Bc. Viktor Hrúz	Simulation Analysis of a Sensor Utilizing Phototransistors for Measuring the Gravitational Constant in a Microsatellite	Ing. Ľuboš Chovanec, PhD.	Robotika a kybernetika
Bc. Matej Kleštinec	Možnosť implementácie DT neutrónového generátora do novobudovaného laboratória	doc. Ing. Branislav Vrban, PhD.	Jadrové a fyzikálne inžinierstvo
Bc. Alexandra Reviľáková	Deciding Primality of Permutation Automata	Dr. Nicolas Adrien Mazzocchi	Aplikovaná informatika

Meno	Názov práce	Vedúci práce	Sekcia
Bc. Dávid Szépvölgyi	Generation of AI Images in Ophthalmology	Ing. Michal Kováč, PhD.	Robotika a kybernetika
Bc. Viktor Štovčík	Napájací zdroj s ultra nízkym šumom a ultra nízkym únikovým prúdom	doc. Ing. Daniel Valúch, PhD.	Elektronika a fotonika, Elektrotechnika, Elektroenergetika
Kateryna Vynnyk	Virtuálne laboratórium pre optické vláknové médium	doc. Ing. Rastislav Róka, PhD.	Informačné a komunikačné technológie
Daniel Žula	Guľôčka na ramene	Ing. Richard Schwarz	Mechatronika II.

Cena Československej sekcie IEEE

Meno	Názov práce	Vedúci práce	Sekcia
Patrik Hakala	Projektovanie VN sietí a zariadení v BIM a analýza siete prostredníctvom integrácie programu ETAP s Revitom	doc. Ing. Peter Janiga, PhD.	Elektronika a fotonika, Elektrotechnika, Elektroenergetika
Filip Kráľ	Riadenie mobilných robotov	prof. Ing. František Duchoň, PhD.	Robotika a kybernetika
Mykola Melnyk	Translation codes constructions supported by computer	prof. Ing. Peter Farkaš, DrSc.	Informačné a komunikačné technológie
Bc. Sabína Ovčiariková	Aplikácia hlbokých neurónových sietí pre analýzu kožných lézií	doc. Ing. Erik Kučera, PhD.	Mechatronika II.
Bc. Tomáš Sedláček	Tvorba modulárneho grafického rozhrania pre zmiešanú realitu	doc. Ing. Erik Kučera, PhD.	Mechatronika I.
Ján Virág	Supravodivosť vs. Prvák FEI STU & vice versa	doc. RNDr. Pavol Valko CSc.	Jadrové a fyzikálne inžinierstvo
Bc. Simona Zlatohlávková	Detekcia phishingových e-mailov	prof. Dr. Ing. Miloš Oravec	Aplikovaná informatika

Cena ZSVTS

Meno	Názov práce	Vedúci práce	Sekcia
Bc. Viktor Hrúz	Simulation Analysis of a Sensor Utilizing Phototransistors for Measuring the Gravitational Constant in a Microsatellite	Ing. L'uboš Chovanec, PhD.	Robotika a kybernetika

Cena HUMUSOFT

Meno	Názov práce	Vedúci práce	Sekcia
Daniel Žula	Guľôčka na ramene	Ing. Richard Schwarz	Mechatronika II.
Bc. Patrik Suhaj	Analýza výsledkov cyklického výkonového testu MOS tranzistora podporená 3D elektrotepelnými simuláciami	doc. Ing. Aleš Chvála, PhD.	Elektronika a fotonika, Elektrotechnika, Elektroenergetika
Bc. Boris Bajla	Simulácie pre optimalizáciu batériového systému v elektrickom autobuse MHD	doc. Ing. Miroslav Mikolášek, PhD.	Elektronika a fotonika, Elektrotechnika, Elektroenergetika
Mykola Melnyk	Translation codes constructions supported by computer	prof. Ing. Peter Farkaš, DrSc.	Informačné a komunikačné technológie

Cena onsemi

Meno	Názov práce	Vedúci práce	Sekcia
Bc. Natalia Kamila Calvo	Memristor Technology for Neural Network Hardware Synapses	Ing. Boris Hudec, PhD.	Elektronika a fotonika, Elektrotechnika, Elektroenergetika

Vybrané práce v jednotlivých sekciách

V nasledujúcich kapitolách sú uvedené všetky ocenené práce ako aj ďalšie vybrané práce v jednotlivých sekciách.

Sekcia 1

Aplikovaná informatika

Implementácia šifrovacieho stroja Condenser PBJ

Bc. Ján Mikulec, Ing. Eugen Antal, PhD.

Ústav informatiky a matematiky, FEI STU v Bratislave

xmikulec@stuba.sk

Condenser PBJ je mechanické šifrovacie zariadenie vyvinuté Pavelom Baráčekom-Jacquierom. Bolo skonštruované v rokoch 1922, až 1924 a bolo využívané Československou diplomaciou až do roku 1934. Zariadenie bolo navrhnuté tak, aby zvyšovalo efektivitu telegrafického prenosu správ a zároveň šifrovalo obsah správy. Využívaný šifrovací algortimus je polyalfabetická substitučná šifra. Konkrétne autoklúč s inicializačným vektorom dlžky 10 znakov.

1 Úvod

V práci sa zamerievame na návrh a implementáciu šifrovacieho algoritmu podľa článku [1] a implementáciu s využitím bigramov. Zariadenie Condenser PBJ bolo do nedávna neznáme. Autori článku [1] na základe spisov v archívoch vytvorili predpokladaný model zariadenia a zrekonštruovali šifrovací algoritmus. Motiváciou tejto práce je na základe ich zistení vytvoriť program, ktorý stroj simuluje a umožňuje prácu s ním. Softvérové riešenie zahŕňa 4 implementácie šifrátora, z čoho jedna tvorí základ pre komponent v študentskej edukačnej aplikácii HCPortal [2][3]. Zvyšné implementácie sa využívaju v desktopovom experimentálnom prostredí (EE). V druhej kapitole sa uvádza popis a fungovanie šifrovacieho stroja a jeho sofvérového prevedenia a opis EE a HCPortal komponentu.

2 Popis šifrovacieho stroja

Mechanický aj softvérový šifrátor Condenser PBJ obsahuje niekoľko hlavných častí a vstup do šifrovacieho algoritmu je podmienený niekoľkými pravidlami, ktoré určujú formát očakávaného vstupu. V nasledujúcich kapitolách popíšeme konkrétnejšie formátovanie textu, aj najpodstatnejšie komponenty šifrátora.

Postup celého procesu od nastavenia šifrovacieho zariadenia, až po dešifrovanie by sa dal zhrnúť nasledovne:

- Skonštruovanie ročného kľúča, denného kľúča a inicializačného vektora
- Aplikovanie denného kľúča na ročný kľúč
- Predformátovanie vstupného textu
- Nastavenie počiatočného vnútorného stavu zariadenia pomocou inicializačného vektora
- Šifrovanie:
    - Iteratívne pripočítavanie blokov otvoreného textu ku vnútornému stavu šifrátora, pričom sa za každým mení vnútorný stav na hodnotu nového zašifrovaného bloku
- Dešifrovanie:
    - Iteratívne odpočítavanie blokov zašifrovaného textu od vnútornému stavu šifrátora, pričom sa za každým mení vnútorný stav na hodnotu nového dešifrovaného bloku

2.1 Formátovanie vstupného textu

Vstupný text podlieha pravidlám formátovania. Text sa musí začínať samohláskou a končiť spoluhláskou. Tieto pravidlá vychádzajú z architektúry šifrovacieho algoritmu. Zariadenie šifruje text po blokoch dlhých 10 znakov a obsahuje dva typy rotorov – samohláskové a spoluhláskové. Tie sú umiestnené v striedavom poradí. Preto musí byť dĺžka vstupného textu deliteľná číslom 10 a musí byť zachovaná striedmosť samohlások a spoluhlások. Stroj umožnuje šifrovať písmená aj čísla. Preto musia byť všetky čísla vo vstupnom texte špeciálne označené, aby ich bolo možné správne dešifrovať. Abeceda zašifrovaného textu nie je totožná s abecedou otvoreného textu, ale tvorí len jej podmnožinu. Preto, aby bolo možné znaky predstavujúce čísla rozoznať, je okrem iného potrebné pred každé číslo otvoreného textu vložiť kľúčové slovo „nume“. To indikuje číslice v texte. Medzi kľúčové slovo a samotné číslo sa ešte vloží jednociferná číslica, ktorá hovorí o počte cifier daného čísla v texte a za číslo sa vloží check-sum. Check-sum je vypočítany ako súčet číslic daného čísla modulo 10. Napríklad číso 12345 by bolo naformátované na „nume5123455“. Viac o formátovaní textu v článku [1].

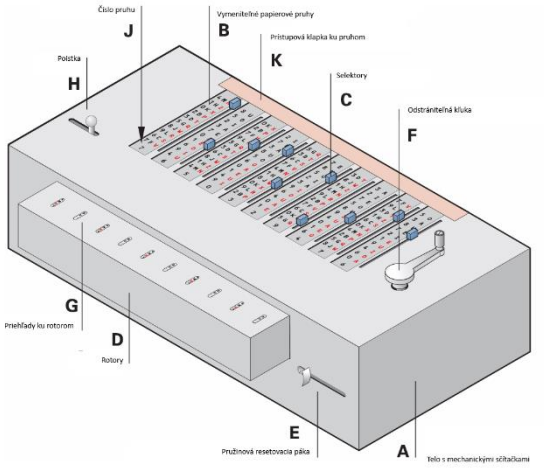
2.2 Inicializačný vektor

Inicializačný vektor (IV) sa používa pri šifrovaní a dešifrovaní na nastavenie počiatočného vnútorného stavu šifrátora. Je dlhý 10 znakov a je špeciálne skonštruovaný.

IV musí obsahovať len písmená abecedy a musí spĺňať rovnaký formát, ako otvorený text s výnimkou pravidla na spracovanie čísel. Tie sú prevedené na písmená abecedy pomocou substitučných tabuliek na základe daných pravidel. Celý popis vyžadovanej štruktúry IV, substitučné tabuľky a spôsob ich použitia, sú bližšie opísané v článku [1].

2.3 Komponenty

Mechanické zariadenie sa skladá z komponentov zobrazených na Obr. 2. [1]



Obr. 1. Grafický model mechanického zariadenia

Najpodstatnejšími pre túto prácu sú:

- (**B**) papierové pruhy
- (**D**) rotory

Každý samohláskový pruh (**B**) sa skladá z rovnakej množiny samohlások. Spoluhláskové pruhy obsahujú zas znaky zo spoločnej množiny spoluhlások. Pruhy v jednotlivých kategóriách (samohláskové/ spoluhláskové) sa rozlišujú len poradím riadkov, ktoré je určené ročným kľúčom. V prípade mechanického zariadenia sa jedná o tajný komponent, pretože ročný kľúč je priamo zakomponovaný v pruhoch. Hodnoty na 10-stranných rotoroch (**D**) sú korešpondujúce ku hodnotám na papierových pruhoch (**B**).

Denný kľúč predstavuje vektor 10 číslic v rozsahu 0 - 9 a určuje poradie pruhov. Aby sa zabezpečili správne pozície jednotlivých pruhov, denný kľúč musí byť v špecifickom formáte. Každé párne číslo sa musí nachádzať na nepárnej pozícii a naopak, pričom indexovanie sa počíta od 0.

Po samostatnom aplikovaní oboch kľúčov na množinu 5 dvojíc spoluhláskových a samohláskových pruhov, dostávame 10 pruhov, ktoré sú použité pri šifrovaní a dešifrovaní.

Znaky na pruhoch sa rozdelujú podľa farebnosti na čierne (v článku [1] sa uvádzajú aj ako biele) a červené. V prípade samohláskových pruhov, sú červené všetky písmená v druhej polovici pruhov. Písmená na prvej polovici týchto pruhov sú čierne. V prípade spoluhláskových pruhov obsahuje každý riadok pruhu dve písmená – prvé čierne a druhé červené. Čísla na všetkých pruhoch majú čiernu farbu. Farebnosť znakov je podstatná vlastnosť zariadenia a má vplyv na šifrovanie a dešifrovanie. Viac o farbách v článku [1].

3 Implementácia PBJ

Vypracované implementácie vieme okrem iného deliť aj podľa prostredia, v ktorom je daná implementácia nasadená. Desktopové prostredie EE obsahuje 3 implementácie a komponent v edukačnej aplikácií HCPortal jednu. Viac o jednotlivých prostrediach v nasledujúcich kapitolách.

Implementácia 1.0.0 slúži ako referenčná. Je to vôbec prvá funkčná softvérová implementácia zariadenia Condenser PBJ. Rýchlosť vykonávanie operácií tejto verzie implementácie nie je optimalizovaná a nepodporuje využitie farebne neusporiadaných pruhov (**VP**).

Táto implementácia obsahuje kontrolu výskytu falošných kľúčových slov „nume“ (**FPN**). Jedná sa o ošetrenie špeciálnych situácií, ktoré v malej miere ovplyvňujú korektnosť výsledkov dešifrovania. Ku tejto situácií môže dôjsť, keď používateľ zadá do textu kľúčové slovo „nume“, aj keď nenasledujú číslice.

FPN ošetruje situácie, kedy false-positive číslo presahuje do ďalšieho bloku textu. Počas dešifrovania sa počíta pozícia dešifrovaného znaku znaku y_i ako $y_i = x_i - k_i \, mod \, 10$, kde x je vstupný zašifrovaný znak a k je znak predošlého dešifrovaného bloku, alebo inicializačného vektora. V prípade, že **FPN** nie je implementované, hodnoty y_i , kde i je z množiny $\{0, ..., 10\}$, po dešifrovaní každého bloku zanikajú – neukladajú sa. Preto ak nume indikuje číslo presahujúce do ďalšieho bloku, nie je takéto číslo možné správne identifikovať.

Autori článku [1] uvádzajú ako možnú alternatívu ku problematike využitie predvypočítaných bigramov (**BG**), ktoré mapujú všetky možné dvojice znakov na ich číselné reprezentácie. Tieto bigramy sa po skonštruovaní používajú na šifrovanie a dešifrovanie. Pri implementácii 1.2.0 bolo toto docielené za pomoci hashovacej tabuľky, ktorá mapuje kľúče na ich hodnoty. V jazyku Java sme pre tento účel využili dátový typ HashMap, ktorý má časovú zložitosť vyhľadávania podľa kľúča O(1).

Toto riešenie výrazne znižuje časovú náročnosť operácií. Samotný proces vypočítania bigramov je ale časovo náročný. Preto by táto implementácia bola vhodná len pre také aplikácie, ktoré šifrujú a/alebo dešifrujú väčšie množstvo textov za pomoci tej istej sady tajných parametrov a nastavení šifrátora.

Verzia	Prostredie	Technológia	Optimalizované	VP	FPN	BG
1.0.0	EE	Java	✗	✗	✓	✗
1.1.0	EE	Java	✓	✓	✓	✗
1.2.0	EE	Java	✓	✓	✗	✓
2.0.0	HCP	Angular	✓	✗	✗	✗

Obr. 2. Rozdelenie implementácií podľa ich charakteristík

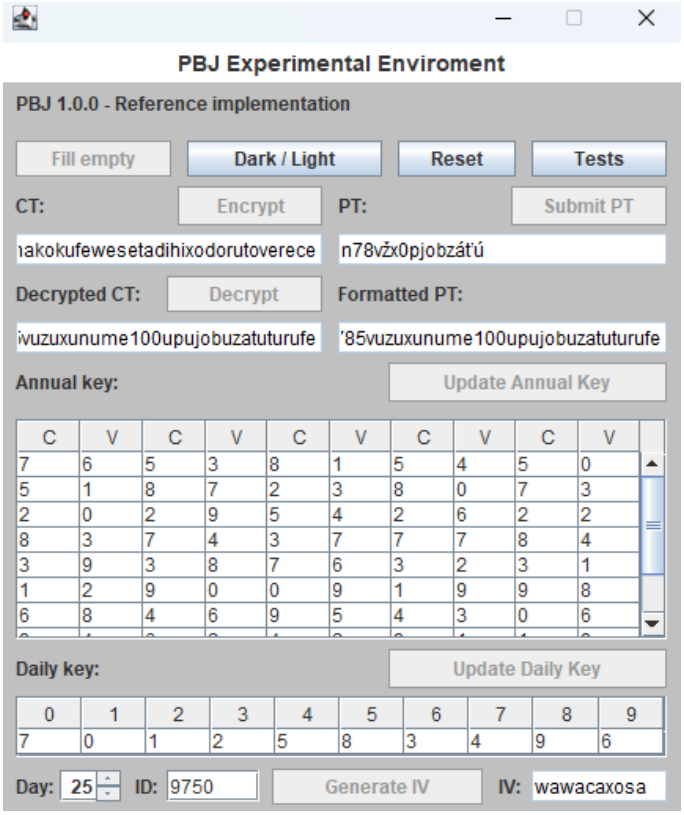
Dataily o ostatných implemtnáciách na Obr. 2.

3.1 Experimentálne prostredie

Experimentálne prostredie (EE) bolo vytvorené pomocou jazyku Java a technológie Java Swing. Predstavuje desktopové grafické rozhranie určené pre prácu s referenčnou implementáciou 1.0.0 (Obr. 3.). Poskytuje tiež testovacie rozhranie so štatistickou analýzou textu (Obr. 4.).

3.1.1 Hlavné okno

Hlavné okno umožnuje používateľovy vyplniť všetky potrebné parametre pre šifrovací a dešifrovací algoritmus. Všetky polia, okrem poľa pre otvorený text, podliehajú validácii, ktorá ošetruje vstupné dáta a upozorňuje používateľa v prípade, ak vložil nevhodný vstup.



Obr. 3. EE - Hlavné okno

Okrem toho poskytuje hlavné okno možnosť náhodného generovania nevyplnených polí, prepínanie tém a kompletné vynulovanie všetkých premenných vrátane vnútorného stavu šifrátora.

Tlačidlo „Tests“ otvára nové okno aplikácie (Obr. 4).

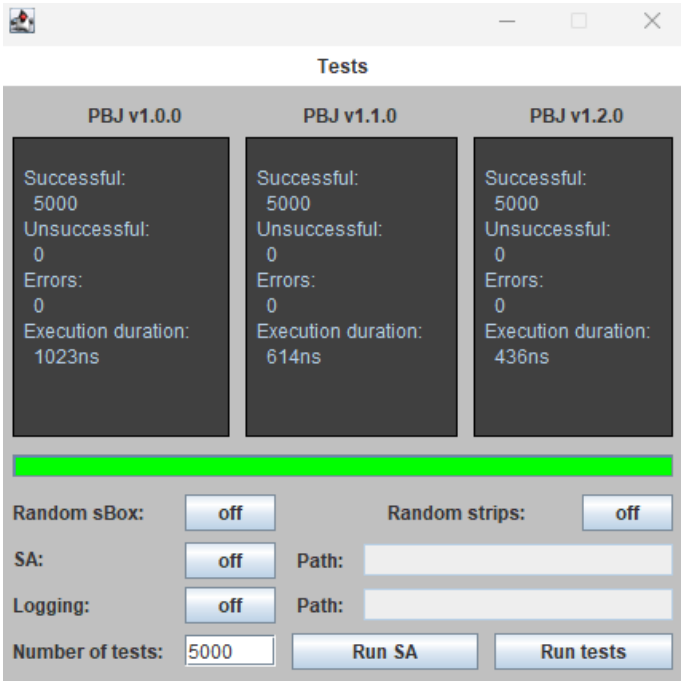
3.1.2 Okno testov

Okno testov poskytuje testovacie rozhranie, umožňujúce opakované generovanie náhodných vstupných parametrov a sledovanie štatistiky úspechu, zlyhaní a chybovosti jednotlivých implementácií šifrátora.

Maximálny počet požadovaných testovacích vzoriek je 9.999.999. Každá vzorka predstavuje náhodné vygenerovanie vstupného textu, inicializačného vektora, denného a ročného kľúča, naformátovanie vstupného textu a šifrovanie a dešifrovanie vstupného textu.

Používateľ môže prepínať medzi zapnutým a vypnutým opakovaným generovaním S-boxov. Toto nastavenie ovplyvňuje iba implementáciu 1.2.0. Generovanie S-boxov zahŕňa aj repetitívne generovanie kľúčov, pretože tie sú použité pri výpočte S-boxov. Implementácia 1.2.0 má zo všetkých verzií najnižší čas vykonania operácií, pričom tento rozdiel narastá s narastajúcim počtom vzoriek.

Aplikácia tiež umožnuje zapnúť opakované generovanie náhodných pruhov s neusporiadanými farbami, čo má vplyv na implementácie 1.1.0 a 1.2.0.



Obr. 4. EE - Okno testov

Po zapnutí logovania a zadaní platnej cesty v systéme, sa všetky chybové a neúspešné testy podrobne zapisujú do logovacích textových súborov na danej ceste. Neúspešný test znamená taký test, kedy sa dešifrovaný text nerovná naformátovanému otvorenému textu. Chybový test predstavuje programovú chybu počas behu programu. Chybové logy slúžili hlavne v procese vývoja na dolaďovanie riešenia.

Okno testov tiež umožňuje vykonanie štatistickej analýzy zašifrovaných textov a sledovanie niektorých vlastností otvoreného textu. Výsledky analýzy sa ukladajú do textových súborov, preto je potrebné zadať validnú cestu v systéme.

Štatistická analýza zahŕňa:

- výpočet indexu koincidencie (ďalej len IC) celého zašifrovaného textu
- výpočet IC jednotlivých blokov zašifrovaného textu
- početnosť výskytu unigramov, bigramov a trigramov
- početnosť výskytu konkrétnych znakov na jednotlivých pozíciách v bloku zašifrovaného textu

Sledované vlastnosti otvoreného textu sú:

- priemerné predĺženie textu po jeho naformátovaní
- maximálne predĺženie
- minimálne predĺženie
- medián predĺženia

Všetky sledované vlastnosti a štatistické výpočty sú robené na 4 skupinách náhodných vzoriek, pričom sa skupiny delia podľa dĺžky naformátovaného otvoreného textu a zašifrovaného textu na: 50, 100, 250 a 500 znakov.

3.1.3 Zlyhania systému

K neúspešným testom dochádza v implementáciach 1.0.0 a 1.1.0 v prípadoch false-positive nume. Sú to situácie, kedy používateľ (v tomto prípade náhodný generátor) zadá vstupný text obsahujúci kľúčové slovo „nume“ ešte pred formátovaním. Vtedy reťazec nemá indikovať číslo v texte a softvér, alebo v prípade mechanického zariadenia obsluhovač, rozozná, že sa jedná o false-positive kľúčové slovo pomocou nesediaceho check-sum. Pre každé písmeno textu existuje jeho číselná reprezentácia a jediný spôsob, ako overiť, či sa skutočne má jednať o číslo, je kontrola práve pomocou check-sum. Existuje ale veľmi malá pravdepodobnosť, že táto kontrolná hodnota bude sedieť aj napriek tomu, že sa jedná o false-positive nume.

Implementácia 1.2.0 pri testovania ukázala vyššiu mieru zlyhaní. Je to spôsobené chýbajúcou implementáciou ošetrenia **FPN**, čo ukázala aj kontrola logových súborov.

3.2 HCPortal

V rámci edukačnej aplikácie HCPortal bola naimplementovaná verzia PBJ 2.0.0 a edukačný komponent, ktorý umožnuje použitie šifrátora a nastavovanie jeho parametrov. Komponent poskytuje stručné vysvetlivky ku jednotlivým funkcionalitám, ktoré informujú používatela o vlastnostiach zariadenia.

Po nasadení tohto riešenia, bude možné daný komponent nájsť pod direktívou */real-ciphers/condenser-pbj*.

Na vytvorenie tejto implementácie a komponentu bol použitý jazyk TypeScript s využitím technológie Angular. Jednou z vymožeností komponentu je 3D interaktívny model mechanického zariadenia Condenser PBJ podľa vzoru z článku [1]. Na vytvorenie modelu bola použitá knižnica Three.js.

Komponent poskytuje používateľovi grafické rozhranie, ktoré sa člení podľa poskytovaných funkcionalít na 4 základné sekcie (Obr. 5.):

- Šifrovací mód
- Dešifrovací mód
- Vizuálny mód
- Nastavenia



Obr. 5. HCPortal – Úvod

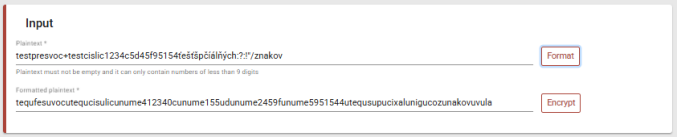
3.2.1 Šifrovací mód

Prvé interaktívne okno v šifrovacom móde umožňuje nastavenie IV. Tlačidlo Generate vygeneruje IV v potrebnom formáte podľa zvolených číselných parametrov dňa a identifikátora (ID). Ak chce používateľ manuálne zadať vlastný IV, vloží ho do textového poľa (Obr. 6.) a použije tlačidlo Validate na overenie správnosti jeho formátu. V prípade nesprávneho formátu je upozornený chybovou hláškou, ktorá presne opisuje konkrétny problém.



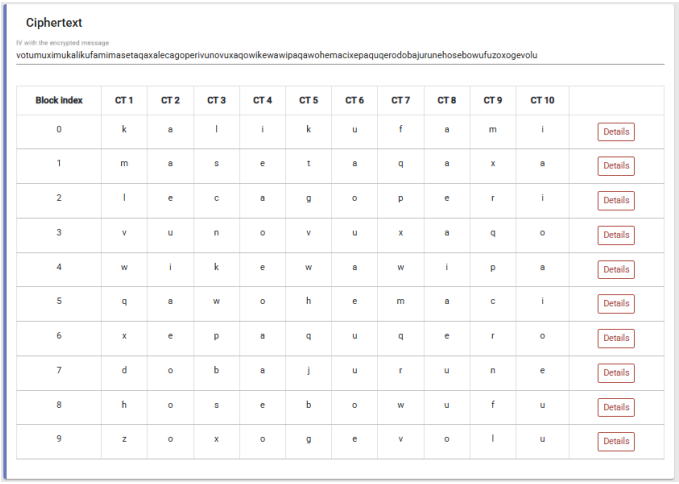
Obr. 6. HCPortal – Šifrovanie – IV

Ďalšie interaktívne okno slúži na naformátovanie vstupného textu a následné šifrovanie tohto textu. Po zadaní textu a použití tlačidla Preformat, sa naformátovaný text zobrazí v textovom poli Formatted plaintex (Obr. 7.). Po stlačení tlačidla Encrypt sa tento text zašifruje.



Obr. 7. HCPortal – Šifrovanie – Formátovanie a šifrovanie

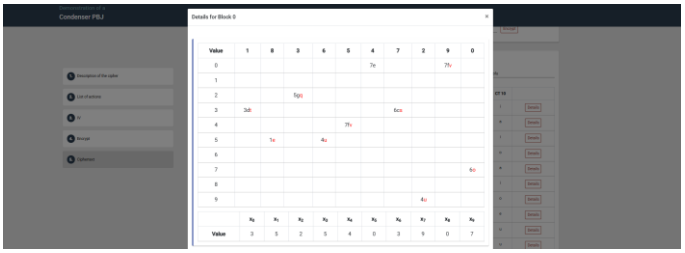
Výsledok šifrovania sa zobrazuje na novej karte. Komponent umožnuje aj detailné zobrazenie výpočtov pre jednotlivé bloky pomocou tlačidiel Details (Obr.8. a Obr. 9.).



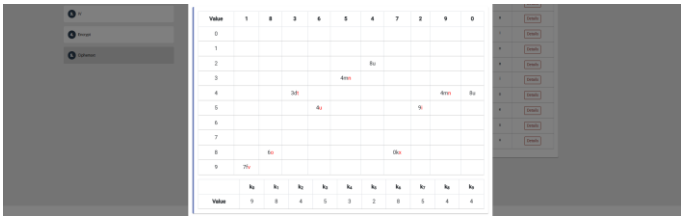
Obr. 8. HCPortal – Šifrovanie – Výsledok šifrovania

V detaile jednotlivých blokov sa nachádzajú výpočty pre každý zo znakov bloku. Konkrétne sa jedná o výpočet

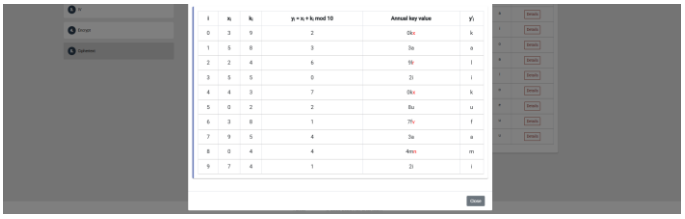
$y_i = x_i + k_i \; mod \; 10$, kde y_i, x_i a k_i sú pozície nového znaku, znaku vnútorného stavu a znaku otvoreného textu na pruhoch, podľa ročného kľúča. Pre dešifrovanie bude platit rovnaký vzorec, v ktorom sa ale hodnoty $x_i - k_i$ budú odpočítavať.



Obr. 9. HCPortal – Šifrovanie – Výsledok šifrovania – Detail 1



Obr. 10. HCPortal – Šifrovanie – Výsledok šifrovania – Detail 2

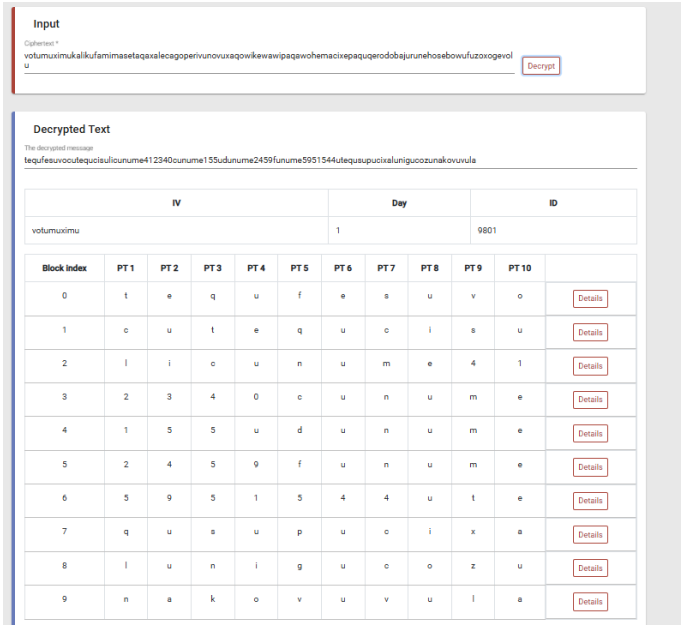


Obr. 11. HCPortal – Šifrovanie – Výsledok šifrovania – Detail 3

Nultý blok na Obr. 8. a Obr. 11. v príklade vidíme ako reťazec „votumuximu“. Šifrovaný text na Obr. 8. hore, ale začína reťazcom „wiwomuzomu“. Tento reťazec je IV, ktorý sa pridáva na začiatok šifrovanej správy a poskytuje nastavenie vnútorného stavu šifrátora pri dešifrovaní.

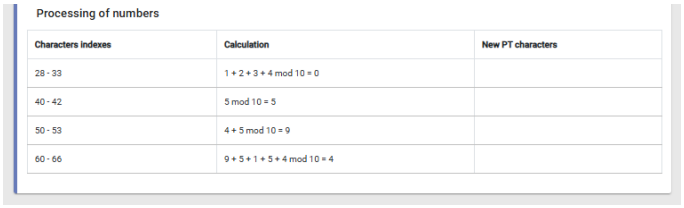
3.2.2 Dešifrovací mód

Dešifrovací mód funguje na rovnakom princípe, ako šifrovací. Používateľ vloží požadovaný text na šifrovanie a stlačí tlačidlo Decrypt. Ak je text v nevhodnom formáte, je používateľ upozornený hláškou o konkrétnom probléme. Ak bolo predtým vykonané šifrovanie, je pole pre vstupný text dešifrovania automaticky vyplnené výstupom zo šifrovacej funkcie, ale hodnotu poľa je možné zmeniť.



Obr. 12. HCPortal – Dešifrovanie – Výsledok dešifrovania

Okrem dešifrovaného textu a možnosti nahliadnutia naň, vidí používateľ aj detail o použitom IV (Obr. 12.). Algoritmus extrahuje hodnoty použitých premenných dňa a ID.

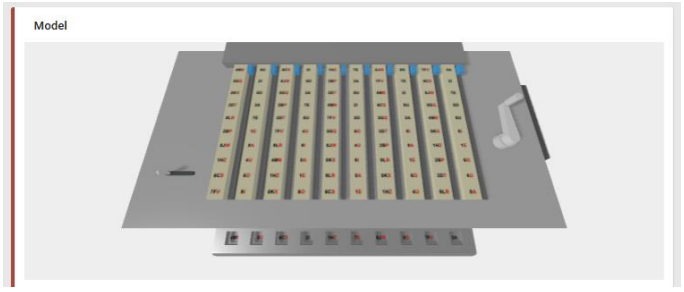


Obr. 13. HCPortal – Dešifrovanie – Detail spracovania čísel

Pod zoznamom blokov dešifrovaného textu je automaticky zobrazený aj detail o konverzii čísel (Obr. 13.).

3.2.3 Vizuálny mód

Po inštruktážnom texte o tom, ako obsluhovať mechanické zariadenie Condenser PBJ, nasleduje samotný 3D model (Obr. 14.).

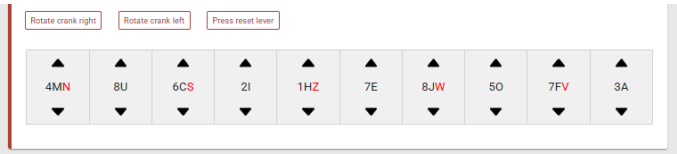


Obr. 14. HCPortal – 3D model

Používateľ môže model otáčať, približovať a odďalovať kameru a vykonávať nasledovné operácie:

- Vertikálny posun selektorov – jednotlivo
- Otáčanie kľukou do oboch smerov
- Potiahnutie resetovacej páky

Vymenované operácie používateľ vykonáva pomocou ovládacích prvkov na Obr. 15.



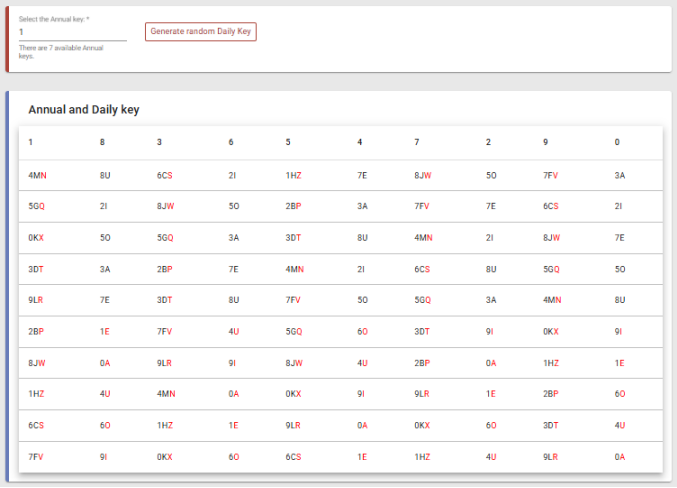
Obr. 15. HCPortal – 3D model – ovládacie prvky

Podľa článku [1] zariadenie obsahovalo aj poistku, ktorej použitie odomklo mechanizmy, ktoré držia kryt rotorov a pruhov na mieste. Vďaka tomu bolo možné vymieňať pruhy a rotory. Tento mechanizmus nebol do 3D modelu pridaný, ale poistka je na modeli zobrazená (Obr. 14 – horná stena zariadenia, vľavo dole).

Poskytovaná funkcionalita umožňuje vykonať úspešné šifrovanie a dešifrovanie podľa inštruktáží za predpokladu, že používateľ zadal korektný vstup.

3.2.4 Nastavenia

Nastavenia slúžia na nastavenie ročného a denného kľúča, prípadne zmenu použitých pruhov. Aplikácia umožňuje vygenerovať náhodný denný kľúč a vybrať si zo 7 predvolených ročných kľúčov (Obr. 16.).



Obr. 16. HCPortal – Nastavenia – Ročný a denný kľúč

Pozície hodnôt denného kľúča je možné kliknutím a pohybom myši preusporiadať (Obr. 17.). Aplikácia neumožňuje používateľovi nastaviť denný kľúč s nekorektným usporiadaním pozícií. Ak používateľ vykonal validnú výmenu dvoch pozícií, zmena sa automaticky aplikuje aj v tabuľke na Obr. 16.



Obr. 17. HCPortal – Nastavenia – Denný kľúč

Akékoľvek zmeny v nastaveniach sa okamžite aplikujú pre všetky procesy, vrátane 3D modelu.

4 Záver

V tejto práci boli naimplementované 4 verzie softvérového prevedenia Condenser PBJ, desktopová experimentálna aplikácia EE a komponent v rámci edukačného webového portálu HCPortal. Práca sa inšpirovala opisom mechanického zariadenia z článku [1], podľa ktorého sa podarilo vytvoriť spomínané implementácie a aplikovať ich v dvoch prostrediach. Práca zahŕňala aj vytvorenie súťažných výziev na prelomenie algoritmum, ktoré budú poskytnuté pre stránku MysteryTwister.

Poďakovanie

Chcel by som vyjadriť vďaku môjmu školiteľovi, doktorovi Eugenovi Antalovi, za profesionálne usmernovanie, odborné rady a podporu počas mojej práce.

Literatúra

[1] ANTAL, Eugen; ZAJAC, Pavol. *The first Czechoslovak cipher machine . Cryptologia, 47:3, 239-260. 2023.* https://www.tandfonline.com/doi/full/10.1080/01611194.2021.1998809

[2] ANTAL, Eugen; ZAJAC, Pavol. HCPortal overview. In: MEGYESI, Beáta. HistoCrypt 2020. Linköping: University Electronic Press, 2020, s. 18–20. ISBN 978-91-7929-827-2.

[3] ANTAL, Eugen; ZAJAC, Pavol. HCPortal modules for teaching and promoting cryptology. In: DAHLKE, Carola. HistoCrypt 2021. Linköping: University Electronic Press, 2021, s. 1–11. ISBN 978-91-7929-026-9.

Deciding Primality of Permutation Automata

Bc. Alexandra Reviľáková

Institute of Computer Science and Mathematics, FEI STU in Bratislava

xrevilakova@stuba.sk

Abstract - Determining whether a deterministic finite automaton (DFA) is composite allows us to keep classical automata problems tractable. A DFA is composite when its language can be expressed as the language intersection of a collection of smaller DFAs. This means that there exists a set of automata, each with fewer states than the original DFA, whose language intersection results in the same language as the original DFA. Conversely, when a DFA cannot be decomposed in this way, it is called prime. This work focuses on primality for permutation DFAs, characterized by transition monoids forming groups. We provide practical techniques for deciding the composition of permutation DFAs efficiently in practice. Building on an existing NP algorithm for permutation DFAs, we introduce two alternative algorithms. The first algorithm improves the state-coverage verification process to optimize acceptance checks, while the second additionally stores explored states, eliminating redundant computations. All three methods - the baseline NP algorithm and our two alternatives - are implemented, and their runtime complexities are empirically evaluated and compared. This study advances decomposition techniques for permutation DFAs and provides insights into optimizing primality-testing algorithms, bridging theoretical foundations with computational efficiency. The Master thesis of the author focuses on the more general topic of *Decomposition of Regular Languages*. The part on permutation DFAs follows the same approach, but the provided optimizations differ from the ones presented in this paper.

1 Introduction

A DFA is composite when there exists a set of automata, each with fewer states than the original, such that the intersection of their language is the language of the original DFA. Otherwise, when a DFA cannot be decomposed in this manner, it is called prime. The fundamental concept of automata decomposition have been introduced in [4]. In practice, decomposition allows complex systems to be broken down into manageable subproblems, improving scalability, enabling parallel processing, and facilitating efficient verification. However, the complexity class of the compositionality problem for DFAs as well as for permutation DFAs, remains unknown.

1.1 Motivating Applications

In formal verification, such as linear temporal logic (LTL) model checking [1], decomposition plays a critical role in mitigating the state explosion problem. For instance, verifying whether a system satisfies a specification $\varphi = \varphi_1 \wedge \varphi_2 \wedge \cdots \wedge \varphi_k$ can be reduced to checking each sub-specification φ_i independently if the language $L(\varphi)$ is decomposable into $\bigcap_{i=1}^{k} L(\varphi_i)$ [4]. This approach avoids constructing a single monolithic automaton for φ, which could have exponentially many states, and instead leverages smaller, parallelizable verification tasks.

In hardware design, DFA decomposition enables efficient circuit implementations. Large DFAs, which model complex control logic, often exceed the physical constraints of embedded systems or FPGA architectures [6]. By decomposing a DFA into k smaller factors, each recognizing a super-language, designers can distribute the computational load across independent modules operating on a shared input bus. For example, a DFA tracking interactions between multiple clients can be split into per-client DFAs [3].

1.2 Related Works

The problem of deciding whether a deterministic finite automaton is prime or composite has been studied across several restricted classes of automata, with varying complexity results.

Kupferman and Mosheiff introduced the problem and established that its decidability lies between NLOGSPACE-hard and EXPSPACE, with the latter bound arising from managing exponentially many candidate factors in the worst case [4]. Their algorithm for general DFAs relies on computing the intersection of all languages recognized by DFAs that include the language of the given DFA and have a smaller index. For the special case of unary DFAs (single-letter alphabets), Jecker et al. later provided a LOGSPACE upperbound by leveraging structural properties unique to unary languages [3].

Permutation DFAs, whose transition function induces a permutation of the states, admit more efficient algorithms. Initially, it was shown that the problem lies in PSPACE [4],

but subsequent work by Jecker et al. [3] improved this result by achieving NP complexity through the analysis of DFA-orbits, which can be seen as a standard subset construction, where the initial superstate can be any set of states that contains the initial states, and not just a singleton. This analysis led to a fixed-parameter tractable (FPT) algorithm, parameterized by the number of rejecting states k, with a time complexity of $O(2^k k^2 \cdot |A|)$. Furthermore, commutative permutation DFAs, where letters act commutatively, reduce the complexity even further: the problem can be solved in NL and even in LOGSPACE for fixed alphabets [3].

Finite languages, recognized by acyclic DFAs, exhibit distinct computational properties. Recent work by Spenner [8] completely characterized the primality of such automata, proving that the decomposition problem of finite languages is NL-complete.

Jecker et al. also analyzed the bounded decomposition problem, which seeks to determine whether a DFA can be decomposed into at most k factors. They demonstrated that this problem is in PSPACE for both general and permutation DFAs. For unary DFAs, this problem can be solved in LOGSPACE. However, for commutative permutation DFAs, where the decomposition problem is relatively easy, the bounded decomposition problem becomes NP-complete. This highlights the trade-off between algebraic structure and computational difficulty [3].

High computational cost of algorithms for general DFAs makes them impractical for large automata, motivating the search for more efficient methods in restricted classes of DFAs. Therefore, in this paper, we focus on permutation DFAs, which are known to be reversible. This property is relevant when modeling hardware since irreversibility causes energy dissipation [5].

1.3 Contribution

In this paper, we take a closer look at the decomposition algorithm for permutation automata. This fixed-parameter tractable algorithm is based on covering all rejecting states using orbit-DFAs. We thoroughly analyze its structure and propose the two following variantes:

- The first algorithm optimizes the state-coverage verification process, reducing the number of acceptance checks.
- The second algorithm enhances efficiency by caching previously explored orbit-DFA states, minimizing redundant computations.

We implement all three methods - the original algorithm and our two alternatives - and empirically evaluate their runtime complexities. The implementation is publicly available on GitHub[1].

[1] https://github.com/AlexxandraR/MasterThesis

1.4 Outline

The second section covers the theoretical foundations necessary for understanding the context. It initially explains concepts from the field of formal languages, then moves on to the specific area of deterministic finite automata, regular languages, and finally introduces the concepts of composite and prime automata.

The third section describes terms related to permutation automata, which are subsequently used in an algorithm that determines whether a permutation automaton is composite. This section includes a detailed example that illustrates the compositionality problem for permutation automata. The last part of the section presents descriptions of three alternative algorithms.

The first half of the fourth section provides information about the implementation - its availability, classes, important methods, and the correct format of inputs for the application. The second half of this section focuses on experiments that compare the execution times of individual algorithms. The measurement method, input data, recorded times, graphical representation, and final evaluation are an integral part of this section.

2 Preliminaries

language theory offers essential concepts like membership and inclusion, which enable us to analyze source code in programming languages and verify if programs meet specified requirements. This section will introduce fundamental concepts from formal language theory and will focus on regular languages and finite automata.

2.1 Automata

An *alphabet*, conventionally denoted by Σ, is defined as a finite, non-empty set of symbols [2]. A *string* (or *word*) is a finite sequence of symbols from an alphabet [2], with its length determined by the number of symbols it contains. The string of length zero is called *empty string* and denoted ε. The set Σ^* represents all possible strings over Σ including ε [2]. In contrast, $\Sigma^+ = \Sigma^* \setminus \{\varepsilon\}$ denotes the set of all non-empty strings [9]. A *language* is formally defined as a subset of Σ^*, consisting of strings constructed from the alphabet according to specific rules or constraints [9].

A *deterministic finite automaton (DFA)* is a classical computational model of abstract machines, heavily used in formal verification. Formally:

Definition 1 (from [2])**.** A deterministic finite automaton (DFA) is a 5-tuple $(Q, \Sigma, \delta, q_0, F)$, where:

- Q is a finite set of states,
- Σ is an alphabet,
- $\delta : Q \times \Sigma \to Q$ is the transition function,
- $q_0 \in Q$ is the initial state, and

- $F \subseteq Q$ is the set of final or accepting states.

The class of languages recognized by DFAs is known as *regular languages*, which are closed under standard operations such as union, intersection, and complementation [7]. A DFA is said to be *complete* if for every state and every input symbol, there exists a defined transition to another state; otherwise, it can be completed by introducing a *dead* state [2]. Furthermore, among all DFAs recognizing the same language, there exists a unique *minimal DFA*, which has the smallest possible number of states [4]. A *permutation DFA* is a deterministic finite automaton whose transitions form a bijective mapping of the state set. It is formally defined in the following definition.

Definition 2 (from [4])**.** Let $A = (Q, \Sigma, \delta, q_0, F)$ be a DFA. If, for every $\sigma \in \Sigma$, the transition function δ_σ forms a *permutation* of the states, then A is a *permutation DFA*.

The algorithm for determining whether a DFA is permutational checks if each transition function forms a valid permutation. It iterates through all states and input symbols, ensuring that each state appears exactly once in the transition mapping for every symbol. The time complexity is $O(|Q| \times |\Sigma|)$.

2.2 Primality and Compositionality

automata can often be analyzed by decomposing them into smaller components, making certain operations, such as language membership queries, more efficient. This motivates the distinction between *composite* and *prime* automata, where a composite automaton can be broken down into simpler parts, while a prime automaton cannot be further decomposed. These notions are formally defined as follows:

Definition 3 (from [3])**.** A DFA A is called *composite* if there exists a family $\{B_i\}_{1 \le i \le k}$ of DFAs such that $|B_i| < |A|$ for all $1 \le i \le k$, and $L(A) = \bigcap_{1 \le i \le k} L(B_i)$. This family $\{B_i\}_{1 \le i \le k}$ is called a *decomposition* of A. Note that each B_i in the decomposition satisfies $|B_i| < |A|$ and $L(A) \subseteq L(B_i)$.

Definition 4 (from [3])**.** A DFA A is called *prime* if it is not *composite*.

In other words, a language of prime DFA cannot be expressed as the language intersection of strictly smaller DFAs, meaning its structure is inherently indivisible. Since every regular language is uniquely represented by its minimal deterministic finite automaton, the notions of primality extend naturally to regular languages based on the structure of their minimal DFA [4].

3 Primality of Permutation DFAs

3.1 Orbit and Covering

Before examining the individual steps of the algorithm, it is important to first introduce key concepts such as orbit-DFAs and orbit cover. These notions serve as the fundamental basis for the algorithm and are essential for its correct implementation.

Definition 5 (from [3])**.** The orbit of U is the set $C_U = \{\delta(U, w) \subseteq Q \mid w \in \Sigma^*\}$, which consists of all subsets of Q that can be reached from U via the transition function δ. If $U \subseteq Q$ includes the initial state q_I of A, we define the orbit-DFA $A^U = \{C_U, \Sigma, \delta, U, C'\}$. Here, the state space C_U is the orbit of U, and the accepting states C' consist of subsets $U' \in C_U$ that contain at least one accepting state of A, meaning $U' \cap F \neq \emptyset$.

Observe that A^U can be interpreted as a standard subset construction, where the initial state is the set $U \subseteq Q$ with $q_I \in U$. By the definition of the state acceptance for orbit-DFA, the language of A is always a subset of the language of A^U, i.e., $L(A) \subseteq L(A^U)$ [3].

Definition 6 (from [3])**.** Given a rejecting state $q \in Q \setminus F$ in A, we say that the orbit-DFA A^U covers q if $|A^U| < |A|$ and if A^U contains a rejecting state $U' \subseteq Q$ such that $q \in U'$. By definition, this means that U' does not include any accepting state of A, i.e., $U' \cap F = \emptyset$.

In other words, an orbit-DFA A^U covers a rejecting state q if it has fewer states than A and contains a state U' that includes q while excluding any accepting states of A. This ensures that A^U preserves the rejecting behavior of A while offering a more compact representation.

Based on the former definitions, we now present the algorithm for determining whether a permutation DFA is composite by checking whether its rejecting states are covered by orbit-DFAs.

Theorem 7 (from [3])**.** *Key results:*

- *A permutation DFA is composite if and only if it can be decomposed into its orbit-DFAs.*
- *A permutation DFA can be decomposed if and only if every rejecting state is covered by at least one orbit-DFA A^U.*
- *The existence of an orbit-DFA covering a rejecting state q can be determined in non-deterministic time $O(|A|^2)$ and deterministic time $O(2^k \cdot k \cdot |A|)$, where k is the number of rejecting states of A.*
- *Let A be a permutation DFA with at least one accepting state and one rejecting state. If the number of states of A is prime, then A is prime.*

Thus, checking whether a permutation DFA is composite reduces to verifying that each rejecting state is covered by some orbit-DFA.

3.2 Illustrative Example

Consider the permutation DFA $D = (\{q_0, q_1, q_2, q_3, q_4, q_5\}, \{a\}, \delta, q_0, \{q_0\})$ where δ is given by Figure 1.

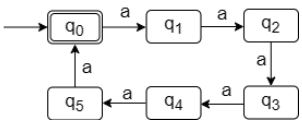


Figure 1: DFA D with the initial state and final state q_0.

Since D has five rejecting states, *orbit-DFAs* can be created by merging subsets of 2 to 5 states and applying the subset construction. In total, D has 30 different *orbit-DFAs*, depicted by Figure 2, but they are only 11 orbit-DFA automata structures. Automata with identical states, transitions, and accepting states but different initial states are represented by a single diagram, with possible initial states (containing q_0) marked by a dashed incoming arrow.

The automaton D has five orbit-DFAs with orbits of size 2, ten with orbits of size 3, ten with orbits of size 4, and five with orbits of size 5. Only the orbit-DFAs depicted by 2.(i), 3.(i), and 4.(i) have less states than D, thus being the candidates for the state-coverage. All orbit-DFAs depicted by 2.(i) and 4.(i) cover states q_1, q_2, q_4, q_5, while the orbit-DFA depicted by 3.(i) covers q_1, q_3, q_5.

If we compute the union of the states covered by each orbit-DFA, we obtain the full set of rejecting states. Since every rejecting state is covered, automaton D is composite.

Examining orbit-DFA 2.(i), we see that it accepts the language $L = \{a^{3i} \mid i \in \mathbb{N} \cup \{0\}\}$. Similarly, orbit-DFA 3.(i) accepts $L = \{a^{2i} \mid i \in \mathbb{N} \cup \{0\}\}$. Orbit-DFA 4.(i) accepts $L = \{a^{3i+2j} \mid i, j \in \mathbb{N} \cup \{0\}\}$ with the initial state $\{q_0, q_1, q_3, q_4\}$, and $L = \{a^{3i+j} \mid i \in \mathbb{N} \cup \{0\}, j \in \{0, 1\}\}$ with $\{q_0, q_2, q_3, q_5\}$. Since their intersection matches the language of D, it follows that D can be decomposed into these orbit-DFAs. In fact, D can be decomposed into two automata: the orbit-DFA depicted by 3.(i) and any orbit-DFA from 2.(i) or 4.(i).

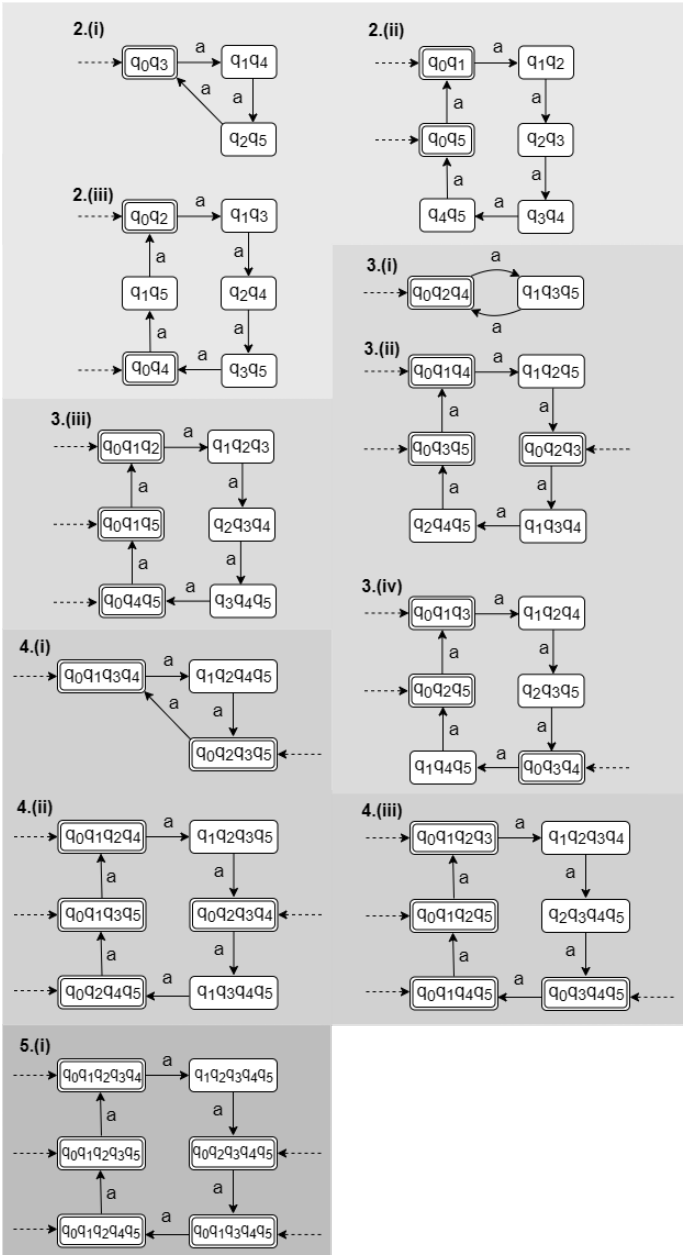


Figure 2: Orbit-DFAs of D where the possible initial states are marked with an incoming dashed arrow and final states have a double border.

3.3 Original Algorithm

The simplest algorithm for determining whether a permutation automaton is composite is based on iterating over all rejecting states of the automaton. For a given state, all states of the orbit-DFAs in which it appears are checked to verify whether the orbit-DFA covers the given rejecting state. If the orbit-DFA covers the state, the algorithm moves on to the next rejecting state. If any rejecting state is not covered by any orbit-DFA, the algorithm terminates with the result that the automaton is not composite. A detailed pseudocode is available [3].

3.4 Algorithm with saved processed states

This algorithm also utilizes orbit-DFAs and the coverage of rejecting states. It differs in how it records the covered rejecting states. Unlike the previous algorithm, this one does not iterate over the rejecting states of the original automaton but instead goes through the rejecting states of all orbits and stores all states of the original automaton covered by a given orbit. If all states are already covered, the algorithm terminates. On the other hand, if some state is not covered, the algorithm must iterate over all orbit-DFAs.

A function dealing with state coverage, instead of returning a boolean value for a single specific state, returns a set of states covered by the given orbit-DFA. Subsequently, the function determining complexity must

perform the union of sets of states covered by individual orbits and check whether all rejecting states are covered. The pseudocode of for this function is provided by Algorithm 1.

Algorithm 1 Function **isComposite**

Input: permutation DFA $A = \{Q, \Sigma, \delta, q_I, F\}$
Output: True/False
1: covered $\leftarrow \{\}$
2: **for all** U where $U \subseteq Q \setminus F$ **do**
3: covered $\leftarrow$ covered $\cup$ **cover**(A, U)
4: **if** covered $= Q \setminus F$ **then**
5: **return** True
6: **end if**
7: **end for**
8: **return** False

3.5 Algorithm with saved processed orbit-DFAs states

This algorithm is the same as the one given *Subsection 3.4*, except that it additionally stores the states of the orbit-DFAs that have already been processed in a set. That is, if we create an orbit-DFA from a certain state, we store its states and obtain the states it covers. If we wanted to process the same orbit from a different state, it would no longer be possible because this state would already be in the set of processed orbit-DFA states.

The function dealing with state coverage must have an additional input parameter - a reference to a set where elements from the resulting set C_U will be stored. In the *isComposite* function, on the second line, there must also be a condition ensuring that U is not in the set containing processed orbit-DFAs states.

4 Experiments

4.1 Implementation

All three of these algorithms were implemented in Java. The source code is available on GitHub[2]. The code is divided into three main classes: *Automaton*, *DFA*, and *FileReader*. The *FileReader* class handles loading an automaton from a text file into an instance of the *Automaton* class. This class provides methods to determine whether an automaton is deterministic, and if it is not, it includes a method for converting it into a deterministic one. The resulting deterministic automaton is represented by the *DFA* class in this code. The *DFA* class contains a large number of methods, but for this paper, the most important ones are:

- *isPermutation()* – checks whether the DFA is a permutation DFA,
- *isComposite()* – checks whether the DFA is composite using the original algorithm,
- *isCompositeTime()* – checks whether the DFA is composite using the algorithm with stored processed states from *Subsection 3.4*,
- *isCompositeMemory()* – checks whether the DFA is composite using the algorithm with stored orbit-DFAs states from *Subsection 3.5*.

The implementation also includes tests for these methods to perform elementary verification of their outputs.

As previously mentioned, the automaton is provided as input in the form of a text file, which must follow a specific structure. The initial state must be specified on the first line. This is followed by any number of transition rules in the format:

```
letter_name,source_state_id->destination_state_id
```

Finally, the accepting states are listed, with each accepting state appearing on a new line.

4.2 Evaluation

The goal of the experiments is to compare the execution time of three alternative algorithms. For the experiments, we used the same type of automata with varying numbers of states. Specifically, we considered an automaton consisting of a single cycle where the unique final state is its initial state and the alphabet is a singleton. An example of such an automaton is the DFA D from the example in *Subsection 3.2*. The experiments were conducted on a laptop running Windows 11 Pro, with an Intel Core i5 2.6 GHz processor and 8 GB of RAM.

In the first experiment (*Figure 3*), we compare exclusively automata that are prime. According to the last point of *Theorem 7*, we know that every permutation automaton with at least one accepting state, one rejecting state, and a total prime number of states is prime. For a prime automaton, all three algorithms must iterate over every orbit because some state will not be covered by any orbit. As expected, execution time gradually increased with the growing number of states of the automaton. Since the differences in execution time between the smallest and largest automaton are very large, a logarithmic scale is used in the graphs to make the comparison of the algorithms more visible even for small automata. For the *isCompositeMemory* algorithm, the execution time is the lowest for the smallest automaton and the highest for the largest automaton compared to the other algorithms. These results can be explained by the increasing amount of stored processed states of orbit-DFAs and the search within this set. For automata with more than five states, the execution time of the *isCompositeTime* algorithm is consistently higher than that of the original *isComposite* algorithm.

[2] `https://github.com/AlexxandraR/MasterThesis`

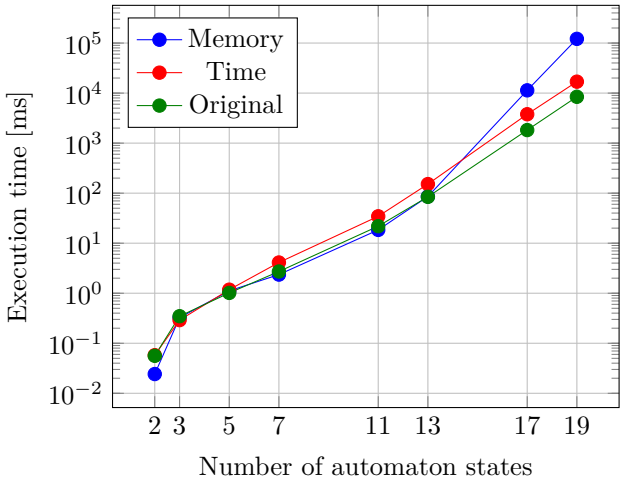


Figure 3: Execution time comparison for different number of prime automaton states

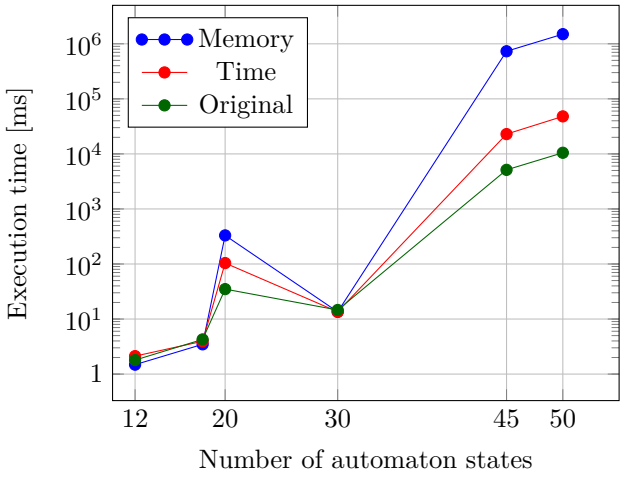


Figure 4: Execution time comparison for different number of composite automaton states

In the second experiment (*Figure 4*), on the contrary, we compare exclusively automata that are composite. We selected automata sizes as the product of three prime numbers, where one of them may appear squared. An interesting observation is that execution time do not always increase with the growing number of states. In particular, we can see a decrease in execution time for an automaton of size 30 in *Figure 4*. A notable finding is that the execution time does not consistently increase with the number of states. Specifically, in Figure 4, we observe that the automaton with 30 states shows a drop in execution time. This is because the runtime is influenced not just by the automaton's size or whether it's prime, but also by how many original states are grouped into a single orbit state. The algorithm constructs orbit states by gradually merging original states - starting with pairs, then triples, and continuing up to one less than the number of rejecting states. Naturally, forming orbit states for larger combinations takes more time. Since it's not possible to anticipate how the original states will be grouped to cover all rejecting states, it's also difficult to predict when the algorithm will run faster. As shown in the graph, the automaton with 30 states is processed faster than the 20-state one. This difference arises because the primality of the 30-state automaton can be checked using orbit states formed from combinations of 2 or 3 original states, while the 20-state automaton requires orbits composed of 2 and 5 original states. When comparing the algorithms, we observe almost identical results as in the previous experiment. It seems that *isCompositeMemory* algorithm is the fastest for small automata, whereas for big automata, it is significantly slower. The *isCompositeTime* algorithm is slower than the *original* for most automata, but for bigger automata, it still outperforms *isCompositeMemory*.

5 Conclusion and Future Works

This paper focuses on deciding the primality of permutation automata. We considered three alternative algorithms: the original algorithm from paper [3], a variant algorithm that saves processed states, and another variant that saves processed orbit-DFA states. We compared experimentally the execution times of these three algorithms separately for automata that are prime and separately for composite automata. In both cases, the original algorithm performed best overall, although, some instances may benefit from the memoization technique. However, these cases occurred only for small automata, and the differences between the algorithms were only on the order of tens of milliseconds. These results may be due to the different order of exploring orbit-DFA states and the storage and exploration of a large number of states, either of the orbit-DFA or the original automaton.

Since the current known algorithm for deciding primality for unrestricted DFAs has a high computational cost, the research community working on primality recently focused on subclasses of DFAs. Identifying efficient algorithms to decide primality for these subclasses contributes to the research effort by providing tools for running empirical experiences. A natural future work is thus to investigate other subclasses, for example DFAs that are both permutation and commutative, or safe DFAs, in order to trigger further memory optimizations. Similarly, implementing the decision algorithm for finite languages may involve different optimization techniques.

Acknowledgments

I would like to thank Dr. Nicolas Adrien Mazzocchi for his invaluable guidance and collaboration in the development of this work.

References

[1] Christel Baier and Joost-Pieter Katoen. *Principles of Model Checking*. MIT Press, 2008.

[2] John E Hopcroft, Rajeev Motwani, and Jeffrey D Ullman. *Introduction to automata theory, languages, and computation*. Pearson, Upper Saddle River, NJ, 3 edition, 2007.

[3] Ismaël Jecker, Nicolas Mazzocchi, and Petra Wolf. Decomposing permutation automata. In *32nd International Conference on Concurrency Theory (CONCUR 2021)*, volume 203 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 18:1–18:19. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2021.

[4] Orna Kupferman and Jonathan Mosheiff. Prime languages. *Information and Computation*, 240:90–107, 2015.

[5] Rolf Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5(3):183–191, 1961.

[6] Volnei A. Pedroni. *Finite State Machines in Hardware: Theory and Design (with VHDL and SystemVerilog)*. MIT Press, 2013.

[7] Michael Sipser. *Introduction to the Theory of Computation*. Cengage Learning, Boston, MA, 3 edition, 2013.

[8] Daniel Alexander Spenner. Decomposing finite languages. In *Mathematical Foundations of Computer Science 2023 (MFCS 2023)*, pages 83:1–83:14. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2023.

[9] Ľubomír Dedera. *Počítačové jazyky a ich spracovanie*. Liptovský Mikuláš, 2014.

Vplyv viackrokového dotrénovania na úspešnosť a vysvetliteľnosť klasifikácie obrazov z optickej koherenčnej tomografie

Bc. Samuel Gibala[1], MUDr. Veronika Kurilová, PhD., FEBO[1,2]

[1]Ústav informatiky a matematiky, FEI STU v Bratislave
[2] Lekárska fakulta, Slovenská zdravotnícka univerzita, Bratislava

xgibala@stuba.sk

Abstrakt – Kvalitu modelov strojového učenia najčastejšie vyhodnocujeme pomocou metrík úspešnosti. Tieto metriky však poskytujú len čiastočný obraz o skutočnej výkonnosti modelu, najmä pri medicínskych dátach, kde je kľúčová aj vysvetliteľnosť. V tejto práci predstavujeme metódu viackrokového dotrénovania modelov strojového učenia, ktorá má potenciál nielen zlepšiť úspešnosť, ale aj zvýšiť mieru vysvetliteľnosti na takú úroveň, ktorá je podobná rozhodovaniu špecialistu (oftalmológa). Tento prístup overujeme na úlohe klasifikácie obrazov z optickej koherenčnej tomografie s využitím metódy GradCAM.

1 Úvod

Optická koherenčná tomografia (OCT) je neinvazívna zobrazovacia metóda, ktorá umožňuje detailnú vizualizáciu štruktúry sietnice oka. Vďaka vysokej priestorovej rozlišovacej schopnosti poskytuje prierezové snímky zobrazujúce jednotlivé vrstvy sietnice, čo je kľúčové pri diagnostike a monitorovaní ochorení [1].

V tejto práci sa zameriavame na oblasť makuly, ktorá predstavuje centrálnu časť sietnice a je zodpovedná za najostrejšie videnie. Makula obsahuje vysokú koncentráciu čapíkov, ktoré sú nevyhnutné pre farebné videnie a detailné rozlišovanie objektov [2]. Zmeny v štruktúre makuly môžu viesť k vážnym poruchám zraku, preto je presná diagnostika týchto zmien pomocou OCT snímok mimoriadne dôležitá.

Automatizovaná analýza OCT snímok pomocou modelov strojového učenia ponúka kvalitné metódy na diagnostiku očných ochorení. V našej práci sa venujeme viactriednej klasifikácii OCT snímok, pričom využívame vision transformer model MedViT, ktorý je špecificky navrhnutý na spracovanie medicínskych obrazových dát [3].

Jedným z kľúčových aspektov praktického využitia takýchto modelov v medicíne je vysvetliteľnosť ich rozhodovania. Pri oftalmologickej diagnostike nie je dostatočné, aby model len správne klasifikoval snímku – dôležité je aj to, aby poskytol informáciu o tom, na ktoré oblasti snímky sa pri rozhodovaní najviac zameral. Táto interpretovateľnosť môže pomôcť oftalmológom lepšie pochopiť závery modelu a zvýšiť ich dôveru v automatizované diagnostické nástroje.

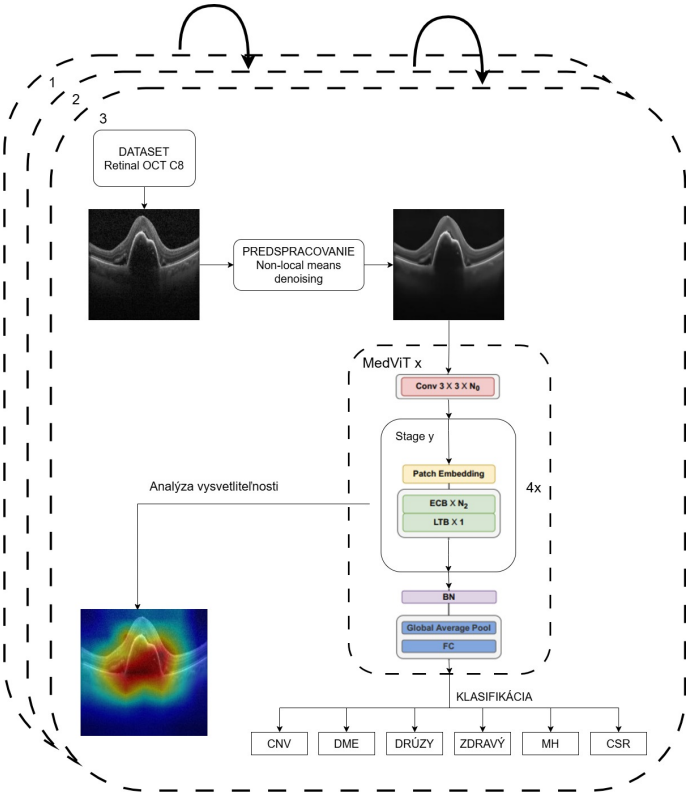
V tejto práci preto skúmame vplyv viackrokového dotrénovania modelov na ich úspešnosť a vysvetliteľnosť. Viackrokové dotrénovanie v porovnaní so štandardným tréningom preukázateľne dosahuje kvalitnejšie výsledky a efektívnu schopnosť transferového učenia. Navyše umožňuje modelom efektívne využívať naučené príznaky a zároveň sa prispôsobovať novým úlohám, čím dosahujú lepšie výsledky aj s obmedzeným množstvom údajov [5] [6] [7]. Na hodnotenie vysvetliteľnosti využívame metódu Grad-CAM , ktorá umožňuje vizualizovať oblasti snímok, ktoré model považuje za najdôležitejšie pri klasifikácii [8]. Naším cieľom je zistiť, či viackrokové dotrénovanie dokáže nielen zlepšiť klasifikačný výkon modelov, ale aj zvýšiť ich interpretovateľnosť pre praktické medicínske použitie.

2 Opis riešenia

Na obrázku 1 môžeme vidieť schému metodiky práce, ktorá znázorňuje jednotlivé kroky spracovania a analýzy obrazových dát v našej práci. Proces začína výberom vhodnej databázy, v našom prípade ide o databázu Retinal OCT C8 [4]. Každý obrázok z databázy je následne predspracovaný metódou Non-local Means Denoising [9], ktorá je použitá na redukciu šumu. Upravené dáta sú potom vstupom do hybridného CNN-Transformer modelu MedViT [3], ktorý vykonáva klasifikáciu snímok do jednotlivých tried. Na analýzu vysvetliteľnosti je využitá metóda Grad-CAM [8], ktorá zvýrazňuje oblasti obrázka, na ktoré sa model pri rozhodovaní zameral. Viaceré vrstvy schémy metodiky práce znázorňujú viackrokové dotrénovanie, kde je vždy dotrénovaný model z predošlého trénovania na pôvodnom datasete a sú osobitne vyhodnotené všetky relevantné metriky trénovania a validácie modelov. Časti diagramu vychádzajú z pôvodnej architektúry MedViT [3] a boli prispôsobené pre potreby tejto práce.

2.1 Retinal OCT C8 dataset

Retinal OCT C8 je dataset, ktorý obsahuje 24 000 vysokokvalitných OCT snímok centra sietnice rozdelených do 8 tried ochorení centra sietnice. Je určený na podporu výskumu a



Obr. 1: Schéma metodiky práce

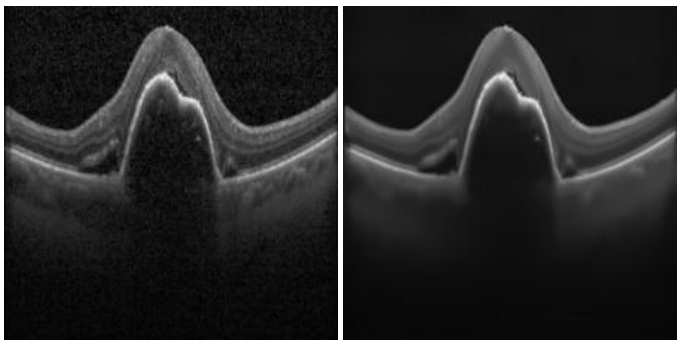
(CNV- choroidálna neovaskularizácia, DME- diabetický edém makuly, MH- diera v makule (z angl. macular hole), CSR- centrálna serózna chorioretinopatia)

tréningu modelov v oblasti klasifikácie ochorení centra sietnice pomocou techník strojového učenia. Snímky každej triedy sú rozdelené na trénovacie, testovacie a validačné. Pre každé ochorenie je k dispozícii 2300 trénovacích, 350 testovacích a 350 validačných snímok [4].

Po analýze snímok a konzultácii s oftalmologičkou MUDr. Veronikou Kurilovou PhD., FEBO sme sa z dôvodu klinického prekrývania jednotlivých tried rozhodli zredukovať dataset na 6 tried. Odstránená bola trieda označená v datasete ako DR, čo predstavuje diabetickú retinopatiu, pri ktorej je predpokladané klinické prekrývanie s triedou DME, čo predstavuje diabetický makulárny edém. Druhou odstránenou bola trieda označená ako AMD, čo predstavuje vekom podmienenú degeneráciu makuly, pri ktorej je predpokladané klinické prekrývanie s triedami CNV (choroidálna neovaskularizácia) a DRUSEN (drúzy), nakoľko ide o vlhkú a suchú formu tohto ochorenia.

2.2 Predspracovanie snímok

Prvým krokom pred samotnou implementáciou riešenia bola analýza datasetu. Z nej sme zistili, že snímky nie sú jednotnej veľkosti, rozlíšenia a najmä vykazujú vysokú mieru šumu. Na základe toho sme sa rozhodli použiť metódy predspracovania tak, aby sme zjednotili snímky na veľkosť požadovanú vybraným modelom a v čo najvyššej miere zredukovali šum. Zmena veľkosti bola vykonaná metódou bilineárnej interpolácie, ktorá plynule škáluje obraz výpočtom hodnôt pixelov na základe váženého priemeru susedných pixelov. Na redukciu šumu sme testovali viacero prístupov (Bilateral Filtering, Non-local Means Denoising, Gaussian Blurring, Median Blurring), najlepšie výsledky klasifikácie sme dosiahli pri použití metódy Non-local Means Denoising algorithm. Jedná sa o metódu, ktorá redukuje šum a zároveň zachováva textúry a jemné detaily snímky. Na rozdiel od tradičných metód, ktoré sa spoliehajú len na lokálne okolie, využíva nadbytočnosť podobných vzorov v celom obraze na dosiahnutie účinnejšej redukcie šumu [9]. Na ľavej strane obrázku 2 môžeme vidieť príklad snímky z datasetu Retinal OCT C8. Na snímke si môžeme všimnúť spomínaný šum najmä v jej tmavých častiach. Na pravej strane obrázku 2 je zobrazená rovnaká snímka s aplikovanou redukciou šumu spomínanou metódou. Na obrázku si môžeme všimnúť, že tmavé plochy sú značne vyhladené a zároveň kľúčové prvky snímky, ktorými sú jednotlivé vrstvy, ostali zachované a výrazné.



Obr. 2: Porovnanie pôvodnej snímky (vľavo) a snímky s aplikovanou redukciou šumu (vpravo)

2.3 MedViT model

MedVit je robustný hybridný CNN-Transformer model určený na klasifikáciu medicínskych snímok z rôznych oblastí. Kombinuje silné stránky konvolučných neurónových sietí (CNN) a vision transformer modelov na riešenie výziev v analýze medicínskych snímok, ako je modelovanie vzdialených vzťahov a odolnosť modelu voči nepriaznivým vplyvom [3].

Najdôležitejšími časťami modelu sú Efficient Convolution Block (ECB) zachytávajúci lokálne príznaky a Local Transformer Block (LTB) určený na zachytávanie viacfrekvenčných signálov. Využitím spomenutých prístupov je dosiahnuté zlepšenie extrakcie globálnych aj lokálnych príznakov. Okrem toho MedViT využíva novú augmentačnú techniku Patch Momentum Changer (PMC), ktorá zvyšuje robustnosť a schopnosť generalizácie kombináciou normalizácie príznakov a augmentácie dát na úrovni tokenov. Model preukazuje vysokú presnosť a efektivitu na rozsiahlych medicínskych obrazových datasetoch, ako je napríklad dataset MedMNIST-2D [3].

MedMNIST 2D je rozsiahla zbierka normalizovaných a predspracovaných 2D biomedicínskych snímok z rôznych zdrojov vrátane röntgenových snímok, ultrazvukových snímok a snímok z CT, ktorá je určená na rôzne, najmä klasifikačné úlohy a je ideálna na porovnávanie algoritmov strojového učenia pri analýze biomedicínskych obrázkov [10].

Model MedViT je k dispozícii v 3 verziách, ktoré sa od seba líšia počtom parametrov:

- MedViT-T (Tiny) - 10.8 milióna parametrov,
- MedViT-S (Small) - 23.6 milióna parametrov,
- MedViT-L (Large) - 45.8 milióna parametrov.

V našej implementácii sme sa rozhodli použiť strednú verziu MedViT-S, ktorá je v implementácii zaužívaná pod označením MedViT_base.

2.4 Trénovanie a validácia

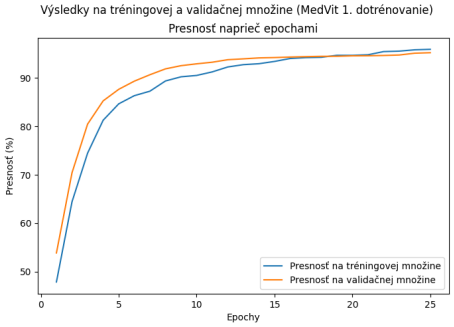
Viackrokové trénovanie modelu, ktoré zvyčajne zahŕňa predtrénovanie na veľkom všeobecnom datasete a následné doladenie na špecifickom datasete, výrazne zvyšuje výkonnosť v rôznych úlohách strojového učenia. Tento prístup využíva silné stránky transferového učenia, pri ktorom sa znalosti získané zo všeobecného datasetu prispôsobujú špecifickým úlohám, čím sa zvyšuje presnosť aj efektívnosť [5].

Ako uvádza Howard a kol. príznaky v konvolučných neurónových sieťach prechádzajú zo všeobecných na špecifické pre danú úlohu, pri prechode z počiatočných do koncových vrstiev siete. Táto vlastnosť umožňuje účinný prenos naučených vlastností zo všeobecných datasetov do špecifických aplikácií, kde môže byť potrebné len jemné doladenie posledných vrstiev, čím sa šetria zdroje a čas. Okrem toho druhá fáza jemného dolaďovania, ktorá využíva nižšiu rýchlosť učenia, umožňuje jemnejšie úpravy modelu, čo mu umožňuje lepšie zachytiť špecifické prvky konkrétnej domény bez toho, aby sa príliš odchýlil od robustných vlastností naučených počas predtrénovania [6] [7].

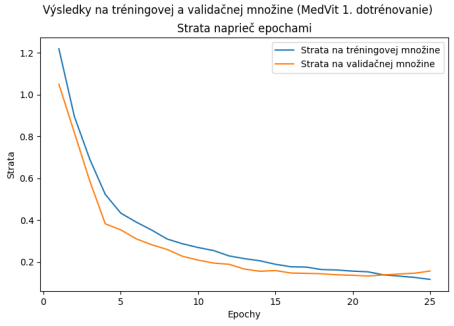
V našej implementácii sme ako základ použili model MedViT predtrénovaný na datasete MedMNIST-2D. Vo všetkých fázach dotrénovania modelu sa používa stochastický gradientný zostup (SGD) ako optimalizátor s využitím momentum techniky. Ako trénovací dataset je použitý spomenutý upravený dataset Retinal OCT C8.

Prvá fáza trénovania prebiehala po viacerých experimentoch 25 epoch s rýchlosťou učenia 10^{-4}, momentom 0.9, keďže model po tomto bode prestal vykazovať významné zlepšenie. Výsledný graf vývoja presnosti počas trénovania môžeme vidieť na obrázku 3. Na grafe je viditeľné, že model mal počas celého trénovania tendenciu postupne zlepšovať presnosť ako na tréningovej, tak aj na validačnej množine a nevykazuje počas tejto fázy známky pretrénovania. Dôkazom je aj graf vývoja straty na tréningovej a validačnej množine zobrazený na obrázku 4. Obe monitorované metriky mali počas trénovania tendenciu vyvíjať sa spoločným smerom a rovnako ani ich hodnoty sa v priebehu trénovania príliš nelíšili.

Výsledné hodnoty skúmaných metrík modelu sú uvedené v tabuľke 1, kde stĺpce predstavujú jednotlivé fázy trénovania. V metrikách pre prvú fázu dotrénovania si môžeme všimnúť, že model vykazuje pomerne vysokú presnosť, pričom hodnoty na všetkých množinách sú si veľmi podobné. Rovnako aj metriky Precision a Recall naznačujú, že je model nielen presný a spoľahlivý, ale aj vyvážený.



Obr. 3: Presnosť na tréningovej a validačnej množine prvej fázy trénovania

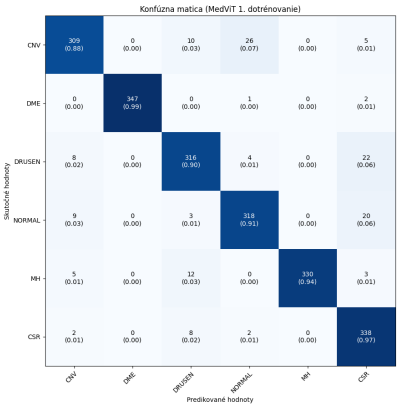


Obr. 4: Strata na tréningovej a validačnej množine prvej fázy trénovania

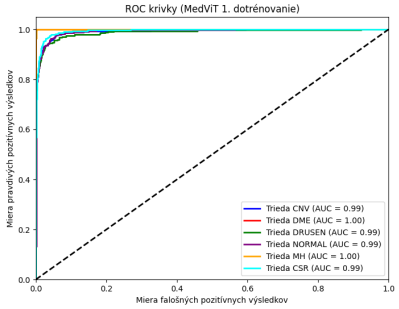
Ďalej sme výsledky vizualizovali pomocou konfúznej matice zobrazenej na obrázku 5 a ROC krivky spolu s metrikou AUC-ROC zobrazenou na obrázku 6. Konfúzna matica nám umožňuje vyhodnotiť presnosť na testovacej množine pre každú triedu samostatne. V našej konfúznej matici vidíme okrem výraznej diagonály, naznačujúcej vysokú presnosť a schopnosť generalizácie, aj pomerne malé odchýlky medzi jednotlivými triedami. Potvrdzuje to aj graf ROC kriviek, ktorých priebeh je na hornej ľavej hranici alebo blízko nej, čo v spojení s hodnotami AUC-ROC medzi 0.99 a 1.00 naznačuje, že model je veľmi presný pri rozlišovaní medzi triedami.

Druhá fáza trénovania prebiehala 10 epoch s rýchlosťou učenia 10^{-5}, momentom 0.9, nakoľko model po tomto bode začal stagnovať. Výsledný graf vývoja presnosti počas trénovania môžeme vidieť na obrázku 7. Z grafu je zrejmé, že model má aj v tejto fáze tendenciu zlepšovať presnosť kontinuálne na trénovacej aj validačnej množine. Dôkazom je aj graf vývoja straty zobrazený na obrázku 8 z ktorého je viditeľné, že sa strata na validačnej množine začína odchylovať až v záverečnej epoche trénovania.

Porovnaním prvej a druhej fázy trénovania v tabuľke 1 si môžeme všimnúť, že bol model schopný vylepšiť svoju presnosť na testovacej množine viac ako 1.9 %. Toto zlepšenie reflektujú aj hodnoty metrík Precision, Recall a F1 score. Dôležité je podotknúť, že touto fázou nedošlo k výraznejšiemu zhoršeniu alebo odchýleniu žiadnej z pozorovaných metrík. Kvalitatívne zlepšenie môžeme pozorovať aj na konfúznej matici zobrazenej na obrázku 9. Jednotlivé triedy vykazujú



Obr. 5: Konfúzna matica prvej fázy trénovania



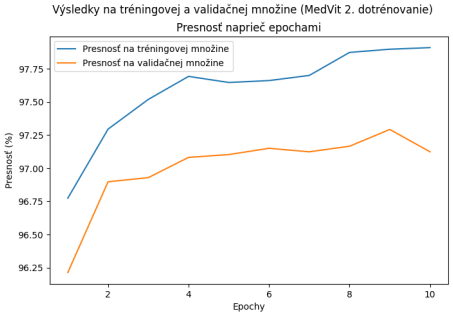
Obr. 6: ROC krivky prvej fázy trénovania

vysoké miery presnosti s progresom napríklad pri triede CNV až o okolo 5% v porovnaní s prvou fázou trénovania. Pri ROC krivkách zobrazených na obrázku 10 taktiež vidíme istý progres, pričom počet kriviek s hodnotou AUC-ROC menšou ako 1 sa znížil na 2 v porovnaní s prvou fázou, kde boli takéto krivky 4.

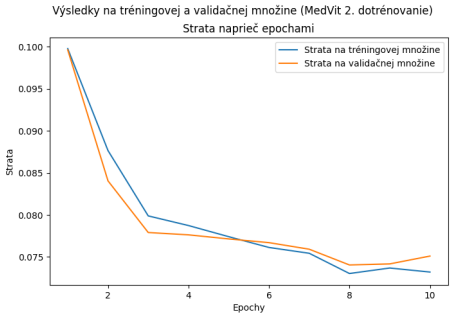
Tretia fáza trénovania prebiehala 10 epoch s rýchlosťou učenia 10^{-6}, momentom 0.9. Výsledný graf vývoja presnosti počas trénovania môžeme vidieť na obrázku 11. Z grafu sú zrejmé výrazné výkyvy presnosti naprieč jednotlivými epochami trénovania, čo zvýrazňuje aj graf straty na tréningovej a validačnej množine na obrázku 12, na základe čoho bolo trénovanie modelu v tomto bode zastavené. Z grafov je zrejmá výrazná stagnácia vývoja presnosti a chyby naprieč jednotlivými epochami. Dôkazom sú aj skúmané metriky, pri ktorých nastali len zanedbateľné zmeny oproti druhej fáze trénovania, a preto môžeme tretiu fázu považovať za neúspešnú a na základe zistených poznatkov skonštatovať, že model pravdepodobne dosiahol maximálnu možnú mieru presnosti na zvolenom datasete.

2.5 Analýza vysvetliteľnosti modelu

Pre potreby analýzy vysvetliteľnosti natrénovaných modelov sme zvolili doplnkovú knižnicu ku knižnici PyTorch, ktorá poskytuje state-of-the-art metódy pre vysvetliteľnosť umelej inteligencie v oblasti počítačového videnia.



Obr. 7: Presnosť na tréningovej a validačnej množine druhej fázy trénovania

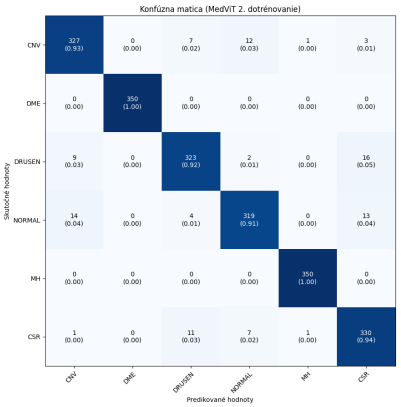


Obr. 8: Strata na tréningovej a validačnej množine druhej fázy trénovania

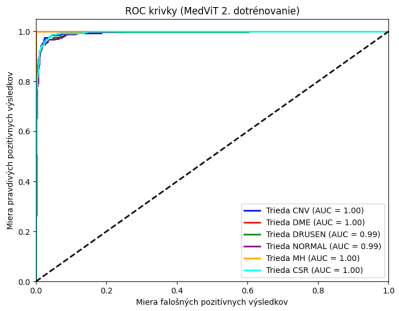
Táto knižnica je navrhnutá pre širokú škálu sietí od tradičných konvolučných neurónových sietí až po vision transformer modely a rovnako aj pre širokú škálu využití ako je klasifikácia, sémantická segmentácia, detekcia objektov a iné. GradCAM je jednou z viacerých metód, ktoré knižnica ponúka. Jedná sa o vykreslenie vážených 2D aktivácií konvolučnej vrstvy priemernou hodnotou gradientu. V súvislosti s vision transformer modelmi by sa však dalo povedať, že sa jedná skôr o vykreslenie vážených 2D reprezentácií patchov (podmnožín snímok, na ktoré je vstupná snímka rozdelená) priemernou hodnotou gradientu [8].

Z dôvodu čo najlepšej miery vysvetliteľnosti sme sa rozhodli vizualizovať výstupy GradCAM metódy ako heatmapu, kde farebná škála od tmavomodrej po tmavočervenú predstavuje mieru, s akou sa model zameral na danú oblasť – od najnižšej (tmavomodrá) po najvyššiu (tmavočervená). Na obrázku 13a s ochorením CNV je zobrazený výstup metódy aplikovanej na pôvodný model predtrénovaný na datasete MedMNIST-2D. Môžeme si všimnúť, že sa model zameral na pomerne veľkú oblasť snímky a pri rozhodovaní pripisoval vyššiu dôležitosť rozsiahlej oblasti v jej spodnej časti. Takéto správanie modelu nie je v súlade s našimi požiadavkami, keďže vo väčšine prípadov vieme presne lokalizovať úzku oblasť, kde sa nachádzajú typické znaky ochorení alebo zdravej snímky.

Obrázok 13b predstavuje výstup metódy aplikovanej na model po prvej fáze trénovania. Vidíme, že oblasť, na ktorú sa model zameriava, je výrazne menšia a vyššia dôležitosť je prikladaná oblasti, na ktorej sa skutočne nachádzajú typické znaky ochorenia zobrazeného na snímke. Po konzultácii s



Obr. 9: Konfúzna matica druhej fázy trénovania

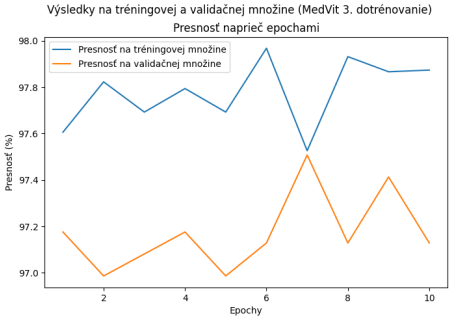


Obr. 10: ROC krivky druhej fázy trénovania

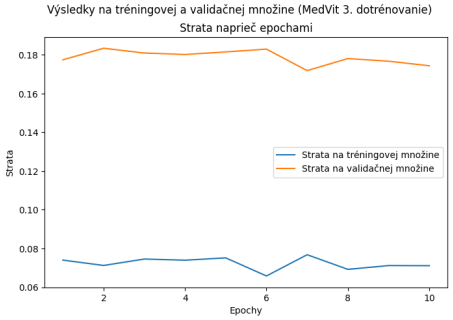
oftalmológom môžeme konštatovať, že model po prvej fáze trénovania dosahuje v oblasti vysvetliteľnosti kvalitnejšie výsledky ako predchádzajúci.

Na obrázku 13c je zobrazený výstup metódy aplikovanej na model po druhej fáze trénovania. Oproti predchádzajúcemu modelu je viditeľné ešte presnejšie zameranie na oblasť, kde sa skutočne nachádzajú typické znaky ochorenia na snímke. Model zároveň prikladá vysokú dôležitosť jednotlivým vrstvám sietnice, ktoré predstavujú kľúčovú štruktúru v snímkach z optickej koherenčnej tomografie. Model po druhej fáze trénovania sa zreteľne presnejšie zameriava na kľúčové časti snímky z pohľadu klasifikácie ochorení. Z hľadiska vysvetliteľnosti model jednoznačne a správne identifikuje dôležitú oblasť snímky.

Obrázok 13d predstavuje výstup metódy aplikovanej na model po tretej fáze trénovania. V porovnaní s predošlým obrázkom treťou fázou trénovania nenastali viditeľné zmeny, čo zodpovedá aj zaznamenanej stagnácii modelu počas tejto fázy trénovania. Z toho vyplýva, že aj z pohľadu vysvetliteľnosti môžeme tretiu fázu považovať za neúspešnú a na základe týchto poznatkov môžeme predpokladať, že model pravdepodobne dosiahol maximálnu možnú mieru vysvetliteľnosti na zvolenom datasete.



Obr. 11: Presnosť na tréningovej a validačnej množine tretej fázy trénovania



Obr. 12: Strata na tréningovej a validačnej množine tretej fázy trénovania

3 Záver

V práci sme skúmali vplyv viackrokového dotrénovania na úspešnosť a vysvetliteľnosť klasifikácie obrazov z optickej koherenčnej tomografie. Ako model bol zvolený hybridný CNN-Transformer model MedViT predtrénovaný na datasete MedMNIST-2D, ktorý preukazuje výborné výsledky na medicínskych dátach. Pre potreby dotrénovania a validácie modelu bolo použitých 6 tried z datasetu Retinal OCT C8, ktorý obsahuje vysokokvalitné OCT snímky centra sietnice.

Z pohľadu vplyvu na úspešnosť klasifikácie môžeme zhodnotiť, že model viackrokovým dotrénovaním vykazoval najlepšie úspešnosti klasifikácie po dvoch dotrénovaniach a následne pri treťom dotrénovaní začal stagnovať a nebol schopný ďalšieho napredovania. Model dosiahol úspešnosť na trénovacej množine 95.1905 % s vyváženými metrikami Precision, Recall a F1 score. Vysokú presnosť potvrdzuje aj konfúzna matica, z ktorej je zreteľné, že sú všetky triedy rovnomerne natrénované. Zo zobrazených ROC kriviek, ktorých priebeh je na hornej ľavej hranici alebo veľmi blízko nej a hodnôt AUC-ROC medzi 0.99 a 1.00 usudzujeme, že model je veľmi presný pri rozlišovaní medzi triedami.

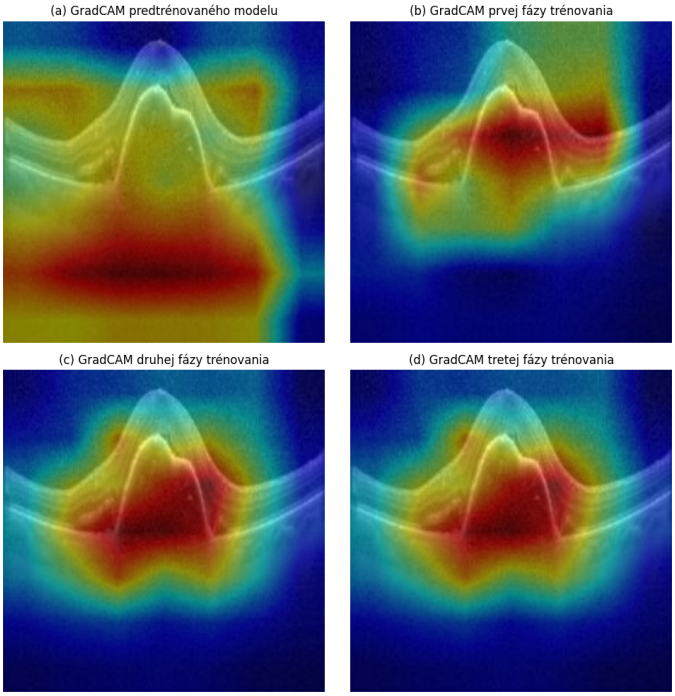
Analýza vizualizácie výstupov metódy GradCAM ako heatmapy ukázala postupné zlepšovanie zamerania modelu na relevantné oblasti snímok v priebehu jednotlivých fáz trénovania. Kým predtrénovaný model sa sústredil na rozsiahlejšie a všeobecné časti snímky, prvá a druhá fáza trénovania viedli k výraznejšiemu zvýšeniu presnosti pri identifikácii oblastí

Tab. 1: Skúmané metriky pre jednotlivé fázy trénovania

dotrénovanie č.	1	2	3
Presnosť na tréningovej množine	95.6742	97.6345	97.9167
Strata na tréningovej množine	0.1275	0.0752	0.0699
Presnosť na validačnej množine	93.7973	94.2235	94.1288
Strata na validačnej množine	0.2011	0.1740	0.1743
Presnosť na testovacej množine	93.2857	95.1905	95.1905
Strata na testovacej množine	0.2017	0.1459	0.1484
Precision	0.9348	0.9520	0.9524
Recall	0.9329	0.9519	0.9519
F1 skóre	0.9332	0.9518	0.9518

obsahujúcich typické znaky ochorení. Druhá fáza trénovania preukázala najvyššiu mieru vysvetliteľnosti, pričom model efektívne identifikoval dôležité časti snímky, vrátane dôrazu na vrstvy sietnice, ktoré sú pri optickej koherenčnej tomografii kľúčové a riadia sa nimi pri diagnostike aj oftalmológovia. Naopak, tretia fáza trénovania už nepriniesla ďalšie zlepšenie, čo potvrdzuje stagnáciu modelu a naznačuje dosiahnutie jeho maximálnej možnej miery vysvetliteľnosti na zvolenom datasete.

Uvedené výsledky naznačujú, že viackrokové dotrénovanie môže byť efektívnym nástrojom na zlepšenie nielen klasifikačnej presnosti, ale aj vysvetliteľnosti modelov v oblasti medicínskeho zobrazovania. V ďalšom výskume plánujeme otestovať tento prístup na širšom spektre modelov strojového učenia s rôznymi architektúrami, počtom dotrénovaní a nastaveniami parametrov. Cieľom je identifikovať optimálny model, ktorý bude kombinovať vysokú presnosť na jednej strane a vysokú mieru vysvetliteľnosti na strane druhej, pričom bude zároveň zohľadňovať potreby a požiadavky medicínskych odborníkov pre jeho praktickú akceptáciu a využitie v klinickej praxi.



Obr. 13: GradCAM vizualizácie predtrénovaného modelu a jednotlivých fáz trénovania modelu

Literatúra

[1] A. H. Kashani, C. Chen, J. K. Gahm, F. Zheng, G. M. Richter, P. J. Rosenfeld, Y. Shi, and R. K. Wang. Optical coherence tomography angiography: A comprehensive review of current methods and clinical applications. *Progress in Retinal and Eye Research*, 60:66–100, 2017. https://doi.org/10.1016/j.preteyeres.2017.07.002.

[2] A. P. Voigt, N. K. Mullin, S. S. Whitmore, A. P. DeLuca, E. R. Burnight, X. Liu, B. A. Tucker, T. E. Scheetz, E. M. Stone, and R. F. Mullins. Human photoreceptor cells from different macular subregions have distinct transcriptional profiles. *Human Molecular Genetics*, 30(16):1543–1558, 2021. https://doi.org/10.1093/hmg/ddab140.

[3] O. N. Manzari, H. Ahmadabadi, H. Kashiani, S. B. Shokouhi, and A. Ayatollahi. MedViT: A robust vision transformer for generalized medical image classification. *Computers in Biology and Medicine*, 157, 106791, 2023. https://doi.org/10.1016/j.compbiomed.2023.106791.

[4] O. S. Naren. Retinal OCT Image Classification - C8 [Data set]. *Kaggle*, 2021. [Data set]https://doi.org/10.34740/KAGGLE/DSV/2736749.

[5] A. Kolesnikov, L. Beyer, X. Zhai, J. Puigcerver, J. Yung, S. Gelly, and N. Houlsby. Big Transfer (BiT): General visual representation learning. *arXiv*, 2020. https://arxiv.org/abs/1912.11370.

[6] J. Howard and S. Ruder. Universal language model fine-tuning for text classification. *arXiv*, 2018. https://arxiv.org/abs/1801.06146.

[7] Z. Yu, D. Shen, Z. Jin, J. Huang, D. Cai, and X.-S. Hua. Progressive transfer learning. *arXiv*, 2020. https://arxiv.org/abs/1908.02492.

[8] J. Gildenblat and contributors. PyTorch library for CAM methods. *GitHub*, 2021. https://github.com/jacobgil/pytorch-grad-cam.

[9] A. Buades, B. Coll, and J.-M. Morel. Non-local means denoising. *Image Processing On Line*, 1:208–212, 2011. https://doi.org/10.5201/ipol.2011.bcm_nlm.

[10] J. Yang, R. Shi, D. Wei, Z. Liu, L. Zhao, B. Ke, H. Pfister, and B. Ni. MedMNIST v2—A large-scale lightweight benchmark for 2D and 3D biomedical image classification. *Scientific Data*, 10(1):41, 2023. Nature Publishing Group UK London.

Detekcia štandardnej kvantizačnej matice JPEG obrázka

Bc. Martin Mydliar[1], Mgr. Ing. Matúš Jókay, PhD.[2]

[1]Ústav informatiky a matematiky, FEI STU v Bratislave

xmydliarm1@stuba.sk

Abstrakt – V našej práci skúmame možnosti detekcie štandardnej kvantizačnej matice v JPEG obrázkoch a jej vplyv na spoľahlivé určenie kvality kompresie. Niektoré pomerne rozšírenejšie knižnice pri kompresii JPEG obrázkov nepoužívajú štandardné kvantizačné matice, ale nahrádzajú ich vlastnými optimalizovanými tabuľkami. Cieľom takých úprav je znížiť viditeľné artefakty alebo zlepšiť výsledný kompresný pomer[1], čo znemožňuje spoľahlivé určenie kvality z kvantizačnej matice obrázka, umiestnenej v hlavičke súboru. Tento faktor kriticky ovplyvňuje praktickú použiteľnosť niektorých aplikácií, napríklad steganografického algoritmu Perturbed Quantization[2]. V rámci tohto algoritmu sa vyžaduje čo najpresnejšie určenie kvality obrázka pre zaručenie bezpečného vloženia údajov. Prínos našej práce však vidíme aj pri iných aplikáciách, ktoré vo všeobecnosti vyžadujú analýzu kompresných parametrov JPEG obrázka.

1 Úvod

Nasledujúca kapitola má za cieľ uviesť čitateľa do všeobecnej problematiky súvisiacej s JPEG štandardom a jeho digitálnym spracovaním, nakoľko článok stavia na týchto princípoch.

Našou motiváciou tejto práce bola úprava a implementácia steganografického algoritmu Perturbed Quantization. Vo všeobecnosti je však možné túto metódu využiť aj v rôznych iných aplikáciách, ktoré vyžadujú vizuálnu analýzu JPEG obrázkov, pričom hlavným príkladom použitia môže byť práve určovanie kvality JPEG obrázkov odvodených od štandardnej kvantizačnej matice. V prípade, že takéto určenie kvality nie je možné (napríklad preto, že pri kompresii bola použitá neštandardná kvantizačná matica), sme chceli nájsť spôsob, ako takúto situáciu detegovať.

Kapitola experimentov sa zameriava na trojicu knižníc, ktoré sa špecializujú na kompresiu obrázkov do JPEG formátu, a to ImageMagick[3], Mozjpeg [4] a Guetzli[5]. Znázorňuje štatisticky vyhodnotené dáta získané spracovaním výstupov z daných knižníc v podobe grafov na zvolenom datasete. Ďalej sa zameriava na zlepšenie dosiahnutých výsledkov prostredníctvom korekčnej metódy a na záver hodnotí dosiahnuté výsledky.

2 JPEG

JPEG je široko používaný formát na zmenšenie veľkosti digitálnych obrázkov prostredníctvom stratovej kompresie. Jeho hlavným cieľom je minimalizovať veľkosť súborov pri zachovaní prijateľnej kvality obrazu pre ľudské oko. JPEG dosahuje tento cieľ využitím dvoch hlavných charakteristík ľudského videnia[6]:

- Vyššia citlivosť na jas než na farbu.
- Nižšia citlivosť na detaily vo vysokých frekvenciách, ako sú jemné textúry.

2.1 Kompresia

Algoritmus JPEG kompresie pozostáva z viacerých krokov, ktoré prispievajú k zníženiu množstva dát, zatiaľ čo sa minimalizuje viditeľný vplyv na kvalitu obrazu. Pred podrobnejším popisom jednotlivých krokov je dôležité spomenúť isté kľúčové vlastnosti[7]:

- *Stratová povaha*: JPEG kompresia obetuje niektoré detaily obrazu, najmä vo vysokofrekvenčných oblastiach, aby dosiahla výrazné zníženie veľkosti súboru.
- *Nastaviteľná úroveň kompresie*: Používatelia môžu kontrolovať kompromis medzi kvalitou obrazu a veľkosťou súboru úpravou pomeru kompresie. Nižšie nastavenia kvality zvyšujú kompresiu, ale môžu viesť k viditeľným artefaktom.
- *Bloková kompresia*: Obraz je rozdelený na nezávislé bloky 8x8 obrazových bodov, čo môže viesť k blokovým artefaktom, ak sa aplikuje príliš vysoká kompresia.
- *Optimalizované pre fotografie*: JPEG najlepšie funguje na fotografických obrázkoch s plynulými prechodmi, ale môže mať problémy s textom alebo ostrými okrajmi, kde sú artefakty viditeľnejšie.

2.1.1 Konverzia farebného priestoru

Prvým krokom JPEG kompresie je konverzia obrazu z farebného priestoru RGB (*Red*, *Green*, *Blue*) do YCbCr. Tento krok sám osebe neznižuje veľkosť súboru, len pripravuje dáta na ďalší krok, kde sa aplikujú samotné kompresné techniky. Je dôležitý, pretože ľudské oko je citlivejšie na zmeny v jase (*luminancia* – zložka Y) než na zmeny vo farbách (*chrominancia* – Cb pre modrú a Cr pre červenú zložku). Oddelenie informácie o jase od farebnej informácie umožňuje agresívnejšiu kompresiu farebných kanálov, pričom kvalita jasu zostáva zachovaná[8].

Vo farebnom priestore YCbCr:

- Y (*luminancia*) ukladá informácie o jase.
- Cb (*chrominancia-modrá*) a Cr (*chrominancia-červená*) ukladajú farebné informácie.

2.1.2 *Downsampling (zmenšenie vzorky)*

Po konverzii obrazu do farebného priestoru YCbCr JPEG kompresia znižuje rozlíšenie farebných kanálov Cb a Cr. Typicky pre tieto kanály dochádza ku zmenšeniu vzorky o faktor 2 alebo 4. Práve tento krok využíva spomínaný fakt, že ľudské oko je menej citlivé na zmeny vo farbe ako na zmeny v jase, takže zníženie rozlíšenia farebných informácií nemá významný vplyv na vnímanú kvalitu obrazu. Napríklad zmenšenie vzorky v pomere 4:2:0 znižuje rozlíšenie chrominancie na štvrtinu pôvodnej veľkosti, čím sa výrazne znižuje množstvo údajov obrazu.

2.1.3 *Delenie obrazu na bloky 8x8 obrazových bodov*

Po zmenšení vzorky sa údaje z jednotlivých kanálov (Y, Cb, Cr) rozdelia na bloky 8x8 obrazových bodov. Každý blok sa následne spracováva nezávisle. Delenie obrazu na menšie bloky umožňuje lokálnu kompresiu, ktorá dokáže využiť opakujúce sa vzory v rámci každého bloku.

2.1.4 *Diskrétna kosínusová transformácia (DCT)*

JPEG aplikuje na každý blok Diskrétnu kosínusovú transformáciu (DCT), ktorá transformuje údaje z priestorovej domény (hodnoty obrazových bodov) do frekvenčnej domény. DCT dekomponuje každý blok na maticu frekvenčných koeficientov. Horný ľavý roh matice predstavuje nízkofrekvenčné komponenty (zachytávajúce väčšinu energie obrazu), zatiaľ čo dolný pravý roh predstavuje vysokofrekvenčné komponenty (zachytávajúce jemné detaily a textúry).

Blok 8x8 obrazových bodov sa transformuje do súboru 64 frekvenčných koeficientov. Táto transformácia je dôležitá pre ďalší krok, kde sa selektívne odstraňujú menej dôležité údaje na základe významu frekvencií pre ľudské vnímanie obrazu.

2.1.5 *Kvantizácia*

Kvantizácia je ďalším dôležitým krokom, v ktorom JPEG dosahuje stratovú kompresiu tým, že znižuje presnosť vo frekvenčnej doméne. Koeficienty DCT sa delia hodnotami z kvantizačnej matice a následne sa zaokrúhľujú na najbližšie celé číslo. Hodnoty prvkov kvantizačnej matice určujú úroveň kompresie, pričom väčšie hodnoty vedú k agresívnejšiemu zmenšovaniu výslednej hodnoty, čo zodpovedá vyššej kompresii ale zároveň aj nižšej kvalite obrazu.

Typicky sa nízkofrekvenčné koeficienty uchovávajú presnejšie (nižšie kvantizačné hodnoty), zatiaľ čo vysokofrekvenčné koeficienty sa silnejšie kvantizujú (vyššie kvantizačné hodnoty). To zodpovedá tomu, že ľudské oko je menej citlivé na vysokofrekvenčné detaily, takže zníženie ich presnosti nemá výrazný vplyv na ľudské vnímanie kvality obrazu.

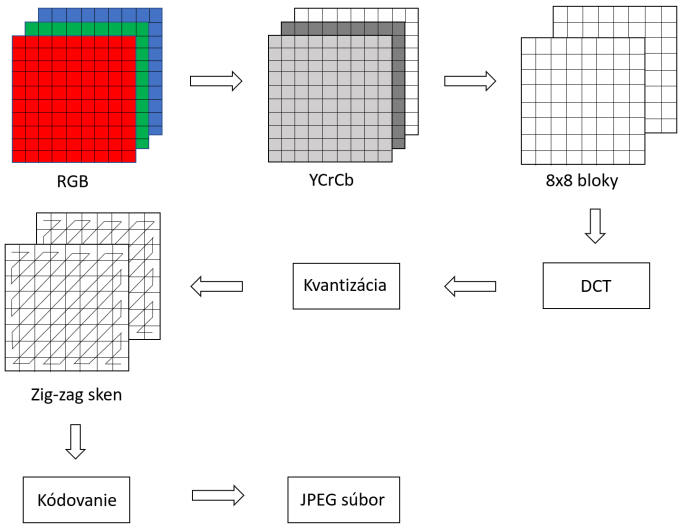
Typická kvantizačná tabuľka pre luminanciu má teda menšie hodnoty v ľavom hornom rohu (zachovanie detailov v jase) a väčšie hodnoty v pravom dolnom rohu (potlačenie jemných detailov). Po kvantizácii je veľa vysokofrekvenčných koeficientov zredukovaných na nulu, čím sa obrazová informácia ešte viac redukuje.

2.1.6 *Zig-zag skenovanie a entropické kódovanie*

Po kvantizácii sa 64 koeficientov z každého bloku usporiada v zig-zag poradí, aby sa nízkofrekvenčné koeficienty (ktoré pravdepodobne obsahujú nenulové hodnoty) dostali na začiatok, zatiaľ čo vysokofrekvenčné koeficienty (ktoré sú pravdepodobne nulové) sa dostali na koniec. Toto usporiadanie uľahčuje ďalší krok kompresie, pri ktorom sa dlhé sekvencie núl môžu efektívne kódovať kratšími binárnymi postupnosťami.

JPEG následne aplikuje Run Length Encoding (RLE) na zakódovanie sekvencií nulových koeficientov a Huffmanovo kódovanie, aby sa dosiahla efektívna bezstratová komprimácia výslednej množiny údajov. Huffmanovo kódovanie funguje tak, že priraďuje kratšie binárne kódy k často sa vyskytujúcim hodnotám, čím sa znižuje veľkosť výsledného súboru.

Obrázok č. 1 znázorňuje zjednodušenú schému JPEG kompresie.



Obr. 1: Zjednodušený proces JPEG kompresie

2.2 Dekompresia

Proces JPEG dekompresie je inverzným ku procesu kompresie:

- Huffmanovo dekódovanie obnoví kvantizované koeficienty DCT.
- Inverzná DCT transformuje údaje z frekvenčnej domény späť do priestorovej oblasti.
- Kanály chrominancie (Cb a Cr) sú procesom zväčšenia vzorky rekonštruované na pôvodné rozlíšenie.
- Nakoniec sa údaje YCbCr prevedú späť do farebného priestoru RGB, čím sa obraz obnoví do pôvodnej reprezentácie.

3 Experimenty

3.1 Dataset a metodológia

Pre experimenty bol použitý dataset[9], ktorý pozostáva zo 100 farebných obrázkov kvetov vo formáte PNG. Tie sme previedli na čiernobiele obrázky. Každý obrázok bol ďalej komprimovaný do formátu JPEG. Pre zvýšenie transparentnosti experimentov uvádzame verzie a účel použitých knižníc:

- ImageMagick (verzia 6.9.12) — reprezentant štandardnej kvantizačnej matice,
- Mozjpeg (verzia 5.0.0) — optimalizovaná pre nižšie artefakty,
- Guetzli (verzia 1.0.1) — zameraná na psychovizuálnu kvalitu.

Experimenty sme vykonali pre jednotlivé kvality v intervale <70; 99>, pričom kvalita 100 bola úmyselne vylúčená kvôli triviálnej kvantizačnej matici (všetky hodnoty sú 1). Pre každú kombináciu knižnice a kvality bolo vygenerovaných 100 JPEG obrázkov, čím celkový počet vzoriek dosiahol 9000 obrázkov (3 knižnice × 30 kvalít × 100 obrázkov).

Výpočet odhadu kvality prebieha pre každý koeficient kvantizačnej matice samostatne. Túto operáciu vyjadrujeme nasledovne:

$$X = 100 - \left(\frac{100 \cdot Q_i - 50}{2 * Q_{50}} \right)$$

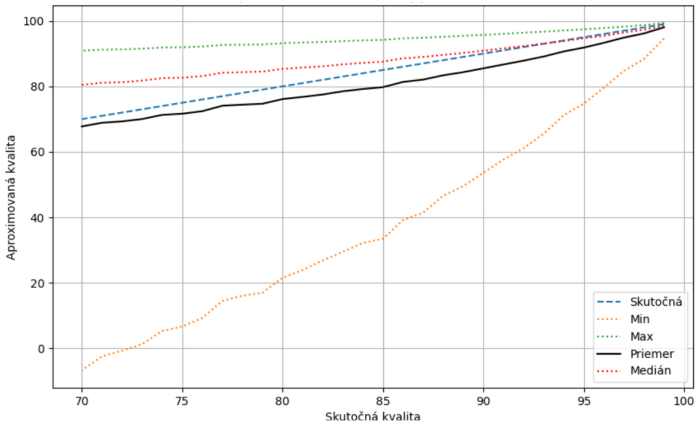
kde X predstavuje odhadovanú kvalitu, Q_i je hodnota kvantizačného koeficientu na i-tej pozícii kvantizačnej matice obrázka a Q_{50} je hodnota kvantizačného koeficientu v rámci štandardnej kvantizačnej matice (uvedená nižšie v texte) na i-tej pozícii. Vzhľadom na jednorozmerný index i uvažujeme usporiadanie prvkov matice do zoznamu postupne v poradí zľava doprava a zhora nadol.

$$Q_{50} = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix}$$

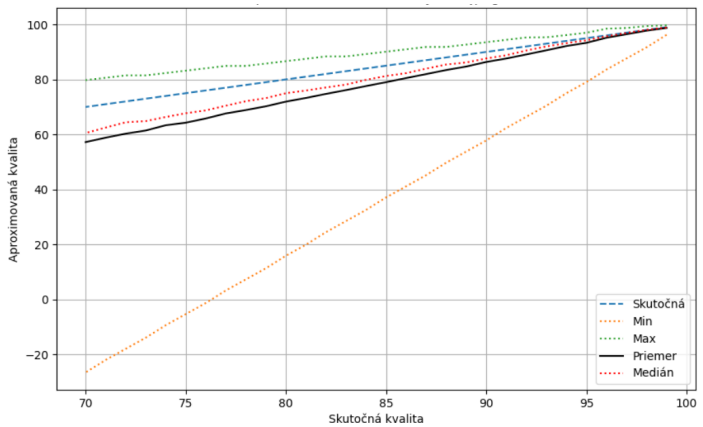
3.2 Štatistické vyhodnotenie

Pre každú knižnicu a kvalitu boli zo všetkých 100 obrázkov vypočítané minimum, maximum, priemer a medián odhadovanej kvality. Vzhľadom na párny počet koeficientov a rozhodnutie zostať na celočíselnej hladine pokladáme za medián prvú z dvojice hodnôt; teda nejedná sa o medián v zmysle štandardnej matematickej definície.

Tieto štatistiky sú vizualizované vo forme grafov pre jednotlivé knižnice na obrázkoch č. 2, 3 a 4. Os X reprezentuje skutočnú kvalitu, os Y odhadovanú kvalitu, pričom legenda k vyobrazeným krivkám sa nachádza v pravom dolnom rohu. Pre spoľahlivý odhad kvality a zároveň určenie, či bol obrázok komprimovaný štandardnou kvantizačnou maticou, je cieľom dosiahnuť čo najmenší rozptyl ľubovoľnej z kriviek od modrej krivky, ktorá reprezentuje skutočnú kvalitu obrázka.



Obr. 2: Aproximácia hodnôt kvality (Guetzli)



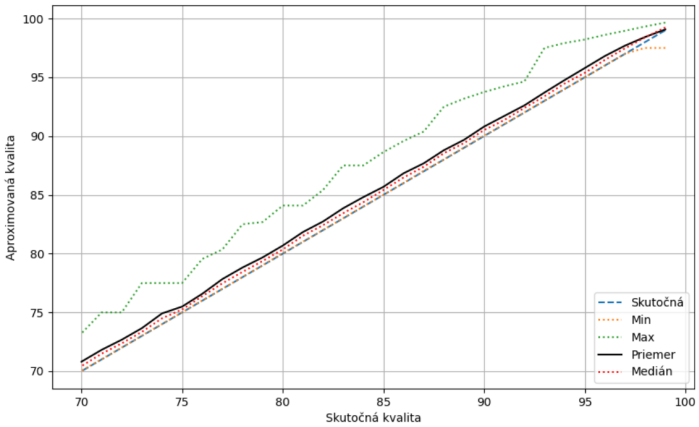
Obr. 3: Aproximácia hodnôt kvality (Mozjpeg)

Z nameraných hodnôt usudzujeme, že medzi knižnicou, ktorá využíva štandardnú kvantizačnú maticu a knižnicami, ktoré využívajú vlastné kvantizačné matice, je už na prvý pohľad zrejmý rozdiel. Najvýznamnejšiu deviáciu môžeme pozorovať na krivke reprezentujúcej minimálne hodnoty odhadovanej kvality. Táto odchýlka stúpa s klesajúcou kvalitou obrázka, zatiaľ čo v prípade štandardnej kvantizačnej matice, s výnimkou vysokej kvality, kopíruje medián.

3.3 Korekčné metódy

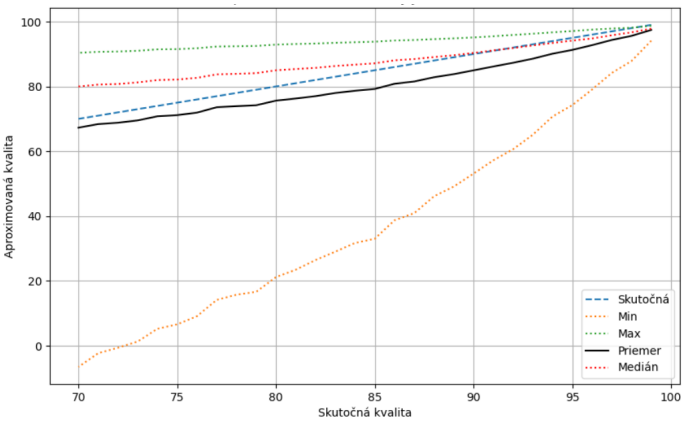
Takmer všetky odhadované hodnoty kvality X predstavujú desatinné hodnoty, preto sme sa za účelom redukcie odchýlok rozhodli otestovať dve nasledovné stratégie:

- Odstránenie desatinnej časti (*floor*),
- zaokrúhlenie hodnoty.



Obr. 4: Aproximácia hodnôt kvality (ImageMagick)

Ukázalo sa, že zaokrúhlenie hodnôt malo síce zlepšujúci charakter, ale rozdiel bol nepatrný. Naopak, odstránenie desatinnej časti prinieslo pri obrázkoch, kde bola použitá štandardná kvantizačná matica pri kompresii, želané výsledky. Medián odhadovanej kvality kopíroval skutočnú kvalitu v celom meranom intervale. Výsledky tejto korekčnej metódy sa nachádzajú na obrázkoch č. 5, 6 a 7.

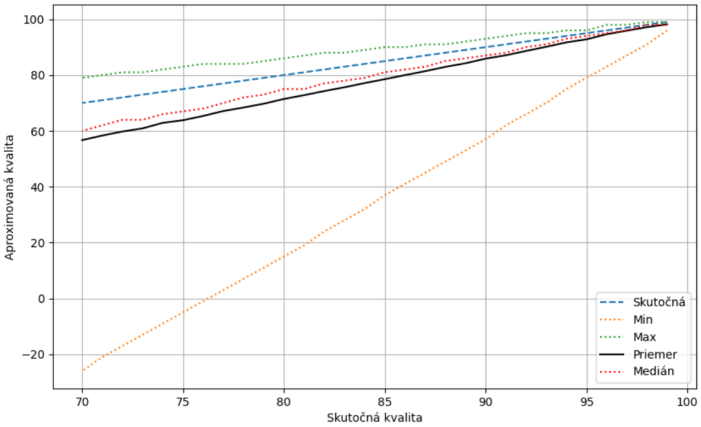


Obr. 5: Aproximácia hodnôt kvality po korekcii (Guetzli)

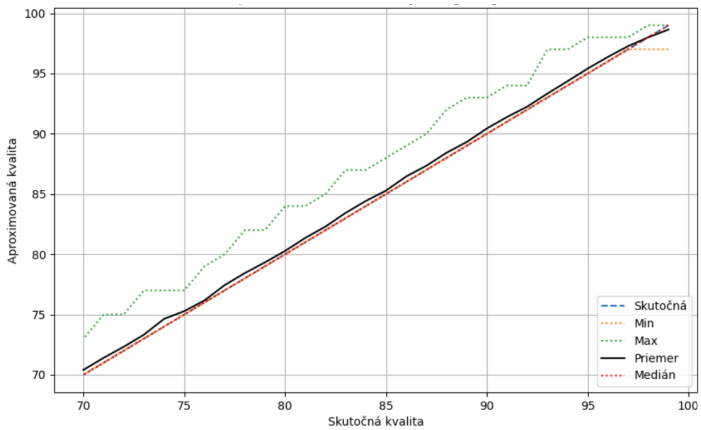
4 Záver

Prostredníctvom experimentov sme porovnali odhady kvality pre JPEG obrázky komprimované knižnicami ImageMagick, Mozjpeg a Guetzli. Výsledky ukázali, že pri neštandardných kvantizačných maticiach vznikajú deviácie medzi zobrazenými štatistickými krivkami, pričom ich veľkosť narastá s klesajúcou kvalitou. Na základe toho usudzujeme, že experimenty potvrdili, že detekcia štandardnej, a teda zároveň aj neštandardnej kvantizačnej matice, je experimentálne možná prostredníctvom štatistickej analýzy odhadovaných hodnôt kvality.

Experimenty ďalej odhalili, že korekcia vo forme odstránenia desatinnej časti hodnôt kvality zvyšuje spoľahlivosť tejto štatistickej metódy, nakoľko umožnila presný odhad kvality v celom pozorovanom intervale, kde prvotná realizácia dosahovala



Obr. 6: Aproximácia hodnôt kvality po korekcii (Mozjpeg)



Obr. 7: Aproximácia hodnôt kvality po korekcii (ImageMagick)

miernu odchýlku od skutočnosti.

Grafická reprezentácia výsledkov názorne ilustruje rozdiely medzi knižnicami a môže slúžiť ako referenčný materiál pre ďalší výskum.

Literatúra

[1] Andrew B Watson. Dct quantization matrices visually optimized for individual images. In *Human Vision, Visual Processing, and Digital Display IV*, volume 1913, pages 202–216. SPIE, 1993.

[2] Jessica Fridrich, Miroslav Goljan, and David Soukal. Perturbed quantization steganography with wet paper codes. *Proceedings of the ACM Workshop on Multimedia and Security*, pages 4–15, 2005.

[3] ImageMagick Studio LLC. *ImageMagick*, 2024.

[4] Mozilla. Mozjpeg: Improved jpeg encoder, 2024.

[5] Robert Obryk and Jyrki Alakuijala. Guetzli: Perceptually guided jpeg encoder. *Google Research*, 2017.

[6] Gregory K. Wallace. The jpeg still picture compression standard. *Communications of the ACM*, 34(4):30–44, 1991.

[7] David Taubman and Michael Marcellin. *JPEG2000: Image compression fundamentals, standards and practice*. Springer Science & Business Media, 2002.

[8] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing*. Prentice Hall, 2008.

[9] Olga Belitskaya. Flower color images, 2018.

Mikroslužbový systém na automatizované spracovanie, testovanie a vyhodnotenie používateľského kódu

Filip Švec, Martin Pajpach

Ústav informatiky a matematiky, FEI STU v Bratislave
Ústav automobilovej mechatroniky, FEI STU v Bratislave

xsvecf@stuba.sk

Abstrakt – Práca prezentuje backend systém gamifikovanej programátorskej platformy navrhnutej na hodnotenie používateľom odoslaného kódu. Jadro systému tvorí mikroslužbová architektúra, ktorá spracováva a vyhodnocuje riešenia programovacích úloh, pričom používateľom poskytuje štruktúrovanú spätnú väzbu. Integráciou hodnotenia kódu do interaktívneho vzdelávacieho prostredia systém podporuje praktickejší a zaujímavejší prístup k výučbe programovania.

1 Úvod

Programovanie je náročnou disciplínou, najmä kvôli množstvu rôznych abstraktných pojmov či zložitej syntaxi. Ak študenti nemajú prístup k spätnej väzbe, môžu mať problém s porozumením danej problematiky. Náročnosť programovania sa často prejavuje aj vo vysokej miere predčasného ukončenia štúdia v odboroch informačných technológií [4][5]. Hodnotenie kódu v reálnom čase tento problém rieši tým, že poskytuje okamžité informácie o odovzdaných riešeniach, čo robí vzdelávanie oveľa praktickejším a interaktívnejším.

Táto práca predstavuje *backend* systém, ktorý spracováva a hodnotí používateľský kód v reálnom čase. Systém je založený na modulárnej mikroslužbovej architektúre, vďaka ktorej môžu používatelia odosielať svoj kód, zobraziť výsledky a postupne svoje riešenia zlepšovať.

Avšak spracovanie používateľského kódu prináša viacero technických výziev. Jedným z najdôležitejších aspektov je bezpečnosť. Spracovanie používateľom zadaného kódu môže ohroziť systém, ak táto operácia nie je dostatočne izolovaná. Dôležitými úlohami sú aj škálovateľnosť, najmä v situáciách, keď systém spracováva veľké množstvo odoslaných riešení naraz alebo zabezpečenie spoľahlivého hodnotenia naprieč rôznymi programovacími jazykmi. Všetky tieto požiadavky je potrebné správne implementovať tak, aby vznikol systém, ktorý bude bezpečný, spoľahlivý a jednoducho rozšíriteľný.

2 Analýza súčasného stavu poznania

Analýza súčasného stavu je venovaná aktuálnym riešeniam v oblasti vzdelávania programovania, no aj všeobecnému prístupu k výučbe a implementácii rôznych metód, ktorých cieľom je motivovať študentov vo vzdelávaní prostredníctvom programovacích úloh a zadaní.

2.1 Gamifikácia

Jednou z najpopulárnejších metód, ktoré motivujú študentov vo vzdelávaní je práve gamifikácia, ktorá implementuje herné prvky, ako hlavnú súčasť vzdelávacích aplikácií [3]. Gamifikáciu je možné dosiahnuť pomocou rôznych nástrojov. Sú nimi napríklad odznaky alebo rebríčky.

1. **Odznaky** sú pridelené používateľom na základe dosiahnutia určitých cieľov. Slúžia, ako vizuálne ocenenie, ktoré posilňuje vnútornú motiváciu a dáva používateľom pocit pokroku. V oblasti programovania sú prideľované napríklad za vyriešenie určitého počtu úloh, za pravidelnú aktivitu alebo umiestnenie v súťažiach. Odznaky majú pozitívny vplyv na motiváciu používania aplikácií a často lákajú používateľa k získaniu čo najväčšieho množstva rôznych úspechov. Okrem toho môžu slúžiť aj ako indikátor úrovne vzdelania v danej oblasti [1].
2. **Rebríčky** predstavujú ďalší z efektívnych gamifikačných prvkov, ktoré dokážu podnietiť používateľov k vzdelávaniu. Porovnávanie výsledkov s ostatnými používateľmi vytvára súťažné prostredie, ktoré môže motivovať k lepšiemu výkonu a častejšiemu riešeniu úloh. V kontexte programovania môžu rebríčky zobrazovať napríklad počet vyriešených úloh, rýchlosť odovzdania riešení či dosiahnuté skóre. Pri rebríčkoch je však dôležitá aj vhodná implementácia, aby sa predišlo práve opačnému efektu a používatelia neboli domotivovaní napríklad ich umiestnením v porovnaní s ostatnými [1][2].

2.2 Programovacie platformy

Existujú rôzne platformy, ktoré ponúkajú štruktúrovaný spôsob vzdelávania založený na vypracovávaní čiastkových úloh zameraných na konkrétny programátorsky problém. Nachádzajú využitie či už v osobnom vzdelávaní, v školských inštitúciách alebo pracovných pohovoroch. Často využívajú aj prvky gamifikácie, ktorými motivujú používateľov k riešeniu úloh, pričom rýchla spätná väzba im pomáha napredovať.

Programovacie platformy majú k dispozícii široké spektrum rôznych úloh či už z oblasti programovania alebo informačných technológií. Z toho dôvodu musia byť navrhnuté

flexibilne a modulárne. Medzi známe programovacie platformy patrí *HackerRank*, *CodeWars* alebo *LeetCode*. Avšak spomínané platformy sú komerčné a spôsob, akým implementujú algoritmy na spracovanie a vyhodnotenie používateľom riešených úloh nie sú voľne dostupné.

2.2.1 Platforma LeetCode

Platforma *Leetcode* je poprednou programovacou platformou. Využíva sa na osobné vzdelávanie alebo pri technických pohovoroch. Momentálne má dostupných približne 3400 úloh z oblasti informačných technológií. Tieto príklady sa členia do rôznych kategórií podľa typu problému, ktorému sa venujú [8].

Platforma poskytuje používateľovi rozhranie, pomocou ktorého môže úlohy zobraziť a vyriešiť. Rozhranie obsahuje okno so zadaním a potrebnými informáciami k jeho vypracovaniu, editor kódu so šablónou k danému príkladu a okno s testovacími dátami a výsledkami spracovania. Cieľom používateľa je doplniť logiku do preddefinovanej metódy, ktorá sa nachádza v šablóne tak, aby dosiahol správny výsledok na základe vstupných argumentov metódy [8].

Spomínaný proces, ktorý má za úlohu spracovať používateľov kód nie je prístupný verejnosti no na základe analýzy *frontend* elementov a spôsobu zadávania príkladov je možné odvodiť niekoľko skutočností. Nasledujúca štatistika a zistenia boli získané z listu *Top 100 Liked Questions*, ktorá obsahuje 100 najlepšie hodnotených príkladov používateľmi [8]. Z nej bolo vybraných prvých 50 úloh, na základe ktorých boli získané dostupné jazyky, formáty šablón na vypracovanie príkladov, typ argumentov vstupných dát a typ výstupov.

1. Zoznam podporovaných **jazykov** bol vo všetkých analyzovaných príkladoch rovnaký. Používateľ si môže vybrať z nasledujúcich možností: *C++, Java, Python, Python3, C, C#, JavaScript, TypeScript, PHP, Swift, Kotlin, Dart, Go, Ruby, Scala, Rust, Racket, Erlang, Elixir* [8].
2. Používateľ má k dispozícii **šablóny** na vypracovanie úloh, ktoré sa líšia podľa vybraného programovacieho jazyka. Úlohou používateľa je doplniť potrebnú logiku do predpripravenej štruktúry, ktorá už zohľadňuje syntaktické požiadavky konkrétneho jazyka. Porovnanie sa ďalej zameriava iba na tieto programovacie jazyky: *Java, C++, C#, Python3, JavaScript a C*. V jazykoch *Java, C++, C#* a *Python3* bola šablóna spravidla založená na triede *Solution* a definícii metódy so špecifickým názvom. V jazykoch *C* a *JavaScript* šablóna neobsahuje triedu *Solution*, iba definíciu metódy. V niektorých prípadoch šablóna obsahovala aj preddefinovanú dátovú štruktúru, ktorú bolo potrebné použiť pri riešení zadania, najčastejšie v úlohách zameraných na prácu s binárnymi stromami [8].
3. **Typy vstupných argumentov** sa medzi úlohami líšili a jedna úloha mohla obsahovať aj viacero parametrov. V analyzovanej kolekcii príkladov však všetky vstupy spadali do týchto kategórií s nasledujúcim počtom [8]:
    a. Primitívne dátové typy a *string* – 21
    b. Dátové štruktúry *array* a *list* – 26
    c. Preddefinovaná dátová štruktúra – 10
4. Aj **výstupné hodnoty** jednotlivých funkcií sa medzi úlohami líšili, no rovnako ich možno zaradiť do týchto kategórií [8]:
    a. Primitívne dátové typy a *string* – 26
    b. Dátové štruktúry *array* a *list* – 14
    c. Návratová hodnota typu *void* – 4
    d. Preddefinovaná dátová štruktúra – 6

Z týchto zistení vyplýva viacero záverov o tom, ako *LeetCode* spracúva a vyhodnocuje používateľov kód [8].

1. Používateľ si môže vybrať spomedzi 19 dostupných jazykov a riešiť ľubovoľné zadania.
2. Každé zadanie musí byť riešiteľné vo všetkých jazykoch, čo obmedzuje použitie len na spoločné funkcie naprieč jazykmi.
3. Vstupné aj výstupné dáta musia byť univerzálne. Úlohy používajú typy, ako *integer, float, string, array, list* či preddefinované štruktúry, ktoré sú podporované všetkými jazykmi alebo ich ekvivalentmi, napr. *char** namiesto *string* v *C*.
4. Výstupy spadajú do rovnakých kategórií ako vstupy, okrem návratovej hodnoty *void*, pri ktorej sa očakáva úprava premennej namiesto výstupu.
5. Z predošlých zistení vyplýva, že platforma používa jednotné testovacie dáta pre všetky jazyky, čo zjednodušuje pridávanie nových úloh aj jazykov a umožňuje jednoduché porovnávanie riešení.
6. Štruktúra šablón naznačuje použitie metód alebo tried, ktoré sú potrebné na správne spracovanie a otestovanie kódu, no detaily vyhodnocovania nie sú známe [8].

2.2.2 Iné programovacie platformy

K popredným alternatívam k *LeetCode* patria ďalšie platformy, ktoré ponúkajú odlišný prístup k vzdelávaniu a hodnoteniu kódu.

HackerRank ponúka širokú škálu úloh od algoritmov cez databázy až po umelú inteligenciu a funkcionálne programovanie, pričom sa zameriava na štruktúrované kurzy pre rôzne úrovne vedomostí [13].

CodeWars kladie dôraz na komunitný prístup k učeniu. Na rozdiel od *LeetCode*, kde úlohy pripravuje prevažne interný tím, na *CodeWars* môžu nové zadania vytvárať samotní používatelia, čo zabezpečuje väčšiu rozmanitosť druhov príkladov. Oproti *HackerRank*, *CodeWars* síce neponúka štruktúrované kurzy, no všetky úlohy sú dostupné bezplatne, bez potreby predplatného, čo ho odlišuje od *LeetCode* aj *HackerRanku* [14].

Sphere Engine API nie je klasická programátorská platforma, ale vývojový nástroj, ktorý slúži na spracovanie a vyhodnocovanie kódu. Ide o systém bez používateľského rozhrania, ktorý poskytuje infraštruktúru pre vykonávanie a hodnotenie riešení v rôznych programovacích jazykoch [15].

3 Ciele a požiadavky práce

Hlavným cieľom práce je vytvoriť flexibilný a modulárny systém na automatizované spracovanie, testovanie a vyhodnotenie používateľského kódu. Na základe analýzy súčasného stavu poznania a analýzy poprednej vzdelávacej

platformy v oblasti programovania *LeetCode*, boli zostavené nasledovné ciele, ktoré by tento systém mal spĺňať:

1. Systém musí byť modulárny a škálovateľný v kontexte pridávania nových programovacích jazykov.
2. Programovacie úlohy musia byť jazykovo nezávislé v rámci použitých dátových typov. Testovacie dáta programovacích úloh musia byť navrhnuté tak, aby sa dali použiť naprieč rôznymi programovacími jazykmi.
3. Systém musí zabezpečiť ukladanie údajov o úlohách, vrátane ich popisov, testovacích dát a šablón konkrétnych jazykov.
4. Systém musí implementovať logiku spracovania používateľovho kódu, ktorá otestuje kód z hľadiska syntaktickej správnosti v kontexte konkrétneho programovacieho jazyka.
5. Systém musí implementovať logiku testovania používateľovho riešenia úlohy v prípade, ak kód spĺňa syntaktické požiadavky daného jazyka a formát preddefinovaných šablón.
6. Systém musí implementovať logiku vyhodnotenia používateľovho riešenia, tak aby bola spätná väzba prehľadná a v prípade nesprávneho vypracovania zobrazila používateľovi konkrétnu chybovú hlášku.
7. Systém musí byť schopný prijímať požiadavky a poskytovať odpovede prostredníctvom štandardizovaného komunikačného rozhrania bez závislosti na konkrétnom používateľskom rozhraní.
8. Požiadavky používateľov na spracovanie, testovanie a vyhodnotenie musia prebiehať paralelne, nesmú blokovať hlavné vlákno.
9. Systém musí byť časovo efektívny a poskytnúť používateľovi včasnú spätnú väzbu.
10. Systém musí implementovať bezpečné a izolované spracovanie používateľského kódu tak, aby následky tohto spracovania nemali dopad na zvyšok systému.
11. Systém musí byť navrhnutý tak, aby bol bezpečný, spoľahlivý a odolný voči útokom.
12. Systém musí byť navrhnutý univerzálne tak, aby umožnil jednoduchú integráciu s ľubovoľným *frontend* rozhraním.

4 Použité technológie a návrh systému

Cieľom tejto kapitoly je všeobecne popísať architektúru navrhovaného riešenia a použité technológie. Na základe definovaných cieľov práce a analýzy existujúcich platforiem bol dôraz kladený na modularitu, škálovateľnosť a prehľadnosť spätnej väzby pre používateľov.

4.1 .NET a C#

Práca bola implementovaná pomocou *.NET* platformy a jazyka *C#*. *.NET* je *open-source* vývojová platforma spravovaná spoločnosťou *Microsoft*, určená na tvorbu rôznych typov aplikácií [10]. *C#* je objektovo orientovaný jazyk vyvinutý taktiež spoločnosťou Microsoft a zväčša je hlavnou voľbou pri vývoji *.NET* aplikácií [11].

Dôležitým faktorom pri výbere platformy *.NET* bola aj jej kompatibilita s rôznymi *frontend* technológiami. Jedným z hlavných cieľov bolo vybudovať systém, ktorý bude možné integrovať s rôznymi formami používateľských rozhraní, či už ide o webovú aplikáciu, alebo napríklad hru. *.NET* ponúka riešenia pre webové aplikácie prostredníctvom technológie *Blazor*, ktorá je jeho súčasťou, no zároveň poskytuje vynikajúcu integráciu s platformou na vývoj hier *Unity*, ktorá tak isto využíva *C#* [10].

4.2 Mikroslužby

Systém využíva mikroslužbovú architektúru, inak nazývanou aj mikroservisová, ktorá na rozdiel od monolitickej architektúry umožňuje rozdelenie systému na menšie, nezávislé služby [7]. Každá mikroslužba zodpovedá za konkrétnu časť systému, čo umožňuje flexibilnejší vývoj a údržbu. Tento prístup zabezpečuje lepšie oddelenie zodpovedností a uľahčuje implementáciu nových funkcií, akou je napríklad podpora ďalších programovacích jazykov.

Mikroslužbová architektúra taktiež zabezpečuje škálovateľnosť v rámci výpočtového výkonu a odolnosť, čo poskytuje lepšiu dostupnosť a stabilitu systému [7]. Keďže jednotlivé mikroslužby sú nezávislé, výpadok alebo údržba jednej služby neovplyvní ostatné komponenty systému. Tento spôsob izolovania zmenšuje riziko širších výpadkov.

4.3 Minimal APIs

Jadrom projektu je rozhranie *API*. *.NET* v kombinácii s konceptom *Minimal APIs* poskytuje prehľadný spôsob budovania komunikačného rozhrania s menším množstvom kódu, ako pri klasických *API endpoints*. Tento prístup je ideálny pre projekty, kde stačí jednoduchá komunikácia medzi systémami bez nadbytočnej zložitosti [9].

4.4 Ocelot API brána

API brána slúži, ako vstupný bod do systému, spravuje prístup a smeruje požiadavky ku konkrétnym službám. Tento nástroj výrazne zjednodušuje správu a komunikáciu medzi klientskym rozhraním a backend mikroslužbami, pretože poskytuje centralizovanú vrstvu na spracovanie všetkých požiadaviek. Ocelot ponúka jednoduchú konfiguráciu a podporuje rôzne bezpečnostné funkcie, ako je autentifikácia alebo *rate limiting* [16].

4.5 MongoDB

V rámci systému je potrebné aj zahrnutie databázy, ktorá uchováva dáta jednotlivých programovacích úloh, testovacích dát a jazykových šablón. Konkrétne bola implementovaná *NoSQL* databáza *MongoDB*.

Hlavný rozdiel medzi *NoSQL* a *SQL* databázami spočíva v spôsobe ukladania a správy dát. *SQL* databázy, ako *MySQL* alebo *PostgreSQL*, udržiavajú dáta v tabuľkách s pevnými vzťahmi medzi nimi. Sú vhodné pre aplikácie, kde je potrebná prísna integrita a transakcie medzi viacerými tabuľkami. *NoSQL* databázy, ako *MongoDB*, sa dajú považovať za flexibilnejšie najmä z hľadiska vzťahov medzi dátami a samotnou štruktúrou dát [6]. *MongoDB* ukladá dáta v dynamických dokumentoch vo formáte *JSON* (*BSON*), čo umožňuje uchovávanie dát so zložitejšou a nepravidelnou štruktúrou [12]. *MongoDB* je vhodná na ukladanie

programovacích úloh aj testovacích dát, práve kvôli ich zložitej štruktúre, ktorá môže obsahovať množstvo vnorených objektov.

4.6 Docker

Docker je platforma zabezpečujúca kontajnerizáciu, ktorá zaobalí aplikáciu spolu so všetkými jej závislosťami do jedného alebo viacerých izolovaných prostredí. Tieto prostredia sa nazývajú kontajnermi a ich hlavnou funkcionalitou je zaistiť rovnaké správanie aplikácie bez ohľadu na to, kde je spustená, čo výrazne uľahčuje nasadzovanie a správu systému [17].

Kontajnerizácia v mikroslužbovej architektúre zjednodušuje infraštruktúru. Jednotlivé časti aplikácie bežia oddelene, čo zlepšuje ich spoľahlivosť a odolnosť. Kontajnery zároveň poskytujú formu izolovaného prostredia pre spracovanie používateľského kódu, čím pomáhajú minimalizovať bezpečnostné riziká a chrániť systém pred možným narušením.

4.7 Seq

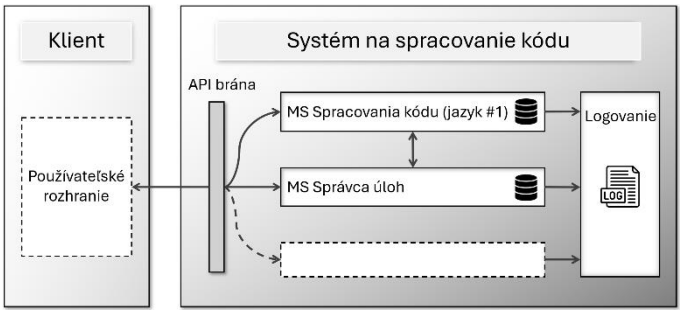
Seq je nástroj na spracovanie štruktúrovaných logov, ktorý umožňuje zhromažďovať a zobrazovať výstupy z aplikácií v reálnom čase. Na rozdiel od klasického logovania do súborov alebo konzoly poskytuje prehľadné webové rozhranie, kde je možné záznamy filtrovať alebo vyhľadávať podľa rôznych kritérií. V kombinácii s knižnicou *Serilog* je možné logovať jednotlivé udalosti priamo do prostredia *Seq* [18].

4.8 Postman

Postman je nástroj určený na testovanie rozhraní *API*, ktorý umožňuje posielať požiadavky na server a sledovať odpovede. V rámci tohto projektu slúži na overovanie správnej funkčnosti jednotlivých rozhraní a komunikácie medzi klientom a serverom.

4.9 Návrh architektúry systému

Hlavnou úlohou systému je spracovať používateľov kód, otestovať ho pomocou príslušných testovacích dát programovacej úlohy a poskytnúť mu spätnú väzbu. Návrh architektúry tohto systému a jednotlivých mikroslužieb (MS) je popísaný na Obr. 1, kde šípky znázorňujú komunikáciu.



Obr. 1 Všeobecný návrh architektúry systému

Spracovanie, testovanie a vyhodnotenie používateľovho kódu zabezpečuje mikroslužba, ktorá realizuje túto logiku pre konkrétny programovací jazyk. Systém aktuálne implementuje mikroslužbu pre jazyk *Python*. Zadania úloh aj testovacie dáta sú uložené v databáze *MongoDB*. Mikroslužba Správca programovacích úloh zodpovedá za sprostredkovanie zadaní používateľovi a komunikuje s mikroslužbami programovacích jazykov, ktoré spravujú vlastné kolekcie šablón potrebných na vypracovanie úlohy v danom jazyku.

Jednotlivé požiadavky na spracovanie kódu, či získanie programovacích úloh sú smerované aj podľa jazyka, v ktorom bola úloha vypracovaná za pomoci *Ocelot API* brány. *API* brána sa taktiež stará o manažovanie prístupu a počtu požiadaviek, ktoré môže používateľ vykonať za určité časové obdobie. Implementovaný je aj systém *Seq* na logovanie rôznych udalostí spojených s behom systému. Jednotlivé mikroslužby, databáza, *API* brána aj *Seq* sú kontajnerizované pomocou *Docker* kontajnerov.

5 Implementácia

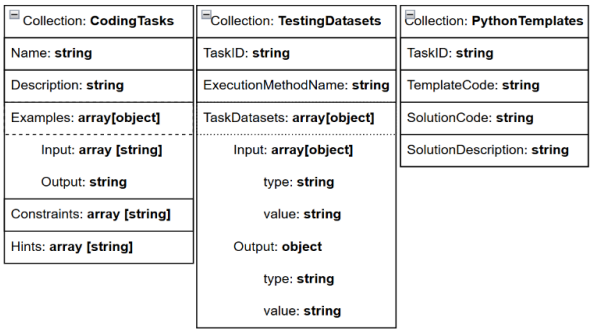
Nasledujúca kapitola približuje implementáciu systému a venuje sa aj zabezpečeniu v kontexte spracovania používateľovho kódu.

5.1 Štruktúra databázy a programovacích úloh

Prvým krokom pri implementácii samotného algoritmu na spracovanie a vyhodnotenie používateľovho kódu bolo definovať štruktúru programovacích úloh, ich testovacích dát a jazykových šablón. Vďaka analýze súčasného stavu a analýze rôznych programovacích platforiem boli odvodené formáty šablón, ktorých úlohou je poskytnúť používateľovi štruktúru, do ktorej bude dopĺňať danú funkcionalitu. V jazyku *Python* sa tieto šablóny skladajú z triedy *Solution* a práve jednej metódy, ktorá je obsiahnutá v tejto triede. Testovacími dátami sú argumenty, respektíve konkrétne hodnoty, ktoré sú počas testovania vložené, ako vstupy do metódy. Tieto dáta sú štruktúrované, ako páry vstupov a výstupov. Vstupné dáta majú vždy práve jednu výstupnú hodnotu.

Databáza je rozdelená na niekoľko kolekcií a ich štruktúra je popísaná na Obr. 2:

1. Kolekcia programovacích úloh (*CodingTasks*)
2. Kolekcia testovacích dát (*TestingDatasets*)
3. Kolekcia šablón (*PythonTemplates*)



Obr. 2. Štruktúra dát v jednotlivých kolekciách

Ako bolo spomenuté aj v kapitole 4.5 *MongoDB*, dáta majú nepravidelnú a zložitú štruktúru, a aj preto bola zvolená práve *NoSQL* databáza, ktorá uchováva údaje v *JSON*

štruktúre. Avšak aj navzdory použitiu *NoSQL* databázy, ktorá striktne neimplementuje vzťahy medzi dátami, dané kolekcie majú nasledovný vzťah. Každá programovacia úloha má práve jedny testovacie dáta a jednu šablónu konkrétneho jazyka. Kolekcie testovacích dát a šablón obsahujú referenciu na programovaciu úlohu, ku ktorej patria vo forme jedinečného identifikátora *TaskID*.

5.2 Spracovanie kódu

Proces spracovania, testovania a vyhodnotenia kódu je špecifický pre každý implementovaný jazyk, a aj preto s každým novo pridaným jazykom pribúda aj príslušná mikroslužba. Požiadavky na spracovanie kódu sú sprostredkované pomocou *POST endpoint* a v tele požiadavky sa nachádza používateľov kód a unikátne *ID* úlohy, podľa ktorého sú z databázy získané testovacie dáta. Ako bolo uvedené v kapitole 2.2.1 Platforma *LeetCode*, z analýzy vyplynulo, že programovacie platformy používajú rovnaké testovacie dáta naprieč rôznymi jazykmi. Tento prístup bol aplikovaný aj pri návrhu tohto systému, kde bolo zabezpečené používanie iba jednoduchých dátových typov a typu *string* a *list*. V jazykoch, akým je napríklad jazyk C, ktorý neimplementuje dátovú premennú *string*, je možné nahradiť tento typ s *char**.

Algoritmus na spracovanie, testovanie a vyhodnotenie kódu sa skladá z niekoľkých pomocných metód. *DriverCodeGenerator* vkladá používateľov kód do inicializačného kódu, ktorý je vykonávaný pomocou *Python3* interpretera. Metóda *ProcessUsersCode* zodpovedá za spustenie vygenerovaného kódu a formátovanie výsledkov do odpovede.

5.2.1 Metóda DriverCodeGenerator

Táto metóda zodpovedá za generovanie pomocného inicializačného skriptu, ktorý slúži na otestovanie používateľského kódu. Zároveň formátuje testovacie dáta z databázy a vkladá ich priamo do tohto skriptu. Výsledkom je reťazec, ktorý sa následne zapíše do dočasného súboru a spustí pomocou metódy *Start* triedy *Process* v hlavnej metóde *ProcessUsersCode*.

Jadrom tohto skriptu je *Python* kód, ktorý obsahuje jednu hlavnú funkciu *main()*. V tejto časti sa do skriptu vloží používateľský kód a pripravené testovacie datasety. Pred spustením sa najskôr overí syntaktická správnosť používateľského kódu pomocou funkcie *compile()*. Ďalej sa skontroluje či kód obsahuje definíciu triedy *Solution* a požadovanej metódy. Ak táto kontrola prebehla úspešne, nasleduje hlavný cyklus tohto skriptu, v ktorom sa postupne vkladajú jednotlivé testovacie datasety do funkcie *getattr()*. Táto funkcia vytvára inštanciu triedy *Solution* a vykonáva konkrétnu metódu, ktorej argumenty sú z príslušného testovacieho datasetu. V každom cykle sa taktiež kontroluje čas behu tejto metódy, typ a aj hodnota výstupu v tomto poradí.

V prípade problémov, či už pri testovaní alebo pri kontrole syntaxi, sú vyvolané výnimky, ktoré zabezpečujú spätnú väzbu o chybe. Celý proces je obalený v *try-except* bloku, ktorý zachytáva potenciálne chyby a zabezpečuje, že každá možná situácia má jasnú výstupnú odozvu.

5.2.2 Metóda ProcessUsersCode

ProcessUsersCode je hlavnou metódou a implementuje špecifický spôsob, ktorým bude kód spracovaný. V prvom kroku vytvára dočasný *Python* súbor *.py* a volá metódu *DriverCodeGenerator*, do ktorej, ako vstupné argumenty, vkladá používateľský kód a testovacie dáta danej úlohy. Táto metóda potom zapíše do dočasného súboru potrebný inicializačný kód.

V ďalšom kroku na spustenie používateľského kódu uloženého v dočasnom súbore je využitá trieda *Process*, ktorá slúži na vytváranie a ovládanie externých procesov. Umožňuje spustiť program, nastaviť parametre jeho spustenia a čítať jeho výstup. V rámci konfigurácie je, ako spustiteľný proces, nastavený *Python3* interpreter, ktorý je nainštalovaný priamo v *Docker* kontajnery tejto mikroslužby. Vstupným argumentom procesu je cesta k dočasnému súboru so skriptom a štandardný výstup z procesu je presmerovaný tak, aby bolo možné zachytiť výsledky a prípadné chyby.

Táto metóda ďalej obsahuje implementáciu *timeout* mechanizmu, ktorý sleduje časový priebeh spusteného procesu a čaká na jeho dokončenie. Priebeh tohto algoritmu je paralelný a neblokuje hlavné vlákno aplikácie. Každé spustenie kódu beží, ako samostatný proces, ktorý sa po dokončení automaticky ukončí alebo sa v prípade prekročenia časového limitu ukončí nasilu.

5.2.3 Formát spätnej väzby

Systém odpovedí je založený na dátovej štruktúre *ResultResponseDto*, ktorá definuje jednotný formát spätnej väzby pre každé vyhodnotenie používateľského kódu.

Každá odpoveď obsahuje tri základné položky. *ResultState*, obsahuje názov chyby respektíve výnimky, ktorá nastala alebo slovo „*Success*“ v prípade, že bolo vypracovanie úlohy správne. *DebugMessage*, obsahuje podrobnejší popis výsledku alebo informácie o konkrétnej chybe. *ResultStatusCode* číselne reprezentuje stav vyhodnotenia kódu. V prípade úspešného vypracovania zadania používateľom sa v správe zobrazuje aj informácia o najdlhšom čase vykonania testu. Ak výstup nezodpovedá očakávanému výsledku, napríklad nesprávny typ alebo hodnota, správa obsahuje porovnanie používateľovho a očakávaného výsledku, vrátane čísla testu, ktorý zlyhal. V pomocnom inicializačnom kóde sú zachytené všetky chyby pomocou *try-except* blokov, čo zabezpečuje, že pri akomkoľvek zlyhaní dostane používateľ jasnú spätnú väzbu o príčine problému.

Systém celkovo pracuje s tromi rôznymi scenármi, ktoré sú reprezentované stavovými kódmi 200, 422 a 500.

1. **200** – znamená, že používateľský kód bol syntakticky v poriadku a prebehol bez technickej chyby, respektíve testovací systém sa dopracoval k aspoň jednému správnemu či nesprávnemu výsledku.
2. **422** – reprezentuje používateľom spôsobenú chybu, pričom táto chyba musela nastať už pri spracovávaní kódu. Môže to byť napr. chyba v syntaxi, neexistujúca trieda *Solution* či prekročený časový limit behu kódu.

3. **500** – popisuje systémovú alebo internú chybu, ktorá nesúvisí s používateľským vstupom. Môže sa jednať napríklad o chybu v komunikácii medzi jednotlivými mikroslužbami alebo databázou.

Tento prístup logicky oddeľuje rôzne scenáre a dáva používateľovi jasnú spätnú väzbu.

5.2.4 Bezpečnosť pri spracovaní kódu

Pri spracovaní kódu je kladený dôraz na bezpečnosť a izoláciu. Používateľov kód beží v *Docker* kontajneri, ktorý zaručuje oddelené prostredie s možnosťou limitácie *RAM* a *CPU*. Ďalej je nastavený *timeout* časový limit, ktorý proces v prípade potreby automaticky ukončí. Zároveň sú zachytávané všetky výnimky pri spracovávaní používateľského kódu, aby bola zabezpečená spätná väzba.

Ďalším bezpečnostným prvkom je *API* braná, ktorá slúži ako jediný vstupný bod z vonkajšieho prostredia. Ostatné mikroslužby sú dostupné výlučne interne. *API* brána tiež aplikuje *rate limiting*, ktorý chráni systém pred nadmerným množstvom požiadaviek a potencionálnymi útokmi. Okrem toho komunikácia a rôzne systémové operácie sú logované a záznamy sú dostupné pre ďalšiu analýzu.

5.2.5 Časová efektivita spracovania kódu

Na úspešnú implementáciu systému je dôležitá aj včasná spätná odozva. Tento čas sa líši v závislosti od počtu testovacích dát, od zložitosti a implementácie úlohy. Nasledujúce údaje boli zaznamenané pomocou softvéru *Postman* a meraním dĺžky vykonávania používateľovho kódu pre jeden testovací dataset na tridsiatich individuálnych spusteniach. Merania boli realizované na implementácii jednoduchej funkcie na otočenie poradia znakov dátovej premennej *string*, v prípade, ak *bool* premenná *flag* má hodnotu *True*. Na vypracovanie danej úlohy bol použitý nasledovný kód:

```
class Solution:
    def reverse(self, word: str, flag: bool) -> str:
        if flag:
            return word[::-1]
        return word
```

V priemere, vykonávanie jednoduchej funkcie na otočenie poriadia písmen dátovej premennej *string* trvalo približne 0.0023 milisekúnd. Tento čas reprezentuje testovací dataset, v ktorom dátová premenná *word* mala hodnotu *"word"* a dátová premenná *flag* mala hodnotu *True*. Celková odozva systému od odoslania požiadavky po získanie odpovedi bola v priemere 35 milisekúnd.

6 Vyhodnotenie cieľov práce

Systém bol navrhnutý a implementovaný na základe stanovených cieľov z kapitoly 3 Ciele a požiadavky práce, podľa ktorých je možné overiť splnenie týchto požiadaviek.

1. Cieľ 1 bol splnený implementáciou mikroslužbovej architektúry, ktorá umožňuje jednoduché pridanie ďalších jazykov v rámci nových mikroslužieb bez zásahu do jadra systému.
2. Ciele 2 a 3 sú splnené prostredníctvom jazykovo nezávislého formátu testovacích dát a implementáciou databázy *MongoDB*, ktorá je schopná uchovávať aj dáta so zložitou a nepravidelnou štruktúrou.
3. Ciele 4, 5 a 6 boli realizované v rámci mikroslužby na spracovanie kódu, ktorý zabezpečuje syntaktickú kontrolu, testovanie riešení a generovanie prehľadnej spätnej väzby.
4. Cieľ 7 je splnený implementáciou *REST API*, ktoré poskytuje jednotný spôsob komunikácie medzi systémom a rôznymi klientskymi rozhraniami.
5. Cieľ 8 je zabezpečený asynchrónnym spracovaním požiadaviek, v ktorom každé testovanie prebieha v samostatnom vlákne.
6. Cieľ 9 je splnený aj vďaka implementácií triedy *Process* a použitiu *Python3* interpretera, ktorého bežný čas spracovania jednoduchej funkcie je pod hranicou jednej milisekundy v testovaných úlohách.
7. Ciele 10 a 11 boli dosiahnuté použitím *Docker* kontajnerov pre izoláciu spracovania kódu, limitovaním systémových zdrojov, logovaním komunikácie a okrem iného aj implementovaním *API* brány s funkciou *rate limiting*.

7 Záver a diskusia

Podľa vyhodnotenia stanovených cieľov bol systém na spracovanie, testovanie a hodnotenie používateľom vypracovaných programovacích úloh úspešne implementovaný.

Systém má pevný základ a poskytuje možnosti pre jeho ďalšie rozširovanie. V oblasti zabezpečenia je to napríklad implementácia autentifikácie a autorizácie používateľov. Z funkcionálneho hľadiska je možné pridávať ďalšie mikroslužby, ktoré môžu, okrem iného, rozšíriť aplikáciu o nové jazyky alebo poskytnúť používateľom možnosť vytvárať ich vlastné programovacie úlohy.

V rámci ďalšieho vývoja platformy je plánované začlenenie tohto riešenia do vzdelávacej aplikácie vytvorenej vo vývojovom prostredí *Unity*. Cieľom bude vytvoriť programovaciu platformu vo forme hry, ktorá by mohla motivovať širšie spektrum používateľov v porovnaní so súčasnými programovacími platformami, ktoré využívajú len určité prvky gamifikácie, akými sú napríklad rebríčky alebo odznaky. Herné prvky môžu zvýšiť záujem a motiváciu riešiť úlohy, nielen kvôli technickému rozvoju, ale aj vďaka postupu v hre, pričom používatelia môžu byť ochotnejší venovať svoj voľný čas aktivite, ktorá spája zábavu so vzdelávaním.

Literatúra

[1] M. -B. Ibáñez, Á. Di-Serio and C. Delgado-Kloos, "Gamification for Engaging Computer Science Students in Learning Activities: A Case Study," in IEEE Transactions on Learning Technologies, vol. 7, no. 3, pp. 291-301, 1 July-Sept. 2014, doi: 10.1109/TLT.2014.2329293. https://ieeexplore.ieee.org/document/6827214

[2] H. Cigdem, M. E. Korkusuz, and C. Karaçaltı, Gamified learning: Assessing the influence of leaderboards on online formative quizzes in a computer programming course, Comput. Appl. Eng. Educ. 32, (2024), e22697.

https://www.researchgate.net/publication/379155563_Gamified_learning_Assessing_the_influence_ofleaderboards_on_online_formative_quizzes_in_acomputer_programming_course

[3] Huesca, G.; Campos, G.; Larre, M.; Pérez-Lezama, C. Implementation of a Mixed Strategy of Gamification and Flipped Learning in Undergraduate Basic Programming Courses. *Educ. Sci.* **2023**, *13*, 474. https://doi.org/10.3390/educsci13050474

[4] Hüsing, T., B. Korte, W. B., Fonstad, N., Lanvin, B., van Welsum, D., Cattaneo, G., Kolding, M.,Lifonti, R. (2013) e-Leadership. e-Skills for Competitiveness and Innovation Vision, Roadmap and Foresight Scenarios. Final Report. https://www.almendron.com/tribuna/wp-content/uploads/2015/03/vision-final-report.pdf

[5] M. Niitsoo, M. Paales, M. Pedaste, L. Siiman, E. Tõnisson (2014) Predictors of informatics students' progress and graduation in university studies, INTED2014 Proceedings, pp. 2521-2529. https://sisu.ut.ee/wp-content/uploads/sites/79/article_predictors_of_informatics_students_progess_and_graduation_in_university_studies.pdf

[6] Capris, Ticiana & Melo, Pedro & Garcia, Nuno & Pires, Ivan & Zdravevski, Eftim. (2023). Comparison of SQL and NoSQL databases with different workloads: MongoDB vs MySQL evaluation. 10.1109/ICDABI56818.2022.10041513.

[7] Vural, Hulya & Koyuncu, Murat & Guney, Sinem. (2017). A Systematic Literature Review on Microservices. 203-217. 10.1007/978-3-319-62407-5_14.

[8] LeetCode, Online Programming Learning Platform (cit. 2025-03-11). Dostupné na: https://leetcode.com

[9] ASP.NET Core Minimal APIs, Microsoft Learn Tutorial (cit. 2025-03-12). Dostupné na: https://learn.microsoft.com/en-us/aspnet/core/tutorials/min-web-api?view=aspnetcore-9.0&tabs=visual-studio

[10] .NET, Free Cross-Platform Development Framework (cit. 2025-03-12). Dostupné na: https://dotnet.microsoft.com

[11] C#, Modern Object-Oriented Programming Language (cit. 2025-03-13). Dostupné na: https://learn.microsoft.com/en-us/dotnet/csharp/

[12] MongoDB, NoSQL Document Database (cit. 2025-03-13). Dostupné na: https://www.mongodb.com

[13] HackerRank, Coding Practice and Interview Preparation (cit. 2025-03-13). Dostupné na: https://www.hackerrank.com

[14] Codewars, Code Challenges for Developers (cit. 2025-03-13). Dostupné na: https://www.codewars.com

[15] Sphere Engine, Code Execution and Assessment API (cit. 2025-03-13). Dostupné na: https://sphere-engine.com

[16] Ocelot, API Gateway for .NET (cit. 2025-03-14). Dostupné na: https://ocelot.readthedocs.io

[17] Docker, Containerization Platform (cit. 2025-03-14). Dostupné na: https://www.docker.com

[18] Seq, Log Management and Structured Logging (cit. 2025-03-14). Dostupné na: https://datalust.co/seq

[19] Postman, API Platform for Building and Testing APIs (cit. 2025-03-14). Dostupné na: https://www.postman.com

Interaktívna vizualizácia 3D modelov

Oskar Mihalik[1], Erik Kučera[2]

[1]Ústav informatiky a matematiky, FEI STU v Bratislave
[2]Ústav automobilovej mechatroniky

xmihaliko@stuba.sk

Abstrakt - Táto práca sa zaoberá návrhom a implementáciou interaktívnej vizualizácie 3D modelov v hernom engine Bevy s využitím programovacieho jazyka Rust. Cieľom práce je vytvoriť aplikáciu, ktorá zefektívňuje komunikáciu medzi 3D dizajnér a klientom, a tým zlepšuje proces návrhu 3D modelov. Vyvinutá aplikácia umožňuje nahrávanie a interaktívne prehliadanie 3D modelov, pričom dizajnér aj klient môžu pridávať poznámky k jednotlivým častiam modelu. Táto funkcionalita podporuje presnejšiu spätnú väzbu a znižuje komunikačné bariéry v procese návrhu. Aplikácia je multiplatformová – môže bežať natívne aj vo webovom prehliadači pomocou webassembly, čo zabezpečuje jej širokú dostupnosť a flexibilitu.

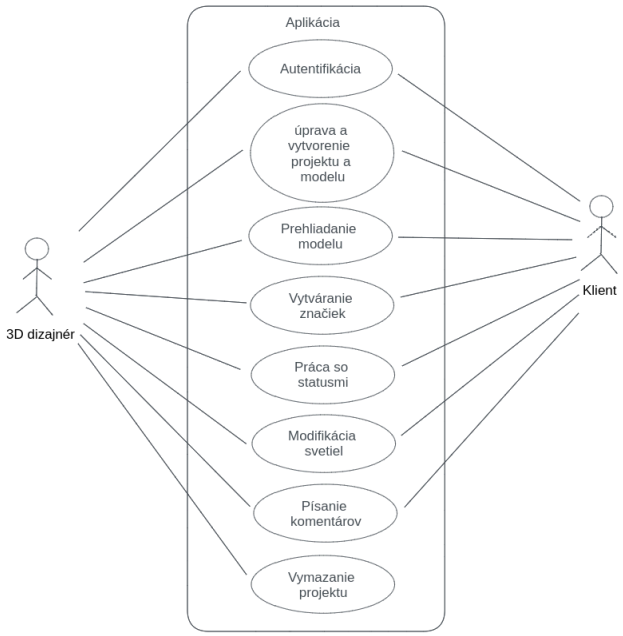
1. Úvod

Digitálna vizualizácia trojrozmerných objektov sa v posledných rokoch stala neoddeliteľnou súčasťou architektúry, dizajnu a inžinierstva. Interaktívne 3D aplikácie umožňujú lepšiu komunikáciu medzi tvorcami modelov a ich klientmi, čím uľahčujú proces návrhu a revízie. Táto práca sa zameriava na vývoj aplikácie pre interaktívnu vizualizáciu 3D modelov, ktorá podporuje spoluprácu medzi dizajnérmi a klientmi prostredníctvom systému komentovania a anotácií.

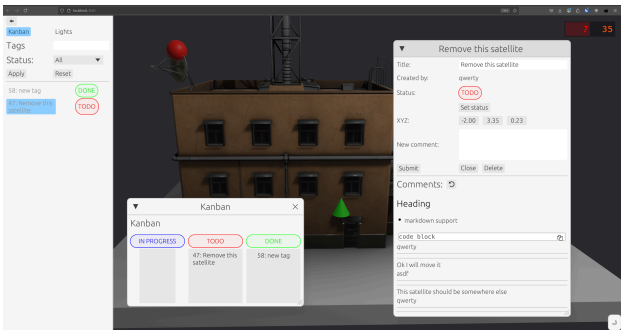
Aplikácia je implementovaná v programovacom jazyku Rust a využíva herný engine Bevy. Jej hlavnou funkcionalitou je možnosť nahrávať 3D modely v glb formáte, pričom dizajnér aj klient môžu pridávať poznámky k jednotlivým častiam modelu. Tieto poznámky sú interaktívne a môžu obsahovať prílohy alebo byť formátované pomocou Markdownu. Používateľské prostredie aplikácie zahŕňa aj nástroje na filtrovanie značiek, osvetlenie modelov a interaktívne ovládanie kamery.

Vývoj aplikácie zahŕňal množstvo výziev, ako je efektívne načítanie a vykresľovanie modelov v prostredí Bevy, implementácia autentifikácie používateľov, správa projektov a integrácia backendu postaveného na frameworku Axum. Aplikácia podporuje viacero platforiem – môže byť spustená natívne aj vo webovom prehliadači, čím sa zvyšuje jej dostupnosť.

Práca sa venuje analýze moderných prístupov a technológií k interaktívnej vizualizácii 3D modelov a popisuje proces návrhu a implementácie aplikácie. Výsledkom je funkčný softvérový nástroj, ktorý zjednodušuje spoluprácu medzi dizajnérmi a klientmi a zároveň poskytuje široké možnosti úprav a interakcie s modelmi.



Obr. 1: Use case diagram



Obr. 2: Ukážka aplikácie

1.1. Existujúce riešenia

1. **ThingLink**: Podporuje vytváranie značiek na 3D a 2D súboroch
   - Podporuje pridávanie značiek vo forme obrázka, videa, webového linku alebo pdf
   - ThingLink je platená aplikácia, najlacnejší plán začína od 500 eur.
   - Nepodporuje vytváranie komentárov alebo manažovanie podľa statusu
2. **Blender, Autodesk**: Tieto programy podporujú vytváranie jednoduchých anotácií alebo poznámok ale nie sú robené pre nie technických používateľov.
3. **Modelo**: Podporuje vytváranie anotácií ale nepodporuje vytváranie komentárov alebo manažovanie podľa statusu

Na záver možno konštatovať, že hoci existuje viacero nástrojov na tvorbu anotácií a prehliadanie 3D modelov, väčšina z nich má určité obmedzenia. ThingLink ponúka širokú škálu značiek, ale chýba mu funkcia komentovania a manažmentu podľa statusu. Blender a Autodesk síce umožňujú anotácie, no sú primárne určené pre technických používateľov. Modelo podporuje anotácie, avšak postráda pokročilé nástroje na spoluprácu.

1.2. Technológie

V tejto časti si opíšeme hlavné technológie, ktoré boli využité na vytvorenie aplikácie, tieto technológie sú relatívne nové a boli zvolené s cieľom preskúmať ich potenciál pri vývoji plnohodnotnej aplikácie.

1.3. Webassembly [1]

WebAssembly (skrátene WASM) je binárny formát nízko úrovňového kódu, ktorý umožňuje spustenie výkonného kódu v prehliadači pri zachovaní blízkej rýchlosti natívnym aplikáciám. Ide o univerzálnu technológiu navrhnutú na spúšťanie kompilovaného kódu v sandboxovanom prostredí, čo umožňuje vývoj aplikácií v jazykoch ako Rust, C, C++, Go alebo AssemblyScript a ich nasadenie na webe vedľa JavaScriptu.

WASM je podporovaný všetkými modernými prehliadačmi (Chrome, Firefox, Safari, Edge), čo z neho robí ideálnu technológiu na vývoj výkonných webových aplikácií bez potreby natívnych knižníc.

Nevýhodou WebAssembly je veľká veľkosť binárnych súborov a ladenie daných binárnych súborov.

1.4. Rust [2]

Rust je moderný systémový programovací jazyk, ktorý kladie dôraz na výkon, bezpečnosť a súbežnosť. Bol vyvinutý s cieľom eliminovať bežné problémy, ako sú pády programov spôsobené dereferencovaním nulových ukazovateľov a pretečenia pamäte.

Rust bol pôvodne vyvinutý v spoločnosti Mozilla a v roku 2015 bol vydaný ako open-source projekt. Dnes je udržiavaný komunitou a podporovaný organizáciou Rust Foundation.

Kľúčové vlastnosti:

1. **Bezpečnosť pamäte bez garbage collectora (GC)**
2. **Systém vlastníctva a životnosti (Ownership & Borrowing)**
3. **Výkon podobný C/C++**

1.5. Bevy [3]

Bevy je moderný open-source herný engine napísaný v jazyku Rust. Využíva Entity-Component-System (ECS) architektúru, vďaka ktorej je rýchly, modulárny a jednoducho rozšíriteľný. Je vhodný na vývoj 2D a 3D hier, vizualizácií a interaktívnych aplikácií.

Kľúčové vlastnosti Bevy

1. **Entity-Component-System (ECS)**: Bevy využíva vlastný ECS systém, ktorý umožňuje efektívne spracovanie veľkého množstva entít.
2. **Výkonný 2D a 3D renderer**: Bevy podporuje vykresľovanie v reálnom čase pre 2D aj 3D grafiku.
3. **Animácie**: Podpora skeletálnej animácie, blendovania a GLTF animácií.
4. **Zvuk**: Možnosť načítať a prehrávať zvukové súbory ako assety.
5. **Cross-platform podpora**: Windows, MacOS, Linux, Web (WASM), iOS, Android.
6. **Rýchla kompilácia**: Kompilujú sa iba najnovšie zmeny.

Na čo sa Bevy hodí?

- **Hry** – Vývoj 2D a 3D hier s moderným ECS prístupom.
- **Interaktívne vizualizácie** – Ideálne na simulácie a 3D modelovanie.
- **Webové aplikácie** – Podpora WebAssembly umožňuje nasadenie na webe.
- **Vývoj experimentálnych aplikácií** – Modulárna architektúra umožňuje jednoduché prototypovanie.

1.6. Axum [4]

Axum je asynchrónny webový framework pre jazyk Rust, ktorý je postavený na tokio a hyper. Poskytuje jednoduché, modulárne a efektívne riešenie na tvorbu webových aplikácií a API.

Kľúčové vlastnosti Axum

1. **Jednoduchý routing**: Definovanie trás a ich spracovanie bez potreby makier.

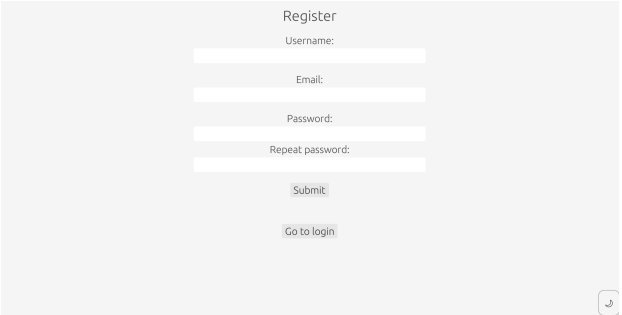
2. **Extrakcia údajov z požiadaviek**: Deklaratívne extraktory umožňujú jednoduché spracovanie parametrov, JSON údajov alebo hlavičiek.
3. Jednoduchá práca s chybami a jednotný spôsob ich spracovania.
4. Jednoduché spracovanie odpovedí a integrácia s rôznymi dátovými formátmi.
5. **Integrácia s Tower ekosystémom**: Využíva tower::Service, čo umožňuje použitie existujúceho middleware (napr. autorizácia, kompresia, tracing).

2. Obrazovky aplikácie

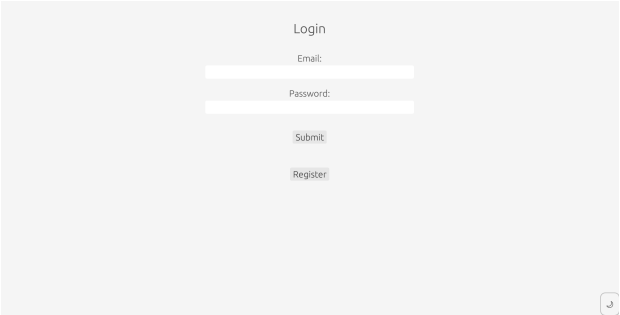
Aplikácia má niekoľko obrazoviek a modálnych okien, ktoré majú svoje vlastné funkcionality.

2.1. Autentifikácia

Používateľ sa môže do aplikácie registrovať:

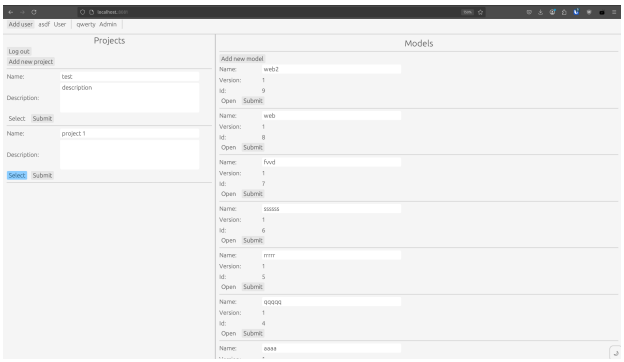


Obr. 3: Register obrazovka
Alebo keď sa už registroval tak sa môže rovno prihlásiť pomocou emailu a hesla.



Obr. 4: Prihlasovacia obrazovka

2.2. Výber projektu a modelu



Obr. 5: Výber projektu a modelu
Na tejto obrazovke je k dispozícii niekoľko dôležitých funkcií, ktoré umožňujú efektívnu správu projektov a modelov:

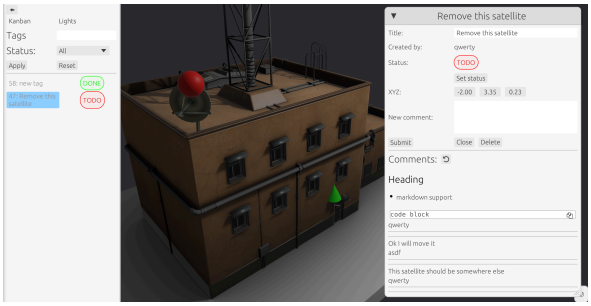
1. **Vytvorenie nového projektu**: Používateľ môže jednoducho vytvoriť nový projekt, ktorý bude obsahovať 3D modely a ďalšie súvisiace údaje. Každý projekt môže predstavovať samostatnú vizualizáciu alebo návrh.
2. **Pridanie novej verzie modelu do projektu**: V rámci existujúceho projektu je možné nahrať a pridať novú verziu 3D modelu. Táto funkcia umožňuje sledovať zmeny a vylepšenia modelu v priebehu času, čím sa uľahčuje iteratívny proces návrhu.
3. **Pridanie používateľa do projektu**: Projekt môže byť zdieľaný s ďalšími používateľmi, ako sú klienti alebo členovia tímu. Táto možnosť umožňuje spoluprácu a interakciu medzi dizajnérom a ostatnými zainteresovanými stranami.
4. **Otváranie verzií modelu**: Používateľ môže vybrať konkrétnu verziu modelu a otvoriť ju na detailné prezeranie a interakciu. To zahŕňa možnosť pridávať komentáre, prezerať poznámky a analyzovať jednotlivé časti modelu.

Tieto funkcie zabezpečujú jednoduchú správu projektov a modelov, čím prispievajú k efektívnej komunikácii medzi dizajnérmi a klientmi.

2.3. Prezeranie modelu

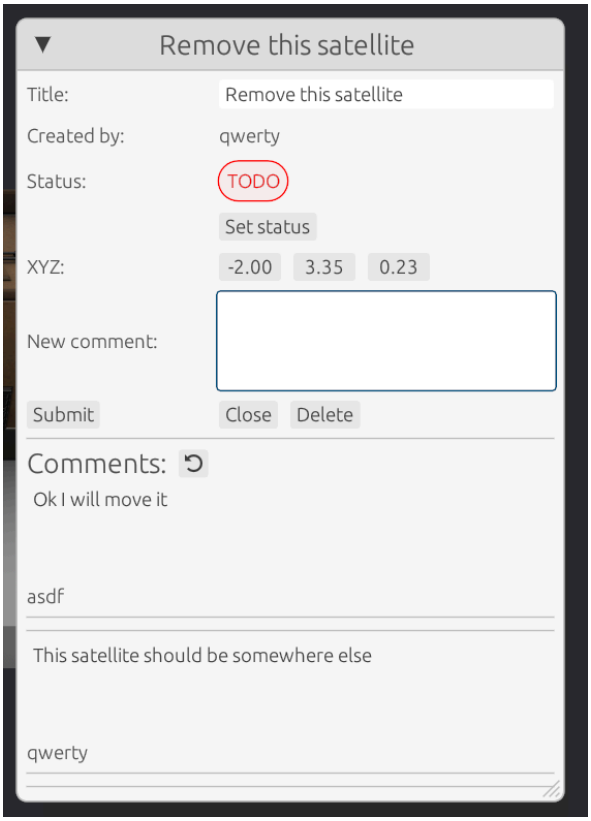
Interaktívna vizualizácia 3D modelu poskytuje množstvo nástrojov na efektívnu komunikáciu medzi používateľmi, ktorí model vytvárajú a tými, ktorí ho schvaľujú alebo pripomienkujú. Hlavné funkcie interaktívnej vizualizácie:

1. **Pridávanie značiek na model**: Kliknutím na nahraný model, môžme pridať novú značku. Značka obsahuje meno, pozíciu, status a komentáre



Obr. 6: Prezeranie modelu

2. **Prezeranie a správa komentárov na značkách**: Používateľ môže vidieť, aké pripomienky alebo komentáre boli pridané k jednotlivým častiam modelu.



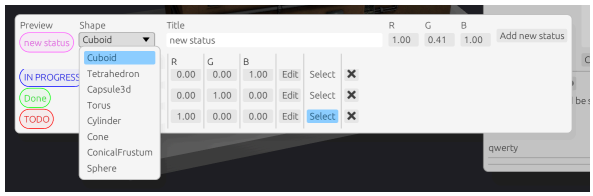
Obr. 7: Detail značky

Komentáre taktiež podporujú Markdown syntax:



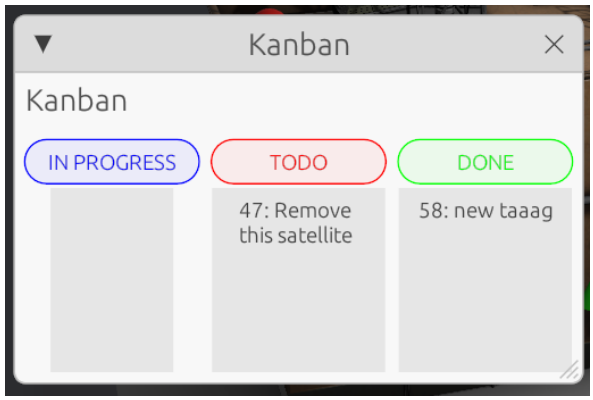
Obr. 8: Podpora Mardownu

3. **Pridávanie a správa statusov na značkách**: Každá značka môže mať priradený status (napr. „V riešení“, „Hotovo“, „Čaká na schválenie“). Statusy pomáhajú organizovať prácu a sledovať postup riešenia jednotlivých pripomienok. Statusu sa dá nastaviť aj typ objektu, či je to kocka, kváder, guľa, ihlan....



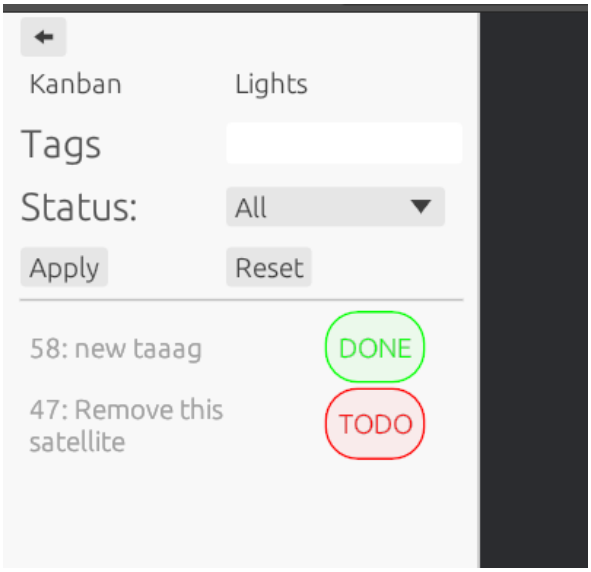
Obr. 9: Správa statusov

4. **Kanban systém pre správu značiek**: Statusy značiek je možné prehadzovať v kanbane, čím sa zefektívňuje pracovný postup. Používatelia môžu jednoducho meniť stav komentárov podľa aktuálneho progresu projektu.



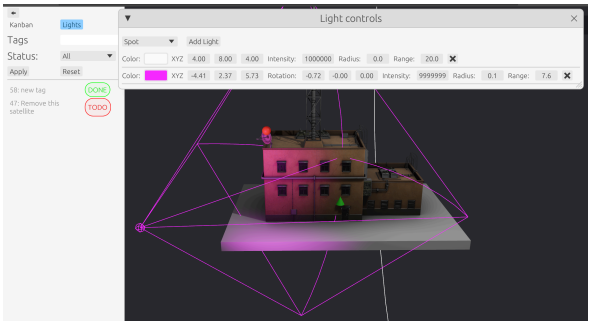
Obr. 10: Kanban

5. **Bočný panel pre správu značiek**: Na bočnom paneli môžme vidieť všetky značky a používateľ dokáže filtrovať podľa mena alebo podľa statusu. Po stlačení tlačidla `Apply` sa filter prejaví aj vizuálne na modeli.



Obr. 11: Ľavý panel

6. **Ovládanie osvetlenia**: Pomocou okna „Light Controls“ je možné vytvárať nové svetelné zdroje dvoch typov – smerové (Directional Light) a bodové (Point Light). Táto funkcionalita slúži na prispôsobenie osvetlenia scény podľa potrieb dizajnéra a klienta, čím umožňuje lepšiu vizualizáciu detailov 3D modelu. Používateľ môže nastavovať intenzitu, farbu a smer svetla, čo pomáha pri simulácii reálnych svetelných podmienok a ich vplyvu na objekt. Možnosť interaktívneho prispôsobenia osvetlenia zvyšuje flexibilitu aplikácie a prispieva k realistickejšiemu zobrazeniu modelov.



Obr. 12: Ovládanie svetla

3. Testovanie

V aplikácii sme vykonali testovanie s viacerými 3D modelmi, ktoré sa líšili svojou veľkosťou, počtom polygónov a komplexnosťou geometrie. Cieľom tohto testovania bolo overiť, či aplikácia dokáže udržať stabilnú snímkovú frekvenciu na úrovni 30 FPS aj pri náročnejších modeloch.

Najväčší 3D model, ktorý sme v rámci testovania použili, mal veľkosť 275,6 MB a obsahoval približne 3,7 milióna polygónov. Napriek vysokej komplexnosti geometrie a značnej dátovej náročnosti aplikácia dokázala tento model úspešne načítať a vykresliť bez zásadného poklesu výkonu.

4. Záver

V tejto práci sme sa venovali návrhu a implementácii interaktívnej vizualizácie 3D budov v hernom engine Bevy s využitím programovacieho jazyka Rust. Hlavným cieľom bolo vytvoriť nástroj, ktorý zjednoduší spoluprácu medzi 3D dizajnérmi a ich klientmi prostredníctvom systému komentovania a anotácií.

Vyvinutá aplikácia umožňuje nahrávanie a interaktívne prehliadanie 3D modelov s možnosťou pridávania poznámok a nastavovania statusov jednotlivým častiam modelu. Implementované funkcionality, ako je vizuálna správa pripomienok, podpora markdownu v komentároch a kanban systém na správu značiek, pomáhajú zefektívniť proces návrhu a revízie modelov. Multiplatformová podpora vrátane WebAssembly navyše zabezpečuje širokú dostupnosť aplikácie.

Porovnanie s existujúcimi riešeniami ukázalo, že dostupné nástroje síce ponúkajú čiastočné možnosti anotácií, avšak vo väčšine prípadov im chýbajú pokročilé nástroje na spoluprácu alebo sú určené predovšetkým pre technických používateľov. Navrhnutá aplikácia rieši tieto nedostatky a poskytuje intuitívne rozhranie pre efektívnu komunikáciu medzi dizajnérom a klientom.

V budúcnosti by bolo možné aplikáciu rozšíriť o podporu ďalších formátov 3D modelov, integráciu s cloudovými úložiskami či vylepšené nástroje na kolaboratívnu editáciu modelov v reálnom čase. Celkovo práca ukázala, že využitie Rustu a Bevy na tvorbu interaktívnych 3D aplikácií má veľký potenciál, a to nielen v oblasti vizualizácie, ale aj v širšom spektre interaktívnych riešení.

Poďakovanie

Ďakujeme za podporu grantom Kultúrnej a edukačnej grantovej agentúry MŠVVaM SR - KEGA 010STU-4/2023 a KEGA 021STU-4/2024.

Literatúra

[1] R. Battagline, *The art of webassembly*. No Starch Press, 2021.
[2] S. Klabnik a C. Nichols, *The rust programming language*. No Starch Press, Inc, 2018.
[3] Bevyengine, “Bevyengine/bevy: A refreshingly simple data-driven game engine built in rust”. Cit: 15. marec 2025. [Online]. Available at: https://github.com/bevyengine/bevy

[4] Tokio, “Axum”. Cit: 14. marec 2025. [Online]. Available at: https://docs.rs/axum/latest/axum/

Detekcia phishingových e-mailov

Bc. Simona Zlatohlávková

Ústav informatiky a matematiky, FEI STU v Bratislave

xzlatohlavkova@stuba.sk

Abstrakt – Cieľom tejto práce je porovnanie efektivity rôznych metód detekcie phishingových e-mailov s využitím moderných jazykových modelov (LLM) a tradičných algoritmov strojového učenia. Skúmané sú tri rôzne dátové množiny – reálna dátová množina z prostredia firmy a dve verejne dostupné dátové množiny. Testované metódy zahŕňajú vlastný model založený na xlm-roberta, ako aj klasifikátory SVM a Random Forest, využívajúce príznaky extrahované z modelu Roberta. Pre porovnanie je zahrnutý aj predtrénovaný model. Výsledky ukazujú, že natrénovanie vlastného modelu môže dosiahnuť vysokú presnosť pri detekcii phishingových e-mailov.

1 Úvod

Phishingové útoky predstavujú jednu z najzávažnejších kybernetických hrozieb v súčasnosti. E-mail je stále najpoužívanejším vektorom útoku, pričom útočníci čoraz častejšie využívajú sofistikované techniky sociálneho inžinierstva. Tradičné prístupy k detekcii phishingu, ako sú blacklisty či heuristiky, často nedokážu efektívne zachytiť nové alebo cielené phishingové kampane.

V posledných rokoch sa do popredia dostali metódy spracovania prirodzeného jazyka (NLP) s využitím veľkých jazykových modelov (LLM), ako sú BERT, Roberta či DistilBERT. Tieto modely umožňujú extrakciu kontextových reprezentácií textu, ktoré môžu výrazne zlepšiť presnosť detekčných systémov. Zároveň však vzniká otázka, či kombinácia hlbokých modelov s tradičnými algoritmami strojového učenia, ako sú SVM alebo Random Forest, môže ponúknuť lepšiu interpretovateľnosť a robustnosť a či natrénovanie vlastného modelu poskytuje spoľahlivé výsledky a dostatočnú presnosť. [1].

Cieľom tejto práce je experimentálne overiť výkonnosť a vhodnosť rôznych kombinácií modelov na detekciu phishingových e-mailov.

2 Metodológia

Vykonané experimenty zahŕňajú testovanie modelov na troch rôznych dátových množinách. Dve množiny (CEAS_08 a Kaggle phishing dataset) obsahujú e-maily rozdelené do dvoch tried (phishingové a legitímne). Naproti tomu firemná dátová množina (AGEL dataset) obsahuje e-maily klasifikované do troch kategórií: phishingové, legitímne a kampaňové. Táto rozdielnosť ovplyvňuje nastavenie klasifikácie – zatiaľ čo pri binárnej klasifikácii model predpovedá iba phishing alebo legitímnu správu, pri multiklasifikácii (firemná dátová množina) je potrebné rozlíšiť aj kampaňové e-maily, čo predstavuje dodatočnú výzvu pre modely.

2.1 Dátová množina

Na experimentálne vyhodnotenie výkonnosti modelov boli použité tri rôzne dátové množiny, pričom každá z nich predstavuje inú formu e-mailovej komunikácie. Tieto dátové množiny obsahujú legitímne aj phishingové e-maily a umožňujú analýzu modelov v rôznych kontextoch.

Použité dátové množiny:

1. CEAS_08 – Táto dátová množina pochádza z verejne dostupných zdrojov, je vhodná na klasickú binárnu klasifikáciu a obsahuje širokú škálu phishingových techník[2,3].

2. AGEL_dataset – Reálna firemná dátová množina obsahujúca e-mailovú komunikáciu spoločnosti. Táto dátová množina je jedinečná, pretože obsahuje e-maily kategorizované do troch skupín. Firemné phishingové e-maily sa často líšia od verejne dostupných phishingových vzoriek – môžu byť sofistikovanejšie a lepšie prispôsobené cieľovej organizácii. Prítomnosť kampaňových e-mailov navyše umožňuje analyzovať prípadné falošné pozitíva (napr. legitímne marketingové e-maily môžu byť omylom označené ako phishing).

3. Kaggle phishing dataset – Verejne dostupná dátová množina, ktorá je užitočná na porovnanie modelov s inými štúdiami a na testovanie generalizácie modelov na rôzne typy phishingových e-mailov[4].

Použitie týchto troch rôznych dátových množín umožňuje lepšie pochopiť, ako sa modely správajú v reálnom firemnom prostredí v porovnaní s verejne dostupnými phishingovými dátami. Výsledky experimentov ukázali, že modely trénované na verejných dátových množinách nemusia byť dostatočne efektívne pri detekcii sofistikovaných firemných phishingových e-mailov.

Tab. 1. CEAS_08 dataset

č.	Označenie legitímnosti	Počet
1	Legitímne	17312
2	Phishingové	21842

Tab. 2. AGEL dataset

č.	Označenie legitímnosti	Počet
1	Legitímne	74
2	Phishingové	197
3	Kampaňové	106

Tab. 3. Kaggle phishing dataset

č.	Označenie legitímnosti	Počet
1	Legitímne	11322
2	Phishingové	7328

2.2 Experimentálne nastavenia modelov

Vzhľadom na rôzny počet tried v dátových množinách boli modely trénované v dvoch nastaveniach:

- Binárna klasifikácia: Pre CEAS_08 a Kaggle phishing dataset modely predpovedajú dve triedy (phishing vs. legitímne).
- Multiklasifikácia: Pre AGEL dataset modely predpovedajú tri triedy (phishing, legitímne, kampaňové).

Architektúra modelov bola upravená podľa typu klasifikácie:

- XLM-RoBERTa-Base bol trénovaný s dvojtriednym výstupom (num_labels=2) pre binárne dátové množiny a s trojtriednym výstupom (num_labels=3) pre firemnú dátovú množinu.
- SVM a Random Forest boli trénované na príznakoch extrahovaných z XLM-RoBERTa samostatne pre binárnu a multiklasovú klasifikáciu.
- DistilBERT model bol aplikovaný len na binárnu klasifikáciu, keďže jeho predtrénovaná verzia nezahŕňa kampaňové e-maily.

2.3 Navrhnutý model XLM-RoBERTa-Base

Navrhnutý model je založený na XLM-RoBERTa-Base, viacjazyčnom jazykovom modeli trénovanom na rozsiahlych textových korpusoch[5].

Model bol inicializovaný pomocou AutoModelForSequenceClassification s tromi výstupnými triedami pre multikasifikáciu a s dvoma výstupnými triedami pre binárnu klasifikáciu.

V tomto experimente bola tokenizácia e-mailov realizovaná pomocou knižnice Hugging Face, konkrétne nástroja AutoTokenizer, ktorý je súčasťou ich ekosystému. Dáta boli spracované do formátu Hugging Face Dátová množina bola rozdelená do tréningovej (80 %) a testovacej (20 %) množiny.

Tréning prebiehal pomocou frameworku Trainer, kde boli použité nasledujúce hyperparametre:

Tab. 4. Nastavenie hyperparametrov XLM-RoBERTa-Base

č.	Parameter	Hodnota
1	Rýchlosť učenia	2e-5
2	Počet epoch	20
3	Veľkosť dávky	10
4	Optimalizátor	AdamW
5	Regularizácia	0.001
6	Kritérium výberu najlepšieho modelu	eval_f1

Model bol trénovaný pomocou early stopping stratégie, pričom bol automaticky uložený najlepší checkpoint na základe F1 skóre.

Pre dátovú množinu AGEL bol aplikovaný proces predtrénovania modelu XLM-RoBERTa-Base na najčastejšie používaných frázach v legitímnych a phishingových e-mailoch v rámci spoločnosti AGEL. Vďaka adaptácii modelu na reálne firemné e-maily je možné výrazne zvýšiť presnosť klasifikácie a efektívnejšie detegovať podvodné správy cielené na konkrétnu organizáciu. Tento prístup minimalizuje riziko neodhalených phishingových útokov.

Tab. 5. Výsledky XLM-RoBERTa-Base

č.	Dátová množina	Precission	Recall	F1 score
1	AGEL	0.9633	0.9605	0.9592
3	CEAS_08	0.9980	0.9980	0.9980
4	Kaggle phishing dataset	0.9811	0.9807	0.9873

2.4 Tradičné metódy SVM a Random Forest

Na porovnanie výkonu pokročilých jazykových modelov s tradičnými metódami strojového učenia boli implementované Support Vector Machine (SVM) a Random Forest (RF) klasifikátory. Tieto modely boli trénované na číselných príznakoch extrahovaných z XLM-RoBERTa-Base, čo umožnilo využiť hlboké jazykové reprezentácie textu bez potreby priameho použitia celého transformerového modelu pri klasifikácii.

Z modelu XLM-RoBERTa-Base sa extrahovali embeddingy CLS tokenu. Tento token predstavuje celkovú reprezentáciu správy a je najčastejšie využívaný pri úlohách klasifikácie textu.

Extrahované príznaky boli následne použité na trénovanie SVM a Random Forest modelov, pričom dátová množina bola rozdelená na 80 % trénovaciu a 20 % testovaciu množinu. Pred samotným trénovaním boli číselné príznaky normalizované pomocou StandardScaler, čím sa zabezpečilo vyvážené spracovanie dát.

Tab. 6. Nastavenie SVM klasifikátora

č.	Parameter	Hodnota
1	Kernel	Lineárny
2	Probability	Áno

Tab. 7. Nastavenie Random Forest modelu

č.	Parameter	Hodnota
1	Počet stromov	150
2	Kritérium rozdelenia uzlov	Giniho index
3	Maximálna hĺbka	Automaticky určená
2	Počiatočný stav (random state)	42

Tab. 8. Výsledky SVM

č.	Dátová množina	Precission	Recall	F1 score
1	AGEL	0.9466	0.9474	0.9459
2	CEAS_08	0.9963	0.9963	0.9963
3	Kaggle phishing dataset	0.9631	0.9627	0.9628

Tab. 9. Výsledky Random Forest

č.	Dátová množina	Precission	Recall	F1 score
1	AGEL	0.8997	0.8816	0.8611
2	CEAS_08	0.9953	0.9953	0.9953
3	Kaggle phishing dataset	0.9628	0.9627	0.9627

Výsledky v Tabuľkách 8 a 9 ukazujú, že model SVM preukázal vyššiu výkonnosť v porovnaní s Random Forest, najmä na dátovej množine AGEL (F1-skóre 0.9459 oproti 0.8611). Pri dátových množinách CEAS_08 a Kaggle phishing dataset boli rozdiely medzi modelmi minimálne, pričom oba modely dosiahli vysoké metriky presnosti a citlivosti. Najlepšie výsledky boli zaznamenané na dátovej množine CEAS_08 (F1-skóre 0.9963 pre SVM a 0.9953 pre Random Forest), čo naznačuje, že táto dátová množina obsahuje dobre separovateľné vzory. Celkovo výsledky indikujú, že SVM je robustnejší pre špecifickejšie dátové množiny, zatiaľ čo Random Forest dosahuje porovnateľnú výkonnosť na väčších a diverzifikovanejších dátových množinách.

2.5 Predtrénovaný model DistilBERT

V rámci experimentov bol testovaný aj predtrénovaný model DistilBERT, konkrétne verzia cybersectony/phishing-e-mail-detection-distilbert_v2.1[6].

Tento model je založený na architektúre DistilBERT, ktorá je optimalizovanou verziou BERT-u s nižším počtom parametrov a rýchlejšou inferenciou. Model bol fine-tunovaný na viactriednu klasifikáciu phishingových e-mailov a URL adries, pričom rozlišuje legitímne a potenciálne phishingové hrozby. DistilBERT je kompaktný transformerový model, ktorý si zachováva 97 % presnosti BERT-u, pričom je o 60 % menší a beží 2× rýchlejšie[6].

Tab. 10. Výsledky DistilBERT

č.	Dátová množina	Precission	Recall	F1 score
1	CEAS_08	0.9116	0.9215	0.9165
2	Kaggle phishing dataset	0.9354	0.9963	0.9649

Tabuľka 10 prezentuje výkonnostné metriky modelu DistilBERT pri klasifikácii phishingových e-mailov na dvoch online dátových množinách: CEAS_08 a Kaggle phishing dataset. Model dosiahol vysokú presnosť a spoľahlivosť na binárnych dátových množinách CEAS_08 a Kaggle phishing dataset. Výsledky ukazujú, že model dokáže efektívne oddeľovať legitímne a phishingové e-maily, pričom na Kaggle phishing datasete dosiahol lepšie metriky.

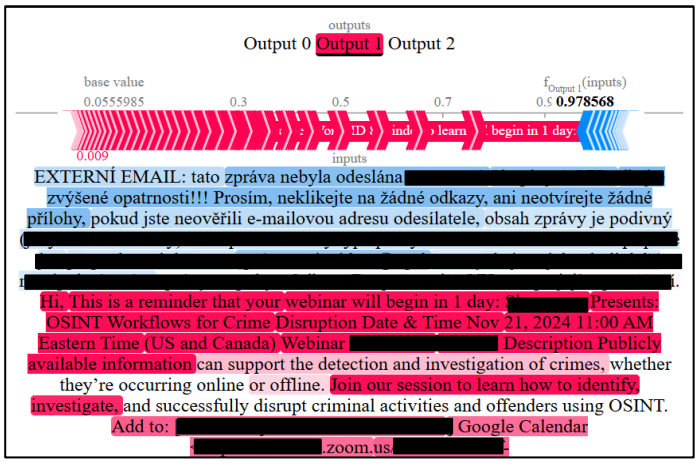
3 Interpretácia modelu XLM-RoBERTa-Base pomocou SHAP

V rámci experimentu boli implementované techniky interpretovateľného strojového učenia s cieľom pochopiť rozhodovacie mechanizmy modelu, ktorý bol trénovaný na klasifikáciu e-mailov. Použitím SHAP (SHapley Additive exPlanations) boli analyzované príspevky jednotlivých slov v e-mailoch k modelovej predikcii.

SHAP je metóda založená na teórii kooperatívnych hier, ktorá umožňuje kvantifikovať, aký vplyv majú jednotlivé znaky (slová) na výstup modelu. V tomto prípade je SHAP využitý na

maskovanie textu, kde každé slovo môže byť postupne „vypnuté“ a meria sa zmena pravdepodobnosti predikcie modelu. Týmto spôsobom môžeme identifikovať kľúčové slová, ktoré model považuje za signifikantné pri klasifikácii e-mailov[7].

V implementácii bol aplikovaný SHAP explainer na dátovú množinu, pričom bol vizualizovaný význam jednotlivých slov prostredníctvom interaktívnych HTML súborov. Tento prístup poskytuje detailný pohľad na to, aké znaky sú kritické pri rozhodovaní modelu a pomáha odhaliť možné biasy alebo nesprávne generalizácie modelu.



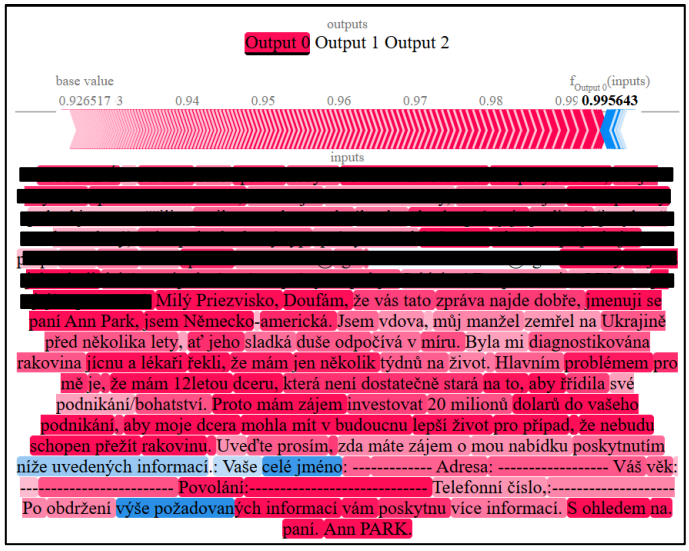
Obr. 1. SHAP vizualizácia legitímneho e-mailu

Model dosiahol pravdepodobnosť 0.978 pre triedu 1 (legitímny e-mail), čo naznačuje, že model s vysokou mierou istoty identifikuje tento e-mail ako legitímny. Červené zvýraznenie textu znamená, že slová zvyšujú pravdepodobnosť predikcie pre legitímny e-mail.

Modré zvýraznenie textu predstavuje slová, ktoré znižujú pravdepodobnosť zatriedenia do triedy pre legitímne e-maily (môžu teda podporovať predikciu zvyšných tried). Model považuje za legitímne časti textu napríklad:

- "Hi, This is a reminder that your webinar will begin in 1 day..."
- "Publicly available information",
- "Join our session",
- "Google Calendar"

Jedným z faktorov, ktorý prispieva k správnej klasifikácii e-mailov ako legitímnych, je analýza hypertextových odkazov obsiahnutých v texte správy. Model využívajúci SHAP analýzu identifikoval, že prítomnosť odkazov smerujúcich na známe a dôveryhodné domény, ako sú Google Calendar, Zoom alebo iné verifikované platformy, významne zvyšuje pravdepodobnosť klasifikácie e-mailu ako legitímneho.



Obr. 2. SHAP vizualizácia phishingového e-mailu

Táto vizualizácia SHAP ukazuje, ako jednotlivé časti e-mailu ovplyvnili modelovú predikciu. Model dosiahol pravdepodobnosť predikcie 0.9956 pre triedu 0 (phishingový e-mail), čo naznačuje, že s vysokou istotou považuje tento e-mail za podvodný. SHAP analýza ukazuje, že text obsahuje manipulatívny príbeh a žiadosť o citlivé údaje, čo model rozpoznal ako klasické znaky phishingu. Tento príklad ukazuje, že model dokáže veľmi efektívne odhaliť podvodné e-maily na základe vzorov v texte.

4 Vplyv značenia kampaní na výkonnosť predtrénovaného online modelu

Pri použití modelu cybersectony/phishing-email-detection-distilbert_v2.1 na dátovej množine poskytnutej firmou Agel som experimentovala s rôznymi spôsobmi mapovania kampaňových e-mailov na binárne kategórie. Kampaňové e-maily oficiálne nepatria medzi phishingové správy, no často obsahujú prvky typické pre phishing, ako sú naliehavé výzvy k akcii alebo odkazy na externé stránky. To môže spôsobovať nejasnosti pri klasifikácii modelmi, ktoré neboli špecificky trénované na ich rozlišovanie.

Pri ich mapovaní ako legitímnych e-mailov model dosiahol presnosť 52,85 %, recall 98,98 % a F1-score 68,90 %, pričom celková presnosť bola iba 53,32 **%**. Keď však tieto e-maily boli mapované ako phishingové, model zaznamenal výrazné zlepšenie s presnosťou 80,76 %, recallom 98,35 % a F1-score 88,69 %, pričom celková presnosť vzrástla na 79,84 **%**. Tento rozdiel poukazuje na význam správneho označovania e-mailov pri trénovaní a hodnotení detekčných modelov a zdôrazňuje potrebu špecifického prístupu ku kampaňovým e-mailom, ktoré môžu modely zavádzať.

5 Záver

Výsledky experimentov ukazujú výrazné rozdiely vo výkonnosti medzi tradičnými modelmi strojového učenia (SVM, Random Forest) a pokročilými transformerovými modelmi (XLM-RoBERTa, DistilBERT) pri detekcii phishingových e-mailov.

Navrhnutý model XLM-RoBERTa-Base dosiahol najlepšie výsledky, pričom na všetkých dátových množinách vykazoval F1 skóre presahujúce 0.98, čím jednoznačne prekonal ostatné prístupy.

Modely SVM a Random Forest boli trénované na príznakoch extrahovaných z XLM-RoBERTa-Base, čo im umožnilo dosiahnuť konkurencieschopné výsledky. SVM sa ukázal ako stabilná alternatíva s vysokou presnosťou a konzistentnými výsledkami, najmä na dátových množinách CEAS_08 (F1 = 0.9963) a Kaggle phishing (F1 = 0.9628), pričom mierne zaostal pri firemnej dátovej množine AGEL (F1 = 0.9459). Random Forest vykazoval mierne horšie výsledky ako SVM, pričom jeho presnosť bola najnižšia pri klasifikácii firemných e-mailov (F1 = 0.8611), čo naznačuje vyšší výskyt falošných pozitív.

DistilBERT bol aplikovaný len na binárne klasifikačné úlohy, kde dosiahol F1 skóre 0.9165 na CEAS_08 a 0.9649 na Kaggle dátovej množine, avšak jeho výkonnosť zaostávala za XLM-RoBERTa-Base aj modelmi trénovanými na jeho extrahovaných príznakoch.

Celkové výsledky naznačujú, že natrénovanie vlastného modelu založeného na XLM-RoBERTa-Base je najvhodnejším modelom na detekciu phishingových e-mailov, pričom jeho príznaky poskytujú dostatočne reprezentatívne informácie na úspešnú klasifikáciu aj v prípade tradičných algoritmov, ako sú SVM a Random Forest.

Poďakovanie

Týmto by som rada vyjadrila úprimné poďakovanie prof. Dr. Ing. Milošovi Oravcovi za jeho cenné rady, odborné vedenie a neoceniteľnú podporu počas vypracovania tohto článku. Zároveň ďakujem tímu z firmy AGEL SK a.s. za pravidelné konzultácie, poskytnutie e-mailových dát a precíznu ručnú klasifikáciu e-mailov, ktorá významne prispela k realizácii experimentov. Ich spolupráca a odborné znalosti boli neoceniteľné pri analýze problematiky detekcie phishingových e-mailov v reálnom prostredí.

Literatúra

[1] ELIOT BAKER - MAXIME CARTIER Phishing Trends Report. In *Phishing Trends Report* [online]. Dostupné na internete: <https://hoxhunt.com/guide/phishing-trends-report>.

[2] CHAMPA, ARIFA I et al. Curated Datasets and Feature Analysis for Phishing Email Detection with Machine Learning. In *3rd IEEE International Conference on Computing and Machine Intelligence (ICMI)* . [s.l.]: IEEE, 2024. s. 1–7. .

[3] CHAMPA, ARIFA I et al. Why Phishing Emails Escape Detection: A Closer Look at the Failure Points. In *12th International Symposium on Digital Forensics and Security (ISDFS)* . [s.l.]: IEEE, 2024. s. 1–6. .

[4] SUBHADEEP CHAKRABORTY [online]. .[s.l.]: Kaggle, 2023. Dostupné na internete: <https://www.kaggle.com/dsv/6090437>.

[5] ALEXIS CONNEAU et al. Unsupervised Cross-lingual Representation Learning at Scale. In *CoRR* [online]. 2019. Vol. abs/1911.02116. Dostupné na internete: <http://arxiv.org/abs/1911.02116>.

[6] CYBERSECTONY [online]. .Machine Learning Model. Dostupné na internete: <https://huggingface.co/cybersectony/phishing-email-detection-distilbert_v2.1>.

[7] ABID ALI AWAN SHAP. In *An Introduction to SHAP Values and Machine Learning Interpretability* [online]. 2023. Dostupné na internete: <https://www.datacamp.com/tutorial/introduction-to-shap-values-machine-learning-interpretability>.

Webová aplikácia na hľadanie spolubývajúcich

Teak Min Guillaume Oh[1], Oľga Stašová[1]

[1]Ústav vedy a výskumu, FEI STU v Bratislave

xoh@stuba.sk

Abstrakt – Cieľom tejto práce je návrh a implementácia komplexnej webovej aplikácie, ktorá slúži na efektívne prepájanie používateľov hľadajúcich spolubývajúcich alebo voľné ubytovanie v zdieľaných bytoch. Aplikácia umožňuje registráciu používateľov, detailné vytváranie profilov a inzerátov, a pokročilé vyhľadávanie potenciálnych spolubývajúcich či ponúk bývania na základe širokého spektra filtrov, ako sú lokalita, rozpočet, záujmy, životný štýl a ďalšie osobné preferencie. Riešenie je postavené na modernom technologickom stacku, ktorý zahŕňa frontend framework Nuxt 3, backend postavený na Node.js s Express.js a platformu Firebase pre autentifikáciu, databázu a úložisko. Na slovenskom trhu predstavuje táto aplikácia unikátny prístup, keďže dostupné alternatívy buď úplne chýbajú, alebo sú výrazne zastarané z pohľadu dizajnu, použiteľnosti a funkcionality. Hlavný dôraz sa kladie na bezpečnosť používateľských dát, intuitívny používateľský komfort, responzívny dizajn a budúcu škálovateľnosť platformy.

1 Úvod

Hľadanie vhodného bývania a spolubývajúcich je pre mnohých ľudí, najmä študentov a mladých profesionálov vo väčších mestách, často náročným procesom. S narastajúcim dopytom po spolubývaní, či už z ekonomických alebo sociálnych dôvodov, rastie aj potreba efektívnych a spoľahlivých nástrojov na vyhľadávanie a prepájanie ľudí s kompatibilnými požiadavkami a životným štýlom. Napriek tomu, že na medzinárodnej úrovni existuje viacero úspešných platforiem (napr. SpareRoom, Roomies), slovenský trh v tomto smere stále zaostáva. Existujúce lokálne riešenia sú často obmedzené na jednoduché inzertné portály, ktoré neposkytujú dostatočné možnosti filtrovania, detailné používateľské profily ani moderné používateľské rozhranie. Mnohé z nich sú navyše vizuálne aj funkčne zastarané a neprispôsobené pre mobilné zariadenia.

Cieľom tejto práce je preto navrhnúť a implementovať modernú, komplexnú webovú aplikáciu špecificky zameranú na slovenský trh. Aplikácia má ambíciu vyplniť medzeru na trhu tým, že používateľom umožní jednoduchým a intuitívnym spôsobom nielen nájsť ponuky bývania, ale predovšetkým nájsť kompatibilných spolubývajúcich podľa detailných kritérií, akými sú lokalita, rozpočet, preferencie týkajúce sa čistoty, hlučnosti, sociálnych návykov, záujmov či životného štýlu. Riešenie je postavené na súčasných a osvedčených technológiách – frontend využíva univerzálny framework Nuxt 3 (postavený na Vue.js), backend je realizovaný pomocou Node.js a frameworku Express.js, a pre správu dát, autentifikáciu a hosting je využitá platforma Firebase.

Hlavným prínosom tejto práce je vytvorenie unikátnej aplikácie, ktorá reflektuje aktuálne používateľské požiadavky na rýchlosť, intuitívnosť, bezpečnosť a estetiku moderných webových služieb. Takéto komplexné riešenie, ktoré integruje inzerciu, profilovanie používateľov, filtrovanie a komunikáciu, zatiaľ na slovenskom trhu chýba, čo zdôrazňuje jeho praktický význam a potenciál. Práca detailne popisuje proces analýzy požiadaviek, návrh architektúry, výber technológií, implementáciu kľúčových funkcionalít a porovnanie s existujúcimi alternatívami.

2 Jadro

Táto časť práce sa podrobne venuje návrhu, implementácii a jednotlivým funkcionalitám vyvinutej webovej aplikácie pre hľadanie spolubývajúcich. Hlavným cieľom bolo vytvoriť moderný, robustný, používateľsky prívetivý a škálovateľný systém, ktorý adresuje identifikovaný nedostatok kvalitných služieb tohto typu na slovenskom trhu.

2.1 Analýza požiadaviek

Na začiatku vývojového procesu bola vykonaná dôkladná analýza trhu a potrieb cieľovej skupiny (študenti, mladí pracujúci). Táto analýza zahŕňala preskúmanie existujúcich slovenských aj zahraničných platforiem a identifikáciu ich silných a slabých stránok. Na základe tejto analýzy boli definované kľúčové funkčné a nefunkčné požiadavky na novú aplikáciu:

- **Registrácia a správa profilu:** Možnosť jednoduchej a bezpečnej registrácie používateľov (email/heslo, prípadne sociálne siete). Vytvorenie detailného používateľského profilu obsahujúceho osobné informácie, fotografie, popis, životný štýl (fajčenie, domáce zvieratá, čistotnosť, denný režim), záujmy a preferencie týkajúce sa hľadaného spolubývania alebo spolubývajúceho.
- **Vytváranie inzerátov:** Možnosť pre používateľov pridávať inzeráty – buď ponuku voľnej izby/bytu na zdieľanie, alebo inzerát typu "hľadám bývanie/spolubývajúceho". Inzeráty by mali obsahovať detaily o lokalite, cene, veľkosti, vybavení, dostupnosti, pravidlách a požadovaných vlastnostiach spolubývajúcich.

- **Pokročilé vyhľadávanie a filtrovanie:** Komplexný systém filtrovania umožňujúci používateľom vyhľadávať inzeráty a profily podľa rôznych kritérií: lokalita (s možnosťou zadania adresy, mesta, oblasti alebo využitia mapy), cenový rozsah, typ ubytovania, dostupnosť, a predovšetkým podľa preferencií životného štýlu a osobnostných charakteristík.
- **Bezpečná autentifikácia a autorizácia:** Zabezpečenie prístupu k používateľským dátam a funkciám prostredníctvom spoľahlivého autentifikačného systému.
- **Intuitívne používateľské rozhranie (UI) a zážitok (UX):** Moderný, čistý a responzívny dizajn, ktorý je ľahko použiteľný na rôznych zariadeniach (desktop, tablet, mobil). Jednoduchá navigácia a prehľadné zobrazenie informácií.
- **Komunikačný modul:** Integrovaný systém na bezpečnú komunikáciu (chat) medzi používateľmi priamo v rámci aplikácie, aby si mohli vymieňať ďalšie informácie a dohodnúť sa na stretnutí.
- **Systém odporúčaní (Matchovanie):** Funkcionalita, ktorá na základe zhody v profiloch a preferenciách aktívne navrhuje používateľom potenciálne vhodných spolubývajúcich alebo ponuky bývania.
- **Správa obľúbených položiek:** Možnosť uložiť si zaujímavé inzeráty alebo profily pre neskorší prístup.
- **Administrátorské rozhranie:** Nástroje pre správcov platformy na manažment používateľov, inzerátov a riešenie prípadných problémov.

2.2 Použité technológie

Pre dosiahnutie stanovených cieľov a požiadaviek bola aplikácia vyvinutá s využitím moderného a efektívneho technologického stacku:

- **Nuxt 3:** Progresívny frontend framework postavený na Vue.js 3. Bol zvolený pre jeho vynikajúce vlastnosti ako Server-Side Rendering (SSR) a Static Site Generation (SSG), ktoré zlepšujú SEO a počiatočné načítanie stránky, automatické importovanie komponentov a modulov, silný ekosystém a skvelý vývojársky zážitok (DX). Umožňuje efektívny vývoj univerzálnych a responzívnych webových aplikácií.
- **Firebase:** Komplexná platforma od Google (Backend-as-a-Service), ktorá bola využitá pre viaceré kľúčové súčasti:
  - *Firebase Authentication:* Pre bezpečnú a jednoduchú správu registrácie, prihlásenia (email/heslo, sociálne siete) a autentifikácie používateľov.
  - *Cloud Firestore:* NoSQL cloudová databáza pre ukladanie používateľských profilov, inzerátov, chatových správ a ďalších aplikačných dát. Poskytuje real-time synchronizáciu a škálovateľnosť.
  - *Firebase Storage:* Pre ukladanie a správu používateľmi nahraných súborov, najmä profilových obrázkov a fotografií k inzerátom.
  - *Firebase Hosting:* Pre jednoduché a rýchle nasadenie (deployment) frontendovej časti aplikácie.
- **Node.js + Express.js:** Backendová vrstva bola implementovaná pomocou Node.js a populárneho minimalistického frameworku Express.js. Táto vrstva slúži ako medzikrok medzi frontendom a Firebase, poskytuje REST API pre komplexnejšie operácie, validáciu na strane servera a zabezpečuje bezpečnejšiu interakciu s Firebase Admin SDK. Tento prístup zvyšuje kontrolu nad dátami a umožňuje implementáciu vlastnej biznis logiky (napr. pre matchovanie).
- **Tailwind CSS:** Utility-first CSS framework, ktorý bol použitý pre rýchly a konzistentný vývoj používateľského rozhrania. Umožňuje vytvárať responzívny dizajn priamo v HTML šablónach bez nutnosti písania vlastných CSS pravidiel, čo výrazne urýchľuje proces návrhu a iterácie.
- **Google Places Autocomplete API:** Služba z platformy Google Maps Platform, integrovaná na inteligentné a presné dopĺňanie adries pri zadávaní lokality vo vyhľadávacích filtroch a formulároch na vytváranie inzerátov. Zlepšuje používateľský komfort a kvalitu geolokačných dát.
- **FilePond:** Moderná JavaScriptová knižnica pre nahrávanie súborov. Poskytuje používateľsky prívetivé rozhranie pre upload obrázkov s podporou drag-and-drop, náhľadmi pred uploadom, indikáciou priebehu a možnosťami validácie (typ súboru, veľkosť).
- **PhotoSwipe:** Responzívna JavaScriptová galéria obrázkov. Slúži na atraktívne zobrazovanie fotografií k inzerátom alebo profilom vo formáte fullscreen lightboxu, ktorý je optimalizovaný pre dotykové ovládanie na mobilných zariadeniach aj pre ovládanie myšou/klávesnicou na desktope.

Tento technologický stack bol zvolený s ohľadom na rýchlosť vývoja, škálovateľnosť, dostupnosť komunity a dokumentácie, a moderné prístupy k webovému vývoju.

2.3 Architektúra systému

Systém je navrhnutý podľa osvedčenej klient-server architektúry, ktorá pozostáva z troch hlavných častí: frontend (klient), backend (serverová logika) a BaaS (Firebase).

Frontend (Klient): Je tvorený univerzálnou aplikáciou postavenou na Nuxt 3. Beží primárne v prehliadači používateľa a je zodpovedný za:

- Zobrazovanie používateľského rozhrania (UI) a interakciu s používateľom.
- Zber používateľských vstupov prostredníctvom formulárov.
- Validáciu vstupov na strane klienta pre rýchlu spätnú väzbu.

- Komunikáciu s backendovým API na získavanie a odosielanie dát.
- Priamu interakciu s niektorými službami Firebase (napr. Authentication pre prihlásenie, Storage pre nahrávanie súborov).

Využitie Nuxt 3 umožňuje efektívne spravovať stav aplikácie, routovanie a renderovanie (SSR alebo CSR podľa potreby).

Backend (Server): Je postavený na platforme Node.js s využitím frameworku Express.js. Jeho hlavné úlohy sú:

- Poskytovanie zabezpečeného REST API pre frontend.
- Obsluha požiadaviek od klienta (napr. filtrovanie inzerátov, spracovanie profilových dát).
- Server-side validácia všetkých prichádzajúcich dát ako dodatočná bezpečnostná vrstva.
- Implementácia komplexnejšej biznis logiky (napr. algoritmus pre matchovanie používateľov).
- Bezpečná komunikácia s Firebase službami (Firestore, Auth) pomocou Firebase Admin SDK. Tento server-to-server prístup je bezpečnejší než priamy prístup klienta k citlivým operáciám v databáze, nakoľko kľúče Admin SDK nie sú vystavené v prehliadači.

Všetky verejné API endpointy sú chránené mechanizmami ako CORS (Cross-Origin Resource Sharing) a citlivé operácie vyžadujú overenie autentifikačného tokenu používateľa.

Backend-as-a-Service (Firebase): Poskytuje infraštruktúru pre:

- **Autentifikáciu:** Správa používateľských účtov a prihlasovacích metód.
- **Databázu (Firestore):** Ukladanie a synchronizácia dát v reálnom čase. Prístup k dátam je ďalej riadený pomocou Firestore Security Rules, ktoré definujú, kto môže čítať a zapisovať konkrétne dáta.
- **Úložisko (Storage):** Ukladanie binárnych súborov (obrázkov). Prístup je riadený pomocou Storage Security Rules.

Táto architektúra poskytuje dobrú rovnováhu medzi rýchlosťou vývoja (vďaka Firebase), flexibilitou a kontrolou (vďaka vlastnému backendu) a výkonom a používateľským zážitkom (vďaka Nuxt 3).

2.4 Kľúčové funkcionality

Aplikácia ponúka širokú škálu funkcií navrhnutých pre čo najefektívnejšie a najpríjemnejšie hľadanie spolubývania:

- **Registrácia a prihlásenie:** Bezpečný proces registrácie a prihlásenia pomocou emailu a hesla, spravovaný cez Firebase Authentication. Možnosť rozšírenia o sociálne prihlásenie (Google, Facebook).
- **Správa rofilov a inzerátov:** Používatelia si môžu vytvoriť detailný osobný profil a tiež pridávať/spravovať inzeráty (ponuka izby alebo hľadanie bývania). Formuláre sú navrhnuté intuitívne s validáciou vstupov.
- **Pokročilé filtrovanie a vyhľadávanie:** Komplexný vyhľadávací nástroj umožňuje filtrovať výsledky na základe lokality (s využitím Google Places Autocomplete pre presnosť a jednoduchosť zadávania), cenového rozpätia, typu ubytovania, dátumu dostupnosti, a najmä podľa detailných preferencií životného štýlu (fajčiar/nefajčiar, domáce zvieratá povolené/nepovolené, preferovaná úroveň čistoty, hlučnosti, sociálnej interakcie, atď.).
- **Interaktívna mapa:** Možnosť zobrazenia ponúk na interaktívnej mape pre lepšiu vizualizáciu lokality. (Poznámka: Toto môže byť budúce rozšírenie, ak nie je implementované).
- **Ukladanie obľúbených:** Funkcia "Pridať medzi obľúbenéümožňuje používateľom jednoducho si uložiť profily alebo inzeráty, ktoré ich zaujali.
- **Integrovaný chat:** Real-time chatovací modul umožňuje priamu a bezpečnú komunikáciu medzi používateľmi v rámci aplikácie, čo uľahčuje dohadovanie detailov alebo osobného stretnutia. Chat využíva Firestore pre ukladanie správ a real-time aktualizácie.
- **Matchovanie (odporúčania):** Systém porovnáva profily používateľov (ich preferencie a životný štýl) a aktívne navrhuje potenciálne kompatibilných spolubývajúcich alebo vhodné ponuky bývania. Algoritmus môže byť založený na jednoduchom porovnávaní zhody v kľúčových atribútoch.
- **Nahrávanie obrázkov:** Moderné rozhranie pre nahrávanie viacerých obrázkov k profilom a inzerátom pomocou knižnice FilePond, s náhľadmi a validáciou.
- **Prehliadanie obrázkov:** Galéria PhotoSwipe pre atraktívne prehliadanie nahraných fotografií.
- **Administrátorské rozhranie:** Samostatná časť aplikácie (alebo sada nástrojov) pre administrátorov na správu používateľských účtov (aktivácia, blokovanie), správu obsahu (schvaľovanie/mazanie inzerátov), monitorovanie aktivity a riešenie nahlásených problémov.

2.5 Bezpečnosť a ochrana dát

Bezpečnosť a ochrana súkromia používateľov boli kľúčovými prioritami pri návrhu a implementácii aplikácie. Boli implementované nasledujúce opatrenia:

- **Bezpečná autentifikácia:** Využitie Firebase Authentication, ktoré poskytuje osvedčené a bezpečné metódy pre správu hesiel a autentifikačných tokenov.
- **Autorizácia a riadenie prístupu:** Detailné nastavenie Firestore Security Rules a Storage Security Rules zabezpečuje, že používatelia môžu pristupovať a modifikovať

iba svoje vlastné dáta alebo dáta, na ktoré majú explicitné povolenie. Backendové API overuje platnosť tokenov pri každej citlivej požiadavke.

- **Validácia vstupov:** Dôsledná validácia všetkých vstupov od používateľov na strane klienta (pre UX) aj na strane servera (pre bezpečnosť), aby sa predišlo chybám a útokom ako napr. Cross-Site Scripting (XSS) alebo NoSQL injection.
- **Ochrana API:** Použitie CORS na obmedzenie prístupu k API len z autorizovaných domén. Zabezpečenie endpointov, aby boli prístupné len pre autentifikovaných a autorizovaných používateľov.
- **HTTPS:** Všetka komunikácia medzi klientom, serverom a Firebase prebieha výhradne cez zabezpečený protokol HTTPS.
- **Ochrana súkromia:** Minimalizácia zbieraných osobných údajov len na nevyhnutné minimum. Používatelia majú kontrolu nad viditeľnosťou svojho profilu a informácií. Zváženie súladu s GDPR (aj keď ide o študentský projekt, princípy sú dôležité).

2.6 Porovnanie s existujúcimi riešeniami

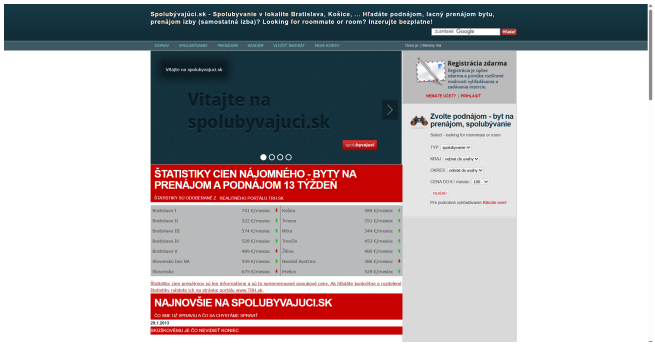
Ako už bolo naznačené v úvode, slovenský trh s aplikáciami na hľadanie spolubývajúcich je pomerne riedky a dominujú mu staršie platformy. Medzi najznámejšie patria portály ako `spolubyvanie.sk`, `spolubyvajuci.sk` a `en.flatmates.sk`.



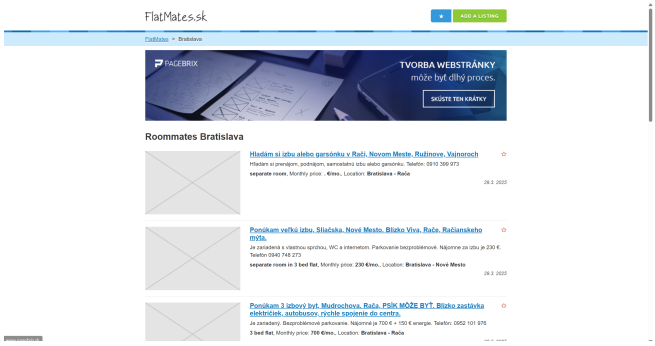
Obr. 1: Stránka spolubyvanie.sk – Príklad tradičného inzerátového portálu bez moderných funkcií a dizajnu, chýbajú detailné používateľské profily.

Tieto existujúce platformy (Obr. 1, 2, 3) majú niekoľko spoločných nedostatkov v porovnaní s navrhovanou aplikáciou:

- **Zastaraný dizajn a UX:** Ich používateľské rozhranie často pôsobí neaktuálne, nie je plne responzívne a celkový používateľský zážitok nezodpovedá moderným štandardom.
- **Obmedzené funkcie profilov:** Väčšinou fungujú ako klasické inzertné weby, kde hlavný dôraz je na inzerát o nehnuteľnosti. Chýbajú detailné používateľské profily zamerané na osobnosť, životný štýl a preferencie, ktoré sú kľúčové pre nájdenie kompatibilného spolubývajúceho.



Obr. 2: Stránka spolubyvajuci.sk – Veľmi jednoduchý systém založený na inzerátoch, bez možnosti vytvárania profilov a pokročilého filtrovania podľa kompatibility.



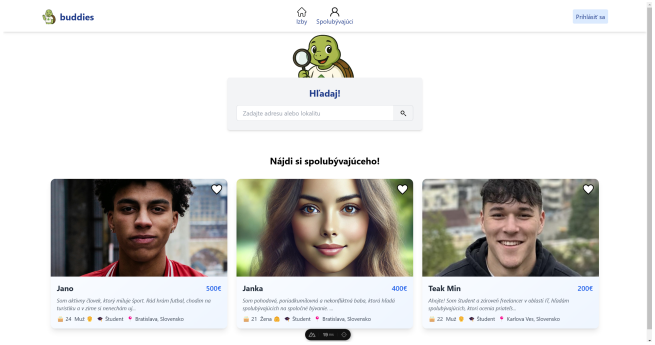
Obr. 3: Stránka en.flatmates.sk – Stránka s podobne zastaraným vzhľadom a funkcionalitou, orientovaná primárne na inzeráty, nie na profily a matchovanie.

- **Nedostatočné filtrovanie:** Možnosti filtrovania sú často obmedzené na základné kritériá ako lokalita a cena. Filtrovanie podľa životného štýlu alebo osobnostných preferencií zvyčajne chýba.
- **Chýbajúce matchovanie:** Neobsahujú žiadne inteligentné funkcie na automatické odporúčanie vhodných kandidátov alebo ponúk.
- **Absencia integrovanej komunikácie:** Komunikácia zvyčajne prebieha mimo platformy (cez email, telefón), čo je menej pohodlné a potenciálne menej bezpečné.

Navrhovaná aplikácia (Obr. 4) sa od týchto riešení zásadne odlišuje práve implementáciou spomenutých moderných prvkov:

- **Moderný UI/UX:** Čistý, intuitívny a plne responzívny dizajn vytvorený pomocou Tailwind CSS.
- **Dôraz na profily a kompatibilitu:** Umožňuje vytvárať detailné profily a využíva tieto informácie pre pokročilé filtrovanie a matchovanie.
- **Inteligentné vyhľadávanie a matchovanie:** Ponúka sofistikované filtre a systém odporúčaní.
- **Integrovaný chat:** Zabezpečuje pohodlnú a bezpečnú komunikáciu priamo v aplikácii.

- **Moderné technológie:** Využitie Nuxt 3, Firebase a Node.js zaručuje vysoký výkon, škálovateľnosť a jednoduchšiu údržbu a ďalší rozvoj.



Obr. 4: Návrh používateľského rozhrania vlastnej aplikácie – Dôraz na moderný vzhľad, prehľadnosť, detailné filtre a zobrazenie profilov.

Tieto vlastnosti robia z navrhovanej aplikácie výrazne konkurencieschopnejšie a pre používateľov atraktívnejšie riešenie problému hľadania spolubývajúcich na Slovensku.

3 Záver

V rámci tejto práce bol úspešne navrhnutý a implementovaný prototyp modernej webovej aplikácie zameranej na vyhľadávanie spolubývajúcich a zdieľaného bývania na Slovensku. Aplikácia bola vytvorená ako reakcia na identifikovanú medzeru na trhu, kde chýbajú komplexné, používateľsky prívetivé a funkčne bohaté platformy tohto typu.

Navrhnuté riešenie efektívne kombinuje intuitívne používateľské rozhranie s robustnou sadou praktických funkcionalít. Medzi kľúčové patrí detailné vytváranie používateľských profilov a inzerátov, pokročilé možnosti filtrovania podľa širokej škály kritérií vrátane životného štýlu a osobných preferencií, systém pre odporúčanie kompatibilných používateľov (matchovanie), integrovaný real-time chat pre bezpečnú komunikáciu a moderné nástroje pre nahrávanie a prehliadanie obrázkov. Použitie Google Places API zjednodušuje prácu s lokalitami.

Architektúra systému, založená na klient-server modeli s využitím moderných technológií ako Nuxt 3 pre frontend, Node.js s Express.js pre backendovú logiku a Firebase pre BaaS (autentifikácia, databáza, úložisko), zabezpečuje nielen dobrý výkon a používateľský zážitok, ale aj vysokú mieru škálovateľnosti, bezpečnosti a relatívne jednoduchú údržbu a ďalší rozvoj. Bezpečnostné aspekty, vrátane ochrany dát a riadenia prístupu, boli zohľadnené počas celého vývojového cyklu.

V porovnaní s existujúcimi slovenskými alternatívami, ktoré sú často len jednoduchými inzertnými portálmi so zastaraným dizajnom a obmedzenou funkcionalitou, predstavená aplikácia ponúka výrazne lepší používateľský zážitok a komplexnejší prístup k problému. Zameriava sa nielen na ponuku a dopyt po bývaní, ale predovšetkým na ľudský faktor – nájdenie vhodného spolubývajúceho na základe vzájomnej kompatibility.

Táto práca tak predstavuje funkčný základ pre platformu, ktorá má potenciál stať sa preferovaným nástrojom pre hľadanie spolubývania na Slovensku. Medzi možné smery budúceho rozvoja patrí ďalšie vylepšenie algoritmu pre matchovanie (napr. s využitím AI), vývoj natívnych mobilných aplikácií pre iOS a Android, implementácia systému hodnotenia používateľov, integrácia s kalendármi pre plánovanie obhliadok, alebo pridanie komunitných funkcií, ako sú fóra alebo skupiny.

Literatúra

[1] Vue.js – The Progressive JavaScript Framework. Dokumentácia. Dostupné na: `https://vuejs.org/guide/introduction.html`

[2] Nuxt 3 – The Intuitive Vue Framework. Dokumentácia. Dostupné na: `https://nuxt.com/docs/getting-started/introduction`

[3] VueFire – Firebase Bindings for Vue.js. Dokumentácia. Dostupné na: `https://vuefire.vuejs.org/`

[4] Express.js – Node.js web application framework. Dokumentácia. Dostupné na: `https://expressjs.com/`

[5] Firebase – Platform Overview. Dokumentácia. Dostupné na: `https://firebase.google.com/docs`

[6] Tailwind CSS – Rapidly build modern websites without ever leaving your HTML. Dokumentácia. Dostupné na: `https://tailwindcss.com/docs/installation`

[7] Node.js – JavaScript runtime built on Chrome's V8 JavaScript engine. Dokumentácia API. Dostupné na: `https://nodejs.org/docs/latest/api/`

[8] FilePond – A JavaScript library that can upload anything you throw at it. Dokumentácia. Dostupné na: `https://pqina.nl/filepond/`

[9] Google Places API Web Service – Overview. Google Maps Platform Dokumentácia. Dostupné na: `https://developers.google.com/maps/documentation/places/web-service/overview`

[10] PhotoSwipe – JavaScript image gallery for mobile and desktop. Dokumentácia. Dostupné na: `https://photoswipe.com/getting-started/`

Sekcia 2

Mechatronika I

Tvorba modulárneho grafického rozhrania pre zmiešanú realitu

Bc. Tomáš Sedláček[1] , doc. Ing. Erik Kučera, PhD.[1]

[1]Ústav automobilovej mechatroniky, FEI STU v Bratislave

xsedlacekt@stuba.sk, erik.kucera@stuba.sk

Abstrakt – Článok predstavuje návrh modulárneho *používateľského rozhrania* (UI) pre aplikácie zmiešanej reality v súlade so štandardmi Industry 4.0. Rozhranie je definované ako submodel v rámci *Asset Administration Shell* (AAS) a vizualizované v prostredí zmiešanej reality pomocou Unity a knižnice *Mixed Reality Toolkit 3* (MRTK3). Výsledkom je systém umožňujúci flexibilné zobrazovanie údajov v aplikácii pre HoloLens 2.

1 Úvod

Industry 4.0 predstavuje súčasný trend digitalizácie a automatizácie výrobných procesov s cieľom vytvoriť inteligentné továrne schopné rýchlej adaptácie, autonómneho rozhodovania a vzájomnej komunikácie bez centrálneho riadenia. S rastúcou dostupnosťou technológií počítačom generovanej reality (*virtuálna realita* (VR), *rozšírená realita* (AR), *zmiešaná realita* (MR)) sa otvárajú nové možnosti pre prepojenie reálneho a digitálneho sveta, najmä v oblasti vizualizácie a interakcie s dátami. Zatiaľ čo tieto technológie sú na trhu prítomné už niekoľko rokov, ich systematické a štandardizované využitie v priemyselných podmienkach je stále predmetom vývoja.

Kľúčovým nástrojom pre zavádzanie štandardizovaných digitálnych reprezentácií zariadení je koncept AAS, ktorý umožňuje formálne popísať štruktúru, stav a funkcionalitu fyzických, ale i softvérových aktív v digitálnej forme. Tento koncept sa v práci využíva na návrh modulárneho používateľského rozhrania pre aplikácie zmiešanej reality.

Cieľom článku je predstaviť architektúru systému, ktorý umožňuje vytváranie a dynamickú vizualizáciu grafických komponentov v MR prostredí, pričom štruktúra týchto komponentov je definovaná v rámci AAS. Riešenie bolo realizované v engine Unity s využitím knižnice MRTK3 a nasadené na zariadení HoloLens 2.

2 Jadro

Táto časť článku sa zameriava na opis kľúčových prvkov navrhovaného riešenia vrátane použitých štandardov, vývojového prostredia, architektúry systému a jeho nasadenia v konkrétnom scenári. Východiskom pre návrh je koncepcia digitálnych komponentov v rámci Industry 4.0 pomocou AAS.

V ďalších podkapitolách je predstavený spôsob, akým je možné definovať používateľské rozhranie ako submodel AAS, jeho následná vizualizácia v aplikácii pre zmiešanú realitu a celková architektúra systému umožňujúca flexibilné a modulárne zobrazenie údajov v MR priestore.

2.1 Teoretické východiská a štandardy

2.1.1 Industry 4.0 a digitálne komponenty

Koncept Industry 4.0 predstavuje moderný pohľad na automatizáciu výrobných procesov s cieľom zvýšiť efektivitu, flexibilitu a transparentnosť výrobných podnikov. Základom tejto vízie je transformácia klasickej výroby na tzv. inteligentné továrne, ktorých komponenty sú schopné autonómneho rozhodovania a vzájomnej komunikácie bez potreby centralizovaného riadenia [3]. Aby bola táto vízia realizovateľná v praxi, boli definované základné princípy Industry 4.0, medzi ktoré patria interoperabilita, virtualizácia, decentralizácia, práca v reálnom čase, orientácia na služby a modularita [4].

Za účelom systematizácie týchto princípov bol navrhnutý referenčný model *Reference Architectural Model Industrie 4.0* (RAMI 4.0), ktorý poskytuje trojrozmerný rámec na popis jednotlivých vrstiev, úrovní a životného cyklu výrobných komponentov [9]. Tento model sa stal základným stavebným kameňom pre vývoj digitálnych reprezentácií v súlade so štandardmi Industry 4.0.

Jedným z hlavných nástrojov pre dosiahnutie interoperability je tzv. Komponent I4.0, ktorý prepája fyzické alebo softvérové aktívum s jeho digitálnym prototypom – v priemyselnej praxi je najvyužívanejšia práve reprezentácia pomocou AAS [8]. AAS predstavuje štruktúrovaný dátový model, ktorý umožňuje zachytiť vlastnosti, funkcionality aktuálny stav aktíva vo forme modulárnych submodelov. Každý submodel môže byť ďalej rozšírený o dynamické väzby, prepojenie na senzorické dáta alebo systémovú logiku.

Na zabezpečenie prístupu k týmto údajom a ich štandardizovaného zdieľania boli definované tri základné spôsoby výmeny dát: statický digitálny popis, rozhranie typu API a peer-to-peer komunikácia s využitím jazyka I4.0 [5]. V tejto práci je štruktúra používateľského rozhrania navrhnutá ako submodel AAS, čo umožňuje jej generovanie a vizualizáciu nezávisle od samotného aplikačného kódu.

2.1.2 Počítačom generovaná realita v priemysle

Technológia *počítačom generovanej reality* (PGR) zohráva v kontexte Industry 4.0 významnú úlohu najmä pri vizualizácii výrobných procesov, návrhu pracovísk a interakcii so zložitými systémami. Pojem PGR zahŕňa technológie virtuálnej, rozšírenej ako aj zmiešanej reality, pričom súhrnne sa tieto technológie v anglickej literatúre označujú ako *extended reality* (XR) [6].

Z hľadiska priemyselného nasadenia má zmiešaná realita najväčší potenciál, keďže umožňuje interakciu reálneho a digitálneho sveta v reálnom čase [2, 7]. V MR aplikáciách dochádza k ukotveniu digitálneho obsahu v reálnom prostredí, čo umožňuje plynulú manipuláciu s hologramami, vizualizáciu dát z digitálnych dvojčiat a prepojenie s údajmi z AAS.

Pre využitie MR v priemysle sú dostupné špecializované zariadenia ako napríklad HoloLens od Microsoftu, ale aj mobilné platformy so softvérovou podporou (Android ARCore, Apple ARKit) [12]. Tieto nástroje napríklad umožňujú vytvárať interaktívne rozhrania nad výrobnými linkami, poskytujúce vizualizáciu dát a navigáciu v priestore.



Obr. 1: Microsoft HoloLens 2 [11]

2.2 Používateľské rozhrania v MR

2.2.1 Výzvy a limity 2D používateľského rozhrania v 3D prostredí

Používateľské rozhrania v zmiešanej realite predstavujú výzvu vďaka potrebe prepájať digitálny obsah s reálnym svetom spôsobom, ktorý je intuitívny a zároveň funkčný. V MR aplikáciách je bežnou chybou používanie klasických 2D používateľských rozhraní prevzatých z desktopových alebo mobilných aplikácií. Tieto rozhrania, hoci známe a rýchlo implementovateľné, nie sú optimalizované pre interakciu v trojrozmernom priestore, čo často vedie k zníženej efektivite a komfortu používania [13].

Experimentálne štúdie ukazujú, že 2D rozhrania v MR prostredí spôsobujú problémy s orientáciou, presnosťou interakcie a znižujú celkovú spokojnosť používateľov. Ovládacie prvky často nie sú vhodne ukotvené v priestore, neberú do úvahy používateľovu perspektívu ani mierku prostredia. Výsledkom je kognitívne náročnejšie ovládanie a nižšia používateľská spokojnosť oproti tradičným zariadeniam ako desktop či tablet [1].

2.2.2 Dizajnové princípy pre MR aplikácie

Efektívny návrh rozhraní pre MR vyžaduje nový prístup zohľadňujúci špecifiká 3D priestoru, ako sú interakcia s hologramami, navigácia v priestore, či použitie gest a hlasových príkazov. V porovnaní s klasickým dizajnom *používateľského rozhrania* (UI) v 2D priestore je potrebné dôslednejšie dbať na adaptabilitu, konzistentnosť, čitateľnosť a jednoduchosť interakcie [10].

Jednou z kľúčových oblastí je vizuálne spracovanie prvkov – nevyhovujúce rozloženie, nízky kontrast textu alebo chaotická organizácia prvkov môžu používateľa frustrovať a znižovať efektivitu práce v MR prostredí [13]. Naopak, využitie modulárneho dizajnu a spätnej väzby (napr. zmeny farby, animácie, potvrdenia akcií) dokáže výrazne zlepšiť používateľský zážitok.

Moderné návrhové prístupy odporúčajú využitie štandardizovaných knižníc, ako je napríklad MRTK3, ktoré ponúkajú optimalizované UI komponenty priamo pre MR aplikácie. Tieto nástroje zároveň zjednodušujú vývoj a znižujú riziko výskytu dizajnových chýb, ktoré by negatívne ovplyvnili používateľskú skúsenosť [1, 10].

2.3 Návrh modulárneho grafického rozhrania

2.3.1 Ciele a požiadavky návrhu

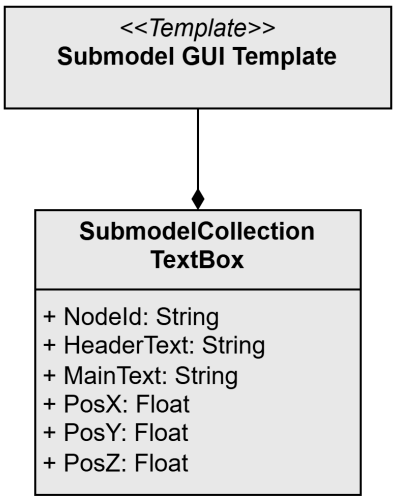
Hlavným cieľom návrhu bolo vytvoriť modulárne, rozšíriteľné a konfigurovateľné používateľské rozhranie, ktoré by bolo kompatibilné s konceptom Asset Administration Shell a zároveň použiteľné v aplikácii zmiešanej reality na zariadeniach typu HoloLens 2.

Navrhované používateľské rozhranie musí spĺňať nasledujúce požiadavky:

- Modularita a škálovateľnosť: systém musí podporovať viacero GUI prvkov definovaných v štruktúre AAS a umožniť ich jednoduché rozšírenie bez potreby zásahov do samotného kódu.
- Zobrazovanie obsahu na základe AAS: jednotlivé komponenty GUI budú reprezentované ako SubmodelCollection objekty v submodeli GUI, čím bude možné dynamicky generovať ich zobrazenie.
- Prehľadnosť a čitateľnosť: obsah jednotlivých dialógových okien musí byť zrozumiteľný a vizuálne konzistentný, s dôrazom na správne umiestnenie v 3D priestore.
- Pripravenosť na dynamické dáta: niektoré prvky rozhrania musia byť schopné zobrazovať živé hodnoty zo senzorov, získavané prostredníctvom OPC UA (nutný ďalší vývoj).
- Integrácia so štandardmi Industry 4.0: návrh musí rešpektovať základné princípy I4.0 a byť postavený na existujúcich štandardoch ako RAMI 4.0, AAS, REST API alebo OPC UA.

2.3.2 Architektúra návrhu

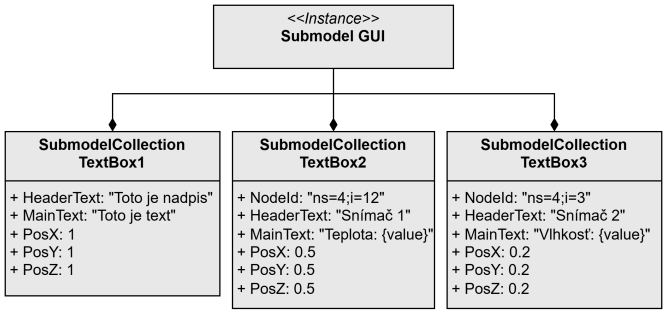
Základným stavebným prvkom návrhu je AAS submodel typu GUI, ktorý obsahuje jednu alebo viacero inštancií prvku typu SubmodelCollection TextBox. Tento textový box predstavuje jednoduchý UI prvok zložený z niekoľkých parametrov: HeaderText, MainText, pozícia v priestore (PosX, PosY, PosZ) a voliteľne NodeId, ktorý slúži ako referenčný bod pre dynamické dáta.



Obr. 2: Navrhnutá štruktúra - AAS Submodel GUI Template s príslušnými parametrami

Pre zvýšenie znovupoužiteľnosti bol definovaný šablónový submodel (Submodel GUI Template), ktorý obsahuje iba základnú definíciu štruktúry TextBoxu. Tento template je možné stiahnuť a použiť ako východisko pre vytvorenie vlastných inštancií GUI komponentov.

Každý SubmodelCollection typu TextBox v submodeli GUI následne predstavuje samostatný vizuálny komponent, ktorý bude v Unity zobrazený ako dialógové okno (Canvas Dialog). Hodnoty parametrov sa čítajú z AAS servera (napr. pomocou BaSyx), pričom systém možno v budúcnosti rozšíriť aj o automatickú aktualizáciu hodnôt na základe NodeId a živých dát z OPC UA servera.



Obr. 3: Príklad konkrétnej inštancie Submodel GUI s viacerými inštanciami SubmodelCollection TextBox

Architektúra systému je navrhnutá tak, aby mohla byť ľahko rozšírená o ďalšie typy UI prvkov (napr. tlačidlá, prepínače), pričom každý nový prvok by bol definovaný ako samostatný template v rámci AAS.

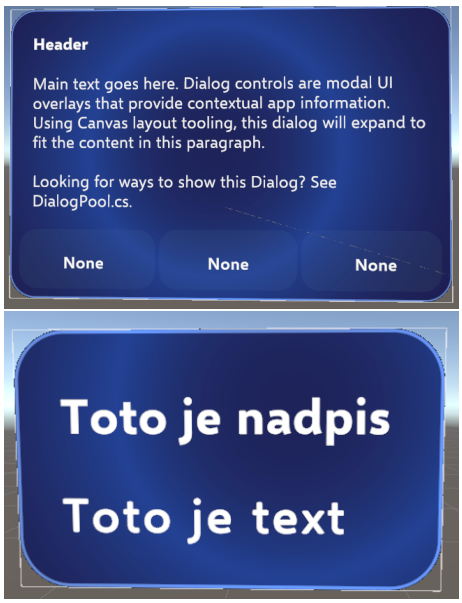
2.3.3 Implementačná platforma

Na vývoj MR aplikácie pre zariadenie Microsoft HoloLens 2 bol zvolený 3D engine Unity v kombinácii s knižnicou MRTK3. Táto dvojica nástrojov predstavuje moderný a stabilný základ pre vývoj aplikácií rozšírenej a zmiešanej reality, pričom zaručuje vysokú kompatibilitu so zariadeniami ekosystému Microsoft.

Unity poskytuje flexibilné prostredie pre tvorbu interaktívnych 3D scén, možnosť dynamickej inštanciácie objektov a rozsiahle možnosti vizualizácie dát. MRTK3 ako open-source framework poskytuje predpripravené komponenty a interakčné prvky optimalizované pre zariadenia MR – vrátane gest, hlasových príkazov, priestorového zarovnania a používateľského rozhrania.

V navrhnutom riešení bol ako základný vizuálny prvok použitý komponent CanvasDialog, ktorý umožňuje zobrazovanie dialógových okien v priestore MR. Tento komponent je súčasťou knižnice MRTK3 a pozostáva z niekoľkých hlavných častí:

- Header Text – zobrazovaný ako nadpis dialógového okna
- Main Text – hlavná textová časť určená na zobrazovanie informácií
- Voliteľné tlačidlá – ktoré je možné zapnúť alebo vypnúť podľa potreby



Obr. 4: CanvasDialog okno z knižnice MRTK3 (hore) a jeho verzia upravená navrhnutou aplikáciou (dole)

Dynamická tvorba používateľského rozhrania je zabezpečená skriptom, ktorý:

- Vytvorí inštanciu komponentu CanvasDialog z pripraveného prefab objektu
- Nastaví obsah hlavičky a hlavného textu podľa údajov zo submodelu typu GUI v AAS
- Umiestni dialógové okno do priestoru na základe súradníc špecifikovaných v AAS (PosX, PosY, PosZ)
- Priradí interakčné komponenty ako Follow a SolverHandler, ktoré zabezpečujú správnu orientáciu a pozíciu okna voči používateľovi

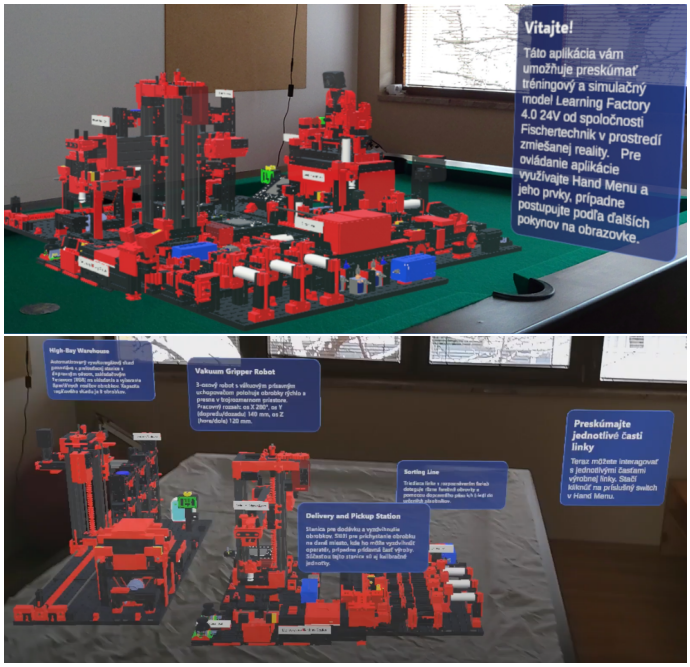


Obr. 5: Architektúra navrhnutého systému pre tvorbu modulárneho grafického rozhrania pre zmiešanú realitu

Týmto spôsobom je vytvorené používateľské rozhranie plne v súlade s konceptom modularity, štandardizácie a dátovej nezávislosti. Štruktúra a obsah rozhrania nie sú definované natvrdo v kóde aplikácie, ale sú odvodené z AAS modelu, čím sa výrazne zvyšuje flexibilita systému a jeho možnosti opakovaného použitia v rôznych priemyselných scenároch.

3 Záver

Navrhnutý systém modulárneho grafického rozhrania pre zmiešanú realitu demonštruje praktickú implementáciu princípov Industry 4.0 v oblasti používateľského rozhrania. Vďaka využitiu štandardu Asset Administration Shell a jeho submodelov je možné vytvárať štruktúrované a opakovane použiteľné definície GUI prvkov, ktoré môžu byť vizualizované priamo v MR aplikácii.



Obr. 6: Príklad aplikácie pre preskúmavanie modelu výrobnej linky Fischertechnik v zmiešanej realite s využitím navrhnutého modulárneho grafického rozhrania

Riešenie bolo nasadené v rámci aplikácie pre preskúmavanie modelu výrobnej linky Fischertechnik a ukázalo svoju flexibilitu, škálovateľnosť a kompatibilitu so štandardmi. Využitím komponentu CanvasDialog v kombinácii s dátovou definíciou v AAS sa podarilo oddeliť dátovú a vizualizačnú vrstvu, čím sa dosiahla vyššia udržateľnosť a rozšíriteľnosť systému.

Navrhované riešenie vytvára základ pre ďalší vývoj rozhraní pre aplikácie v zmiešanej realite s podporou štandardizovanej digitálnej reprezentácie objektov v rámci Industry 4.0.

Poďakovanie

Ďakujeme za podporu grantom Kultúrnej a edukačnej grantovej agentúry MŠVVaM SR - KEGA 010STU-4/2023 a KEGA 021STU-4/2024.

Referencie

[1] Marianna Ciccarelli et al. “Interface and interaction design principles for Mixed Reality applications: the case of operator training in wire harness activities”. In: *Procedia Computer Science* 204 (2022). International Conference on Industry Sciences and Computer Science Innovation, s. 540–547. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2022.08.066. URL: https://www.sciencedirect.com/science/article/pii/S1877050922008043.

[2] Sanika Doolani et al. "A Review of Extended Reality (XR) Technologies for Manufacturing Training". In: *Technologies* 8.4 (2020). DOI: 10.3390/technologies8040077.

[3] Ed. "Industry 4.0". In: *Automa* (11/2015), s. 43. URL: https://automa.cz/cz/casopis-clanky/industry-4-0-2015_11_54253_6867/ (cit. 26. 02. 2025).

[4] Nikola Ilankovic et al. "Smart factories - The product of Industry 4.0". In: *Prosperitas* 7.1 (2020), s. 19–30. DOI: https://doi.org/10.31570/Prosp_2020_01_2. (Cit. 28. 02. 2025).

[5] Industrial Digital Twin Association. *Specification of the Asset Administration Shell Part 2: Application Programming Interfaces*. Tech. spr. IDTA, 2023. URL: https://industrialdigitaltwin.org/wp-content/uploads/2023/04/IDTA-01002-3-0_SpecificationAssetAdministrationShell_Part2_API.pdf (cit. 02. 03. 2025).

[6] Erik Kučera. *Digitálne a inteligentné technológie pre Industry 4.0*. SPEKTRUM Publishing, 2024. ISBN: 9788022754316. DOI: 10.61544/gysj2849. URL: http://dx.doi.org/10.61544/GYSJ2849.

[7] Erik Kučera. *Virtuálna, rozšírená a zmiešaná realita - slajdy z prednášok Virtuálna a zmiešaná realita pre Industry 4.0*. URL: https://vzri.mechatronika.cool/sites/default/files/Prednaska%207%20VR.pdf (cit. 18. 03. 2025).

[8] OPC Foundation. *Introduction to the Asset Administration Shell*. 15. jún 2021. URL: https://reference.opcfoundation.org/I4AAS/v100/docs/4.1 (cit. 03. 03. 2025).

[9] Karsten Schweichhart. *Reference Architectural Model Industrie 4.0 (RAMI 4.0)*. URL: https://ec.europa.eu/futurium/en/system/files/ged/a2-schweichhart-reference_architectural_model_industrie_4.0_rami_4.0.pdf (cit. 15. 01. 2025).

[10] Gangseok Suh et al. "Description of usability problems in augmented reality user interfaces". In: *ICIC Express Letters, Part B: Applications* 13.5 (2022), s. 527–533. ISSN: 2185-2766. DOI: 10.24507/icicelb.13.05.527.

[11] Synergiz. URL: https://www.synergiz.com/wp-content/uploads/2023/01/Hololens-2-mixed-reality-headset-by-microsoft.png (cit. 26. 03. 2025).

[12] Petra Valiauga. *Ako rozšírená a zmiešaná realita zapadajú do Priemyslu 4.0*. 17. aug. 2020. URL: https://www.atpjournal.sk/rubriky/prehladove-clanky/ako-rozsirena-azmiesana-realita-zapadaju-dopriemyslu4.0.html?page_id=31011 (cit. 12. 01. 2025).

[13] Yongkang Xing, Conor Fahy a Jethro Shell. "Assessing web 2D user interface experiences in mixed reality". In: *Heliyon* 10.11 (2024). ISSN: 2405-8440. DOI: https://doi.org/10.1016/j.heliyon.2024.e31916. URL: https://www.sciencedirect.com/science/article/pii/S2405844024079477.

Rozpoznávanie gest pre úlohy ovládania mechatronického systému

Bc. Miloš Ilovský

Ústav informatiky a matematiky

xilovsky@stuba.sk

Abstrakt – Tento príspevok sa zameriava na ovládanie mechatronických systémov na základe gest, s dôrazom na prirodzenú a intuitívnu interakciu človek-robot (HRI). Pomocou technológií snímania a neurónových sietí (LSTM) má finálny systém poskytnúť praktické, efektívne a presné rozpoznávanie gest vhodné pre využívanie v reálnom svete. Článok popisuje metódy zahrňujúce senzory Azure Kinect na zber kostrových dát, klasifikáciu gest na báze LSTM a simulácie PyBullet na ovládanie robotického ramena. Štúdia poskytuje základ pre budúce vylepšenia, ako je klasifikácia gest v reálnom čase pri ovládaní mechatronických zariadení.

1 Úvod

V posledných rokoch sa interakcia medzi človekom a robotom (HRI) stáva čoraz významnejšou oblasťou výskumu v dôsledku rýchleho rozvoja a prijatia technológií Industry 4.0. Cieľom HRI je, aby interakcie medzi ľuďmi a robotmi boli prirodzené, intuitívne a efektívne. Tradičné metódy riadenia robotov, ktoré sa často spoliehajú na klávesnice, ovládače alebo vyhradené rozhrania, môžu byť pre operátorov, ktorí takéto technológie nepoznajú, neintuitívne alebo náročné. Rozpoznávanie gest predstavuje alternatívu, ktorá ponúka intuitívne rozhranie, ktoré môže výrazne zlepšiť použiteľnosť a pohodlie v aplikáciách priemyselnej robotiky. [5]

2 Jadro

Vyvinutý systém spája mechatronické robotické rameno navrhnuté na ovládanie pomocou intuitívnych ľudských gest a neurónové siete typu LSTM. Tieto gestá sú zachytené a následne spracované pomocou senzorov Azure Kinect. Zariadenie Azure Kinect, známe pre svoje presné funkcie sledovania kostry, zachytáva presné údaje o polohe ľudských kĺbov a končatín a efektívne mapuje pohyby používateľa do zrozumiteľného digitálneho formátu. Tieto podrobné kostrové údaje potom fungujú ako základný vstupný zdroj pre rozpoznávanie gest, ktoré sa vykonáva pomocou neurónových sietí s dlhou krátkodobou pamäťou (LSTM).

2.1 Použitie gest

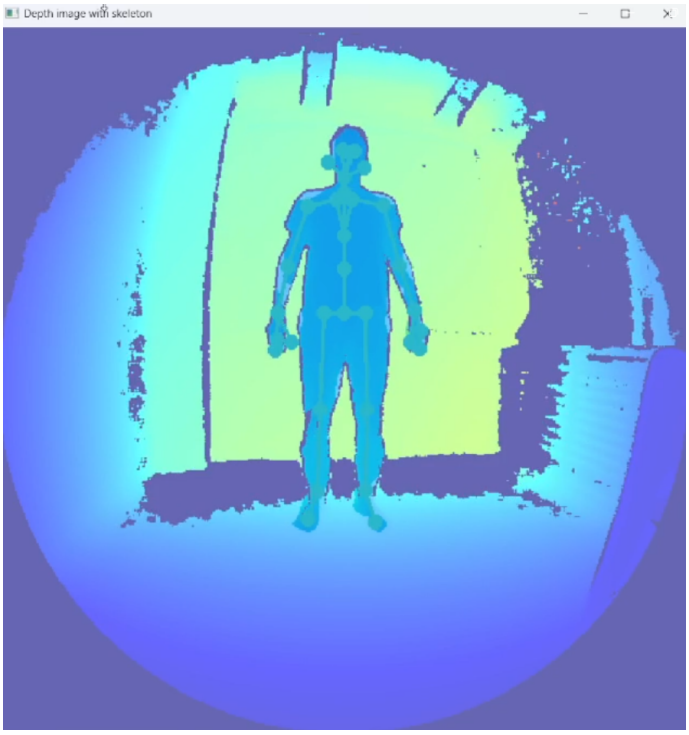
Pri návrhu sady gest pre ovládanie mechatronického systému sa kládol dôraz na prirodzenosť, jednoduchosť pohybu a zároveň spoľahlivosť rozpoznania dát zo senzora Azure Kinect. Gestá museli byť dostatočne odlišné z pohľadu dynamiky aj tvaru trajektórií, aby ich neurónová sieť vedela rozlíšiť aj pri malej vzorke tréningových dát. Zároveň museli byť bezpečné, teda nevytvárať pri výkone žiadne nežiaduce robotické pohyby v prípade falošnej detekcie. [8]
Hlavný model rozpoznáva štyri základné gestá:

- **Gesto doprava:** Ľavá ruka je spustená pri tele, následne sa plynulo posunie doprava a vráti späť. Tento pohyb bol zvolený kvôli jeho jednoznačnej horizontálnej trajektórii a jednoduchému vykonaniu. Slúži na aktiváciu pohybu ramena doprava.
- **Gesto doľava:** Pravá ruka je spustená pri tele a posunie sa smerom doľava a späť. Zrkadlový pohyb k predošlému gestu umožňuje symetrické rozpoznávanie pohybov v oboch smeroch. Slúži na otočenie ramena opačným smerom.
- **Gesto nahor:** Pravá ruka ide pomaly hore a následne prudko smerom dole. Toto gesto má špecifickú rýchlostnú charakteristiku, ktorú dokáže LSTM sieť dobre zachytiť. Slúži na zodvihnutie robotického ramena.
- **Gesto nadol:** Ľavá ruka ide prudko hore a následne pomaly klesá späť. Tento pohyb má opačnú dynamiku ako gesto nahor, čím vytvára jasne rozlíšiteľný časový vzor. Využíva sa na spustenie ramena nadol.

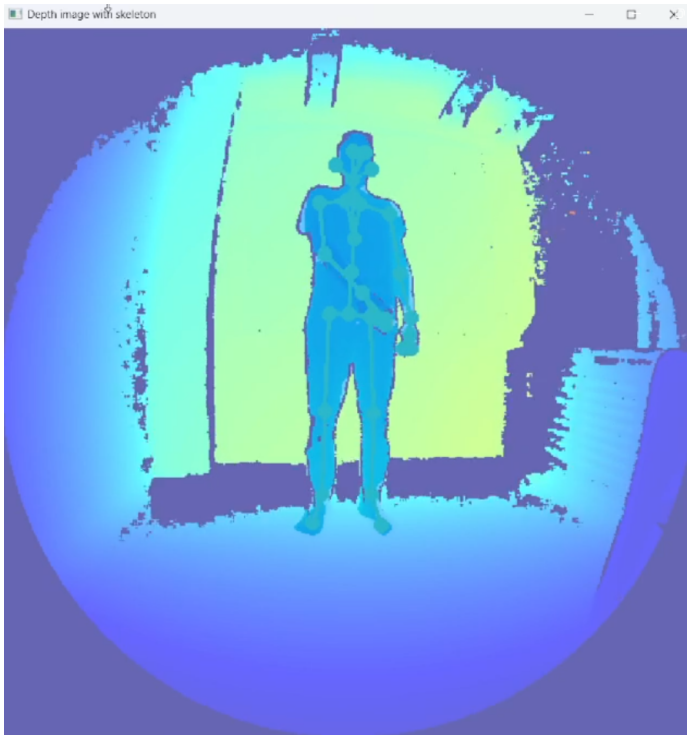
Druhý model, ktorý je samostatne natrénovaný, rozpoznáva binárne potvrdenie alebo zamietnutie predchádzajúceho gesta:

- **Potvrdenie:** Ruky sa vysunú mierne pred telo. Tento pohyb je nenáročný, ľahko realizovateľný a má malú pravdepodobnosť výskytu mimo úmyselného vykonania.
- **Zamietnutie:** Ruky sa vo výške hrudníka prekrížia. Tento pohyb bol zvolený kvôli jednoznačnému pohybu oboch rúk.

Výber gest prešiel viacerými testovaniami. Overovalo sa, ako ich rozpoznáva Azure Kinect, či dochádza k výrazným chybám v sledovaní kĺbov, a ako presne ich dokáže odlíšiť LSTM model. Dôležité bolo aj ergonomické hľadisko – gestá



Obr. 1: Ukážka gesta doľava - začiatok a koniec



Obr. 2: Ukážka gesta doľava - stred

nesmú byť fyzicky náročné, aby bolo možné systém používať opakovane aj počas dlhších časových úsekov.

2.2 Vytváranie datasetu

Správny dataset tvorí základ pre efektívne trénovanie neurónovej siete LSTM. Pre dosiahnutie vysokej presnosti a spoľahlivosti je dôležité získať kvalitné údaje o gestách. Táto sekcia popisuje podrobnosti týkajúce sa metodiky zberu dát, spracovania a ukladania, ako aj predbežné úpravy, ktoré zabezpečujú, že výsledné dáta sú vhodné na trénovanie modelov neurónových sietí.

Na získavanie dát bol použitý senzor Azure Kinect, ktorý poskytuje detailné údaje o pohyboch ľudského tela prostredníctvom kostrového sledovania. Azure Kinect zachytáva trojrozmerné súradnice jednotlivých kĺbov v reálnom čase, pričom tieto súradnice sú následne ukladané do súborov formátu .npy. Proces zberu dát je riadený skriptom napísaným v jazyku Python. Tento skript inicializuje Azure Kinect senzor, ktorý následne zachytáva údaje o jednotlivých kĺbov používateľa. Pri každom snímaní, Azure Kinect poskytuje dáta o polohe všetkých relevantných kĺbov tela, pričom skript zaznamenáva tieto informácie pre každý snímok. Súbor údajov pre jedno gesto pozostáva z pevne stanoveného počtu snímkov (100 snímkov na súbor), čo umožňuje jednotné spracovanie a trénovanie neurónovej siete.

Výstupné súbory sú systematicky ukladané vo formáte .npy do určeného priečinka, pričom každý súbor reprezentuje konkrétnu sekvenciu pohybov pre jedno gesto. Tento prístup zabezpečuje jednoduchú správu dát a ich následnú dostupnosť pre tréningovú časť.

2.3 Augmentácia datasetu

Techniky rozširovania údajov zohrávajú zásadnú úlohu pri zvyšovaní robustnosti modelov neurónových sietí, najmä pri trénovaní s unikátnymi súbormi údajov. V kontexte rozpoznávania gest môže byť získavanie obrovského množstva rôznorodých a realistických vzoriek iba pomocou manuálneho nahrávania nepraktické a časovo náročné. Využitie metód zväčšovania údajov sa preto stáva dôležitou stratégiou na umelé zvýšenie objemu množiny trénovacích údajov.

V tejto práci bolo vybraných niekoľko metód rozšírenia datasetu. Tieto metódy boli vybrané na riešenie bežných odchýlok a neistôt, ktoré sa prirodzene vyskytujú pri vykonávaní ľudských gest, ako sú malé pozičné posuny, meniace sa rýchlosti pohybu a šum snímača. Začlenenie týchto rozšírení zaisťuje, že sa trénovaný model stane odolným voči týmto situáciám v reálnom svete, čo výrazne zlepší jeho schopnosť predikovať počas reálneho použitia.
Metódy augmentácie použité v tejto práci zahŕňajú:

- **Pridanie šumu**: K súradniciam kĺbov je pridaný malý náhodný šum, aby sa simulovali prirodzené nepresnosti snímača a menšie mimovoľné pohyby používateľov. Toto rozšírenie robí model tolerantným voči malým chybám merania alebo poruchám, ktoré sú časté v reálnom prostredí.

- **Posun kĺbov**: Súradnice kĺbov sa posúvajú o náhodný posun. Táto metóda rieši variabilitu polohy používateľa vzhľadom na snímač, čo umožňuje modelu rozpoznať gestá vykonávané mierne odlišne z hľadiska polohy.

- **Mierka kĺbov**: Celková mierka súradníc kĺbov je upravená malým náhodným faktorom. Škálovanie predstavuje fyzické rozdiely medzi používateľmi alebo variácie pri vykonávaní gest. To umožňuje modelu prispôsobiť sa rôznym používateľom s rôznymi fyzickými vlastnosťami.

- **Časový posun** : Časové trvanie sekvencií gest je náhodne natiahnuté alebo stlačené. Táto augmentačná technika sa zameriava na prirodzenú variabilitu rýchlosti gest, čo umožňuje modelu presne klasifikovať gestá vykonávané rôznymi rýchlosťami.

Implementovaný skript uľahčuje systematické rozširovanie a organizáciu údajov datasetu. Každý pôvodný súbor vzorky gest prechádza týmito zmenami, čím sa vytvorí viacero nových rozšírených vzoriek. Skript efektívne zvláda rozširovanie údajov pomocou knižnice NumPy.
Pre ilustráciu sú augmentačné funkcie definované takto:

```
import numpy as np

def add_noise(data, noise_level=0.01):
    return data + np.random.uniform(-noise_level,
                        noise_level, data.shape)

def shift_joints(data, shift_amount=1):
    return data + shift_amount

def scale_joints(data, scale_factor=1.1):
    return data * scale_factor

def time_stretch(data, stretch_factor=1.1):
    indices = np.linspace(0, len(data) - 1,
                int(len(data) * stretch_factor))
    return np.array([data[int(i)]
                    for i in indices])
```

Tieto funkcie sa aplikujú postupne na každú vzorku údajov, čím sa systematicky generujú rôzne rozšírené variácie. Každá rozšírená vzorka je individuálne uložená v korešpondujúcom adresári, čo zaisťuje prehľadnú organizáciu a ľahký prístup počas fázy tréningu neurónovej siete. Štruktúrovaný prístup tiež zjednodušuje budúce rozširovanie súboru údajov a opakované vylepšenia.

Využitím týchto techník rozšírenia sa počiatočný dataset, ktorý pozostával z približne 50 zaznamenaných vzoriek na gesto, výrazne rozšíril na približne 350 rôznych rozšírených vzoriek na gesto. Toto podstatné zvýšenie zaisťuje, že sa neurónovej sieti zvyšuje schopnosť zovšeobecňovať a presne klasifikovať nové, zatiaľ neznáme údaje o gestách v praktických scenároch. Systematická aplikácia týchto rozširujúcich metód zvyšuje rozmanitosť a robustnosť datasetu a účinne rieši bežné problémy, s ktorými sa stretávame pri úlohách rozpoznávania gest. Prostredníctvom starostlivého výberu a aplikácie rozširujúce techniky použité v tomto výskume podstatne zlepšujú celkový výkon, presnosť a spoľahlivosť neurónovej siete.

2.4 Neurónová sieť

Neurónové siete LSTM boli vybrané pre tento projekt z dôvodu ich schopnosti spracovávať sekvenčné údaje. Sú užitočné pri rozpoznávaní ľudských gest, ktoré sú vo svojej podstate časovými vzormi. Gestá nie sú definované jednou pózou, ale skôr tým, ako sa spoločné pozície časom vyvíjajú. Preto musí model analyzovať trajektórie kĺbov a nie iba statické polohy. Siete LSTM poskytujú pamäť v priebehu času, čo je nevyhnutné na pochopenie takéhoto pohybu. [6]

Hlavnou výhodou LSTM siete je, že udržiava vnútorný stav, ktorý sa aktualizuje a prenáša cez časové kroky. Vďaka tomu je ideálny na analýzu údajov časových radov. V našom prípade je každá sekvencia gest reprezentovaná ako časový rad 3D súradníc. Napríklad, ak je vstupná sekvencia dlhá 100 snímok a každá snímka obsahuje 3D polohu viacerých spojov, model sa musí naučiť nielen priestorovú štruktúru, ale aj prechody a vzory pohybu cez týchto 100 snímok. Aby to bolo možné implementovať, údaje gest zo súborov .npy boli predspracované do jednotného vstupného formátu. Každý súbor obsahoval 100 snímok údajov, pričom každý kĺb mal 3 hodnoty (x, y, z). Toto pole sa potom odovzdalo ako vstup do LSTM.

Architektúra modelu začína sekvenčným modelom. Prvá vrstva LSTM obsahuje 64 neurónov a vracia sekvencie, čo znamená, že celý výstup pre všetky časové kroky je odovzdaný ďalšej vrstve. Je to dôležité, pretože umožňuje druhej vrstve LSTM spracovať nielen konečný stav, ale aj celý časový kontext.

Po tejto prvej vrstve nasleduje dávková normalizácia (batch normalization). Dávková normalizácia pomáha stabilizovať a zrýchliť tréning normalizáciou aktivácií v rámci malej dávky. Nasleduje vrstva vypadávania (dropout) s mierou vypadávania 0,2. Dropout pomáha predchádzať preťaženiu náhodným prerušením niektorých spojení počas tréningu, čím sa zaisťuje, že sieť sa príliš nespolieha na špecifické neuróny. [7]

Druhá vrstva LSTM s 32 neurónmi a return_sequences=True pridáva hĺbku do časového modelovania siete. Zahŕňa regularizáciu L1 a L2, ktorá penalizuje veľké váhy, aby sa podporili jednoduchšie, všeobecnejšie predpovede.

```
model.add(LSTM(32, return_sequences=True,
kernel_regularizer=regularizers.l1_l2(0.1)))
```

Nasleduje ďalšia vrstva výpadávania, tentoraz s nižšou mierou 0,1, pretože sme už znížili riziko nadmerného preťaženia neurónov spôsobeného skorším výpadkom a regularizáciou. Nakoniec je pridaná tretia vrstva LSTM s return_sequences=False, ktorá zobrazuje iba konečný stav. V tomto momente ide už o kompaktnú reprezentáciu celého gesta.

Tento konečný stav prechádza do dense vrstvy s aktiváciou softmax. Táto vrstva obsahuje jeden neurón na triedu gest, v našom prípade päť, a vydáva rozdelenie pravdepodobnosti pre triedy gest. Softmax zabezpečuje, že výstup siete možno interpretovať priamo ako klasifikačnú pravdepodobnosť.

Na optimalizáciu siete sa používa optimalizátor Adam s rýchlosťou učenia 0,0001. Zvolili sme nízku rýchlosť učenia aby sme predišli pretrénovaniu.

```
model.compile(optimizer=Adam(learning_rate=0.0001),
loss='categorical_crossentropy', metrics=['accuracy'])
```

Model LSTM slúži ako jadro systému rozpoznávania gest. Jeho vrstvená architektúra spracováva údaje o gestách spôsobom, ktorý zachytáva pohyb kĺbov na nízkej úrovni, aj kontext gest na vysokej úrovni. Prostredníctvom starostlivo vyváženej kombinácie hĺbky, regularizácie a sekvenčného modelovania systém efektívne prevádza ľudský pohyb do strojovo interpretovateľných príkazov. To otvára cestu pre citlivú a intuitívnu interakciu medzi človekom a robotom v reálnych scenároch, kde prirodzená komunikácia prostredníctvom gest zohráva rozhodujúcu úlohu.

2.5 Prostredie pyBullet

Integrácia technológie rozpoznávania gest s realistickým prostredím robotickej simulácie je nevyhnutná na zabezpečenie presnej a spoľahlivej interakcie človek-robot (HRI). Prostredie fyzikálnej simulácie PyBullet poskytuje vhodnú platformu na bezpečné testovanie robotov pred nasadením systému do fyzického prostredia.

Na začiatku sa simulačné prostredie v PyBullet nastaví pripojením k jeho grafickému rozhraniu. Prvý krok zahŕňa definovanie základných parametrov simulácie, ako je gravitácia, trenie a kolízne vlastnosti, ktoré sa veľmi podobajú skutočným podmienkam.

Po načítaní robotického ramena sa stanovia parametre kĺbov a motory sa nakonfigurujú na riadenie polohy. To umožňuje robotickému ramenu presne reagovať na príkazy odvodené od gest.

Ďalej sa model mechanického ramena, konkrétne robotické rameno UR3e, načíta do prostredia simulácie PyBullet. Tento robotický model definovaný súborom Unified Robot Description Format (URDF) presne popisuje geometriu ramena a konfiguráciu kĺbov. Zabezpečenie toho, že URDF robot presne odráža skutočné špecifikácie robota, umožňuje simulácii presne napodobňovať fyzické rameno.

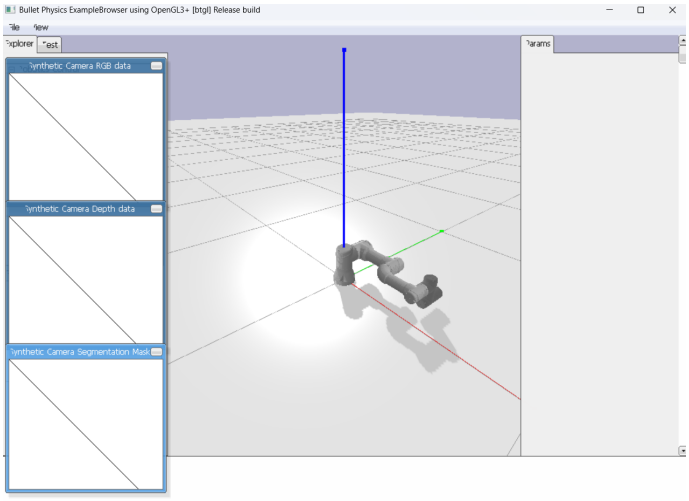
Počas vykonávania simulácie zabezpečujú nepretržité aktualizácie odozvu v reálnom čase. PyBullet neustále prechádza simuláciami, aktualizuje polohu, rýchlosť a interakcie robotického ramena na základe príkazov gest. [9]

2.6 Integrácia pyBullet

Integrácia Azure Kinect do systému hrá rozhodujúcu úlohu pri zachytení a interpretácii ľudských gest s vysokou presnosťou a spoľahlivosťou. Na rozdiel od tradičných RGB alebo 2D senzorov poskytuje Azure Kinect podrobné sledovanie kostry pomocou hĺbkovej kamery a sofistikovaných algoritmov na sledovanie tela založených na umelej inteligencii. Vďaka tomu je mimoriadne vhodný na rozpoznávanie gest v trojrozmernom priestore a v rôznych podmienkach prostredia.

Samotný senzor Azure Kinect je schopný sledovať používateľov v rôznych podmienkach vrátane slabého osvetlenia a čiastočnej oklúzie, vďaka čomu je celý systém vhodný aj do prostredí kde nie je ideálne osvetlenie.

Balík SDK Azure Kinect sa jednoducho integruje s programovacím jazykom Python prostredníctvom knižnice pykinect_azure, čo umožňuje bezproblémové získavanie a



Obr. 3: Prostredie pyBullet

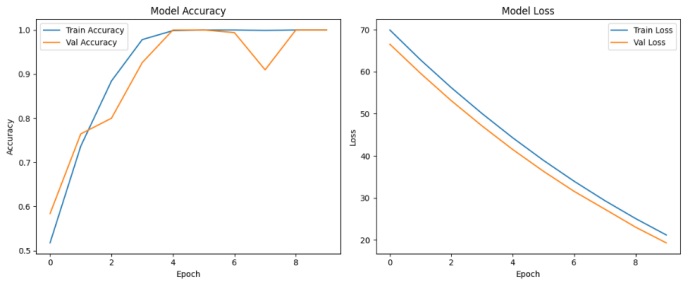
transformáciu údajov v reálnom čase. V každej snímke sa získavajú údaje, extrahujú sa kĺby a pokračuje sa v sledovaní, kým osoba zostáva v zornom poli snímača. Je možné sledovať viacerých používateľov, aj keď pre jednoduchosť, súčasný systém spracováva iba prvé telo.

Spočiatku sa metóda inverznej kinematiky, ktorá je prístupná v prostredí pyBullet, považovala za metódu, ktorá sa bude používať pri prevode pohybov rúk na akcie robotického ramena. Avšak vzhľadom na jej nestabilitu, zložitosť a nepresnosti, ktoré vznikli pri ovládania na základe gest v reálnom čase, bol implementovaný alternatívny prístup. Konečný systém sa spolieha na interpretáciu pozičných rozdielov priamo medzi kĺbmi (napr. rameno k ruke), pričom tieto rozdiely mapuje na uhlové hodnoty pre robotické kĺby v simulačnom prostredí PyBullet. Toto rozhodnutie výrazne zvýšilo stabilitu, odozvu a predvídateľnosť robotického riadiaceho systému.

Snímač Azure Kinect je inicializovaný s nastavením hĺbky s vysokým rozlíšením a sledovaním umožňujúcim rozpoznanie kostry celého tela. Extrahované údaje o kostre zahŕňajú trojrozmerné polohy 32 kĺbov. Pre naše účely sa používa podmnožina kĺbov (ramená a ruky). Tie sa extrahujú a naformátujú do sekvencií, ktoré sa prenesú do neurónových sietí typu LSTM na klasifikáciu.

Rozpoznávanie gest v reálnom čase funguje vo fázach. Prvá fáza zhromažďuje údaje gest so 100 snímok a posiela ich do natrénovaného modelu na klasifikáciu. Po rozpoznaní gesta s dostatočnou istotou sa vyžaduje druhé gesto na potvrdenie akcie. Gestá na potvrdenie sa rozpoznávajú pomocou druhého modelu. Tento dvojfázový proces zabraňuje falošným pozitívam a poskytuje používateľovi mechanizmus na overenie pred spustením pohybu robota.

Ak je rozpoznané gesto „Úplné ovládanie“, systém prejde do kontinuálneho režimu ovládania. Tu sa v reálnom čase sleduje poloha pravej ruky vzhľadom na rameno. Tento rozdiel sa používa na odvodenie smeru a veľkosti pohybu. Na rozdiel od inverznej kinematiky, ktorá by sa pokúšala prispôsobiť polohu ruky v 3D priestore vyriešením nelineárnej rovnice pre uhly



Obr. 4: Graf presnosti a straty modelu

kĺbov, tento systém zjednodušuje úlohu tým, že konvertuje rozdiely polohy na priamo mapované ciele uhlov jednotlivých kĺbov.

Hodnoty kĺbov v PyBullet sú ohraničené na preddefinované bezpečné rozsahy, aby sa zabezpečilo, že príkazy odvodené od extrémneho pohybu používateľa nebudú mať za následok neprirodzené alebo nebezpečné konfigurácie pre robotické rameno. Okrem toho sú hodnoty zosilnenia rýchlosti a polohy vyladené tak, aby vyvážili plynulosť a rýchlosť.

Režim úplného ovládania zahŕňa jednoduché gesto na ukončenie tohto režimu: ak používateľ zdvihne ľavú ruku nad hlavu (zistené pomocou Y-súradníc ramena a ruky), systém tento režim okamžite opustí a vráti sa do pôvodného režimu rozpoznávania gest.

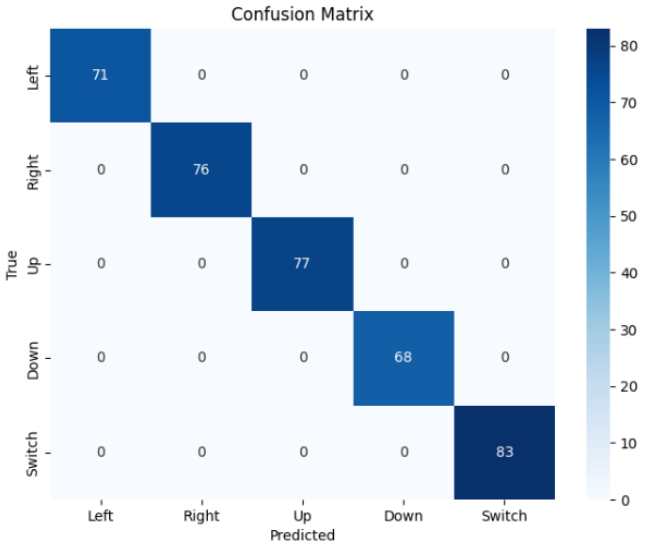
2.7 Vyhodnotenie

Výsledky trénovania hlavného modelu ukazujú veľmi vysokú presnosť na tréningovej aj validačnej množine. Ako vidíme na grafe presnosti (Obr. 4), model dosahuje takmer 100% presnosť už po 5 epochách, pričom validačná presnosť kopíruje tú tréningovú bez výrazných známok pretrénovania. Podobne, na grafe straty (Obr. 4) môžeme pozorovať postupné znižovanie hodnoty loss funkcie počas tréningu, čo naznačuje dobrú konvergenciu modelu.

Konfúzna matica (Obr. 5) potvrdzuje, že model bol schopný správne klasifikovať všetkých päť gest bez jediného falošného rozpoznania na testovacích dátach. Každá trieda má zaznamenaný počet správnych predpovedí medzi 68 až 83, čo zodpovedá vysokej robustnosti modelu v rámci známych dát.

Napriek týmto veľmi pozitívnym výsledkom na testovacej množine však model nedosahuje úplne rovnaký výkon pri reálnom použití. Počas testovania v reálnom čase sa ukázalo, že niektoré gestá sú v určitých prípadoch nesprávne klasifikované, prípadne sa vôbec nedetegujú. Táto odchýlka je spôsobená nedostatočnou diverzitou v trénovacom datasete, konkrétne obmedzeným množstvom augmentácií, ktoré by model pripravili na rôzne variácie gest medzi rôznymi používateľmi a prostrediami.

Aktuálne augmentácie boli zamerané na základné transformácie, ako je šum, mierne posuny a škálovanie. V budúcnosti sa plánuje rozšírenie techník augmentácie o zložitejšie priestorové a časové zmeny. Cieľom je zvýšiť robustnosť modelu a znížiť výskyt chýb pri reálnom použití. Aj napriek týmto obmedzeniam sú výsledky modelu veľmi sľubné a poskytujú pevný základ pre



Obr. 5: Konfúzna matica modelu

ďalší vývoj.

2.8 Použitie v praxi

Riadiaci systém založený na gestách vyvinutý v tomto projekte ponúka široký potenciál pre nasadenie v reálnom svete, najmä v priemyselných prostrediach v súlade s princípmi Industry 4.0. Keďže priemyselné odvetvia sa viac automatizujú a prepájajú, potreba intuitívnej interakcie HRI je čoraz potrebnejšia. Tradičné rozhrania, ako sú fyzické ovládacie panely, diaľkové ovládače alebo zložité programovacie jazyky, môžu byť limitujúce najmä pre netechnických operátorov. Systémy ovládania založené na gestách poskytujú prirodzenú alternatívu, ktorá môže znížiť zložitosť ovládania a zároveň zvýšiť flexibilitu, bezpečnosť a produktivitu pracovníkov.

Priemysel 4.0 kladie dôraz na kybernetické fyzické systémy, výmenu údajov v reálnom čase a inteligentnú automatizáciu. V tejto súvislosti mechatronické systémy ovládané gestami umožňujú ľudským pracovníkom viesť, dohliadať a ovládať automatizované systémy pomocou prirodzenej reči tela. To otvára množstvo prípadov použitia, od asistencie vo výrobe až po vzdialenú správu robotiky v nebezpečných prostrediach.

Jedna možnosť aplikácie je v inteligentných výrobných prostrediach, kde roboty a ľudia koexistujú v priestoroch spolupráce. Pomocou gest môžu operátori zadávať príkazy robotickým ramenám na premiestnenie, spustenie, zastavenie alebo úpravu pohybov bez prerušenia pracovného toku. To môže byť užitočné najmä pri montážnych linkách, kde môže technik potrebovať ovládať viacero robotických ramien v rôznych fázach bez toho, aby potreboval prístup k vyhradenému ovládaciemu terminálu pre každé z nich.

Vezmime si napríklad automobilovú výrobnú linku, kde pracovníci musia niekedy zasahovať do procesov automatizovaného zvárania alebo umiestňovania komponentov. Namiesto zastavenia výrobnej linky alebo prechodu k ovládaciemu panelu môže pracovník zdvihnúť ruku alebo vykonať definované gesto

na dočasné pozastavenie stroja, premiestnenie komponentov alebo zmenu režimov. To vedie k minimalizácii prestojov a lepšej prirodzenej odozve.

Ďalšia aplikácia spočíva v diaľkovo ovládaných priemyselných prostrediach, ako sú elektrárne, bane, ropné plošiny na mori alebo jadrové zariadenia. Tieto prostredia často vyžadujú robotické systémy na kontrolu, manipuláciu alebo údržbu. V takýchto scenároch poskytuje ovládanie gestami ľahký a intuitívny spôsob, ako ľudským operátorom nasmerovať robotické systémy z bezpečnej vzdialenosti, najmä v kombinácii s kamerami alebo rozhraniami rozšírenej reality. Keďže systém možno trénovať tak, aby rozpoznal viacero odlišných príkazov a overil každý z nich prostredníctvom gesta schválenia, minimalizuje sa riziko neúmyselných akcií – čo je v prevádzkových kontextoch veľmi dôležité.

3 Záver

Tento príspevok hovorí o komplexnom prístupe ovládania mechatronických systémov pomocou rozpoznávania ľudských gest, čím sa vytvára intuitívne a prirodzené rozhranie medzi človekom a strojom. Systém bol navrhnutý, implementovaný a otestovaný s cieľom overiť, či je možné pomocou senzorických údajov z Azure Kinect a sekvenčnej analýzy pomocou neurónových sietí typu LSTM dosiahnuť presné, stabilné a spoľahlivé ovládanie robotického ramena v reálnom čase. Jadro systému tvorí robustná dvojúrovňová architektúra rozpoznávania gest – hlavné gesto, ktoré reprezentuje akciu, a potvrdzovacie gesto, ktoré znižuje pravdepodobnosť falošnej aktivácie. Tento dvojfázový model prispieva k vyššej spoľahlivosti celého riešenia, čo je kľúčové najmä pri nasadení v reálnych priemyselných alebo kritických prostrediach. Implementácia bola realizovaná v prostredí PyBullet, ktoré poskytlo fyzikálne realistickú simuláciu pohybu robotického ramena. Vďaka tomu bolo možné bezpečne testovať odozvu systému na jednotlivé rozpoznané gestá bez nutnosti okamžitej práce so skutočným hardvérom. Zvolený súbor gest bol starostlivo navrhnutý tak, aby zohľadňoval faktory ako jednoduchosť, prirodzenosť pohybu, zreteľnosť trajektórie a ľahká rozpoznateľnosť na základe skeletových údajov. Zároveň bola zabezpečená realizovateľnosť každého gesta aj počas dlhodobejšieho používania. Z pohľadu praktického nasadenia má systém obrovský potenciál v oblasti Industry 4.0, kde môže slúžiť ako efektívne rozhranie pre riadenie robotických a mechatronických zariadení v reálnom čase. Vyvinutý systém spája moderné technológie počítačového videnia, hlbokého učenia a robotiky do funkčného riešenia, ktoré môže slúžiť ako základ pre budúce rozšírenie intuitívnych rozhraní medzi človekom a strojom. Výsledky ukazujú potenciál pre ďalšie praktické nasadenie už v reálnom prostredí.

Poďakovanie

Chcel by som sa poďakovať doc. Ing. Otovi Haffnerovi, PhD., za odborné rady pri vývoji systému.

Literatúra

[1] Zhu, Guangming and Zhang, Liang and Shen, Peiyi and Song, Juan. Multimodal Gesture Recognition Using 3-D Convolution and Convolutional LSTM. *IEEE Access*, 2017.

[2] M. Baccouche, F. Mamalet, C. Wolf, C. Garcia, and A. Baskurt. Sequential deep learning for human action recognition. *HBU,* , 2011.

[3] T. -M. Tai, Y. -J. Jhang, Z. -W. Liao, K. -C. Teng and W. -J. Hwang. Sensor-Based Continuous Hand Gesture Recognition by Long Short-Term Memory. *IEEE Sensors Letters*, 2018.

[4] Haffner, Oto and Kučera, Erik and Beňo, Lukáš and Pribiš, Rudolf and Pajpach, Matin and Janecký, Dominik. Gesture Recognition System Based on Motion Capture Suit and LSTM Neural Network. *2023 International Conference on Modeling, Simulation & Intelligent Computing*, 2023.

[5] Roda-Sanchez, Luis and Olivares, Teresa and Garrido-Hidalgo, Celia and Luis de la Vara, José and Fernández-Caballero, Antonio Human-robot interaction in Industry 4.0 based on an Internet of Things real-time gesture control system. 2021

[6] S. Hochreiter and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, 1997.

[7] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *Mach. Learn. Res.*, 2014.

[8] Xihao Wang, Hao Shen, Hui Yu, Jielong Guo, Xian Wei. Hand and Arm Gesture-based Human-Robot Interaction: A Review. 17. Sep. 2022

[9] Christopher E. Mower, Theodoros Stouraitis, João Moura, Christian Rauch, Lei Yan, Nazanin Zamani Behabadi, Michael Gienger, Tom Vercauteren, Christos Bergeles, Sethu Vijayakumar. ROS-PyBullet Interface: A Framework for Reliable Contact Simulation and Human-Robot Interaction. 13. Oct. 2022

Študentská vedecká a odborná činnosť
Fakultná prehliadka prác, 8. apríl 2025
Fakulta elektrotechniky a informatiky STU v Bratislave Sekcia 2. Mechatronika I

Vzdelávacia aplikácia pre počítačové videnie

Dominik Kysel

[1]Ústav vedy a výskumu, FEI STU v Bratislave
[2]FEI STU

xkyseld@stuba.sk

Tento článok sa zameriava na problematiku vzdelávacej aplikácie v oblasti Počítačového videnia (PV), uvediem existujúce riešenia a navrhnem vlastné riešenie.
V prvej kapitole sa analyzujú základné pojmy počítačového videnia a spracovania obrazu a zdôrazňujú sa metódy použité v aplikácii na vizualizáciu rôznorodých filtrov a detekciu hrán a kruhov. Ďalej sa budeme venovať existujúcim riešeniam a uvedieme si ich nedostatky a následne navrhneme vlastné riešenie.
V závere si zhrniem existujúce riešenia a porovnám ich s uvedeným návrhom.

1 Úvod

Počítačové videnie vzniklo, ako reakcia na potrebu počítačov rozumieť a interpretovať vizuálne informácie podobne ako ľudia. Jeho korene siahajú do polovice minulého storočia, keď sa začali objavovať prvé experimenty v oblasti rozpoznávania obrazov. Avšak, až s nástupom výkonnejších výpočtových systémov a algoritmov v posledných desaťročiach sme dosiahli skutočný pokrok v tejto oblasti. Dnes počítačové videnie dosahuje nové hranice. Jednou z hlavných tém, ktorá zaujíma výskumníkov a odborníkov, je Rozšírená realita (AR) a inteligentné systémy detekcie. AR ponúka užívateľom bohatší a interaktívnejší zážitok tým, že virtuálne prvky sú skombinované so skutočným prostredím. Zároveň, inteligentné systémy detekcie umožňujú rýchle a presné rozpoznávanie objektov, čo má široké využitie od priemyslu po zdravotníctvo. Široké využitie v priemyselnej technike, ale neznamená, že sa s technológiami využívajúcimi Počítačové videnie (PV)(z angl. computer vision - CV) nestretávame aj v bežnom živote. Najbežnejším príkladom sú rýchle filtre či už priamo v aplikáciách na fotografovanie bežných smartfónov alebo ako súčasť filtrov v sociálnych sieťach. S rastúcou popularitou používania umelej inteligencie, bola pochopiteľne použitá aj v oblasti PV, čo vyvolalo rozruch, nakoľko niektoré spracovávané obrazy a videá boli získané bez autorských práv a na základe nich boli generované nové diela, ktoré boli následne distribuované bez súhlasu alebo kompenzácie autorov pôvodných diel.

Každopádne je jasné, že počítačové videnie je dôležitou súčasťou našich životov a bude hrať ešte dôležitejšiu úlohu v budúcnosti, a preto je dôležité, aby sme rozumeli jeho princípom a fungovaniu.

S ohľadom na súčasné trendy a inovácie, je jasné, že počítačové videnie bude hrať ešte dôležitejšiu úlohu v našom každodennom živote. Očakáva sa, že sa bude ďalej rozvíjať v oblastiach ako umelá inteligencia, rozšírená realita a presnejšie rozpoznávanie objektov.

Na trhu existuje množstvo aplikácií, ktoré demonštrujú základné funkcie počítačového videnia. Ich častým problémom je ale to, že často obsahujú iba jednu alebo dve funkcie, čo znamená, že pre demonštráciu viacerých funkcií je potrebné mať nainštalovaných viacero aplikácií, často z rôznych zdrojov a od rôznych vývojárov. Toto môže byť nepraktické a neefektívne. Ďalšou nevýhodou je, že tieto aplikácie často nemajú voľne prístupný zdrojový kód, čo znamená, že nie je možné prispievať k ich rozvoju a rozširovaniu a taktiež nie je možné zistiť, ako fungujú a čo presne robia.

Cieľom tejto práce je priblížiť problematiku vzdelávacích aplikácii v oblasti počítačového videnia, ukázať existujúce riešenia a načrtnúť návrh riešenia pre lepšie spracovanie daných aplikácii.

2 Spracovanie obrazu a počítačové videnie

Spracovanie obrazu je kľúčovou súčasťou mnohých technologických a vedeckých aplikácií. Jeho cieľom je zlepšiť vizuálnu informáciu pre ľudské vnímanie alebo poskytnúť efektívne kódovanie pre úložné a prenosové účely. V tejto časti sa zameriame na základné koncepty a operácie, ktoré sú základom pre pokročilejšie techniky spracovania obrazu, konkrétne si uvedieme filtre a detektory hrán [1] [2].

Digitálny obraz je reprezentácia vizuálnej informácie vo forme mriežky pixelov, kde každý pixel obsahuje hodnotu intenzity farebného kanála, respektíve hodnoty farebných kanálov. Základné vlastnosti digitálneho obrazu zahŕňajú rozlíšenie, ktoré určuje počet pixelov v obraze a bitovú hĺbku, ktorá určuje počet bitov použitých na reprezentáciu intenzity každého pixelu [3][4].

2.1 Obraz

Obraz môžeme vnímať ako vnem zachytený na sietnici ľudského oka alebo obraz spracovaný kamerou. Obraz vieme zadefinovať aj matematicky, a to ako funkciu f dvoch $f(x,y)$, pre statický obraz, alebo troch $f(x,y,z)$ premenných, pre dynamické alebo trojrozmerné obrazy. Hodnoty týchto funkcií zodpovedajú určitým fyzikálnym vlastnostiam, ako je jas, teplota

alebo absorpcia žiarenia. Pre dvojrozmernú obrazovú funkciu sa používa aj termín intenzitný obraz [5].

Pixel (z anglického "picture element") je časť digitálne spracovaného obrazu, často označovaná ako bod, ktorá v matici reprezentuje jeden prvok. Je najmenšou nedeliteľnou časťou digitálne spracovaného obrazu. Pixel predstavuje množstvo svetelnej energie v danom bode [5].

V rámci digitalizácie obrazu sa vykonáva vzorkovanie a kvantovanie spojitého obrazu. Vzorkovaním sa spojitý obraz rozdelí na diskrétne body, ktoré sú usporiadané do mriežky, čo znamená, že digitálny obraz je dátová štruktúra, ktorou je zvyčajne matica. Prakticky sa dá povedať, že najmenšou jednotkou v rastri je pixel, pričom raster má oproti mriežke aj zadefinované susedné vzťahy medzi jednotlivými bodmi. Vzorkovanie sa vykonáva do mriežky, ktorej veľkosť je definovaná podľa kvality digitálneho obrazu, ktorú chceme dosiahnuť. Kvantovanie obrazu zabezpečuje, že jas alebo funkcia jasu nadobúda v digitalizovanom obraze celočíselné hodnoty. Kvantovanie je prechod hodnôt spojitej obrazovej funkcie a ich digitálnych ekvivalentov. Počet kvantizačných úrovní, respektíve úrovní jasu, musí byť dostatočne vysoký, aby sa predišlo vzniku falošných obrysov, ktoré vznikajú, ak je počet úrovní nižší, ako je počet, ktorý dokáže ľudské oko rozlíšiť. Počet úrovní jasu je daný počtom b bitov použitých na určenie hodnôt jasu pixela a to vzťahom $k = 2^b$, pričom bežná hodnota je 8 bitov na pixel pre každý kanál, čiže pre červenú, zelenú a modrú farbu. Samozrejme, existujú systémy, ktoré používajú iný počet bitov, napríklad 16 [5].

2.2 Konvolúcia

Konvolúcia sa v rámci spracovania obrazu využíva vo forme konvolučných jadier, ktoré môžeme chápať ako masku, ktorá sa postupne aplikuje na jednotlivé body obrazu. Konvolúcia je matematická operácia, kde výstupom je súčet vstupných hodnôt prenásobených váhami, ktoré sú určené ako koeficienty konvolučného jadra. Môže nastať aj situácia, kedy konvolučné jadro má nejakú časť mimo hraníc obrázka. V tomto prípade sú hodnoty prenásobené koeficientami jadra doplnené buď nulami, alebo krajnými hodnotami z najbližšieho okraja obrazu [6]. Lineárna konvolúcia je matematicky popísaná ako

$$g(x,y) = \sum_{(m,n)\in O} \sum h(x-m, y-n) f(m,n) \tag{1}$$

kde g je výstupný obraz, h je konvolučná maska a O je okolie [5].

2.3 Filtre

Filtre v PV sú nástroje, ktoré sa používajú na zlepšenie kvality alebo vzhľadu obrazu, prípadne aj ako predspracovanie pre ďalšie účely. Filtre fungujú na princípe aplikovania matematických operácií na jednotlivé pixely obrazu. V praxi sa využíva množstvo filtrov, ktoré sa líšia svojou funkciou a výsledkom. V tejto práci sa budeme zaoberať niekoľkými základnými filtrami, ktoré sú použité v aplikácii.

V rámci tejto práce si uvedieme niekoľko filtrov, ktoré sú použité v aplikácii. Spomenuté bude ich fungovanie, implementácia, využitie a výsledok ich použitia. Pri výbere a aplikácii nižšie uvedených filtrov je nevyhnutné zvážiť ich vplyv na obraz a nastaviť ich parametre tak, aby boli dosiahnuté požadované výsledky, či už ide o zvýraznenie textúry, zlepšenie kontrastu, alebo iné úpravy obrazu [7]. V rámci tejto práce sa bude venovať iba nasledujúcim filtrom:

- **Gaussov filter**
- **Mediánový filter**
- **Bilaterálny filter**

2.3.1 Gaussov filter

Gaussov filter, pomenovaný po slávnom matematikovi Johannovi Carl Friedrich Gauss, je základným nástrojom pre rozmazanie obrazu a redukciu šumu, využívajúc konvolúciu s Gaussovým jadrom. Patrí medzi lineárne vyhladzovacie filtre. Jadro Gaussovho filtra je nastavené tak, aby výstup bol vážený priemer okolia jednotlivého pixelu, pričom hodnoty koeficientov sú vyššie k stredu jadra, čím je zabezpečené jemnejšie rozmazanie pri odstránení šumu. Gaussov filter je daný vzťahom:

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{2}$$

Kde G(x, y) predstavuje hodnotu Gaussovho filtra na pozícii x,y a σ predstavuje intenzitu rozmazania, pričom vyššia hodnota σ vyžaduje väčšie konvolučné jadro, aby bolo dostatočne presne reprezentované [8].

2.3.2 Mediánový filter

Mediánový filter, na druhej strane, je výnimočný svojou schopnosťou odstrániť tzv. ‘salt-and-pepper’ šum bez toho, aby došlo k rozmazaniu dôležitých hrán obrazu, čo je kľúčové pre zachovanie detailov. Filter funguje na základe výberu vhodnej veľkosti masky, najčastejšie 3x3 alebo 5x5, ktorá postupne prechádza obrazom. Maska obsahuje hodnoty intenzity jednotlivých pixelov, ktoré sa následne zoradia od najmenšej po najväčšiu a vyberie sa stredná hodnota. Bod, ktorého hodnota sa aplikáciou tohto algoritmu upraví, sa nachádza v strede masky [9]. Konvolučné jadro veľkosti 3x3 tohto filtra je [5]

$$h = \frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \tag{3}$$

2.3.3 Bilaterálny filter

Bilaterálny filter rozširuje koncept priestorového váhovania zavedením intenzitného váhovania, čím umožňuje selektívne rozmazanie obrazu s ohľadom na zachovanie hrán. Základom tohto filtra je už spomínaný Gaussov filter. Matematicky je tento filter reprezentovaný ako:

$$I^B(x) = \frac{\sum_{y\in N}(g_{\sigma S}(x,y) g_{\sigma R}(I(x), I(y)) I(y))}{\sum_{y\in N} g_{\sigma S}(x,y) g_{\sigma R}(I(x), I(y))} \tag{4}$$

Kde $I^B(x)$ je hodnota filtra na pozícii x. $I(x)$ určuje intenzitu pixelu. N je okolím pixelu, na ktorý je filter práve aplikovaný. Ďalej, $g_{\sigma S}$ a $g_{\sigma R}$ sú jadrá Gaussovho filtra a určujú mieru vyhladenia a zachovania hrán, a to tak, že ak $g_{\sigma R}$ je veľké, aplikuje sa práve Gaussov filter, a ak je naopak nízke, tak sa filter neaplikuje vôbec [10].

2.4 Detekcia hrán

Detekcia hrán je kritickým prvkom v oblasti počítačového videnia, umožňujúcim identifikáciu hraníc objektov v obraze. Táto technika je často využívaná na získanie dôležitých informácií o štruktúre a tvare objektov v obraze. Hrany predstavujú zmeny intenzity pixelov a môžu byť identifikované pomocou rôznych algoritmov. V rámci spracovania obrazu sa detekcia hrán často využíva na segmentáciu obrazu, čo umožňuje lepšie rozlíšiť objekty od pozadia. Medzi populárne metódy detekcie hrán patrí Cannyho detektor, Sobelov operátor, ktorý sme si uviedli už v rámci filtrov, alebo Prewittov operátor. Detekcia hrán má široké využitie v rôznych odvetviach počítačového videnia, vrátane rozpoznávania objektov, medicínskeho obrazovania, autonómnych vozidiel a mnohých ďalších aplikácií [11] [12].

V rámci tejto práce sa budeme zaoberať niekoľkými metódami detekcie hrán, ktoré sú použité v aplikácii. Spomenuté bude ich fungovanie, implementácia, využitie a výsledok ich použitia.

Použité metódy detekcie hrán:

- **Sobelov detektor**
- **Cannyho detektor**
- **Houghova transformácia na detekciu čiar**
- **Houghova probabilistická transformácia**
- **Houghova transformácia na detekciu kruhov**

2.4.1 Sobelov detektor

Sobelov filter, ktorý je výsledkom práce Irwina Sobela a Garyho M. Feldmana, je ďalším kľúčovým nástrojom používaným na detekciu hrán. Využíva dve 3x3 jadrá na výpočet aproximácií derivátov obrazu, čím zvýrazňuje hrany a pomáha identifikovať dôležité štruktúry v obraze. Samotný detektor má smer a teda, pri využití Sobelovho filtra ako detektora, je potrebné využiť filter v smere osi x a filter v smere osi y. Konvolučné jadro tohto filtra pre osu x je následovné:

$$G_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \tag{5}$$

a v smere osi y je definovaný ako:

$$G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \tag{6}$$

[13]

2.4.2 Cannyho detektor

Cannyho detektor je populárny algoritmus detekcie hrán, ktorý kombinuje viacero krokov, vrátane aplikácie Gaussovho filtra, výpočtu gradientu a potlačenia nevýznamných hrán. Aplikovaný Gaussov filter slúži na potlačenie šumu, následne je použitý Sobelov filter na detekciu vertikálnych a horizontálnych hrán. Cieľom je z možných hrubých hraníc vytvorenie ostrých hraníc hrúbky 1 pixel. V gradiente sú zachované len lokálne maximá. Následne je aplikované dvojité prahovanie, ktoré určuje potenciálne hranice oblastí, ktoré budú následne spojené [9].

2.4.3 Houghova transformácia

Houghova transformácia je metóda, ktorá sa využíva na detekciu tvarov v obraze, ktoré je možné parametricky opísať. Pôvodná metóda bola vyvinutá na detekciu priamok v obraze v roku 1962 pánom Paulom Houghom. Obraz musí byť najprv spracovaný hranovým filtrom, keďže body, ktoré budeme opisovať v tzv. Houghovom priestore ležia práve na týchto hranách. V Houghovom priestore sú jednotlivé body obrázka reprezentované sinusoidou. Pre využitie Houghovej transformácie na detekciu čiar je daná rovnica:

$$r = x\cos\theta + y\sin\theta \tag{7}$$

kde x a y reprezentujú súradnice bodu obrázka a r a θ sú neznáme, pričom r je vzdialenosť normály od počiatku súradnicového systému a θ je uhol medzi normálou a kladnou časťou x-ovej osi. Hodnoty r a θ sú následne naplnené do akumulátora, ktorý si môžeme predstaviť ako maticu:

$$\begin{bmatrix} r_{min}, \theta_{min} & \dots & r_{min}, \theta_{max} \\ \dots & \dots & \dots \\ r_{max}, \theta_{min} & \dots & r_{max}, \theta_{max} \end{bmatrix} \tag{8}$$

a následne extrémy hodnôt akumulátora reprezentujú parametre hľadanej priamky, respektíve priamok. Všeobecná verzia Houghovej transformácie sa dá použiť na detekciu všeobecných parametrických kriviek v obraze. Táto forma Houghovej transformácie je v našom prípade použitá pre detekciu kruhov. Je daná rovnicou:

$$y - y_0 = a(x - x_0)^2 \tag{9}$$

V tomto prípade vidíme, že premenné máme 3, a to y_0, x_0 a a. To znamená, že akumulátor je 3-rozmerný, a teda sa výrazne zvýši náročnosť algoritmu. V praxi sa teda využíva obmedzenie parametru a, čo znižuje výpočtovú náročnosť.

Houghova probabilistická transformácia je variantom Houghovej transformácie, ktorý namiesto hlasovania pre všetky potenciálne priamky v obraze vyberá náhodné podmnožiny bodov a hlasuje len pre priamky, ktoré prechádzajú týmito bodmi. Týmto spôsobom je možné detegovať krivky na kratších úsekoch a zároveň je efektívnejší, pretože nevyžaduje výpočet Houghovej transformácie pre všetky body v obraze. Keďže sa ale jedná o náhodný výber a užší rozsah môže nejaké priamky/krivky "prehliadnuť"[9] [14].

3 Existujúce riešenia

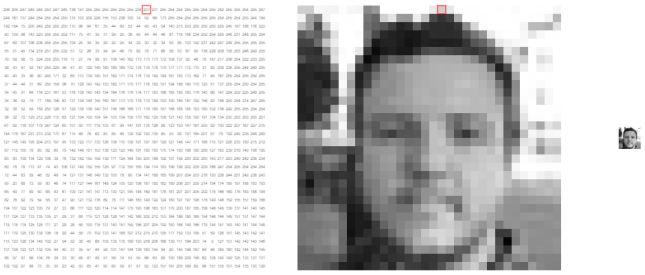
Ako už bolo spomenuté v úvode, na trhu existujú riešenia poskytujúce demonštráciu použitia techník spracovania obrazu v oblasti počítačového videnia. Množstvo z nich sú však iba výsledky aplikácie filtrov, detektorov hrán, prípadne aplikácie iných algoritmov v oblasti spracovania obrazu, ale tými sa v tejto práci zaoberať nebudeme. Keďže sa jedná iba o výsledok aplikácie, používateľ, ktorým by bol napríklad študent, s daným príkladom nemôže nijako interagovať mimo použitia iných aplikácií, prípadne vytvorenia vlastnej.

Ďalším typom príkladov sú interaktívne aplikácie, ktoré ponúkajú vyskúšanie modifikácie aplikovaných algoritmov, sledovanie zmien a to bez použitia aplikácií tretích strán a bez nutnosti vytvorenia vlastného riešenia. Toto by bolo aj cieľom nášho návrhu. Problémom ale pri týchto interaktívnych riešeniach býva, že sú často zamerané iba na jednu konkrétnu úlohu, prípadne ponúkajú iba jeden typ modifikácie. Zároveň je ich zdrojový kód, ktorý by mohol tiež pomôcť pri pochopení problematiky spracovania obrazu, často neprístupný.

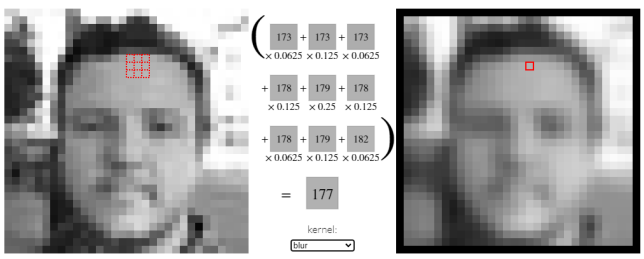
3.1 Konkrétny príklad

Ako príklad si uvedieme demonštračné riešenie aplikovania konvolučných jadier na obrázok spracované pánom Victorom Powellom [15].

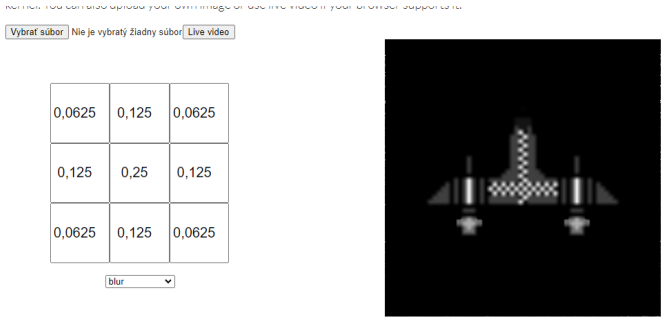
V tomto riešení je uvedených pár demonštračných príkladov, ktorými sú aplikovanie jadra na používateľom zvolený obrázok (obrázok 3), náhľad na hodnoty pixelov preddefinovaného obrázka (obrázok 1) a náhľad na hodnoty obrázka po aplikácii konvolučného jadra (obrázok 2)



Obr. 1: Náhľad na hodnoty pixelov preddefinovaného obrázka [15]



Obr. 2: Náhľad na hodnoty pixelov preddefinovaného obrázka po aplikácii jadra [15]



Obr. 3: Aplikovanie jadra na používateľom zvolený obrázok [15]

V týchto príkladoch vidíme, ako je aplikované konvolučné jadro na obrázok, a to pomocou náhľadu na hodnoty pixelov na obrázku 1 a následne na hodnoty pixelov na obrázku 2. Následne nám je ponúknutá aj možnosť aplikácie jedného z preddefinovaných jadier, prípadne doplnenie vlastných hodnôt, a jeho aplikácie na buď preddefinovaný alebo nami vybraný obrázok 3. Problémom je, že pri aplikácii na vlastný obrázok už nemáme možnosť sledovať meniace sa hodnoty obrázku a zároveň náhľad do zdrojového kódu, síce je možný, ale bez manuálu k danému zdrojovému kódu nie je jasné jeho fungovanie.

4 Riešenie

V predošlej kapitole boli spomenuté nedostatky existujúcich riešení a teda obsahom tejto kapitoly bude návrh riešenia, ktoré by sa dané nedostatky vyriešilo. Pre návrh aplikácie je potrebné mať jasne určené požiadavky, ktoré budú obsahovať zvolené technológie, prostredie alebo podporné materiály. Po zadefinovaní požiadavok sa môžu vytvoriť prípady použitia, ktoré určia funkčnosť jednotlivých častí aplikácie a testovacie prípady, ktoré danú funkčnosť overia.

4.1 Softvérové požiadavky

V tejto časti si uvedieme požiadavky, podľa ktorých by boli následne vybrané technológie, v ktorých by sa aplikácia vyvinula.

Softvérové požiadavky teda sú:

- **Programovací jazyk** - programovací jazyk v ktorom je práca vyvíjaná musí mať dostatok materiálov pre efektívny vývoj a zároveň aj dostatočnú podporu knižníc pre oblasť spracovania obrazu.
- **Knižnica pre spracovanie obrazu/počítačové videnie** - vybraná knižnica musí obsahovať dostatočnú škálu algoritmov v oblasti počítačového videnia a zároveň dostatočnú dokumentáciu, pre efektívny vývoj a podporu pre zaručenie dlho trvácnosti tejto práce.
- **Knižnica pre grafické užívateľské rozhranie** - vybraná knižnica musí obsahovať dostatok nástrojov pre vývoj intuitívneho používateľského prostredia

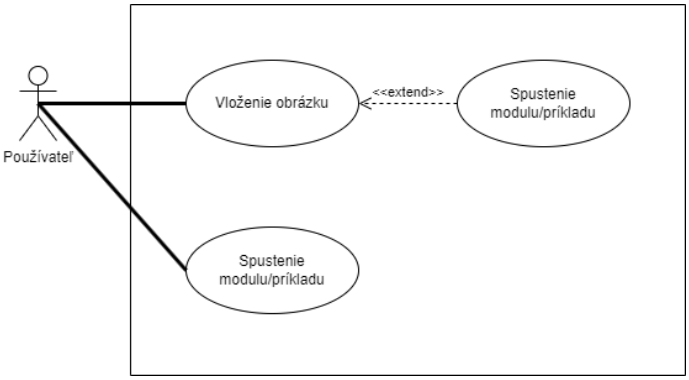
- **Závislosti a doplnkové knižnice** - závislosti a knižnice ktoré sú potrebné pre správne fungovanie vyššie uvedených knižníc

- **Manuál** - vysvetlenie funkcionality a použitia aplikácie.

4.2 Prípady použitia

Prípady použitia slúžia na opis správania systému a teda popisujú očakávania používateľov od systému. Tým sa popíšu aj požiadavky funkčnosti jednotlivých častí, čo pomôže vývojárovi pri efektívnom vývoji.[16]

Ako príklad uvediem následovný prípad použitia.



Obr. 4: Prípad použitia

4.3 Testovacie prípady

Testovacie prípady slúžia na overenie správnosti produktu, v tomto prípade aplikácie. Testujú funkčnosť aplikácie ale aj jej vzhľad. Testovacie prípady sú často spájané s prípadmi použitia, pričom pre jeden prípad použitia môže byť viacero testovacích prípadov a to preto, lebo každý testovací prípad sa zaoberá špecifickými vstupmi. Testovacie prípady zabezpečia efektívne overenie správneho spracovania a funkčnosti aplikácie. Na vytváranie testovacích scenárov, teda súborov testovacích prípadov, je možné použiť množstvo aplikácii, prakticky je ale vhodné všetko čo umožňuje štrukturovaný a prehľadný zápis kľúčových informácii, ako názov testovacieho prípadu, jeho popis, kroky testovacieho prípadu, očakávaný výsledok a reálny výsledok. Príklad ako by mohol vyzerať testovací prípad je nižšie.

ID	1	**Názov**	Načítanie obrázka v hlavnom okne		
Prípad použitia		UC1	**Úroveň splnenia testu**	Musí – ~~Mal by~~ – ~~Mohol by~~	
Rozhranie		používateľ/aplikácia			
Vstupné podmienky					
Výstupné podmienky					
Krok	**Akcia**		**Očakávaná reakcia**	**Skutočná reakcia**	
1	Používateľ klikne na tlačidlo "Výberte obrázok"		Otvorí sa okno v ktorom môže používateľ zvoliť obrázok	Okno sa otvorilo a bolo možné zvoliť obrázok	
2	Zobrazí sa možnosť výberu obrázka a používateľ vyberie ľubovoľný obrázok formátu .jpg/.png/.jpeg		Obrázok sa nahrá do aplikácie a v hlavnom okne sa zobrazí jeho názov	Obrázok sa načítal a v aplikácia sa zobrazil jeho názov	

Obr. 5: Testovací prípad

4.4 Návrh

Z toho čo je uvedené vyššie v tejto kapitole by som ako riešenie navrhol aplikáciu, skladajúcu sa zo samostatných modulov, ktoré by bolo možné spúšťať nezávislo od seba. Jednalo by sa teda o jednoduché hlavné rozhranie, ktoré by slúžilo na spúšťanie samostatných modulov, ktoré by boli sústredené na konkrétne úlohy. Toto by zabezpečilo jednoduché integrovanie nových modulov, čo by bolo možné aj vďaka dôkladnej dokumentácii. Pre aplikáciu by bolo rozumné zvoliť rozsiahlu knižnicu v oblasti počítačového videnia, ako napríklad OpenCV, ktorá je dostupná v jazykoch ako Python, C++ a Java. Ako programovací jazyk by som zvolil Python, kvôli jednoduchosti práce v danom jazyku. Pre podporné knižnice by bolo potrebné zvoliť také, ktoré by obsahovali všetky nástroje pre výpočty, ako napríklad NumPy, knižnica Pillow pre načítanie a prácu s obrázkami. Knižnica TKInter by mohol byť zvolený na vytvorenie grafického používateľského rozhrania.

5 Záver

V tejto práci sme sa venovali priblíženiu problematiky aplikácie na demonštráciu filtrov a detektorov v oblasti strojového videnia a spracovania obrazu, pričom boli uvedené aj poznatky potrebné na porozumenie jednotlivých častí, ktoré aplikácia má demonštrovať. Ďalej boli uvedené aj existujúce riešenia, kde sme spomenuli, že veľa z nich má problémy v oblasti rozsahu, čiže malé množstvo príkladov, ktoré skôr pôsobia ako statické obrázky než ako interaktívne demonštrácie, problémy v oblasti jednoduchosti pochopenia, a to kvôli chýbajúcej dokumentácií alebo nemožnosti sa dostať k funkčnosti algoritmov. Nakoniec bolo uvedené aké rozhodnutia je potrebné spraviť v rámci softvérových požiadaviek, uviedlo sa čo sú prípady použitia a čo sú testovacie prípady, a navrhlo sa riešenie formou hrubého návrhu aplikácie.

Referencie

1. RUSS, John C. *Computer-assisted microscopy : the measurement and analysis of images*. Computer-assisted microscopy : the measurement and analysis of images. New York: Plenum Press, 1990. ISBN 0306434105.

2. K. SUMITHRA S. Buvana, R. Somasundaram. A Survey on Various Types of Image Processing Technique. *INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT)* [online]. 2015 [cit. 2024-05-05]. ISSN 2278-0181. Dostupné z : https://www.ijert.org/research/a-survey-on-various-types-of-image-processing-technique-IJERTV4IS030552.pdf.
3. MURÁŇ, Juraj. Počítačové videnie v kocke III. - Spracovanie videozáznamu: Spracovanie videozáznamu. *Umelá inteligencia* [online]. 2019 [cit. 2024-05-05]. ISSN 2232-332X. Dostupné z : https://umelainteligencia.sk/pocitacove-videnie-v-kocke-iii-spracovanie-videozaznamu/.
4. MARČAN, Peter. Úvod do strojového videnia (Machine Vision). *Daily Automation* [online]. 2024 [cit. 2024-05-05]. ISSN 2453-8175. Dostupné z : https://www.dailyautomation.sk/01-strojove-videnie-uvod.
5. HAFFNER, Oto. *Strojové videnie pre Industry 4.0*. (nepublikovaná pracovná verzia). [B.r.].
6. DAVIES, E Roy. *Machine Vision: Theory, Algorithms and Practicalities*. Third. Elsevier, 2004. ISBN 0-12-206093-8.
7. KOVÁČ, Ondrej (zost.). *Filtrácia obrazu - Číslicové spracovanie obrazov - KTE Confluence: Filtrácia obrazu* [online]. Fakulta Elektrotechniky a Informatiky, Technická Univerzita v Košiciach, 2019-10-25 [cit. 2024-05-05]. Dostupné z : https://kte-server.fei.tuke.sk/confluence/pages/viewpage.action?pageId=10059893.
8. LINDA G. SHAPIRO, George C. Stockman. *Computer Vision*. Prentice Hall, 2001. ISBN 9780130307965.
9. JARMILA PAVLOVIČOVÁ, Marek Loderer. *Číslicové spracovanie obrazu*. First. FELIA s.r.o., 2016. ISBN 978-80-89824-04-5.
10. AURICH V., WEULE J. *Non-Linear Gaussian Filters Performing Edge Preserving Diffusion*. First. Springer Berlin Heidelberg, 1995. ISBN 978-3-642-79980-8.
11. JAIN, Ramesh; KASTURI, Rangachar; SCHUNCK, Brian. *Machine Vision*. 1995. ISBN 978-0-07-032018-5.
12. HAFFNER, Oto. Detekcia hrán v obraze: Detekcia hrán v obraze. *posterus* [online]. 2011 [cit. 2024-05-05]. ISSN 1338-0087. Dostupné z : https://www.posterus.sk/?p=11234.
13. SOBEL, Irwin. An Isotropic 3x3 Image Gradient Operator. *Presentation at Stanford A.I. Project 1968*. 2014.
14. DAVIES, E Roy. *Computer and machine vision : theory, algorithms, practicalities*. Fourth. Elsevier, 2012. ISBN 9780123869913.
15. POWELL, Victor (zost.). *Image Kernels: Explained Visually* [online]. Setosa, 2015-01-29 [cit. 2024-05-08]. Dostupné z : https://setosa.io/ev/image-kernels/.
16. COCKBURN, Alistair. *Use Cases : jak efektivně modelovat aplikace*. Vyd. 1. CP Books, Brno, 2005. ISBN 9788025107218.

Využitie výpočtovej inteligencie pre úlohy kontroly kvality

Eunika Farkašová

Ústav automobilovej mechatroniky, FEI STU v Bratislave

xfarkasovae@stuba.sk

Abstrakt – Táto práca sa venuje využitiu výpočtovej inteligencie, konkrétne systémov strojového videnia, v oblasti kontroly kvality vo výrobných procesoch. Detailne analyzuje technológie, komponenty a typy úloh, ktoré strojové videnie dokáže riešiť, ako je detekcia objektov, chýb, meranie objektov či rozpoznávanie znakov a vzorov. Praktická časť práce predstavuje aplikáciu smart kamery Cognex In-Sight 2800, ktorá bola použitá na riešenie problému manuálneho rozpoznávania sortimentu v predajni, čím sa eliminovala ľudská práca a chybovosť. Výsledkom je automatizovaný systém, ktorý využíva detekciu vzorov a jasové hodnoty objektov pre spoľahlivú identifikáciu a počítanie produktov, doplnený intuitívnym používateľským rozhraním.

1 Úvod

Implementácia automatizácie do výrobných procesov priniesla so sebou veľké zmeny. Robustné a komplikované procesy museli byť rozdelené na menšie a jednoduchšie inštrukcie, ktoré dokážu byť rýchlo a spoľahlivo vykonané strojom. Po úprave procesov vznikla potreba zaviesť do automatizovaných procesov kontrolu, ktorú najskôr vykonávali ľudia. Práve takáto forma kontroly poukázala na obmedzenia a problémy spojené s manuálnou kontrolou, čo zosilnilo potrebu zavedenia strojového videnia aj do tejto oblasti. Systémy so strojovým videním sú považované za inteligentné riešenia, ktoré sú schopné riešiť aj náročnejšie úlohy merania a kontroly [2].

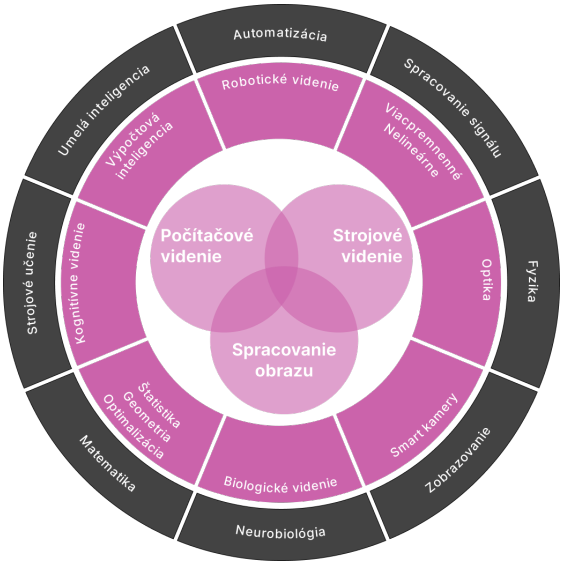
V dôsledku rastúcich štandardov kvality vo výrobe, značným výhodám v porovnaní s ľudským videním a manuálnej obsluhe sa strojové videnie zaraďuje medzi hlavné technológie v oblasti automatizácie[1] .

2 Strojové videnie

Strojové videnie je technológia, ktorá umožňuje strojom snímať, rozpoznávať a lokalizovať objekty. Dokáže zachytávať ich obrazy, ukladať informácie o ich materiáli, orientácii alebo štruktúre a vykonávať analýzu dát [1].

Cieľom strojové učenia je poskytnúť strojom schopnosti videnia, ktoré sú podobné ľudským vlastnostiam, pričom využívajú algoritmy pre spracovanie obrazu. V priemysle sa strojové videnie často označuje aj ako "priemyselné videnie"[1], [3].

Pri strojovom videní sa spájajú viaceré oblasti za účelom automatizácie úloh, zvýšenie efektivity vo výrobe a vizuálnej kontroly, viď. Obrázok 1. Pri spracovávaní informácií sa aplikujú techniky spracovania obrazu, pričom sa nejedná o tú istú vednú oblasť. V praxi často dochádza k zámene týchto pojmom. Ďalším pojem, s ktorým je strojové videnie často zamieňané je počítačové videnie.



Obr. 1: Prienik oblastí: strojové videnie, počítačové videnie a spracovanie obrazu [3]

V nižšie uvedenej Tabuľke 1 sú definované jasné rozdieli medzi strojovým videním, spracovaním obrazu a počítačovým videním [3].

2.1 Komponenty strojového videnia

Od vizuálneho systému strojového videnia sa očakáva, že dokáže spoľahlivo vykonávať operácie, procesy ako je získavanie a analýza obrazu, detegovanie prvkov alebo objektov v rámci nasnímaného obrazu. Na vykonávanie týchto operácií sú však potrebné nasledovné komponenty: zdroj svetla, objektív, snímače obrazu, vizuálne spracovanie - kamera a komunikácia. Na zobrazenie výstupu sa často používajú aj doplnkové komponenty ako je monitor alebo dotyková obrazovka [4].

Všetky uvedené komponenty majú svoje vlastné funkcie ale musia navzájom spolupracovať aby mohli byť využité v procesoch strojového videnia [3].

Procesy sa začínajú kontrolou výrobku, následne sa aktivuje získavanie obrazu ak je výrobok na správnom mieste.

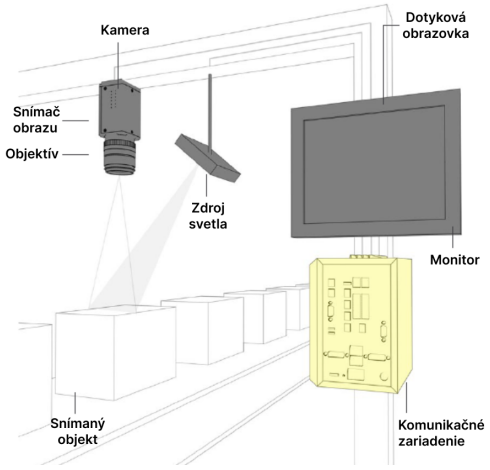
	Vstup	Výstup
Spracovanie obrazu	Obraz je spracovávaný pomocou algoritmov na úpravu alebo vylepšenie. Výsledkom je nová a lepšia verzia pôvodného obrazu.	Je vrátený upravený obraz.
Počítačové videnie	Obraz je analyzovaný pomocou algoritmov v prostrediach, ktoré sú nepredvídateľné alebo neovládateľné.	Porozumenie obrazu, predikcia a učenie sa pre ďalšie získanie informácií ako je napríklad segmentácia, rozpoznávanie, rekonštrukcia.
Strojové videnie	V priemyselnom prostredí za predvídateľných okolností sa na analýzu snímok používajú kamery alebo videokamery.	Porozumenie obrazu a učenie sa na získavanie informácií o výrobných procesoch a kvalite.

Tab. 1: Rozdiely medzi strojovým videním, spracovaním obrazu a počítačovým videním [3]

Výberom vhodného zdroju svetla sa osvetlí snímaný objekt, čo zabezpečí zvýšenú kvalitu obrazu a svetelných podmienok pre jeho zachytenie objektívom. Objektív následne obraz zaostrí a pretransformuje do formy svetla pre snímač. Snímač svetlo preloží do formy digitálneho obrazu a podľa dostupných možností zrealizuje predspracovanie obrazu čím zabezpečí, že vyniknú správne obrazové prvky. Ďalším krokom je vizuálne spracovanie, počas ktorého sa vykoná úpravu obrazu pre potreby analýzy. Z obrazu sa získajú požadované informácie. V poslednom kroku sa spúšťajú automatické udalosti, ako je napríklad prijatie alebo zamietnutie výrobku, pri ktorých sa využívajú komunikačné zariadenia. Tie využívajú diskrétne vstupno-výstupné signály, ktoré boli odoslané sériovým pripojeným do zariadenia, ktoré informovanie môže používať alebo ich zaznamenávať [3], [5], [7].

2.2 Typy úloh strojového videnia

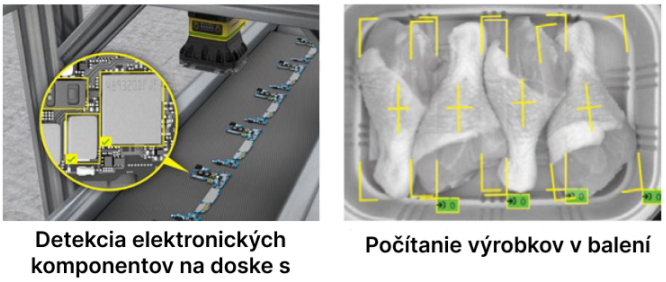
V priemyselnom sektore sú často systémy strojového videnia špecializované a navrhnuté pre plnenie jedinečných úloh, ktoré sa v závislosti od pracovných prostredí od seba výrazne odlišujú. Nie je možné definovať jednu všeobecnú kategóriu, ktorá by sa dala využiť pre rôzne typy úloh, nakoľko aj pri manuálnej práci sú využívané iné pracovné postupy pri meraní rozmerov výrobku a iné pri kontrole chýb na výrobkoch. Z tohto dôvodu sú zadefinované rôzne typy samostatných kategórii, ktoré sú špecializované a prispôsobené na vykonávanie určitých typov úloh [8].



Obr. 2: Komponenty strojového videnia [5]

Medzi hlavné kategórie patrí [8], [9], [10]:

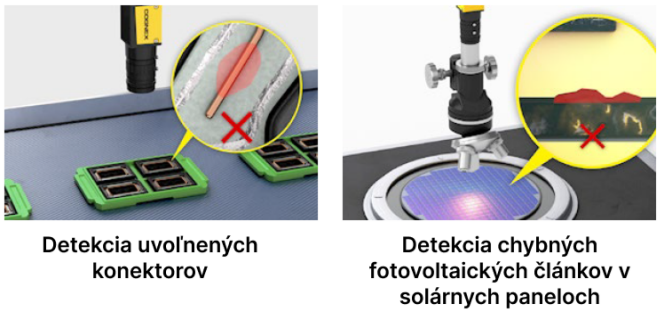
Detekcia objektov - Pri detekcii objektov sa systémy špecializujú na identifikáciu objektov podľa definovaných charakteristických vlastností ako je tvar, rozmer farba a textúra. Na základe týchto vlastností sa systém dokáže zamerať len na relevantné objekty, aj keď sa v jeho zornom poli nachádzajú aj iné typy predmetov. Následne vyhodnotí prítomnosť objektu v zornom poli kamery. Rozpoznanie obrazu sa radí medzi najpoužívanejšie kategórie nakoľko sa využíva aj pri počítaní objektov, triedení výrobkov alebo kontrole kvality. Náročnosť úloh detekcie objektov závisí od definovaných charakteristických vlastností objektu.



Obr. 3: Príklady detekcie objektu [5]

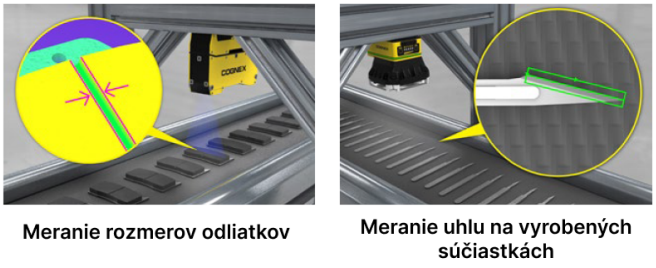
Detekcia chýb - Chyby môžu vzniknúť takmer v každom kroku výrobného procesu. Môže sa jednať o chybu súvisiacu s umiestnením komponentu alebo mechanického poškodenia. Aplikáciou tejto kategórie je možné rýchlo odhaliť vzniknuté chyby na veľkých ale aj malých vzorkách. Na správnu detekciu chýb je potrebné najskôr systém naučiť, ako vyzerá bezchybný objekt. Následne systém nasnímaný objekt analyzuje a porovná ho s bezchybným objektom. Ak sú medzi nimi odlišnosti, systém identifikuje rozdielnosť ako chybu.

Meranie objektov - Na bezkontaktné meranie sa využívajú nástroje na analýzu rozmerov snímaného objektu a presné výpočty geometrických rozmerov. Presnosť takéhoto merania je



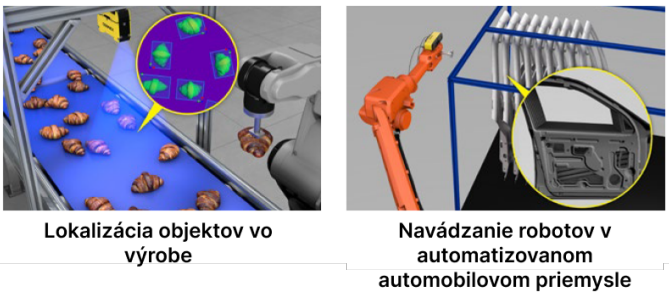
Obr. 4: Príklady detekcie chyby[5]

na úrovni mikrónov aj keď sa objekty vo výrobných linkách pohybujú vysokou rýchlosťou. Strojové videnie nám umožňuje odmerať dĺžku objektu, jeho plochu či objem, presné meranie vzdialeností, priemerov a mnoho ďalších parametrov.



Obr. 5: Príklady merania objektov[5]

Navádzanie robotov a lokalizácia objektov - Spojením robotiky a strojového videnia sa navádzanie robotov rýchlo rozvíja. Systémy strojového videnia určia presnú polohu objektov, nájdu záchytné body a poskytnú robotovi údaje o polohe objektu v reálnom čase. Navádzané roboty sa využívajú pri montáži, vyberaní, umiestňovaní alebo triedení objektov a umožňujú flexibilný prístup k podávaniu objektov,čím eliminujú potrebu presných prípravkov vo výrobe. Presnosť takto navádzaných robotov sa pohybuje na úrovni mikrónov.



Obr. 6: Príklady lokalizácie objektov a navádzania robotov[5]

Rozpoznávanie čiarových kódov - Pri rozpoznávaní čiarových kódov sa identifikujú objekty pomocou značiek, ktoré sa na nich nachádzajú. Medzi takéto značky patria štandardizované čiarové kódy a kódy dátovej matice. Snímače na čítanie kódov uchovávajú obrazy načítaných kódov, ktoré využijú pri analýze kódov, ktoré nie je možné správne prečítať . Vďaka tomu dokážu detegovať problémy s poškodeným kódom alebo nevhodným osvetlením. Na rozpoznávanie využívajú metódy zo všetkých oblastí spracovania obrazu ako je určovanie polohy, detekcia hrán a filtrovanie. Vďaka tejto technológii je umožnené sledovanie, identifikácia výrobkov a rozpoznávanie objektov priamo vo výrobe.



Obr. 7: Príklady rozpoznávania čiarových kódov[5]

Rozpoznávanie znakov - Rozpoznávanie znakov je založené na technológiách optického rozpoznávania znakov (OCR) a optického overovania znakov (OCV). Systémy OCR interpretujú alfanumerické znaky, bez vopred nastavených znakových údajov a systém OCV potvrdzujú prítomnosť špecifických znakových reťazcov. Extrahovaný text je následne spracovaný do formátu, ktorý je čitateľný pre stroj. Takéto systémy rozpoznávajú znaky s 99,99% presnosťou za niekoľko milisekúnd. Využívajú sa najmä vo farmaceutickom a potravinárskom priemysle, pri identifikácii dátumu výroby, dátumu spotreby, číslo šarže, názov výrobku a pri určení výrobcu.



Obr. 8: Príklady rozpoznávania znakov[5]

Rozpoznávanie vzorov - Systémy strojového videnia umožňujú zatrieďovanie objektov na základe rozpoznania vzorov, ktoré sa nich nachádzajú. Vzory je potrebné vopred určiť a definovať ich veľkosť a tvar. Kamera zanalyzuje všetky objekty v jej zornom poli ak sa tam nachádza jeden alebo viac vzor, identifikuje ich. Následne je možné pomocou pridania vhodných nástrojov určiť aj počet identifikovaných vzorov.

3 Priemyselné kamerové systémy

Na dosiahnutie čo najlepších výsledkov pri úlohách strojového videnia je kľúčovým faktorom výber vhodného priemyselného



Obr. 9: Príklad rozpoznávania vzorov s pridaným nástrojom na určenie ich počtu

kamerového systému. Tento výber je často finančnej aj časovo náročný a výsledky nemusia byť vždy optimálne.

Na trhu je veľké množstvo kamier, pričom každá kamera má niekoľko špecifikácií, ktoré sú založené na rôznych predpokladoch. Preto sa odporúča vykonať prieskum trhu. Avšak treba rátať s tým, že veľké množstvo informácií o kamerách môže proces hľadania vhodnej kamery ešte viac skomplikovať. Ešte pred vykonaním prieskumu trhu je potrebné dôsledne si určiť podmienky, ktoré musí kamera spĺňať pre úspešné splnenie danej úlohy [6].

3.1 Parametre priemyselných kamerových systémov

Pri definovaní kritérií kamery, je potrebné zamerať sa aj na parametre priemyselných kamier, ktoré sú pre každú kameru iné a dokážu ovplyvniť náš vyber [11].

Jedná sa o tieto parametre:

1. **Zorné pole** - Definuje oblasť, v ktorej je kamera schopná zosnímať objekt. Zorné pole sa označuje v stupňoch a je priamo ovplyvnené ohniskovou vzdialenosťou objektívu kamery, ktorá je vyjadrená v milimetroch. Ak je ohnisková vzdialenosť dlhšia, zorné pole je užšie a dochádza k zväčšeniu snímaného obrazu. Samozrejme, platí to aj naopak, čím je ohnisková vzdialenosť kratšia, tým je zorné pole širšie a zväčšenie je menšie.[11]

2. **Pracovná vzdialenosť** - Je to vzdialenosť od objektívu po snímaný objekt. V rámci tejto vzdialenosti dokáže kamera vykonávať funkcie strojového videnia s vysokou presnosťou. Pri nedodržaní rozsahu pracovnej vzdialenosti môžu byť výsledky funkcií neuspokojivé avšak aj nemusia, záleží od druhu kamery, typu úlohy a vplyvu pracovného prostredia [11].

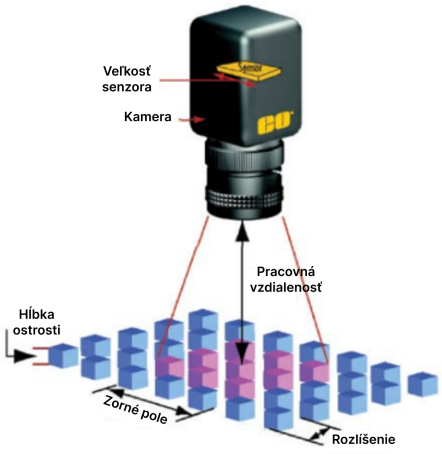
3. **Rozlíšenie** - Určuje do akej miery kamera dokáže zachytiť detaily snímaného objektu. V kamerových snímačoch sú pixely rozdelené na dve časti - horizontálne a vertikálne. Veľkosť každého pixela je definovaná v milimetroch, vďaka čomu je možné stanoviť rozlíšenie daného snímača a určiť, či je vhodný pre našu úlohu. V práci sa potvrdilo pravidlo: čím je snímaný objekt menší o to väčšie rozlíšenie je potrebné pre jeho nasnímanie [11].

4. **Hĺbka ostrosti** - Je označenie pre maximálnu hĺbku objektu, ktorú dokáže daná kamera zaostriť. Je to dôležitý parameter pri úlohách rozpoznávania čiarových kódov. Pri tomto type úloh sú objekty v zornom poli kamery rôznej veľkosti čo má za následok, že čiarové kódy objektov sú často v inej pracovnej vzdialenosti z hľadiska ohniska objektívu. Hĺbku ostrosti pre zaostrovací bod ovplyvňuje clona objektívu, ktorá sa označuje písmenom *f*. Ak je v špecifikácii kamery pri objektíve číslo pri *f* väčšie, teda clona je menšia, k snímaču je prepustené menšie množstvo svetla čoho následkom je väčšie hĺbka ostrosti. Menšie číslo pri písmene f označuje väčšiu clonu a menšiu hĺbku ostrosti [11].

5. **Veľkosť senzora** - Kamerové senzory obrazu sú vyrábané v rôznych veľkostiach od 6,35 milimetra po 25,4 milimetra. Pomocou horizontálnej veľkosti senzora a požadovaného zorného poľa je možné určiť primárne aké zväčšenie obrazu je potrebné pre konkrétne úlohy [11].

6. **Primárne zväčšenie** - Je to pomer medzi veľkosťou senzora a zorným poľom kamery. Ako sme spomenuli v predchádzajúcich definíciách, ohnisková vzdialenosť stanovuje zorné pole a mieru zväčšenia obrazu. Toto zväčšenie sa označuje ako primárne zväčšenie alebo PMAG. Na výpočet primárneho zväčšenia sa využíva nasledovný vzorec [11].

$$PMAG = \frac{\text{Horizintálna veľkosť senzora (mm)}}{\text{Veľkosť zorného poľa (mm)}} [11]$$



Obr. 10: Parametre priemyselných kamerových systémov[11]

3.2 Typy priemyselných kamier podľa aplikačnej úrovne

Možnosť nasadiť algoritmus strojového videnie do rôznych priemyselných kamerových systémov priniesla so sebou vznik nových typov kamerových systémov, ktoré spĺňajú iné aplikačné požiadavky ako už doteraz spomenuté kamery [4].

Rozlišujú sa na základe ich aplikačnej úrovne:

Smart kamery - sú to komplexné zariadenia, ktoré v jednom kompaktnom celku spájajú obrazový senzor, výpočtovú jednotku a softvér pre spracovanie obrazu. Ich hlavnou výhodou je schopnosť spracovávať obraz priamo na mieste, bez potreby pripojenia na externý počítač. Typickým využitím smart kamier sú aplikácie, kde je potrebná presná kontrola kvality výrobkov, detekcia chýb, rozpoznávanie čiarových kódov či optické rozpoznávanie znakov (OCR). V praxi sa tieto kamery často používajú na kontrolu správnosti montáže, meranie rozmerov produktov alebo identifikáciu a triedenie objektov na výrobných linkách [4].

Vizuálne senzory - predstavujú jednoduchšiu alternatívu k smart kamerám. V porovnaní so smart kamerami majú obmedzené výpočtové možnosti a využívajú sa na špecifické, menej náročné úlohy ako je napríklad: detekcia prítomnosti objektov, základné farebné analýzy, rozpoznávanie jednoduchých vzorov a kontrolu správnosti montáže v menej komplexných aplikáciách. Ich výhodou je jednoduchá inštalácia, nižšie náklady a spoľahlivá prevádzka pri úlohách, ktoré nevyžadujú pokročilé spracovanie obrazu [4].

Zabudovaný vizuálny systém - Zabudované vizuálne systémy sú určené pre komplexnejšie aplikácie, pri ktorých je potrebná vysoká presnosť a integrácia s inými zariadeniami. Tieto systémy pozostávajú z kamery, ktorá je súčasťou väčšieho celku, ako sú robotické pracoviská, výrobné linky či autonómne vozidlá. Vďaka ich pokročilým algoritmom strojového videnia sú schopné vykonávať náročné úlohy, akými sú 3D merania, pokročilá kontrola kvality, detekcia chýb a navigácia robotických zariadení. Zabudované vizuálne systémy umožňujú vysoký stupeň automatizácie a presnosti v moderných výrobných procesoch, pričom sú kľúčové najmä v oblastiach, kde súčasné trendy smerujú k vyššej autonómii a efektivite [4].

4 Aplikácia pre úlohy kontroly kvality

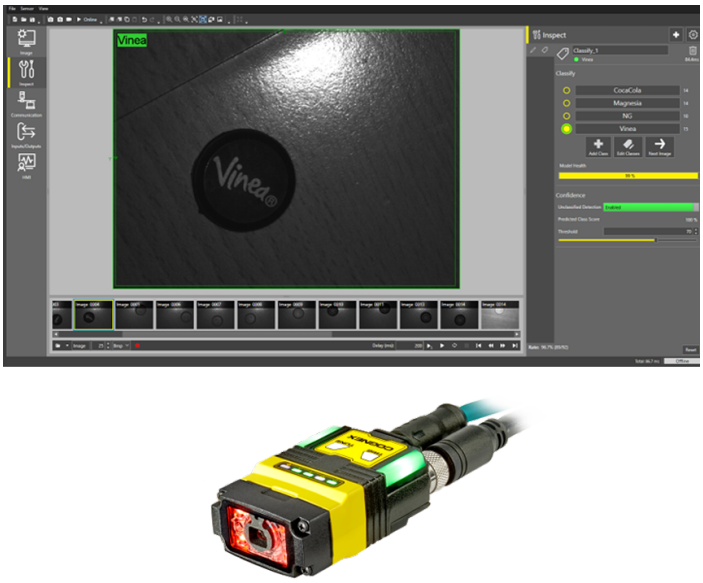
Cieľom práce je pomocou cenovo dostupnej kamery vyriešiť problém z praxe a eliminovať ľudskú prácu a chybovosť.

Ako prvý krok sme vykonali analýzu problému a následne sme definovali problém: V danej predajni majú zákazníci možnosť samoobsluhy v sklade a v rámci jedného nákupu si vyberú rôzne druhy tovaru. Človek pri pokladni musí všetky tieto druhy rozlíšiť, spočítať presný počet kusov z daných druhov a následne nablokovať. Tento manuálny proces počítania a kontroly predlžuje čas pri pokladniach a zvyšuje chybovosť. Riešim je automatické rozpoznávanie vybraného sortimentu pomocou smart kamery.

Na realizáciu riešenia sme sa rozhodli pre smart kameru Cognex In-Sight 2800 pre jej cenovú dostupnosť a ľahkú implementáciu a intuitívne prostredie Easy Builder.



Obr. 11: Fotografia z predajne a vybraného sortimentu pre automatické rozpoznávanie v debničke



Obr. 12: Ukážka použitého nástroja Easy Builder a kamera Cognex In-Sight 2800

Na Obrázku 11 je debnička z predajne, ktorá má veľké rozmery vzhľadom na zorné pole kamery a na to aby sme ju dokázali nasnímať celú by kamera musela byť umiestnená vo výške 230 centimetrov, čo nám nedovoľujú aktuálne pracovné podmienky. Z tohto dôvodu sme sa rozhodli pre tvorbu 3D modelu, ktorý slúži ako náhrada skutočnej debničky. 3D model sme vytvorili pomocou nástroja Blender a vytlačili na 3D tlačiarni, viď. Obrázok 13.

4.1 Návrh riešenia

V zmysle definovaných požiadaviek riešenie spočíva v:

- Rozpoznávaní objektov pomocou vhodnej funkcie
- Definovaní počtu rozpoznaných objektov
- Tvorbe používateľského rozhrania

Rozpoznávanie objektov pomocou vhodnej funkcie

Nakoľko sa v definovanom sortimente sa nachádzajú objekty, ktoré sú od seba navzájom odlíšené vzorom alebo farbou bolo



Obr. 13: Porovnanie objektov (vľavo) a reálny 3D model debničky (vpravo)

potrebné zadefinovať vo funkciách strojového videnia presné parametre, na základe ktorých ich budeme vedieť od seba odlíšiť.

Rozpoznávané objekty sa nachádzajú na fixných pozíciach rozhodli sme sa použiť funkciu klasifikácie. Pri použití tejto funkcie, sme mali na každú pozíciu zadefinovanú klasifikáciu objektov, v rámci ktorej sme rozpoznávali viacero tried. Definícia väčšieho počtu tried (7) a funkcií pre klasifikáciu (12) kameru spomaľovali veľmi a práca s ňou bola náročná. Preto sme sa rozhodli pre riešenie pomocou detekcie vzorov.



Obr. 14: Pohľad na skutočné objekty (vľavo) a rozpoznané vzory na objektoch pomocou Cognex kamery (vpravo)

Definovanie počtu rozpoznaných objektov

Po správnej detekcií vzorov, sme využili funkciu na určenie počtu, ktorú sme naviazali na definované vzory. Pri objektoch s rovnakým vzorom ale inou farbou toto riešenie nebolo postačujúce, preto sme do riešenia doplnili funkciu pre určenie hodnoty úrovne jasu daného objektu. Pre každú farbu sme definovali rozsah hodnôt podľa ktorých sme potom objekty na základe ich jasovej hodnoty zaraďovali, viď Obrázok 15.

Tvorba používateľského rozhrania

V rámci riešenia sme vytvorili vlastné používateľské rozhranie, webovú stránku, na ktorej sa nachádza počet rozpoznaných objektov a snímka, ktorú urobila kamera. Postup pri tvorbe používateľského rozhrania bol nasledovný:

1. Tvorba FTP serveru cez minipočítač Raspberry Pi, na ktorý smart kamera odošle výsledky použitých funkcií strojového videnia spolu so snímkou celého zorného poľa



5	Svijany_zele	Brightness_	Brightness_
6	2.000	106.751	119.369

Obr. 15: Pohľad na skutočné objekty (vľavo) a rozpoznané vzory na objektov s určením jasových hodnôt a počtu rozpoznaných objektov (vpravo)

2. Tvorba webovej aplikácie cez Node-RED
3. Zobrazenie údajov z FTP servera na webovej stránke

5 Záver

Implementácia systému strojového videnia prostredníctvom smart kamery Cognex In-Sight 2800 úspešne vyriešila praktický problém rozpoznávania a počítania produktov v predajnom prostredí. Použitie výpočtovej inteligencie výrazne znížilo chybovosť a časovú náročnosť pôvodného manuálneho procesu. Vytvorením používateľského rozhrania sa zabezpečila jednoduchá a efektívna komunikácia medzi kamerovým systémom a používateľom. Pre ďalší rozvoj je možné zvážiť aplikáciu tohto prístupu v rôznych oblastiach automatizácie a kontroly kvality s cieľom ďalšieho zvyšovania efektivity výrobných procesov.

Literatúra

[1] Valavani.Anand, Sheila and Loganathan, Priya. A Guide for Machine Vision in Quality Control. isbn = 9781003002826, December 2019

[2] Golnabi, H. and Asadpour, A. Design and application of industrial machine vision systems. Robotics and Computer-Integrated Manufacturing, doi=10.1016/j.rcim.2007.02.005, September 2007.

[3] Industrial Vision Systems Ltd. The vision to automate url = chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.industria content/uploads/2020/05/The-Vision-to-Automate-Industrial-Vision-Systems.pdf, 2020.

[4] Hornberg, A. Handbook of Machine Vision and Computer Vision. isbn = 9783527413430, 2017

[5] Cognex Corporation Introduction to Vision System Components ://www.cognex.com/blogs/machine-vision/introduction-to-vision-system-components.

[6] Zebra What Is a Machine Vision Camera? ://www.zebra.com/ap/en/resource-library/faq/what-is-machine-vision-camera.html.

[7] Labudzki, Remigiusz and Legutko, Stanislaw and Raos, Pero The essence and applications of machine vision Tehnicki Vjesnik, August 2014.

[8] Davies, E.R. Computer and Machine Vision: Theory, Algorithms, Practicalities. isbn = 9780123869081, 2012

[9] EAM EAM blog ://www.eaminc.com/blog/.

[10] Meaxpert (MXP) 4 Basic Types of Machine Vision Systems ://meaxpert.com/blog/basic-types-of-machine-vision-systems.

[11] Cognex The Six Fundamental Parameters of Machine Vision Optics ://www.cognex.com/blogs/machine-vision/the-six-fundamental-parameters-of-machine-vision-optics.

EKV - Elektronická kontrola vstupu

Andrej Marián Drugda[1], Felix Jozef Juraj Vítek[2] a Katarína Žáková[1]

[1]Ústav automobilovej mechatroniky, FEI STU v Bratislave
[2]Ústav aplikovanej informatiky, FEI STU v Bratislave

xdrugdaa@stuba.sk

Jadrom daného projektu je realizácia elektronického zabezpečovacieho systému. Celý projekt je realizovaný pre STU FEI v Bratislave a vznikol na podnet pani prof. Ing. Kataríny Žákovej, PhD. Našou úlohou bola kompletná realizácia zabezpečovacieho systému do laboratória, ktorý bude slúžiť ako kontrola prístupu do miestnosti na základe hardwarovej autentizácie. V danej dokumentácií sú prehľadne popísané všetky náležitosti ku návrhu, funkcionalite a sprevádzkovaniu systému.

1 Úvod

Bezpečnosťou sa ľudia zaoberajú od nepamäti. Strata nadobudnutého imania alebo blízkej osoby je pre ľudí vždy citlivá záležitosť. Preto sa človek proti vzniku takýchto nežiaducich udalostí prirodzene chráni. Prvotné zabezpečenia boli výlučne materiálneho charakteru na štýl fyzických prekážok/zábran. Zmena nastala príchodom elektrickej energie. Jeden z prvých zabezpečovacích systémov skonštruoval Američan A.Pope v roku 1853. Princíp poplachového zabezpečovacieho systému spočíval na paralelnom zapojení spínacích kontaktov v oknách a dverách. Zopnutím niektorého z kontaktov bol uzatvorený obvod s elektrickým zvončekom, ktorý spustil poplach.

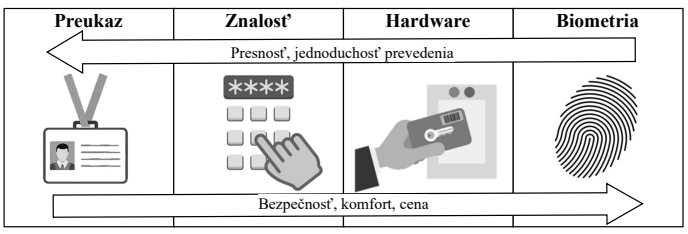
Aktuálne sa elektronické zabezpečovacie systémy (ďalej už len EZS) delia na 4 podkategórie podľa špecializácie a účelového charakteru. Jednotlivé bezpečnostné odvetvia sa medzi sebou častokrát kombinujú, aby sa dosiahla čo najväčšia miera zabezpečenia.

- PZS – Poplachové Zabezpečovacie Systémy poskytujú ochranu pred nežiadúcim vniknutím do akéhokoľvek objektu.
- EKV – Elektronická Kontrola Vstupu umožňuje spoľahlivo sledovať, evidovať a riadiť priechod osôb prostredníctvom snímačov/terminálov a prístupových mechanizmov.
- DVS – Dohľadové Video Systémy alebo inak aj CCTV je komplexné riešenie vizuálno-akustickej ochrany/kontroly pomocou sledovacích zariadení a úložných medií, ktoré ponúkajú pre užívateľa prehľad o celkovom stave objektu.
- EPS – Elektronický Požiarny Systém zabezpečuje ochranu objektu pred požiarom alebo vznikom iných nežiadúcich okolitých situácií (únik plynu).
- ADO – Anti-Dron Offense predstavuje jednu z nových oblastí, kde bolo potrebné zaviesť bezpečnostné systémy. Ako z názvu vyplýva drony predstavujú rastúce riziko pre verejné služby ale aj väzenia, letiská či prístavy. Preto musia prevádzkovatelia zvážiť a riešiť túto vznikajúcu hrozbu zneužitia komerčných bez-pilotných lietadiel.

1.1 EKV

Z vyššie vymenovaných EZS sa naša téma radí do Elektronickej Kontroly Vstupu (ďalej len EKV). Ako bolo spomenuté, jedná sa o prístupový systém určený na kontrolu priechodu osôb do miestnosti/do chránenej zóny na základe identifikácie. Identifikáciu (autentifikáciu) je možné vykonávať 4 spôsobmi, ktorých rozdiely sú znázornené šípkami(viď. obr.1). Smer šípky určuje zlepšenie:

1. Preukazom – Identifikačná karta s autentifikačnými údajmi (Preukaz poistenca, OP)
2. Znalosťou – Sústreďuje sa na pamäť autorizovanej osoby (Číselné heslo, vzor)
3. Hardvérová – Čipová karta, kľúčenka alebo iný druh jedinečného, autentifikačného zariadenia priradeného autorizovanej osobe.
4. Biometrická – Nenapodobiteľné biologické črty konkrétnej osoby, ktoré je možné rozpoznať a porovnať s inou predlohou (odtlačky prstov, geometria tváre, geometria celej dlane).



Obr. 1: Spôsoby autentifikácie

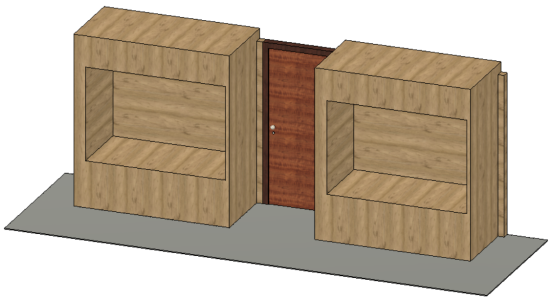
1.2 Východiskové parametre

1.2.1 Myšlienka projektu

Dôvodom vzniku témy daného projektu, bolo zjednodušenie prístupu do výskumného laboratória. Dvere na laboratóriu sú z vonkajšej strany opatrené guľou. Ak niekto chcel vstúpiť do laboratória, bolo potrebné, aby bol k dispozícii kľúč od miestnosti. Keďže zodpovedná osoba za laboratórium nedokáže byť vždy k dispozícii (pedagogika, schôdze, obedová prestávka, ...), tak na vstup do laboratória bolo potrebné čakať. Riešením by mohlo byť vytvorenie systému EKV, ktorý umožní nielen vstup do miestnosti, ale aj ďalšie výhody. Prístup do laboratória bude monitorovaný nepretržite a softvérová aplikácia umožní prehľadne vidieť časy príchodov a odchodov z laboratória, riadiť povoľovacie práva pre konkrétne osoby alebo dokonca kontrolovať dochádzku.

1.2.2 Situačné prostredie realizácie

Ako prvá miestnosť obstaraná systémom EKV, bude laboratórium pre výučbu mechatronických predmetov napríklad VKARM. Laboratórium sa nachádza na prvom poschodí v bloku-D. Po ľavej strane sa pozdĺž celej steny nachádzajú drevené skrine s pravidelnými medzerami pre dvere, ktoré slúžia ako vstupy do jednotlivých tried. Pre lepšiu predstavu sme vytvorili situačný 3D model (viď. obr.2,3).



Obr. 2: Dvere laboratória zo strany chodby



Obr. 3: Dvere laboratória z vnútra miestnosti

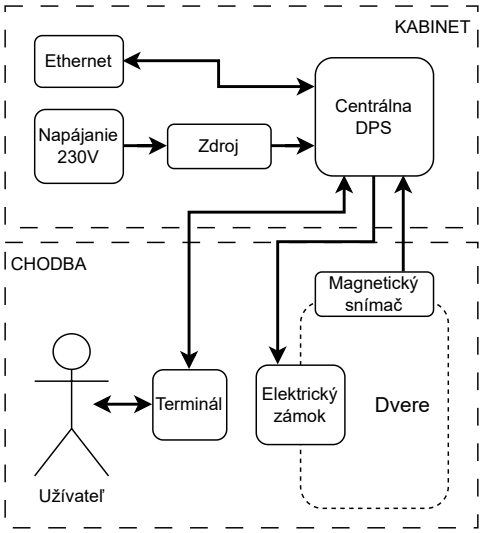
2 Jadro

Každý nový projekt sa začína konceptom a detailným návrhom logistiky. Chronológia jednotlivých krokov návrhu sa odvíja od spôsobu realizácie a účelu zariadenia. V našom prípade sme postupovali následovne -> Bloková schéma -> Výber komponentov -> Návrh DPS -> Návrh púzdra pre terminál -> Oživenie zariadenia -> Návrh architektúry systému -> Návrh databázy -> Implementácia firmvéru -> Implementácia TCP servera -> Implementácia webovej aplikácie.

2.1 Hardvérová realizácia

2.1.1 Bloková schéma

Koncept navrhnutého EKV pozostáva z dvoch celkov. Prvý obsahuje súčasti uložené vo vnútri laboratória a druhý priamo aplikované komponenty na dverách (viď. obr.č.4). Šípky znázorňujú všeobecné prepojenie a smer toku signálov medzi jednotlivými komponentami. Signály sú iba orientačné a nejedná sa o presnú reprezentáciu prepojení. Počty káblov a ich funkcie (dátový, napájací) sú popísané v ďalších bodoch.



Obr. 4: Bloková schéma systému

2.1.2 Súpis použitých komponentov

Použité komponenty môžeme rozdeliť do dvoch kategórií:

1. Hotové zariadenia/moduly
   - Elektrický zámok
   - Magnetický spínač
   - Ethernet modul W5100
   - NFC/RFID modul PN532
   - Napájací adaptér
2. Jednotlivé súčiastky
   - ATmega328
   - Pasívne súčiastky - rezistory, kondenzátory,diódy
   - Tranzistory (unipolárne/bipolárne)
   - Stabilizátor AMS1117-5.0
   - Kryštál oscilátor

- Tlačidlo, DC Jack
- Bzučiak, LED
- Konektory, Svorkovnice

Elektrický zámok

Bez elektrického zámku sa nezaobíde žiadny systém EKV. Plní zásadnú funkciu otvárania dverí. Na trhu existuje veľa typov elektrických zámkov s rozdielnymi parametrami. Odlišujú sa v pracovnom napätí, prítlačnej sile, prúdovom odbere alebo dĺžke pracovného času.

My sme do našej práce zvolili elektrický zámok MINIORNO s aretáciou (viď. obr.č.5). Elektrický zámok je vybavený cievkou nízkeho prúdu so širokým rozsahom napätia 9-24V, spotreba prúdu je 280 mA pre 12 V/DC a 250 mA pre 12 V/AC [3].



Obr. 5: Elektrický zámok



Obr. 6: Magnetický spínač

Magnetický spínač

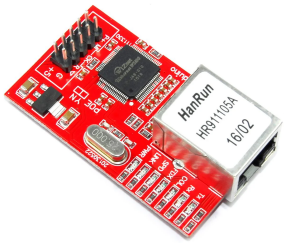
Stav otvorenia/zatvorenia dverí sa v bezpečnostných systémoch zvykne detegovať pomocou magnetických snímačov. Priblížením magnetického telesa ku spínaču sa kontakty pritiahnu a spojením uzatvoria elektrický obvod. Celý magnetický snímač teda tvoria dve časti (viď.obr.č.6). Pevný spínač s kontaktami (rám dverí) a pohyblivý magnet (dvere).

Ethernet modul W5100

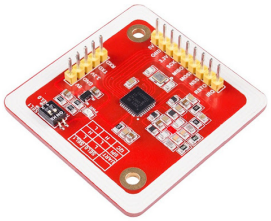
Internetové pripojenie a sieťovú komunikáciu sme zabezpečili použitím ethernetového modulu W5100. Prvotným návrhom bola implementácia čipu W5100 formou embedded priamo na DPS. Avšak počas návrhu sme narazili na komplikácie spojené s neistotou správnosti komponentov, prípadne osadenia externou firmou. Preto sme stavili na istotu a využili hotový ethernet shield s rovnakým čipom (viď.obr.č.7). Napájacie napätie modulu je 5V a pri plnom vyťažení má odber prúdu niečo nad 200 mA. Modul podporuje internetové protokoly: TCP, IP Ver.4, UDP, ICMP, ARP a Ethernetové protokoly: DLC, MAC. Taktiež podporuje operabilitu v half/full duplex móde a na komunikáciu s MCU využíva rozhranie SPI.

NFC/RFID čítačka

Požiadavkou zabezpečenia hardwarovej autentizácie v našom systéme, bola schopnosť čítania čipových kariet ISIC. Z existujúcich čítačiek sme vybrali kombinovaný modul PN532 (viď. obr.č.8) z funkciou RFID aj NFC. Je to prídavná hodnota projektu a určite ju ocení nejeden študent keď si zabudne ISIC. Takto bude postačovať uloženie karty do mobilu a pred vstupom sa študent preukáže priložením mobilu o terminál. Modul pracuje na napätí 5V a v závislosti od režimu prevádzky (čítanie, zápis) má odber prúdu okolo 150 mA. Výhodou daného modulu je jeho flexibilita v spôsobe komunikácie. Disponuje totiž tromi komunikačnými rozhraniami: SPI, I2C a HSU (vysokorýchlostný UART). Karty dokáže prečítať na vzdialenosť až 7 cm a podporuje takmer všetky štandardy kariet vrátane: Mifare 1k/4k, ISO/IEC 14443-4 alebo FeliCa [2].



Obr. 7: Ethernet modul W5100



Obr. 8: NFC/RFID čítačka

Napájací adaptér

Stabilný zdroj je kľúčom k stabilnému systému. Napájací adaptér sme vyberali na základe potrebného prúdu a napätia. Ako sme už spomenuli, najväčším pracovným napätím v celom systéme disponuje elektrický zámok 12V, preto sme adaptér dimenzovali na 12V. Odber prúdu sme následne stanovili po súčte spotreby predošlých dvoch modulov, zvyšnej elektroniky, elektrického zámku + miernej rezervy:

$$I_{M1}+I_{M2}+I_{EL}+I_{EZ}+I_R = 200+200+150+250+200 = 1000mA$$

Pre našu realizáciu sme teda vybrali 12W zdroj, ktorý disponuje na výstupe 12V a 1A.

ATmega328P

Srdcom každej výpočtovej techniky je riadiaci procesor. Existuje množstvo typov ako napríklad CPU, MCU, µP, DSP, FPGA, ktoré sú určené pre rôzne oblasti aplikácií. Pre náš vytváraný systém najviac vyhovovalo využitie Mikrokontroléra (MCU), nakoľko disponuje vopred integrovanými, internými obvodmi, ako sú ADC/DAC, časovače, pamäte alebo komunikačné rozhrania, čo z veľkej časti uľahčuje prácu pre dizajnéra. Na spojazdnenie stačí MCU osadiť do embedded riešenia DPS, pripojiť k nemu príslušné periférie a nahrať bootloader (konfiguračný program). Nevýhodou môže byť menšia flexibilita využitia oproti Mikroprocesorom, no v dnešnej dobe už nie je problém vybrať z rád existujúcich MCU ten správny s potrebnými parametrami pre špecializované použitie.

Pri výbere MCU sme sa zamerali na svetoznámu firmu Atmel (súčasť Microchip Technology) a u nej konkrétne na radu AVR AT, ktoré sú bohato využívané aj v platforme Arduino.

Zo širokej škály dostupných MCU sme vybrali konkrétne čip ATmega328P. Jedná sa o 8-bitový mikrokontrolér s architektúrou RISC. Prípona -P označuje verziu PicoPower čo z neho robí čip vhodný pre low-power aplikácie s nízkou spotrebou energie (zariadenia napájané z akumulátora). V tabuľke č.1 sú rozpísané ďalšie parametre mikrokontroléra ATmega328P [1].

Pasívne súčiastky

Súčasťou každého elektronického obvodu sú pasívne súčiastky založené na odbore, kapacite a indukcií. V našom projekte sme využili iba dva druhy a to kondenzátory a

Tab. 1: Parametre MCU ATmega328P

Napájanie	Jadro	Frekv.	Flash
1.8V - 5V	8-bit RISC	20 MHz	32 KB
EEPROM	GPIO	ADC	Kom.rozhranie
1 KB	23	8 (10-bit)	I2C,SPI,UART

rezistory (viď. obr.č.9,10). Pre všetky pasívne súčiastky sme sa snažili zvoliť jednotný typ púzdra SMD na povrchovú montáž s veľkosťou 0805 ($2x1,25mm$). Výnimkami sú iba 4 THT keramické kondenzátory (viď. obr.č.11) a dva THT fóliové kondenzátory (viď. obr.č.12).



Obr. 9: SMD rezistor 0805



Obr. 10: SMD kondenzátor 0805



Obr. 11: Keramický kondenzátor



Obr. 12: Fóliový kondenzátor

Tranzistory

Táto nenápadná súčiastka vniesla život novodobej výpočtovej technike. Tranzistor disponuje množstvom výhod akými sú veľkosť, úspora energie či cena a dnes už skoro neexistuje zariadenie kde by sa nenachádzal v jednej zo svojich podôb.

Tranzistory rozdeľujeme podľa princípu činnosti na unipolárne (riadené elektrickým poľom, viď.obr.14) a bipolárne (riadené elektrickým prúdom, viď.obr.13). V našej práci sme využili obidve varianty ako spínacie prvky. Bipolárne slúžia na menej zložité úkony (spínanie LED), zatiaľ čo unipolárne slúžia na spínanie elektrického zámku. Hlavnými sledovanými parametrami pri výbere tranzistorov boli: Maximálny priepustný prúd, maximálne zaťažovacie napätie a spínacie napätie. Z rád existujúcich tranzistorov sme sa preto rozhodli vybrať jedny z najpoužívanejších. *Bipolárny BC337* a *Unipolárny IRFZ48N*. Podrobné parametre sú zapísané v tabuľke č.2.

Stabilizátor

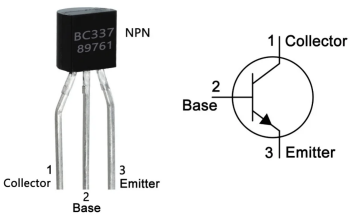
Elektronické výpočtové obvody, pracujú rádovo na mili až mikro ampéroch. Pre docielenie malých prúdov potrebujeme pracovať s malými vstupnými napätiami.

Na túto úlohu je určený Stabilizátor. Dokáže stabilizovať výstupné napätie v stanovenom rozsahu určenom konkrétnym typom stabilizátora alebo jeho prevedenia. Tak isto je potrebná znalosť jeho maximálneho výstupného prúdu, ktorý môže stabilizátor dodať. Stabilizáciu napätia je možné realizovať množstvom iných spôsobov ako napríklad napäťovým deličom pomocou rezistorov a diód alebo step-down pulzným meničom. Pre našu realizáciu sme sa však rozhodli zvoliť komplexné riešenie pomocou jednej súčiastky kvôli jednoduchosti prevedenia.

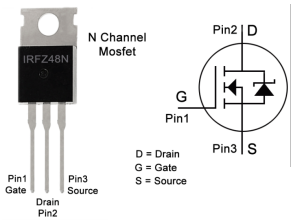
Vybrali sme stabilizátor *AMS1117-5.0* s výstupným napätím 5V, ktorý dokáže stabilizovať vstupné napätie v rozsahu 6-12V a dodať maximálny výstupný prúd 800 mA. Je v prevedení SMD s púzdrom SOT223 (viď.obr.15).

Tab. 2: Parametre tranzistorov

BC337	Parameter	Hodnota
V_{CEO}	Collector-Emitter U ($I_B = 0$)	45 V
V_{EBO}	Emitter-Base U ($I_C = 0$)	5 V
I_C	Collector current	0.5 A
IRFZ48N	Parameter	Hodnota
V_{GS}	Gate-to-Source range. U	± 20 V
$V_{GS(th)}$	Gate Threshold U	2 - 4 V
V_{DSS}	Drain-to-Source max.U	55 V
I_D	$T = 25°C$ $V_{GS} = 10V$	64 A



Obr. 13: BC337



Obr. 14: IRFZ48N



Obr. 15: Stab. AMS1117-5.0



Obr. 16: Oscilátor 16Mhz

Oscilátor

Funkciu merania času zabezpečujú v MCU oscilátory. Oscilátor je zariadenie, ktorého parametre sa istý čas periodicky menia po privedení vstupného impulzu. V elektronickom sektore sa jedná o premenlivé napätie. Pre vygenerovanie kmitavého priebehu napätia existuje množstvo spôsobov. Medzi základné typy oscilátorov patria LC a RC články. V zložitejších oscilátoroch môžeme nájsť začlenený aj tranzistor či operačný zosilňovač. Taktiež sa dajú pre vytvorenie oscilátorov využiť rôzne integrované obvody typu Timer (NE555).

Nevýhodou zostavených oscilátorov je ich veľkosť, ktorú zaberajú na DPS, navyše keď v prípade mini aplikácií rozhoduje každý milimeter.

Pre náš projekt sme si vybrali preto kryštálový oscilátor 16 MHz s púzdrom typu HC49/S (viď. obr.16). Oproti alternatívnym riešeniam zaberá omnoho menšiu plochu na DPS, je lacný a extrémne presný čo je potrebné pre správnu funkciu komunikačných rozhraní (SPI,UART,I2C).

Tlačidlo, DC Jack

V našom projekte sme použili štvor-vývodové tlačidlo typu SMD 6x6 mm (viď.obr.17), ktoré zohráva funkciu externého resetu pre potrebu pri programovaní. Napájací konektor DC Jack

sme zvolili typu 2.5 mm namiesto 2.1 mm (viď.obr.18) z dôvodu odolnejšej prúdovej zaťažiteľnosti.



Obr. 17: SMD tlačidlo



Obr. 18: DC Jack 2.5 mm

Bzučiak, LED

Bzučiaky, LED-ky a tak isto aj rôzne zobrazovacie zariadenia (LCD displeje), zohrávajú podstatnú rolu v nadviazaní a spríjemnení interakcie medzi človekom a počítačom. Preto sme pri ich výbere paralelne pracovali na terminálovom púzdre, aby sme mohli ich rozmery/parametre navzájom prispôsobovať pre docielenie čo najlepšieho estetického aj praktického výsledku.

Bzučiak (alebo aj Buzzer) je akustické zariadenie využívajúce sa na signalizačné funkcie. Z hľadiska menšej veľkosti sme sa rozhodli využiť piezoelektrickú pasívnu verziu (viď.obr.21). Dôvodom pasívnej varianty bzučiaku je schopnosť generovania širokého spektra frekvencií.

LED (Light-Emitting Diode) je luminiscenčná dióda, ktorá po privedení napätia v správnej polarite, vyžiari úzky pás svetla. Pre účely teminálu sme zvolili valcovú THT Ledku s priemerom 5mm a polguľatou špičkou (viď.obr.19). Do centrálnej DPS sme potrebovali šetriť miestom, preto sme zvolili variantu SMD chipled s veľkosťou púzdra 0805 (viď.obr.20).



Obr. 19: LED 5 mm



Obr. 20: SMD LED 0805



Obr. 21: Pasívny buzzer

Konektory, svorkovnice

Konektory sú štandardizované spojnice, ktoré slúžia na prepojenie zariadení prostredníctvom káblov, alebo priamo vzájomným prepojením konektorov. S konektormi sa bežný človek stretáva na dennej báze pri nabíjaní mobilného telefónu, ale vo veľkom sú využívané aj v priemysle.

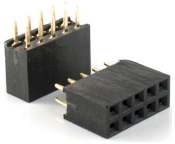
V našom projekte sme potrebovali vyriešiť prepojenie medzi centrálnou DPS a terminálom. Ako vhodné kontaktné médium sa nám ukázalo využitie IDE kábla pre jeho tenký dizajn (šírka jedného vodiča viď.obr.24). Kábel musí byť vedený popri vnútornej stene skrine, preto by nám guľaté alebo hrubšie varianty káblov mohli spôsobovať prekážku. S IDE káblom je možné manipulovať aj poza police, preto bude postačovať vyvŕtať maximálne 2 diery pri vstupe a výstupe zo skrine. Pre IDE kabel existuje štandardný konektor IDC. Preto sme zvolili podľa počtu pinov verziu IDC 2x3 s kolíkovou roztečou 2.54 mm (viď.obr.22). Ďalšie konektory použité v centrálnej DPS boli typu header pin, využívané tak isto v platforme arduino (viď.obr.23). V našom prípade ich používame pre nahrávanie bootloadera, programovanie a pre pripojenie ethernet modulu.

Svorkovnice ako z názvu vyplýva obsahujú vodivé svorky. Tie sú najčastejšie realizované predpruženými, skrutkovacími alebo zaklapávacími mechanizmami. Rozdiel oproti konektorom je v spôsobe pripájania. Zatiaľ čo do samice konektora stačí vsunúť jeden konzistentný proti kus, do svorkovnice vchádzajú jednotlivé žily kábla osobitne.

Do našej centrálnej DPS sme využili konkrétne 6 skrutkovacích svorkovníc s roztečou 3.5 mm (viď.obr.25) pre pripojenie ele.zámku, mag.snímača a odchodového tlačidla.



Obr. 22: IDC konektor 2x3



Obr. 23: Header pin 2.54 mm



Obr. 24: IDC kabel



Obr. 25: Svorkovnica 3.5 mm

2.1.3 Návrh a popis DPS

Doska Plošných Spojov (ďalej len DPS) tvorí kostru celého elektronického zariadenia. Od nej sa odvíja výsledná spoľahlivá funkčnosť systému, zložitosť programovania, ale aj celkový dizajn systému. Návrh prešiel od prvého konceptu množstvom modifikácií no zďaleka to nie je finálny produkt. Stále sa tu nachádza množstvo oblastí na vylepšenie a optimalizáciu, avšak pre základný požadovaný účel to zatiaľ bohate postačuje.

Ako je možné vidieť na blokovej schéme obr.4, systém je rozdelený na 2 časti. Podľa navrhnutej štruktúry je pochopiteľné, že sme potrebovali vytvoriť dve DPS-ky. Jednu pre centrálnu logiku (ústredňu) a druhú, ktorá bude zohrávať úlohu terminálu na interakciu s užívateľmi.

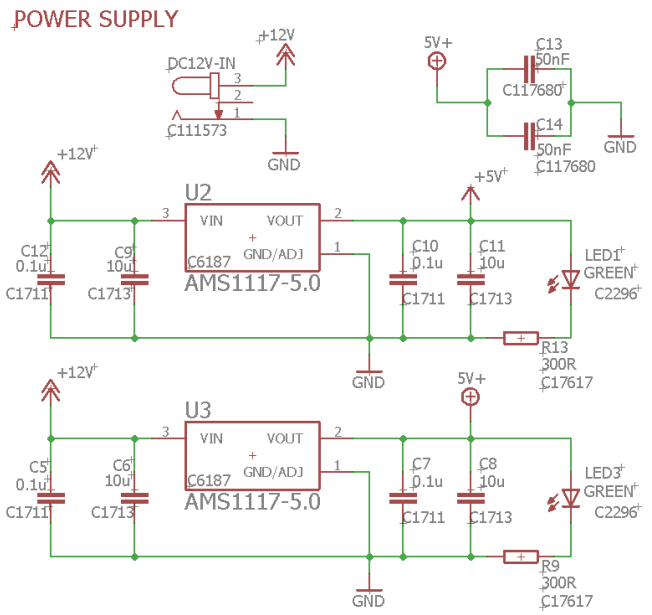
Prvotný konceptom bola realizácia celého systému EKV na jednej DPS. Kvôli nadrozmerným komponentom a neestetickosti prevedenia sme od tohto návrhu upustili. Avšak pri opätovnom návrhu druhej verzie by sme tento koncept určite použili.

Treba podotknúť, že systém sme navrhovali ako *decentralizovaný*. Aj napriek tomu, že sa uvádza v blokovej schéme "Centrálna DPS", nejedná sa o centrálnu v pravom slova zmysle. Jedna centrálna DPS je určená pre kontrolu prístupu jediných dverí. V prípade potreby kontroly prístupu do viacerých laboratórií, bude potrebné nad každé dvere osadiť jednu centrálnu DPS. Decentralizácia je zároveň jedným z bezpečnostných prvokov systému. V prípade centralizovaného

systému stačí vyradiť hlavnú riadiacu jednotku aby padol celý systém. Preto sa zabezpečovacie systémy v praxi vyrábajú na decentralizovanom princípe ako je použitý aj v našej realizácií.

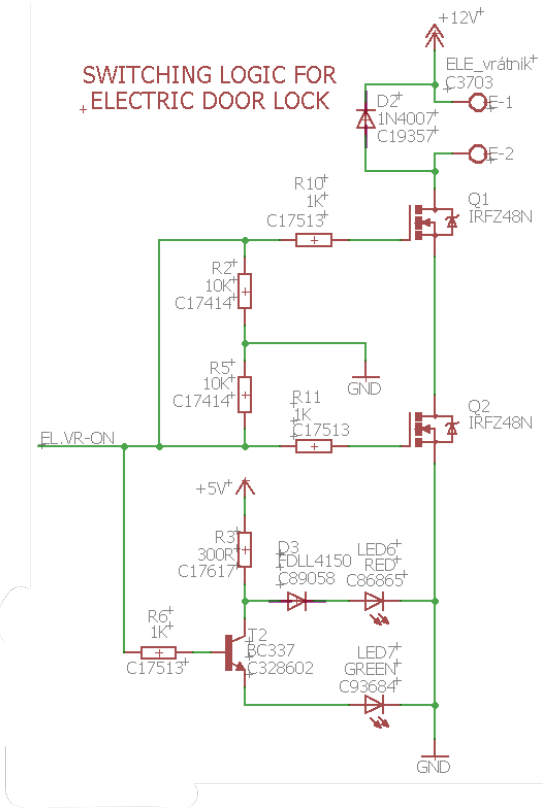
CENTRÁLNA DPS

Napájanie je prvým bodom návrhu pre zabezpečenie činnosti elektrického zariadenia. Bolo potrebné vypočítať aký maximálny odber prúdu bude zariadenie vykazovať a tak isto aké jednotlivé napätia si súčiastky vyžadujú. V našom návrhu sme stanovili maximálne potrebné napätie na 12V, ktoré vychádza z parametrov elektrického zámku. Pre zvyšnú elektroniku postačuje 5V. O transformáciu z 12V na 5V sa starajú dva stabilizátory AMS1117-5.0. Po prepočte prúdových odberov sme zistili, že by postačoval jeden no z dôvodu prvého návrhu sme nechceli doplatiť na prípadné poddimenzovanie systému. Stabilizátory sme na vstupe a výstupe opatrili filtračnými kondenzátormi pre vyrovnávanie prípadných napäťových špičiek a na výstupe každého stabilizátora sme osadili jednu informačnú LED, ktoré signalizujú aktívny stav zariadenia. Vstup do dosky je zabezpečený pomocou DC napájacieho jacku 2.5 mm. (Schéma zapojenia viď.obr.26).



Obr. 26: Schéma napájania

Logika pre spínanie elektrického zámku je založená na dvoch unipolárnych MOSFET tranzistoroch IRFZ48N. Názorné zapojenie do série je ďalším bezpečnostným prvkom, ktorý eliminuje stav SFC (Single Failure Condition). Eliminovanie prípadu jedného zlyhania znamená, že obvod je navrhnutý tak, aby v prípade výpadku jednej súčiastky/komponentu, bol schopný pokračovať v plynulej prevádzke alebo prešiel do bezpečného stavu. V reálnej interpretácií to znamená, že každý dôležitý komponent musí byť poistený duplikovaním alebo obídený náhradným riešením. Pre udržanie bezpečnosti v našom systéme EKV je dôležitá správna funkcia spínania elektrického zámku (ELE.Z). Kľúčové komponenty sú preto spínacie tranzistory. Nečakaná napäťová špička by mohla spôsobiť prieraz tranzistora, čo by malo za následok vytvorenie vodivého spoja aj bez aktívneho hradla. To by viedlo ku permanentne aktivovanému ELE.Z a voľne otvárateľným dverám. Preto sme spínaciu časť posilnili prídavným tranzistorom Q2 (viď. obr.27). ELE.Z predstavuje v našom obvode indukčnú záťaž. Preto je na jeho svorkách zapojená dióda D2, ktorá eliminuje vypínací zostatkový prúd. Zvyšná časť obvodu s tranzistorom BC337 slúži pre svetelnú signalizáciu aktivity ELE.Z.



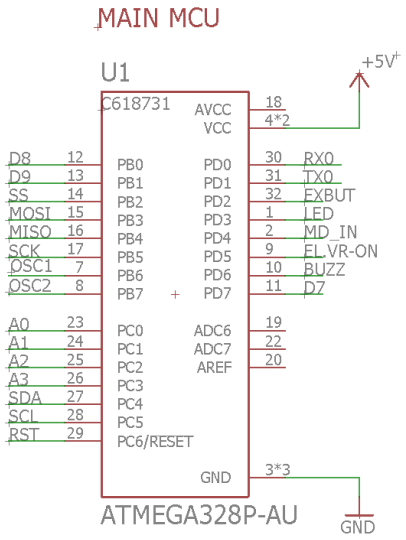
Obr. 27: Spínacia logika pre elektrický zámok

Mikrokontrolér a jeho zapojenie je realizované štandardným spôsobom. Na piny XTAL1 a XTAL2 sú privedené kontakty oscilátora, napájanie je brané z 5V stabilizátora a zvyšné GPIO sú zapojené podľa rozhodnutia programátora F.J.J.Víteka. Samozrejme sme dbali na ich generálne definovanú funkciu (RX,TX,D,A,MOSI...). Kompletný prehľad napojenia mikrokotroléra je zobrazený na obrázku č.28.

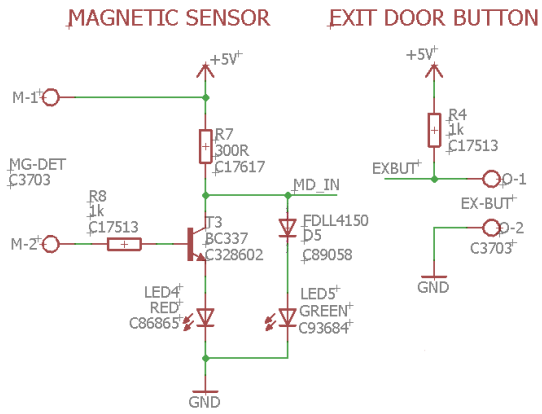
Vstup z magnetického snímača sme zapojili podobným spôsobom pomocou tranzistora BC337 a signalizačnými LED ako tomu je pri ELE.Z (viď.obr.29). Pre odchodové tlačidlo sme signalizáciu neriešili.

Terminálové LED sú riešené na princípe pusch-pull. Ide o paralélne, vzájomne záverné zapojenie dvoch LED (červená, zelená), ktoré signalizujú kladnú alebo zápornú odozvu terminálu na priloženie karty. V prípade potreby rozsvietenia červenej LED, stačí vygenerovať na pridelenom pine MCU napätie 5V. V opačnom prípade rozsvietenia zelenej LED, je potrebné znížiť napätie na 0V. Vypnutie obidvoch LED sa realizuje zmenou konfigurácie pinu z výstupu na vstup (high Z).

Ako posledné ukážeme prehľad zvyšných zapojení pomocných obvodov pre MCU a pripojenia konektorov



Obr. 28: Schéma zapojenia MCU ATmega328P



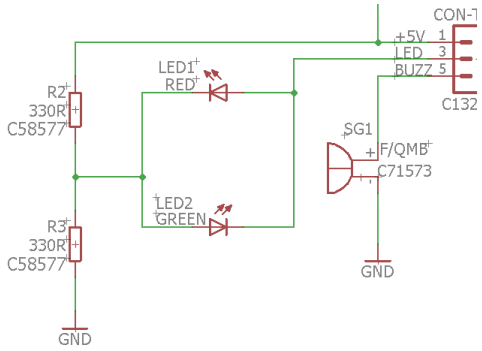
Obr. 29: Vstupy magnetického snímača a odchodového tlačidla

(viď.obr.31,32,33). Po návrhu elektrických schém zapojení prišiel na rad konkrétny návrh DPS (rozmiestnenie súčiastok, vytvorenie vodivých ciest). Celý návrh sme realizovali v program EAGLE 7.6.0. Všetkým súčiastkam sme pridelili popisné číslo LCSC, kvôli rýchlejšej identifikácií externou firmou JLCPCB, do ktorej sme dosky dali vyrobiť a osadiť. Pre výrobu sme zvolili štandardný materiál FR-4 s hrúbkou medi 35 μm.

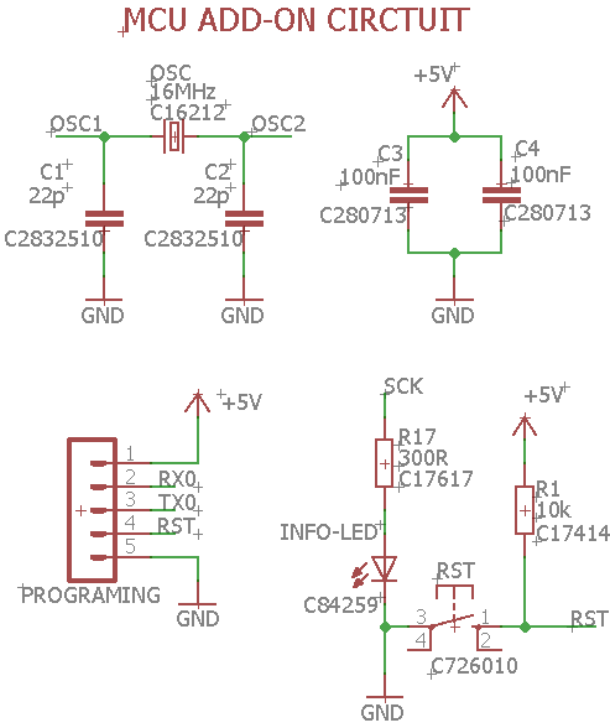
2.1.4 Návrh krytu terminálu

Každé elektronické zariadenie uvedené na verejný trh musí byť vybavené krytom pre bezpečnú manipuláciu a pracovnú činnosť. Vhodné púzdro musí prekrývať všetky živé časti, s ktorými by omylom mohli laici prísť do kontaktu a zároveň musí poskytovať sofistikovaný ergonomický dizajn a user friendly interakčné prvky.

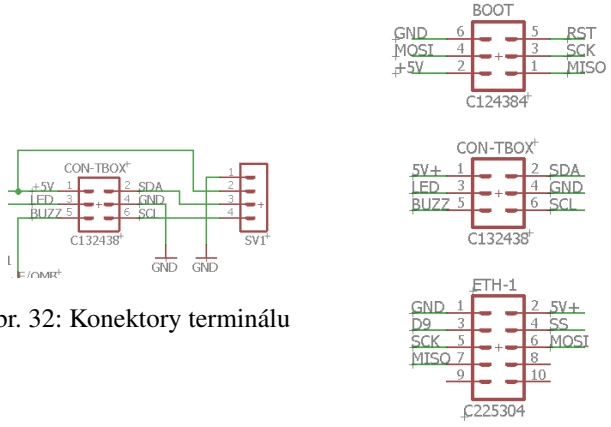
Na základe takto stanovených pravidiel sme sa snažili navrhnúť kryt pre terminál. Model sme vytvárali na 2x v programe Fusion360 (viď. obr.34,35,36,37). Výsledný 3D výtlačok je z materiálu PLA.



Obr. 30: Terminálová schéma zapojenia LED a bzučiaka



Obr. 31: Doplnkové obvody pre MCU

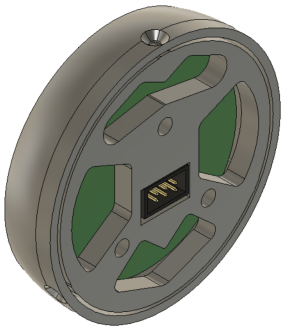


Obr. 32: Konektory terminálu

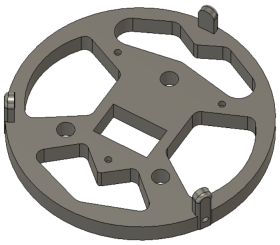
Obr. 33: Konektory centrály



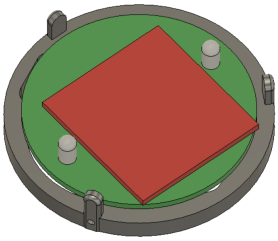
Obr. 34: Vrchný kryt



Obr. 35: Spodná strana



Obr. 36: Základná doska



Obr. 37: Osadená doska

2.1.5 Spojazdnenie EKV

Po zapájkovaní všetkých súčiastok a následnom zapojení periférií prišiel na rad prvý programový zásah vypálením zavádzača. Bootloader je konfiguračný program, ktorý nastaví hlavné parametre MCU ako konfiguráciu pinov, minimálne pracovné napätie atď. Pre nahranie bootloadera sme vyhradili osobitný konektor s názvom BOOT (viď.obr.33). Samotné vypálenie zavádzača sme realizovali pomocou platformy Arduino UNO s nastavením Arduino as ISP v programovacom prostredí Arduino IDE 1.8.18. Vypálenie bootloadera neprebehlo na prvýkrát úspešne. Narazili sme totiž na chybu spojenú so zámenou čipu. V JLCPCB zle diagnostikovali informáciu a zvolenom MCU. Namiesto ATmega328P na dosku osadili ATmega328. Jedna malá zmena v písmenku P spôsobila, že popisný konfiguračný súbor nerozpoznal signatúru daného čipu a preto doň nemohol vypáliť zavádzač. Vyriešili sme to mechanickým prepísaním signatúry v konfiguračnom súbore a následnom stiahnutí rozširujúcich knižníc pre špecifickú verziu čipu ATmega328.

Následne už bolo možné do čipu vypáliť zavádzač a programovať ho pomocou rozhrania UART (RX,TX).

2.2 Softvérová realizácia

Softvérová časť sa skladá z dvoch hlavných častí:

1. Centrálny server na správu prístupov
2. Firmvér zariadenia

Firmvér zariadenia je nahratý do každého zariadenia, ktoré otvára určitú miestnosť. Centrálny server na správu prístupov je iba jeden jediný. Všetky zariadenia sa pripájajú na tento server.

2.2.1 Centrálny server na správu prístupov

Súčasťou centrálneho serveru je:

1. Frontend - webová aplikácia
2. Backend - REST API
3. TCP server

Frontend je SPA (single-page application) vo framework-u Vue.js pre administrátorov systému a správcov miestností. Komunikuje s backend-om cez protokol HTTPS. Backend je REST API servis vo framework-u FastApi. TCP server je súčasťou backend procesu a komunikuje výlučne len so zariadeniami cez vlastný protokol popísaný v podkapitole 2.3. TCP server je potrebný, pretože knižnice implementujúce protokol HTTPS/TLS nie je možné na vybraný hardvér nahrať kvôli nedostatku flash pamäte.

2.2.2 Firmvér zariadenia

Firmvér pre zariadenie bol písaný v Arduino jazyku čo je v podstate parciálna implementácia štandardných jazykov C a C++. Najbližšie má tento špecializovaný jazyk ku štandardu C++ 11 [4]. Firmvér je relatívne jednoduchý, jeho úlohou je komunikácia s centrálnym serverom, čítanie kariet a otváranie zámku podľa odpovedí zo servera. Všetka logika kontrolujúca prístupy študentov je implemetovaná na centrálnom serveri.

Pri spustení sa zariadenie pripojí na server a autentifikuje sa pomocou unikátneho id zariadenia a symetrického kľúča. Po verifikácii serverom všetka nasledovná komunikácia je šifrovaná náhodne vygenerovaným symetrickým kľúčom. Zariadenie používa na komunikáciu so serverom protokol presnejšie popísaný v podkapitole 2.3. Ak chce študent odomknúť zámok priloží jeho ISIC kartu, z ktorej zariadenie prečíta id danej karty a pošle serveru požiadavku na otvorenie. Keď server dostane takúto požiadavku prečíta id karty a nájde študenta, ktorý má kartu pridelenú, overí či má daný študent do miestnosti prístup v aktuálnom čase a pošle odpoveď naspäť zariadeniu. Ak mal študent korektne pridelený prístup v aktuálnom čase zariadenie otvorí zámok na dobu desať sekúnd.

2.3 Komunikačný protokol

Aby sme zabránili MITM (man-in-the-middle) útokom a podobným hrozbám sme chceli pôvodne použiť štandardný protokol HTTPS/TLS a JSON ale ako už bolo spomenuté v predošlej podkapitole knižnice na HTTPS sú príliš veľké pre náš hardvér. Podobne aj knižnice na prácu s JSON sú príliš veľké a nezmestia sa do flash pamäte hardvéru. Preto sme navrhli a implementovali jednoduchý binárny protokol s podporou šifrovania.

Každý paket obsahuje na začiatku tzv. hlavičku (viď tabuľku 3). Verzia protokolu je v súčasnej implementácii TCP servera a firmvéru vždy nastavená na 1. Tento údaj sme pridali hlavne kvôli budúcnosti, keby sa protokol zásadne zmenil, týmto údajom je možné špecifikovať, ktorú verziu protokolu klient/server podporuje. Typ šifrovania určuje, či

a akým spôsobom sú dáta v pakete zašifrované. V súčasnej implementácii existujú dva spôsoby šifrovania:

1. `NONE` - dáta nie sú zašifrované
2. `CHACHA20_POLY1305` - dáta sú zašifrované pomocou šifry ChaCha20 a autentifikované pomocou kryptografickej MAC funkcie Poly1305

Tab. 3: Štruktúra paketov

Offset	Názov	Popis
0 B	`version`	verzia protokolu
1 B	`encryption`	typ šifrovania
2 B	`type`	typ paketu
3 B	`packet_length`	veľkosť celého paketu

Typ paketu hovorí o tom, aké dáta paket obsahuje. Veľkosť paketu udáva, koľko bajtov je potrebné na uloženie paketu do pamäte. Po tejto hlavičke nasledujú dáta paketu, ktorých štruktúra záleží od už spomínaného typu paketu.

2.4 Správa prístupov

Dôležitou súčasťou systému je aj správa prístupov do jednotlivých miestností. Táto časť je implementovaná ako webová aplikácia, do ktorej majú prístup len administrátori systému a študenti, ktorí dostali prístup od administrátora. Bez prihlásenia sa nedá vykonať žiadny úkon.

Hlavnou funkcionalitou aplikácie je správa prístupov a oprávnení do jednotlivých miestností. Najzaujímavejšia stránka pre administrátora je teda zoznam a detail miestností. V detaile miestnosti dokáže administrátor spravovať nielen prístupy študentov do miestnosti ale aj oprávnenia na správu danej miestnosti. Prístupy definujú, kto a v akom čase má dovolené vojsť do miestnosti, t.j. kedy sa mu pri priložení ISIC karty odomknú dvere. Existujú tri druhy prístupov:

1. jednorázový,
2. opakujúci sa,
3. neobmedzený.

Jednorázový prístup udeľuje možnosť vstupu do miestnosti študentovi v konkrétnom dátume a časovom okne napr. 29. 3. od 8 do 10 hodiny. Podobne funguje aj opakujúci sa prístup s tým rozdielom, že sa dá definovať ako často sa časové okno opakuje napr. každý týždeň. Neobmedzený prístup udeľuje študentovi prístup v ľubovolnom dátume a čase. Pri opakujúcich sa a neobmedzených prístupoch je možné definovať aj platnosť prístupu napr. pri pridelení neobmedzeného prístupu vieme definovať, že bude platný iba od začiatku do konca letného semestra.

Oprávnenia udeľujú konkrétnym osobám práva na úprava alebo na náhľad do danej miestnosti. Existujú dva druhy oprávnení: “dozorca” a “zodpovedná osoba”. “Dozorca” má oprávnenie len “na náhľad” do miestnosti teda vidí ktorí študenti majú do miestnosti prístup ale nedokáže študentovi prístup prideliť ani upravovať atribúty miestnosti. “Zodpovedná osoba” má oprávnenie prideľovať študentom prístupy do miestnosti ale nedokáže prideľovať oprávnenia iným osobám.

Okrem prístupov a oprávnení administrátori a osoby s oprávneniami vidia v detaile miestnosti aj históriu prístupov v ktorej sú údaje o tom kto/kedy sa snažil odomknúť dvere do miestnosti a či sa dvere odomknúť podarilo.

3 Záver

V rámci tohto projektu bol úspešne navrhnutý, implementovaný a oživený kompletný systém elektronickej kontroly vstupu (EKV) pre potreby laboratória na STU FEI v Bratislave. Spojením na mieru vyvinutého hardvéru, pozostávajúceho z centrálnej riadiacej jednotky a prístupového terminálu s NFC/RFID čítačkou a softvérového riešenia zahŕňajúceho firmvér zariadenia, centrálny server s webovou aplikáciou pre správu prístupov a vlastný zabezpečený komunikačný protokol, sa podarilo naplniť požiadavku na zjednodušenie a monitorovanie prístupu oprávnených osôb. Výsledný systém tak predstavuje funkčné a bezpečné riešenie pôvodného problému, pričom ponúka flexibilnú správu prístupových práv a detailný prehľad o vstupoch do zabezpečeného priestoru.

Literatúra

[1] Microchip Technology Inc. ATmega328/P Datasheet Complete. DS40001984A. [online]. 2018. [cit. 2025-03-28]. Dostupné z: `https://ww1.microchip.com/downloads/aemDocuments/documents/MCU08/ProductDocuments/DataSheets/ATmega48A-PA-88A-PA-168A-PA-328-P-DS-DS40002061B.pdf`

[2] ELECHOUSE PN532 NFC RFID Module User Guide [online]. 2013. [cit. 2025-03-28]. Dostupné z: `https://www.elechouse.com/elechouse/images/product/PN532_module_V3/PN532_%20Manual_V3.pdf`

[3] ORNO GROUP Sp. z o. o. Symmetric electric strike with lock, MINI, low current coil [online]. 2025. [cit. 2025-03-28]. Dostupné z: `https://files.orno.pl/product_sheet/en/or-ez-4025_product_sheet_2218.pdf`

[4] Arduino Language Reference [online]. [cit. 2025-03-28]. Dostupné z: `https://docs.arduino.cc/language-reference/`

Sekcia 3

Mechatronika II

Referenčný systém merania teploty pre teplotné vláknové segmentové interferometrické senzory

Andrea Kuňáková[1]

[1] Ústav automobilovej mechatroniky, Fakulta elektrotechniky a informatiky, Slovenská Technická Univerzita Ilkovičova 3, 81219 Bratislava, Slovenská Republika

xkunakovaa@stuba.sk

Abstrakt – Optické vláknové senzory na meranie teploty si nachádzajú čoraz širšie uplatnenie v priemyselných a výskumných aplikáciách. Kalibrácia a testovanie týchto senzorov si zvyčajne vyžaduje špecializované laboratórne vybavenie, ktoré je finančne nákladné a ťažko prenosné. V bežnon softvérovom vývoji tak často chýba nástroj, pomocou ktorého by bolo možné senzory efektívne overiť napríklad v kancelárskych podmienkach – počas vývoja aplikácie, ktorá zaznamenáva ich namerané dáta. Moja práca sa zaoberá konštrukciou prototypu prenosného referenčného systému, ktorý umožňuje testovanie optického vlákna v reálnom čase aj mimo laboratórnych podmienok.

1 Úvod

Teplota je jednou z najčastejšie sledovaných fyzikálnych veličín vo výskume aj priemysle. Kým tradičné senzory poskytujú spoľahlivé výsledky vo väčšine bežných aplikácií, náročnejšie prostredia vyžadujú inovatívnejšie riešenia. Jedným z nich je aj meranie pomocou optických vláknových senzorov, ktoré vďaka svojej odolnosti, presnosti a necitlivosti na elektromagnetické rušenie nachádzajú čoraz širšie uplatnenie.

S rastúcim záujmom o túto technológiu však prichádzajú aj výzvy. Jednou z hlavných je overenie správnosti merania a kalibrácia týchto senzorov. V ideálnom prípade sa takéto testovanie vykonáva v laboratórnych podmienkach – no v praxi to nie je vždy možné. Vývojové tímy, testeri či firmy často potrebujú rýchly, prenosný a dostupný nástroj, pomocou ktorého by vedeli senzory otestovať aj mimo laboratória.

Na trhu je momentálne minimum riešení, ktoré by takýto testovací systém ponúkali za rozumnú cenu. Často ide o drahé a robustné zariadenia určené pre veľké prevádzky. Pre menších vývojárov, výskumníkov alebo firmy to predstavuje problém – nie pre technickú náročnosť, ale pre dostupnosť.

Práve tu sa ukazuje potenciál jednoduchých riešení, ktoré dokážu využiť lacné a dostupné technológie ako Peltierove články či Arduino. Tieto komponenty, bežne využívané aj v hobby projektoch, sa pri správnom návrhu dokážu premeniť na účinný a spoľahlivý nástroj pre testovanie a porovnávanie výstupov teplotných senzorov – a to bez potreby zložitých zariadení či špecializavaného laboratória.

2 Parametre a vlastnosti systému

Pred samotným návrhom zariadenia boli stanovené základné požiadavky, ktoré definoval inovačný tím v mojej práci. Systém mal spĺňať niekoľko kľúčových kritérií: mal obsahovať teplotné senzory s viditeľným výstupom pre používateľa, umožňovať prepínanie režimov Peltierovho prvku medzi chladením, ohrevom a vypnutím, a zároveň mal byť celý systém prenosný a napájaný z vlastného zdroja bez potreby pripojenia do siete, aby mohol samostatne fungovať.

Na základe týchto požiadaviek som si ešte pred samotnou realizáciou spracovala technický prieskum, ktorého cieľom bolo navrhnúť čo najjednoduchšie a zároveň funkčné riešenie. V rámci neho som si určila zoznam potrebných komponentov a definovala ich vzájomné prepojenia.

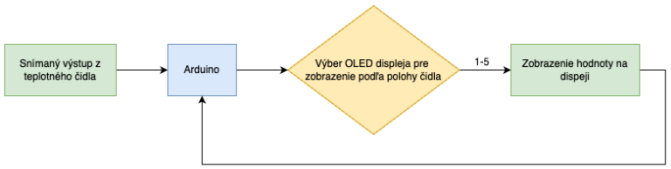
Mozgom celého systému sa stalo Arduino Nano. Dôvod jeho výberu spočíva v jeho nízkej cene a jednoduchej programovateľnosti.

Pre meranie teploty som si zvolila modul s digitálnym teplotným čidlom DS18B20, ktorý mi poskytuje dostatočnú presnosť merania teploty pre účely verifikácie funkčnosti vlákna. Výhodou tohto senzora je aj použitie OneWire protokolu, vďaka ktorému môžem, aj pri viacerých senzoroch, posielať dáta do Arduina prostredníctvom jediného dátového vodiča.

Na zobrazenie hodnôt z teplotného čidla som použila 0.96″ OLED displej, ktorého rozmery sú postačujúce pre dobrú čitateľnosť nameraných hodnôt. Zároveň ponúka množstvo možností na úpravu spôsobu zobrazovania.

Keďže som vo svojom riešení plánovala použiť viacero Peltierových článkov – konkrétne päť kusov – a zároveň ku každému priradiť vlastné teplotné čidlo a OLED displej, narazila som na problém, že všetky displeje majú z výroby rovnakú I2C adresu. Aby som zabezpečila ich individuálnu komunikáciu s Arduinom, rozhodla som sa použiť 8-kanálový multiplexer TCA9548A. Ten umožňuje pomocou I2C protokolu adresovať displeje jednotlivo, aj napriek tomu, že majú rovnaké IP adresy.

Pre napájanie týchto častí obvodu som sa rozhodla použiť UPS 18650 battery shield v konfigurácii 1P, do ktorého som vložila nabíjateľnú batériu s kapacitou 2800 mAh. Toto riešenie mi umožňuje spoľahlivo napájať túto časť systému ktorej komponenty pracujú pri napätí 5V.



Obr. 1 – Bloková schéma zobrazovania na displeji

Po výbere komponentov na meranie teploty som pristúpila k výberu tých, ktoré budú slúžiť na ovládanie stavu Peltierovho prvku.

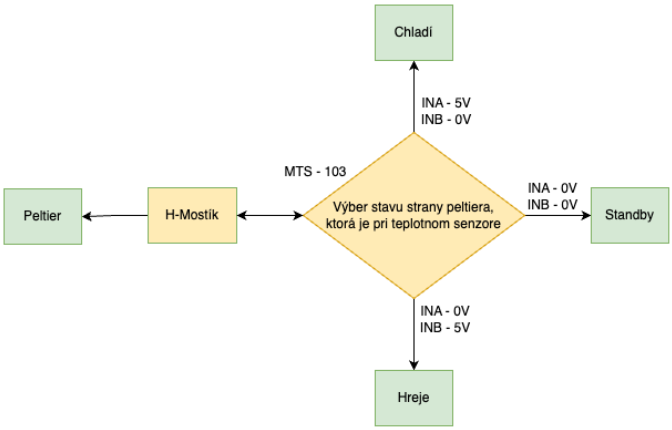
Jedným z kľúčových rozhodnutí bol výber napájacieho zdroja, keďže systém mal byť prenosný, nezávislý od elektrickej siete a zároveň sa očakávalo, že zvládne nepretržitú prevádzku minimálne počas jednej hodiny. Po porovnaní viacerých možností som sa rozhodla pre 12 V olovený akumulátor, a to namiesto často používaných modelárskych Li-Po batérií.

Aj keď je olovený akumulátor výrazne ťažší než Li-Po alternatívy, v praxi to nepredstavuje zásadný problém – zariadenie bude počas používania väčšinou umiestnené na pevnom podklade, napríklad na stole. V prípade potreby presunu medzi pracoviskami je možné akumulátor jednoducho preniesť v rukách, takže podmienka prenosnosti zostáva splnená.

Výhodou zvoleného akumulátora je, že v kombinácii s použitými komponentmi umožňuje systému pracovať bez prerušenia viac než jednu hodinu, čím sa zároveň spĺňa aj požiadavka na výdrž. Jeho celková kapacita je 240 Wh, pričom predpokladaný výkon systému dosahuje približne 218 W.

Ďalším dôležitým krokom bol výber komponentu, ktorý umožní zmenu polarity prúdu, čo je nevyhnutné pre správne prepínanie teplotného režimu Peltierovho článku medzi chladením a ohrevom. Ako riešenie som zvolila modul H-mostíka VNH2SP30, pomocou ktorého možno cez vstupy INA a INB meniť smer prúdu pretekajúceho cez článok. Zmenu polarity dokážem zabezpečiť pripojením trojpolohového spínača MTS-103 na vstupy INA a INB. V závislosti od polohy spínača sa aktivuje jeden z kanálov H-mostíka, čím dôjde k zmene smeru prúdu a tým aj k zmene teplotného režimu Peltierovho prvku.

Pre svoj systém som si zvolila Peltierov článok TEC1-017080S, ktorý má kompaktné rozmery 15 × 15 mm. Na zníženie napätia privádzaného na článok som použila step-down menič HX-M404. Tento menič má maximálny výstupný prúd 8 A, čo zodpovedá maximálne povolenému prúdu pre zvolený Peltierov článok. Vďaka regulácii výstupného napätia môžem riadiť intenzitu zmeny teploty a zároveň znížiť energetickú spotrebu, čím predĺžim životnosť samotného Peltierovho prvku.



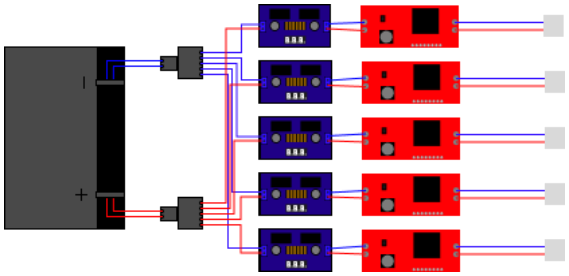
Obr. 2 – Bloková schéma ovládania polarity prúdu na Peltierovom prvku

3 Konštrukcia

V tejto kapitole predstavím jednotlivé časti navrhovaného systému z pohľadu hardvérového riešenia, konštrukcie, komunikácie senzorov a zobrazovacích jednotiek.

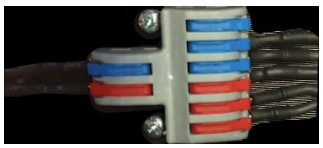
3.1 Návrh hardvérového riešenia

V úvodnej fáze návrhu systému som vypracovala schému zapojenia komponentov na ovládanie Peltierovho prvku, ktorá zachytáva základnú predstavu o usporiadaní a prepojení jednotlivých komponentov.



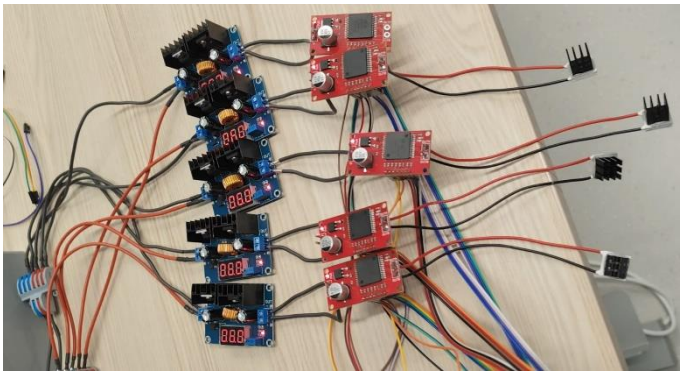
Obr. 3 – Schéma zapojenia komponentov na ovládanie Peltierovho prvku

Na distribúciu 12 V a spoločnej zeme ku všetkým komponentom som využila klon WAGO svorky, ktorý umožňuje jednoduché a bezpečné rozvetvenie napájania. Vďaka tomu môžu byť všetky step-down meniče napájané paralelne z jedného zdroja.



Obr. 4 – Wago svorka s napájacími káblami

Následne som káblami prepojila step-down meniče s H-mostíkmi, pričom na výstupy H-mostíkov som napojila Peltierove články.



Obr. 5 – Zapojené komponenty na ovládanie Peltierovho prvku

Po tom, čo som overila funkčnosť jednotlivých komponentov, som pristúpila k pripájaniu trojpolohových prepínačov MTS-103 k H-mostíku s cieľom otestovať prepínanie polarity prúdu pre Peltierov článok.

Následne som zistila, že nemôžem H-mostík len prepojiť s prepínačom, ale je potrebné správne nastaviť všetky jeho vstupy, aby fungoval spoľahlivo. Keďže som chcela zachovať manuálne ovládanie, nepoužila som Arduino na ovládanie logiky, ale len ako zdroj stabilného napätia 5 V a GND.

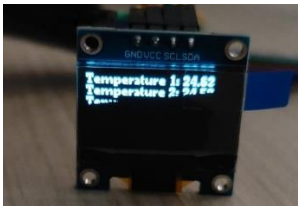
Zapojenie H-mostíka som riešila nasledovne. Pin 5V na module som pripojila na výstup 5 V. Pin GND som pripojila na výstup GND. Pin EN (enable), ktorý zabezpečuje aktiváciu H-mostíka, som pripojila na 5 V, aby bol H-mostík stále aktívny. Pin CS (current sense), ktorý slúži na monitorovanie prúdu, som nepotrebovala, a preto som ho pripojila na GND. Vstupy INA a INB som pripojila na výstupy trojpolohového prepínača MTS-103. Každá krajová poloha spínača zabezpečuje logickú úroveň HIGH (5 V) pre jeden z týchto vstupov, zatiaľ čo druhý zostáva v logickej nule. V strednej polohe sú oba vstupy v stave LOW, čo spôsobí zastavenie prúdu. Pin PWM, ktorý riadi výkon (vstup pre pulzne šírkovú moduláciu), som nastavila napevno na 5 V, čím som zabezpečila trvalý výstupný výkon bez potreby PWM riadenia.

Pri takto navrhnutom zapojení zmena polarity prúdu prebiehala podľa očakávaní, čím bola táto časť konštrukcie úspešne dokončená.

3.2 Komunikácia senzorov s Arduinom a zobrazovanie údajov

Po zostavení základnej hardvérovej časti systému som začala s pripájaním teplotných senzorov DS18B20 k Arduino Nano. Následne som vytvorila jednoduchý testovací program, pomocou ktorého som overila ich funkčnosť a správnosť výstupných hodnôt.

Po úspešnej verifikácii funkcie senzorov som pokračovala s implementáciou výstupu nameraných údajov na OLED displej. V prvej fáze som pripojila jeden displej, aby som otestovala spôsob zobrazovania dát a overila správnosť komunikácie.



Obr. 6 – Test výpisu teploty na displej

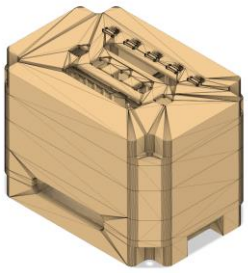
Následne som upravila program tak, aby dokázal zbierať dáta zo všetkých piatich senzorov. Zatiaľ, predtým než bude sprevádzkovaná komunikácia displejov prostredníctvom multiplexera, som zabezpečila výpis nameraných hodnôt aspoň na dvoch OLED displejoch. Hoci majú tieto displeje z výroby rovnakú I2C adresu, tento problém sa dá dočasne vyriešiť prepájaním pájkovacích bodov na zadnej strane displeja, čím sa zmení jeho adresa na druhú preddefinovanú možnosť, ktorú displej podporuje.



Obr. 7 – Výpis teploty z dvoch teplotných čidiel na OLED displeje

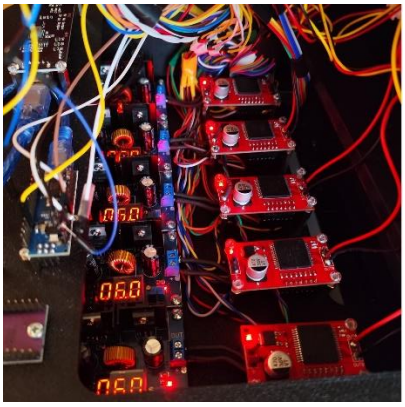
3.3 Konštrukcia boxu a osádzanie komponentov

Prvý model boxu, do ktorého budú umiestnené súčiastky, mi pomohol navrhnúť kolega. Spoločne sme vytvorili úvodné nákresy s cieľom navrhnúť riešenie, ktoré by umožnilo pohodlne umiestniť celú elektroniku. Pri návrhu sme kládli dôraz na to, aby bol box čo najkompaktnejší, no zároveň poskytoval dostatočný priestor na ventiláciu komponentov a zabránil ich prehrievaniu. Kolega následne pripravil niekoľko počiatočných modelov, ktoré som schválila, a finálnu verziu sme následne odoslali do 3D tlače.

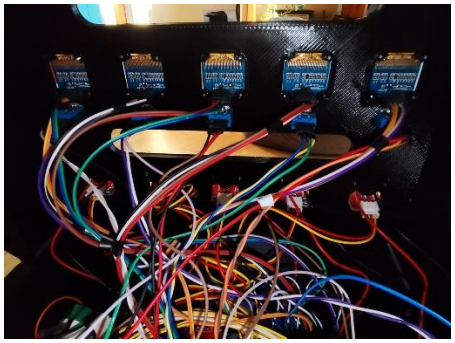


Obr. 8 – Hotový model boxu

Po doručení finálne vytlačeného boxu som začala s osádzaním elektronických súčiastok do jeho vnútornej konštrukcie.



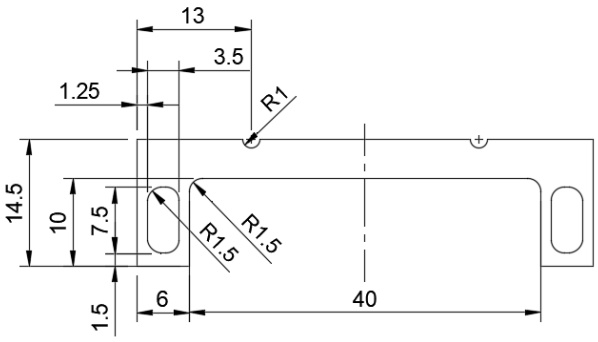
Obr. 9 – Step-Down meniče a H-mostíky v boxe



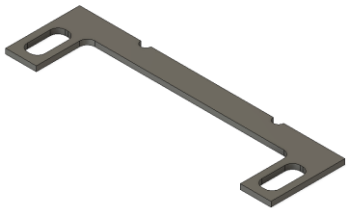
Obr. 10 – Zapojené OLED displeje, trojpolohové spínače a teplotné čidlá

3.4 Držiaky pre Peltierove prvky

Na stabilné uchytenie Peltierových prvkov ku boxu som v programe Autodesk Fusion 360 navrhla držiaky. Na základe 3D modelu som vytvorila technický výkres, ktorý som odovzdala na výrobu.

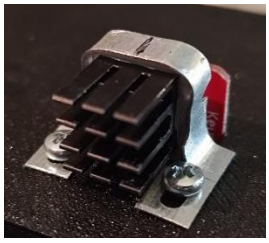


Obr. 11 – Parametre držiaku



Obr. 12 – 3D model držiaku

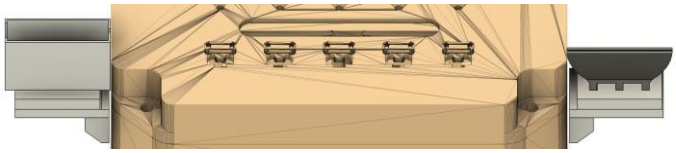
Keď boli hotové tak som ich poohýbala a pripevnila na box.



Obr. 13 – Pripevnený držiak

3.5 Držiak pre optické vlákno

Keďže meranie pomocou optických senzorov je citlivé na geometriu vlákna, bolo nevyhnutné navrhnúť držiaky, ktoré zabezpečia jeho správne napnutie a polohu. Vytvorila som návrh držiakov, ktoré umožňujú stabilné uchytenie koncov vlákna. Jeden koniec je pripevnený ku boxu, z ktorej vlákno vystupuje, a druhý je zachytený v samostatnom držiaku tak, aby bolo vlákno dostatočne napnuté v súlade s požadovanou geometriou.



Obr. 14 – Modely držiakov na vlákno



Obr. 15 – Vytlačený držiak v ktorom je umiestnený box z ktorého je vedené vlákno

3.6 Aktuálny stav

V súčasnosti je prvotný prototyp systému takmer dokončený. Aktuálne je možné zobrazovať teplotu na dvoch z piatich OLED displejov, pričom na rozšírení zobrazovania na všetky displeje intenzívne pracujem a očakávam jeho skoré

dokončenie. Z hľadiska funkčnosti už systém plnohodnotne ovláda všetkých päť Peltierových článkov prostredníctvom H-mostíka, vrátane prepínania medzi režimami chladenia a ohrevu. Rovnako tak všetkých päť teplotných senzorov úspešne sníma teplotu.

Posledným krokom v rámci funkčného hardvéru je spojazdnenie I2C multiplexera, ktorý umožní zobrazovanie meraných hodnôt zo všetkých senzorov na samostatných OLED displejoch súčasne.

Paralelne s tým momentálne pracujem na návrhu a výrobe prvého prototypu dosky plošných spojov, ktorá bude použitá v tomto systéme. Jej implementáciou sa výrazne zníži priestorová náročnosť kabeláže a zároveň sa zjednoduší celkové zapojenie, ktoré je v súčasnosti riešené cez tri nepájivé polia.



Obr. 16 – Aktuálny obrázok boxu

3.7 Plány na vylepšenie systému

Jedným z hlavných aspektov, ktoré plánujem v budúcnosti optimalizovať, je hmotnosť systému. Určite by bolo možné použiť menší a ľahší akumulátor, ktorý by spĺňal rovnaké napájacie požiadavky za podobnú cenu. Alternatívou môže byť aj vlastná výroba batériového zdroja zapojením viacerých článkov do série.

Zmenšenie napájacieho zdroja by zároveň umožnilo zmenšiť rozmery boxu, v ktorej je aktuálne celý systém osadený, čo by výrazne zlepšilo jeho prenosnosť. Hoci už súčasný prototyp bol navrhnutý s dôrazom na jednoduchý prenos, do budúcnosti plánujem ďalšie zníženie hmotnosti a celkových rozmerov zariadenia.

S touto úpravou úzko súvisí aj potreba návrhu dosky, prípadne viacerých dosiek plošných spojov, ktorá by umožnila efektívnejšie usporiadanie komponentov pri minimálnych priestorových nárokoch. Zároveň by sa tým eliminoval problém s nadmerným množstvom prepojovacích káblov. Plánujem vytvoriť obojstranný návrh DPS, ktorý by umožnil lepšie využitie vnútorného priestoru nového, kompaktnejšieho riešenia.

4 Záver

V tomto článku som sa venovala návrhu a realizácii referenčného systému na overovanie presnosti merania optických vláknových senzorov. Projekt bol náročný nielen po technickej, ale aj po organizačnej stránke, no zároveň mi priniesol veľa nových skúseností. Počas jeho realizácie som si prešla celým procesom od prvotného návrhu, výberu komponentov, cez programovanie až po finálnu montáž a testovanie. Výsledkom je funkčný, skoro hotový, kompaktný a prenosný systém, ktorý umožňuje simulovať rôzne teplotné podmienky, porovnať výstupy z optického vlákna so štandardnými senzormi a overiť tak správnosť merania aj nastavenie geometrie vlákna. Verím, že získané poznatky budem môcť využiť aj pri ďalších technických výzvach.

Poďakovanie

Chcela by som v prvom rade poďakovať kolegovi Petrovi Polákovi, ktorý mi pomohol so zabezpečením všetkých potrebných komponentov. Veľká vďaka patrí aj kolegovi Danielovi Chromekovi za pomoc pri prototypovaní modelu boxu. Ďakujem tiež doc. Ing. Richardovi Baloghovi, PhD. za rady a pomoc pri riešení technických otázok, ktoré počas práce vznikli. A nakoniec by som chcela poďakovať aj doc. Ing. Vladimírovi Gogovi, PhD. za 3D tlač držiakov na vlákno a výrobu oceľových držiakov na Peltierove články.

Literatúra

[1] Arduino Nano Datasheet [online]. Dostupné na odkaze: https://docs.arduino.cc/resources/datasheets/A000005-datasheet.pdf

[2] TEC1-017080-15X15 W150mm Datasheet [online]. Dostupné na odkaze: https://img.gme.cz/files/eshop_data/eshop_data/10/601-053/dsh.601-053.1.pdf

[3] Automotive fully integrated H-bridge motor driver datasheet [online]. Dostupné na odkaze: https://www.st.com/content/ccc/resource/technical/document/datasheet/group2/66/b8/f5/2c/9a/66/41/c7/CD00043711/files/CD00043711.pdf/jcr:content/translations/en.CD00043711.pdf

[4] DS18B20 datasheet [online]. Dostupné na odkaze: https://www.analog.com/media/en/technical-documentation/data-sheets/ds18b20.pdf

[5] WA12200 datasheet [online]. Dostupné na odkaze: https://img.gme.cz/files/eshop_data/eshop_data/1/540-522/dsh.540-522.2.pdf

[6] Guide for DS18B20 Temperature Sensor with Arduino [online]. Dostupné na odkaze: https://randomnerdtutorials.com/guide-for-ds18b20-temperature-sensor-with-arduino/

[7] Tutorial for VNH2SP30 Monster Motor Module (Single Channel) [online]. Dostupné na odkaze: https://www.instructables.com/Tutorial-for-VNH2SP30-Monster-Motor-Module-Single-/

[8] Guide for I2C OLED Display with Arduino [online]. Dostupné na odkaze: https://randomnerdtutorials.com/guide-for-oled-display-with-arduino/

Guľôčka na ramene

Daniel Žula

Ústav automobilovej mechatroniky, FEI STU v Bratislave

xzula@stuba.sk

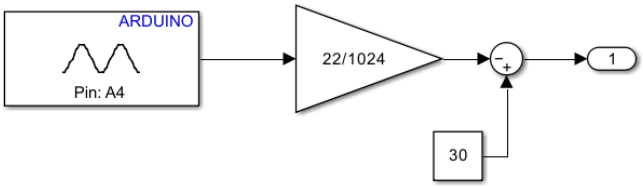
Abstrakt – Cieľom článku je zhotovenie a implementácia riadenia na reálnom laboratórnom zariadení Guľôčka na ramene (Ball and Beam). Článok prezentuje stavbu zariadenia, experimentálne navrhnutý PID regulátor na overenie jeho funkčnosti, porovnanie matematického modelu na základe parametrov zariadenia s nameranou prechodovou charakteristikou, využitie tohoto modelu pri návrhu stavového regulátora pomocou metódy rozmiestňovania pólov a následné otestovanie týchto regulátorov na reálnom zariadení a ich porovnanie so simuláciou.

1 Úvod

Riadenie dynamických systémov patrí medzi kľúčové oblasti automatizácie a mechatroniky, pričom cieľom pri tomto laboratórnom zariadení bolo overenie a porovnanie teoretických poznatkov v praxi. Motiváciou tejto práce bolo vytvoriť lacné, ľahko ovládateľné a vizuálne rýchlo pochopiteľné zariadenie, ktoré demonštruje základné princípy regulácie polohy. S využitím bežne dostupných komponentov, sme skonštruovali systém "Guľôčka na ramene" – jednoduché zariadenie, kde poloha guľôčky na naklápajúcej sa koľajnici je riadená pomocou signálu ovládajúceho servomotor. Výstupom systému je poloha guľôčky, meraná infračerveným senzorom. Tento prístup umožnil nielen overenie teoretických konceptov, ale aj praktickú implementáciu riadenia reálneho laboratórneho zariadenia v prostredí MATLAB-Simulink.

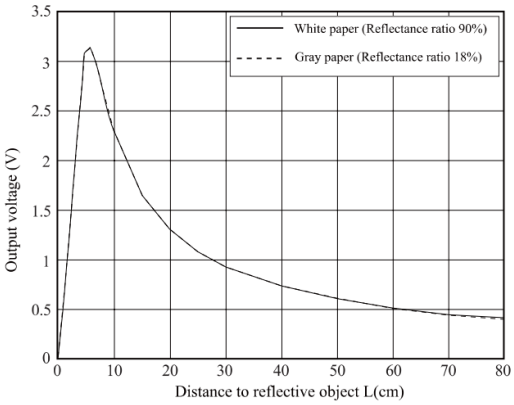
2 Laboratórne zariadenie

Laboratórne zariadenie sa skladá z infračerveného Time-of-Flight senzora, servomotora, mikropočítaču Arduino Nano 3.0 a posuvného potenciometra, ktorým nastavujeme žiadanú polohu guľôčky pomocou posuvného kontaktu. Výstupom tohto potenciometra je napätie ktoré Arduino prijíma v rozsahu $0 - 5\,V$, čo v analógovom vstupe zodpovedá hodnotám $0 - 1023$[1]. Aby žiadaná hodnota zodpovedala žiadanej polohe guľôčky v centimetroch, upravili sme napätie z potenciometra podľa obrázku 1.



Obr. 1: Zapojenie prevodníka pre vstupnú požadovanú hodnotu.

Arduino Nano 3.0 je mikrokontrolér s čipom ATMega328P, ktorý pracuje pri napätí 5 V a vie snímať digitálne aj analógové hodnoty vstupov. Jeho výstupom môže byť digitálny signál alebo analógový signál vytvorený pomocou PWM[2]. Arduino programujeme pomocou softvérového prostredia MATLAB-Simulink, ktoré pomocou rozšírenia Simulink Support Package for Arduino Hardware dokáže vygenerovať kód a nahrať ho do zariadenia priamo z blokovej schémy. Komponenty nášho zariadenia sme s blokovou schémou v prostredí Simulink prepojili pomocou analógových vstupných blokov, ktoré tieto komponenty po ich pripojení umožnili používať priamo v Simulinkovej schéme. Náš senzor vzdialenosti, ktorým snímame polohu guľôčky je Sharp GP2Y0A21YK0F. Tento senzor pracuje na princípe odrazu infračerveného svetla a je zložený z diódy, ktorá vyšle infračervené svetlo a zo snímača, ktorý meria odraz tohto svetla. Výhodou tohoto senzora je, že môže byť priamo napájaný z Arduina a zároveň má analógový výstup ktorý je možné pomocou Arduina jednoducho spracovať. Rozsah senzora je 10 – 80 centimetrov, kvôli čomu senzor musel byť umiestnený tak, aby v prípade, keď je guľôčka najbližšie, bola medzi nimi vzdialenosť aspoň 8 centimetrov, pretože ak by bol bližšie tak napätie by mohlo zodpovedať dvom rôznym vzdialenostiam, ako môžeme vidieť na obrázku 2. Na tomto obrázku zároveň môžeme vidieť, že odrazivosť svetla od objektu nemá takmer žiadny vplyv na výslednú hodnotu[3]. V praxi sa ale ukázalo, že biela penová guľôčka je lepšia voľba oproti lesklej kovovej, pretože pri kovovej guľôčke bol výstup zo senzora od určitej vzdialenosti rovnaký a nevedeli sme využiť plný rozsah koľajnice.



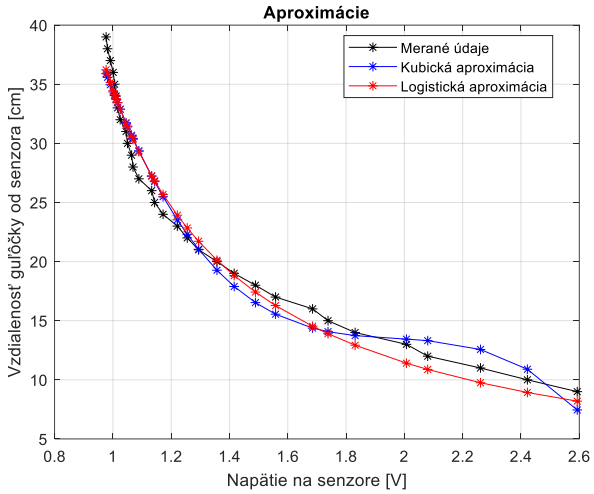
Obr. 2: Závislosť vzdialenosti objektu od výstupného napätia.[3]

Výstupom zo senzora bol veľmi zašumený signál ktorý mal v pravidelných intervaloch šum ktorý bol niekoľkonásobne väčší ako samotný signál. Ako riešenie tohoto problému sme medzi senzor a Arduino paralelne zapojili kondenzátor ktorý sa medzi pulzami senzora priebežne dobíja a umožňuje senzoru využiť väčší okamžitý prúd, než aký môže dodať Arduino. Medzi Arduino a kondenzátor sme pridali rezistor aby sme zabránili vysokým elektrickým prúdom počas toho ako sa kondenzátor nabíja. Tento kondenzátor pomohol výrazne znížiť veľkosť tohoto šumu, ale nedokázal ho úplne vyhladiť. Preto sme navrhli dolnopriepustný filter štvrtého rádu. Pre tento filter sme experimentálne odhadli jeho rád aj časovú konštantu $T_f = 0.011$ a zadefinovali sme ho pomocou funkcie *tf* v MATLABE. Následne sme náš filter pomocou funkcie *c2d* pre našu vzorkovaciu periódu $T_{VZ} = 0.002$ prepočítali na jeho diskrétnu formu, ktorú sme v Simulinku aplikovali pomocou bloku Discrete Filter.

Výstup zo senzora je napätie, no my chceme vyjadriť polohu guľôčky v centimetroch. Preto sme pomocou aproximácie vytvorili prevodník, ktorý prekonvertuje namerané napätie na centimetre. Hodnoty potrebné na aproximáciu sme zistili tak, že sme guľôčku posúvali po koľajnici po centimetrových krokoch a pre každú vzdialenosť v centimetroch sme odčítali napätie zo senzora. Toto sme zopakovali na celej dĺžke koľajnice na rozsahu 8 – 35 centimetrov vzdialenosti od senzora. Skúšali sme vhodnosť viacerých druhov aproximácie tejto závislosti. Prvá aproximácia bola polynóm 3. rádu.

$$y = ax^3 + bx^2 + cx + d \tag{1}$$

Ako môžeme vidieť na obrázku 3, približne od $1{,}8\,V - 2{,}2\,V$ sa aproximovaná vzdialenosť pri zmenách napätia výrazne nemenila. Toto pásmo necitlivosti spôsobilo, že aproximácia pre náš regulačný obvod nebola vhodná. Ak by sme ju použili vznikla by na koľajnici približne 5cm dlhá zóna s minimálnou citlivosťou. Preto sme použili logistickú aproximáciu, ktorej parametre sme vypočítali pomocou funkcie *lsqcurvefit*. Na to aby sme ju vedeli spraviť sme potrebovali približne odhadnúť východiskový vektor parametrov $P0$. Z týchto parametrov sme následne získali optimálne parametre P.



Obr. 3: Porovnanie kubickej a logistickej aproximácie.

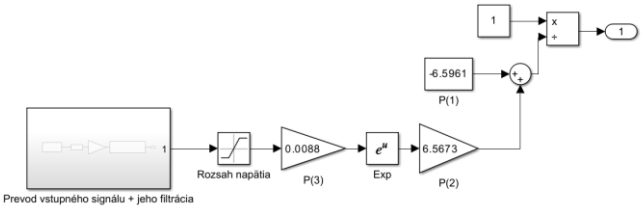
Táto aproximácia pre náš účel vhodnejšia lebo vzdialenosť s rastúcim napätím klesá na celom intervale a nie je tam žiadne pásmo necitlivosti. Logistická funkcia má tvar:

$$y = \frac{1}{p_1 + p_2 * e^{p_3 * x}} \tag{2}$$

Výsledný vektor parametrov $P = [-6.5961 \quad 6.5673 \quad 0.0088]$ sme využili v Simulinku v prevodníku, ktorý spolu s diskrétnym filtrom odhaduje z napätia polohu guľôčky. Na to, aby sme vedeli pracovať s výstupom analógového vstupu, musíme pomocou bloku Data Type Conversion previesť tento výstup z dátového typu Int16 na Double, čo nám umožní pracovať s reálnymi číslami. Signál z analógového vstupu má rozsah $0 - 1023$. Túto hodnotu prekonvertujeme na napätie v rozsahu $0 - 5\,V$. Ďalší problém, ktorý sme museli vyriešiť je, že výstup z diskrétneho filtra je v prvej vzorke 0 a teda jeho výstup z prevodníka je nekonečno, čo narúša funkčnosť obvodu a preto je potrebné medzi filter a prevodník vložiť saturáciu kde maximum a minimum udáva dosiahnuteľný rozsah nášho napätia, s ktorým pracujeme, od $0{,}9766\,V$ do $2{,}593\,V$.

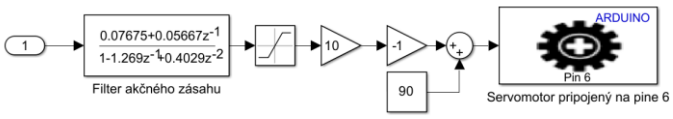


Obr. 4: Dátová konverzia, prevod na napätie a filtrácia vstupného signálu zo senzora.



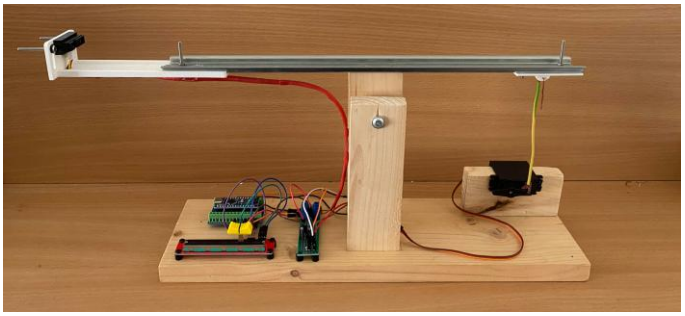
Obr. 5: Prevodník medzi napätím (V) a vzdialenosťou (cm).

Posledný komerčný komponent v našom regulačnom obvode je servomotor MG996R. Tento servomotor slúži na naklonenie ramena. Jeho presnú polohu vieme nastaviť podľa riadiaceho PWM signálu. Tento aktuátor kvôli jeho výkonovej náročnosti nie je napájaný priamo z Arduina ale je napájaný zo samostatného externého zdroja. Tento servomotor dokáže operovať v uhloch $0° - 180°$. Tento rozsah vieme ovládať pomocou bloku Standart Servo Write, ktorý očakáva vstup v rovnakom rozsahu. Aby servomotor pri nulovej požadovanej zmene polohy držal koľajnicu vo vodorovnej polohe, potrebovali sme v Simulinku nastaviť prevodník tak, aby pri nulovom vstupe bol servomotor natočený o 90°, čo je vodorovná poloha koľajnice. Zároveň sme nechceli aby pri vysokých zmenách polohy pracoval servomotor na plnom rozsahu aj kvôli tomu, že prepojenie medzi ním a koľajnicou je urobené pomocou rovného drôtu a v prípade maximálneho vytočenia by došlo k jeho pokriveniu. Optimálne riešenie bolo nastaviť rozsah zmeny od $-40°$ do $40°$, čiže pracovný rozsah servomotora bol od $50°$ do $130°$. Po odmeraní sme zistili, že natočenie koľajnice o 1 stupeň vykonáme otočením servomotora o 10 stupňov. Aby sme to vedeli implementovať ako prevodník, tak sme pred vstupom do servomotora umiestnili saturáciu akčného zásahu od $-4°$ do $4°$, čo zodpovedá uhlu natočenia koľajnice. Pri prevode želaného akčného zásahu (sklon koľajnice) na uhol natočenia servomotora bolo potrebné hodnotu akčného zásahu zo saturácie prenásobiť číslom 10. Akčný zásah je ešte potrebné prenásobiť hodnotou -1 pretože prenosová funkcia tohto modelu je záporná a bez tejto zmeny by regulačný obvod s regulátorom navrhnutým pre kladnú prenosovú funkciu reagoval naopak. Vzhľadom na to, že sme pri nulovom želanom akčnom zásahu chceli mať servomotor v polohe 90°, sme k výstupu z tohoto prevodníka pripočítali 90. Pre hladší priebeh sme akčný zásah filtrovali pomocou diskrétneho filtra 2. rádu, kde sme experimentálne odhadli časovú konštantu $Tf = 0.12$.



Obr. 6: Zapojenie filtra akčného zásahu s prevodníkom pre servomotor.

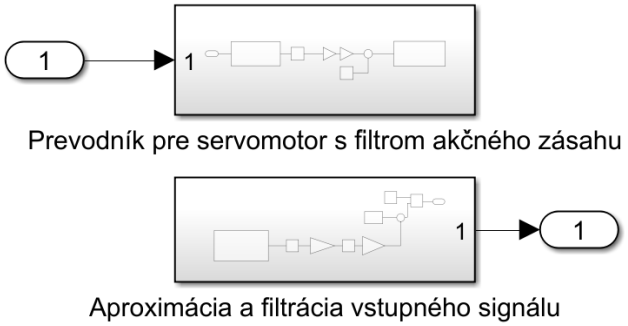
Zhotovený prototyp laboratórneho zariadenia sa nachádza na obrázku 7.



Obr. 7: Laboratórne zariadenie.

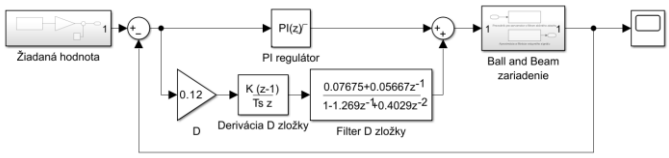
3 Experimentálne navrhnutý PID regulátor

Návrh všetkých prevodníkov a filtrov umožnil vytvorenie uzavretého regulačného obvodu so spätnou väzbou, do ktorého sme umiestnili PID regulátor. To sme realizovali pomocou bloku PID controller v ktorom sme nastavili PI regulátor a D zložku sme pridali ako paralelnú vetvu, aby sme vedeli vyfiltrovať samostatne aj akčný zásah D zložky. Tento regulačný obvod je na obrázku 9. Tu sme ešte raz použili rovnaký filter ako pri filtrovaní akčného zásahu. Pri vzorkovacej frekvencii 200 vzoriek za sekundu sa vzhľadom na rýchlosť systému v uzavretom regulačnom obvode negatívne neprejavilo ani oneskorenie vytvorené tromi filtrami. Podobne ako pri diskrétnom filtri, toolbox pre Arduino nepodporuje spojité bloky ani v prípade PID regulátora ale iba jeho diskrétnu verziu. Vďaka našej veľmi malej perióde vzorkovania ho môžeme navrhnúť ako keby bol spojitý. Kvôli zjednodušeniu Simulinkovej schémy sme všetky prevodníky a filtre skryli do subsystémov ako na obrázku 8.



Obr. 8: Blok „Ball and Beam“ obsahujúci prevodník pre servomotor a prevodník signálu zo senzoru s filtráciou signálu ako jeden blok.

Tieto dva bloky boli následne vložené do spoločného bloku určeného na komunikáciu so zariadením.



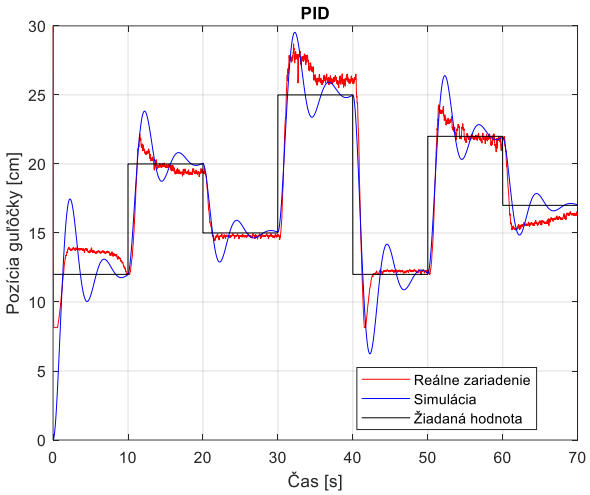
Obr. 9: Celkové zapojenie pomocou bloku PID Controller.

Parametre P, I, D sme určili experimentálne heuristickou metódou. Tieto zložky sme ladili dovtedy, kým sme nedosiahli želané správanie obvodu.

Tab. 1. Parametre bloku PID regulátora

č.	Názov parametra	Hodnota parametra
1	P	0,3
2	I	0,05
3	D	0,12

Simuláciu aj meranie na reálnom zariadení sme urobili na sérii rovnakých kladných a záporných skokov. Schéma pre simuláciu PID regulátora je zhodná so schémou na obrázku 9, ale namiesto bloku „Ball and Beam“ je použitá identifikovaná prenosová funkcia z rovnice 7[4].



Obr. 10: Porovnanie regulácie pomocou bloku PID Controller v simulácií a na reálnom zariadení.

4 Identifikácia modelu

Matematický model sme vypočítali na základe parametrov systému podľa všeobecných fyzikálnych vzorcov [6]. Konkrétne hodnoty sú v tabuľke 2.

Tab. 2: Parametre

č.	Parameter	Hodnota
1	m Hmotnosť guľôčky	$0{,}013\ kg$
2	R Polomer guľôčky	$0{,}025\ m$
3	g Gravitačné zrýchlenie	$-9{,}81\ ms^{-2}$

Moment zotrvačnosti guľôčky vieme vypočítať podľa všeobecného vzorca z našich parametrov.

$$J = \frac{2}{5}mR^2 = \frac{2}{5} * 0{,}013kg * (0{,}025m)^2 = 3{,}25 * 10^{-6}\ kgm^2 \quad (3)$$

Následne vieme zistiť linearizovanú prenosovú funkciu nášho modelu na základe fyzikálnych zákonov ako v [5].

$$\left(\frac{J}{R^2} + m\right)\ddot{s} = -mg \quad (4)$$

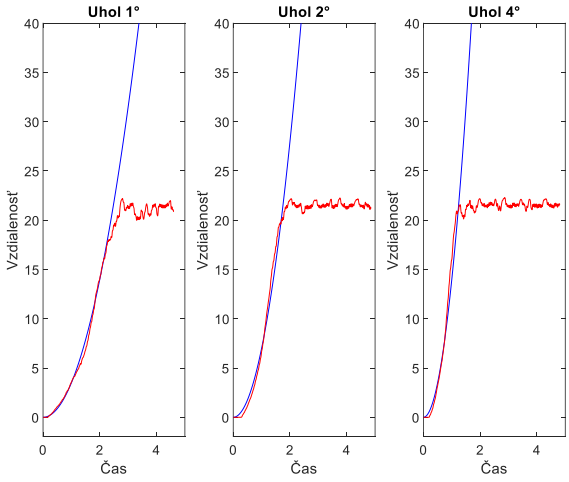
Po použití Laplaceovej transformácie vznikne finálna prenosová funkcia.

$$G(s) = -\frac{mg}{\frac{J}{R^2} + m}\frac{1}{s^2} \quad (5)$$

Pre overenie našej matematicky získanej prenosovej funkcie sme systém identifikovali aj priamo z jeho reakcie na skokovú zmenu akčného zásahu. Tvar prenosovej funkcie je dvojitý integrátor a všetky parametre systému spolu tvoria jednu konštantu, ako vidíme v rovnici 5. Túto konštantu nahradíme parametrom H, ktorý budeme identifikovať.

$$G(s) = \frac{H}{s^2} \quad (6)$$

Prechodová charakteristika prenosovej funkcie v rovnici 6 bude pri konštantnom akčnom zásahu rásť smerom do nekonečna. Skokovú zmenu akčného zásahu sme na našom reálnom zariadení dosiahli tak, že sme vyrovnali koľajnicu do vodorovnej polohy, na jej začiatok umiestnili guľôčku a postupne sme koľajnicu naklonili o 1, 2 a 4 stupne a zaznamenali sme jej pohyb. Do tejto nameranej prechodovej charakteristiky sme sa experimentálnym dosadzovaním rôznych hodnôt za konštantu H snažili nájsť prenosovú funkciu s podobnou prechodovou charakteristikou pri všetkých uhloch naklonenia.



Obr. 11: Porovnanie prechodových charakteristík identifikovanej prenosovej funkcie s prechodovými charakteristikami guľôčky na reálnom zariadení pri rôznych uhloch pre H = 7.[4]

Keď porovnáme priebehy identifikovanej prenosovej funkcie (modrá) a priebehy reálneho zariadenia (červená) tak môžeme vidieť, že pri všetkých uhloch nám sedela hodnota H = 7. To znamená že naša identifikovaná prenosová funkcia má tvar:

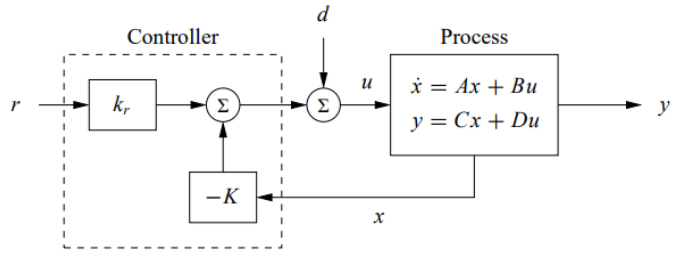
$$G(s) = \frac{7}{s^2} \quad (7)$$

Pri porovnaní prechodovej charakteristiky reálneho systému a charakteristiky získanej prenosovej funkcie môžeme pozorovať, že sú zhodné. To znamená že matematický model je správny a zhoduje sa s našim identifikovaným modelom.

5 Stavový regulátor

Ako druhý spôsob ovládania obvodu sme zvolili stavový regulátor. Princíp stavového riadenia je založený na využití spätnej väzby z výstupu, podobne ako pri iných typoch regulátorov. Na rozdiel od regulátorov typu PID však spätná väzba nepracuje priamo s výstupom, ale s vektorom stavových premenných x. Tieto premenné sa pomocou vektora spätnoväzobných zosilnení K transformujú do jednej hodnoty, ktorá sa spolu so žiadanou hodnotou w použije na vytvorenie akčného zásahu ovládajúceho systém.

$$u = -Kx + w \tag{8}$$



Obr. 12. Schéma stavovej spätnej väzby.[7]

V tomto systéme sa modelujú stavy (x), vstup (u) a výstup (y). Rovnice stavového priestoru sú:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + Bu(t) \\ y(t) &= Cx(t) + Du(t) \end{aligned} \tag{9}$$

V našom prípade sú stavy nášho systému nasledovné:

$$\begin{aligned} x_1 &- (poloha\ guľôčky) \\ x_2 &- (rýchlosť\ guľôčky) \end{aligned} \tag{10}$$

Po dosadení budú rovnice pohybu vyzerať nasledovne:

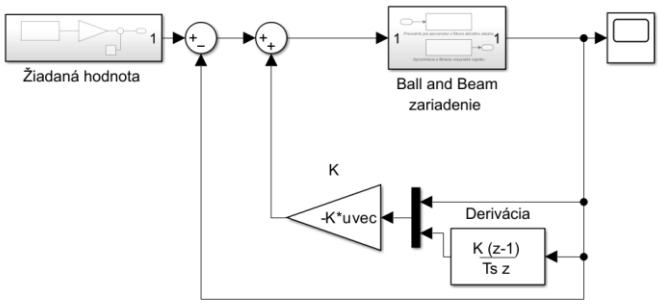
$$\begin{aligned} \dot{x}_1 &= x_2 \\ \dot{x}_2 &= -H \end{aligned} \tag{11}$$

Ak dosadíme rovnice 11 do rovníc 9 získame naše matice systému.

$$A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 \\ H \end{bmatrix} \tag{12}$$

$$C = \begin{bmatrix} 1 & 0 \end{bmatrix}$$



Obr. 13: Schéma zapojenia stavového regulátora bez I zložky.

Nedostatok stavového regulátora je že neodstraňuje trvalú regulačnú odchýlku, čo si vyžiadalo zavedenie integračnej (I) zložky. Táto zložka monitoruje regulačnú odchýlku, ktorá je do stavového systému začlenená ako dodatočný stav.[7]

$$\begin{pmatrix} \dot{x} \\ \dot{z} \end{pmatrix} = \begin{pmatrix} Ax + Bu \\ y - w \end{pmatrix} = \begin{pmatrix} Ax + Bu \\ Cx + w \end{pmatrix} \tag{13}$$

Stav z predstavuje integrál regulačnej odchýlky. V ustálenom stave platí, že $\dot{z} = 0$, pretože $y = w$. Spätná väzba stavového PI regulátora sa navrhuje podobne ako pri štandardnom stavovom regulátore, pričom na výslednú veľkosť akčného zásahu vplývajú nielen stavy obsiahnuté vo vektore x, ale aj stav z.

$$u = -Kx - Ki\, z \tag{14}$$

Rozšírením stavového systému o integrál odchýlky z, získame naše rozšírené stavové matice. Pomocou týchto matíc navrhneme stavový PID regulátor.

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}$$

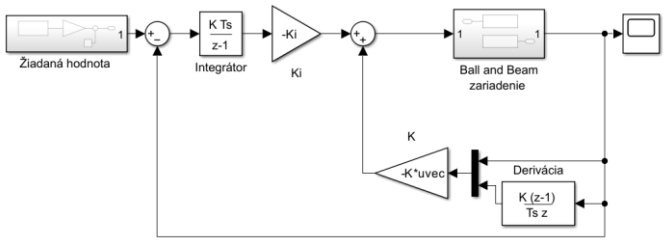
$$B = \begin{bmatrix} 0 \\ H \\ 0 \end{bmatrix} \tag{15}$$

$$C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Po pridaní I zložky sme pomocou metódy rozmiestňovania pólov pre naše zvolené póly pomocou funkcie *acker* získali parameter spätnoväzobného zosilnenia *Kroz*.

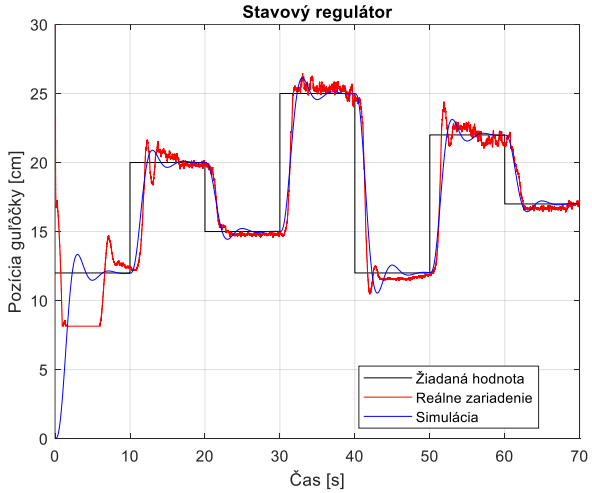
$$\begin{aligned} p_1 &= -1{,}2 \\ p_2 &= -0{,}4 + 1{,}5i \\ p_3 &= -0{,}75 - 1{,}5i \end{aligned} \tag{16}$$

Parameter *Kroz* je vektor s rozmerom 3, kde prvé 2 hodnoty sú hodnoty pre *K* a posledná hodnota matice je hodnota pre *Ki*, teda pre I zložku. Túto zložku sme následne pridali do schémy ako blok Gain spolu s blokom Discrete Integrator.



Obr. 14: Schéma zapojenia s pridaním I zložky.

Simuláciu pre stavový regulátor sme uskutočnili na sérií rovnakých skokov ako pri experimentálne navrhnutom PID regulátore, pričom schéma pre simuláciu je z zhodná so schémou na obrázku 14 ale namiesto bloku „Ball and Beam“ je použitá naša identifikovaná prenosová funkcia z rovnice 7.



Obr. 15: Porovnanie regulácie stavového regulátora v simulácií a na reálnom zariadení.

6 Záver

Na základe uskutočnených simulácií a experimentov na reálnom zariadení "Guľôčka na ramene" môžeme konštatovať, že navrhnuté regulátory – experimentálne navrhnutý PID regulátor aj matematicky navrhnutý stavový regulátor dosiahli dobré výsledky. Porovnanie priebehov regulácie v simulácii a na reálnom zariadení, ako je znázornené na obrázkoch 10 a 15, ukazuje, že oba regulátory dokázali efektívne riadiť polohu guľôčky na ramene pri kladných aj záporných skokoch, pričom oba regulátory poskytli ustálenie polohy guľôčky v prijateľnom čase a neobsahovali vysoké preregulovania. Integračná zložka zavedená do oboch regulátorov úspešne eliminovala trvalú regulačnú odchýlku. Zhoda medzi simulovanými a reálnymi prechodovými charakteristikami potvrdzuje správnosť identifikovaného matematického modelu systému. Vzhľadom na rozsah tejto práce sme sa pre stavový regulátor zamerali na metódu rozmiestňovania pólov , ktorá sa ukázala ako efektívnejšia v porovnaní s experimentálne navrhnutým PID regulátorom, pretože dosahovala rýchlejšie nastavenie polohy guľôčky a to s menšími preregulovaniami. Zariadenie bolo úspešne implementované a otestované prostredníctvom Simulinku, pričom zadávanie žiadanej hodnoty pomocou potenciometra je jednoduché a intuitívne. Napriek tomu, že identifikácia modelu a návrh regulátorov vyžadovali experimentálne ladenie a riešenie špecifických problémov, ako šum zo senzora, výsledky potvrdili funkčnosť navrhnutého riešenia ale aj potenciál pre ďalšie vylepšenia.

Literatúra

[1] Aspencore, “Electronic Tutorials”, Accessed: Mar. 26, 2025. [Online]. Available: https://www.electronics-tutorials.ws/resistor/potentiometer.html

[2] Arduino, “Arduino® Nano ESP32 Datasheet.”

[3] Sharp, “Sharp GP2Y0A21YK0F Datasheet”.

[4] P. V. M. Maalini, G. Prabhakar, and S. Selvaperumal, “Modelling and control of ball and beam system using PID controller,” in *Proceedings of 2016 International Conference on Advanced Communication Control and Computing Technologies, ICACCCT 2016*, Institute of Electrical and Electronics Engineers Inc., Jan. 2017, pp. 322–326. doi: 10.1109/ICACCCT.2016.7831655.

[5] M. F. Rahmat, H. Wahid, and N. A. Wahab, “APPLICATION OF INTELLIGENT CONTROLLER IN A BALL AND BEAM CONTROL SYSTEM.”

[6] University of Michigan, “Ball & Beam: System Modeling.” Accessed: Mar. 24, 2025. [Online]. Available: https://ctms.engin.umich.edu/CTMS/index.php?example=BallBeam&section=SystemModeling

[7] J. , M. R. Aström, “Feedback Systems,” 2009.

Aplikácia hlbokých neurónových sietí pre analýzu kožných lézií

Sabína Ovčiariková[1], Erik Kučera[1]

[1]Ústav automobilovej mechatroniky FEI STU v Bratislave

xovciarikova@stuba.sk

Diagnostika kožných lézií pomocou umelej inteligencie predstavuje významný krok k včasnej detekcii závažných ochorení. V tejto práci sa zameriavame na klasifikáciu kožných lézií pomocou konvolučných neurónových sietí (`CNN`). Skúmame vplyv rôznych techník na zlepšenie presnosti modelu, vrátane predspracovania obrázkov, vyvažovania datasetu metódami resamplingu, `SMOTE` a váhovania tried ako aj použitia pokročilých architektúr ako DenseNet a ResNet.

1 Úvod

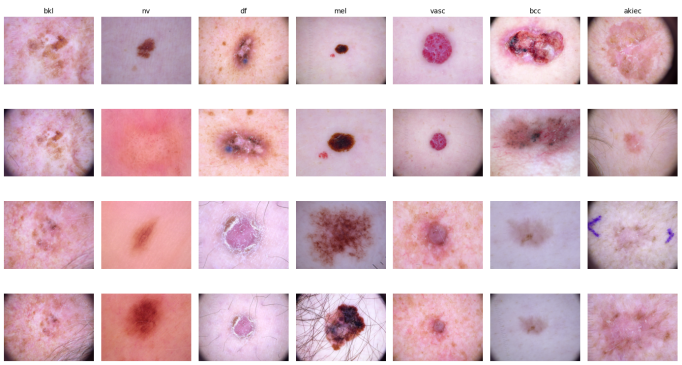
Kožné ochorenia predstavujú významný zdravotný problém, pričom včasná a presná diagnostika zohráva kľúčovú úlohu pri efektívnej liečbe.

V súčasnosti sa diagnostika kožných lézií opiera predovšetkým o vizuálne vyšetrenie dermatológom, dermatoskopiu a histopatologickú analýzu biopsií. Napriek pokroku v zobrazovacích technikách a vyššej dostupnosti dermatoskopov však stále existuje riziko subjektívneho hodnotenia a chybnej diagnostiky. Z tohto dôvodu sa čoraz viac uplatňujú metódy umelej inteligencie, ktoré môžu podporiť lekárov v rozhodovacom procese a zvýšiť presnosť diagnostiky. Okrem asistencie odborníkom vzniká aj potenciál na využitie umelej inteligencie v samodiagnostike, kde by si pacienti mohli prostredníctvom mobilných aplikácií predbežne vyhodnotiť riziko kožných ochorení na základe analýzy obrázkov lézií.

Táto práca sa zameriava na automatickú klasifikáciu kožných lézií pomocou konvolučných neurónových sietí (`CNN`) a pokročilých modelov hlbokého učenia ako sú DenseNet a ResNet. Využitý dataset HAM10000 obsahuje snímky siedmich rôznych typov kožných lézií. V práci analyzujeme výkonnosť rôznych modelov pri rozlišovaní jednotlivých kategórií lézií a porovnávame ich schopnosť správne klasifikovať snímky v kontexte reálnych diagnostických požiadaviek.

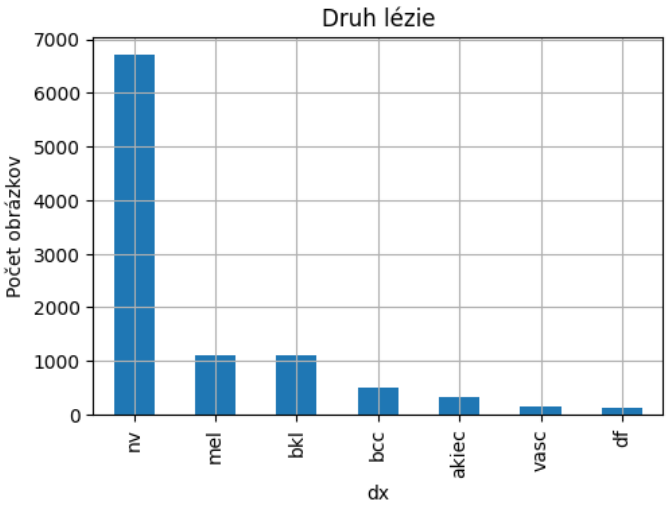
1.1 Výber datasetu

Na klasifikáciu kožných lézií sme použili dataset HAM10000, ktorý obsahuje 10 015 obrázkov rozdelených do 7 tried: Melanocytové névy (`nv`), Melanóm (`mel`), Benígna keratóza (`bkl`), Bazocelulárny karcinóm (`bcc`), Aktinická keratóza (`akiec`), Vaskulárne lézie (`vasc`) a Dermatofibrom (`df`). Dataset bol vytvorený pre ISIC Challenge 2018, súťaž zameranú na zlepšenie detekcie rakoviny kože pomocou umelej inteligencie [1]. Ukážka tried datasetu je zobrazená na obrázku č. 1 Okrem



Obr. 1: Ukážka tried datasetu

obrázkov obsahuje aj metadáta ako vek, pohlavie pacienta a polohu lézie. Keďže dataset je nevyvážený (obrázok č. 2), využili sme metódy na jeho vyváženie. Nerovnomerné zastúpenie tried spôsobuje, že modely uprednostňujú dominantnú triedu, čím dochádza k skresleniu výsledkov [2].



Obr. 2: Rozdelenie kožných lézií v datasete

1.2 Metódy vyvažovania datasetu

Aby sme vyrovnali distribúciu tried a predišli zaujatosti modelu, využili sme tri metódy, a to resampling, `SMOTE` a metódu váhovania tried. Resampling slúži na vyrovnanie distribúcie tried v datasete, aby model neuprednostňoval početnejšie triedy. Využili sme oversampling, ktorý sme implementovali pomocou *sklearn.utils()*, čím sme zvýšili počet vzoriek menšinových tried. Počet vzoriek v triedach `bcc, akiec, bkl, df a vasc` sme upravili na 2000, pre triedu `mel` na 3000, pričom počet vzoriek triedy `nv` zostal nezmenený. Tento prístup sme zvolili na základe výsledkov tréningu predchádzajúcich modelov.

Ďalej sme použili metódu `SMOTE` (*imblearn*), ktorá generuje syntetické vzorky na základe najbližších vzoriek, čím zlepšuje generalizáciu modelu. Po aplikácií tejto metódy bol počet vzoriek vo všetkých triedach vyrovnaný na úroveň najpočetnejšej triedy (`nv`), teda 6705 vzoriek.

Okrem toho sme aplikovali váhovanie tried pomocou metódy inverzného počtu vzoriek (`INS`), kde menej početné triedy dostali vyššiu váhu, čím sa zmiernila ich chybná klasifikácia. Implementovali sme ju pomocou *sklearn*.

2 Implementácia modelov hlbokého učenia pre analýzu kožných lézií

V tejto sekcii sa zameriame na implementáciu a využitie modelov hlbokého učenia pri analýze kožných lézií. Špecificky sa budeme venovať konvolučným neurónovým sieťam (`CNN`), ako aj architektúram DenseNet a ResNet. Tieto modely dokážu automaticky extrahovať relevantné črty z obrázkov a efektívne sa učiť z rozsiahlych datasetov, čo ich predurčuje na úlohy ako je detekcia a klasifikácia kožných lézií. V nasledujúcich podsekciách detailne opíšeme ich návrh, implementáciu a proces trénovania.

2.1 Konvolučná neurónová sieť

V tejto sekcii sa zameriavame na návrh a trénovanie konvolučnej neurónovej siete `CNN`. Rovnako ako pri ostatných architektúrach aj tu využívame knižnicu *TensorFlow*. *TensorFlow* je open-source framework vyvinutý spoločnosťou Google, ktorý sa vo väčšej miere používa na implementáciu a trénovanie modelov strojového učenia.

2.1.1 Architektúra konvolučnej neurónovej siete

Architektúra `CNN` pozostáva zo štyroch konvolučných vrstiev (*3×3, ReLU*) s postupne klesajúcim počtom filtrov (*1024, 512, 256, 128*). Prvé vrstvy extrahujú základné črty (hrany, textúry), neskoršie vrstvy identifikujú zložitejšie vzory. Každá konvolučná vrstva je doplnená *MaxPooling (2×2)* na redukciu dimenzie a *Dropout (0,3)* na prevenciu overfittingu.

Po extrakcii príznakov sa dáta sploštia (*Flatten*) a prechádzajú cez dve *Dense* vrstvy – prvá so 64 neurónmi, druhá výstupná so 7 neurónmi, pričom aktivačná funkcia *softmax* generuje pravdepodobnosti tried. Model je kompilovaný so stratovou funkciou *categorical_crossentropy* a optimalizátorom *Adam*, ktorý kombinuje výhody *momentum* a *RMSprop*.

2.1.2 Trénovanie konvolučnej neurónovej siete

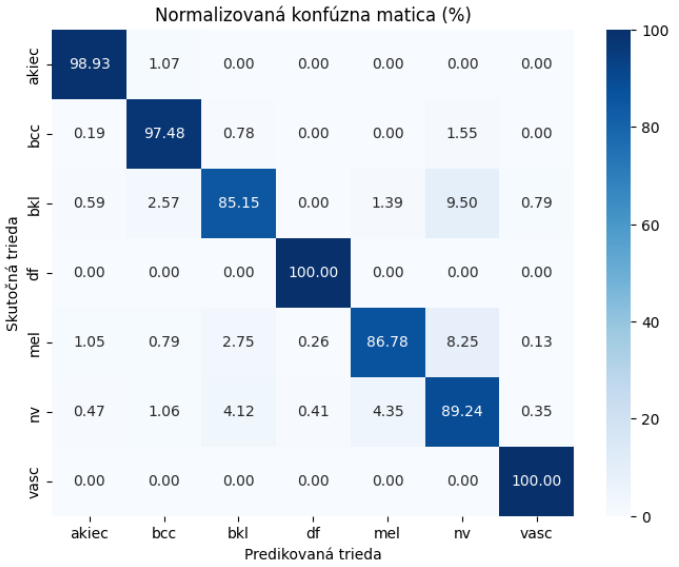
Trénovanie modelu spočíva v postupnej úprave jeho váh na základe spätnej väzby zo vstupných dát s cieľom minimalizovať stratu a zlepšiť generalizáciu na neznáme dáta.

Dáta sú rozdelené na tréningovú (75 %) a validačnú (25 %) množinu pomocou *train_test_split()*. Tréningová množina slúži na učenie modelu, zatiaľ čo validačná množina na hodnotenie výkonnosti na nových dátach.

Pri trénovaní sme nastavili *batch size* na 16, čo predstavuje kompromis medzi výpočtovou náročnosťou a stabilitou učenia.

2.1.3 Vyhodnotenie modelu konvolučnej neurónovej siete

Pri tomto modeli sme testovali dve metódy vyvažovania datasetu: resampling a `SMOTE`. Najprv hodnotíme výsledky metódy resampling. Tréning bol ukončený early stoppingom na 99. epoche. Konfúzna matica modelu konvolučnej neurónovej siete s využitím metódy resampling je zobrazená na obrázku č. 3. Na diagonále sú správne klasifikované prípady (*True Positives*).



Obr. 3: Konfúzna matica modelu konvolučnej neurónovej siete s využitím metódy resampling

Triedy `df` a `vasc` boli predikované so 100 % presnosťou bez chýb.

Mimo diagonály vidíme nesprávne klasifikácie. Pri `bkl` bolo 2,57 % prípadov predikovaných ako `bcc`, 1,39 % ako `mel` a 9,50 % ako `nv`, čo naznačuje problémy s odlíšením týchto tried. Pri `mel` bolo 2,75 % prípadov nesprávne klasifikovaných ako `bkl` a 8,25 % ako `nv`.

Niektoré triedy (`df, vasc`) model klasifikuje presne, zatiaľ čo pri iných dochádza k zámenám, najmä medzi `bkl, nv a mel`, čo môže byť spôsobené podobnosťou ich čŕt. V tabuľke č 1 je uvedený klasifikačný report modelu `CNN` s využitím metódy resampling. Pre triedu `akiec` model dosiahol precision

Tab. 1: Klasifikačný report modelu CNN s využitím metódy resampling

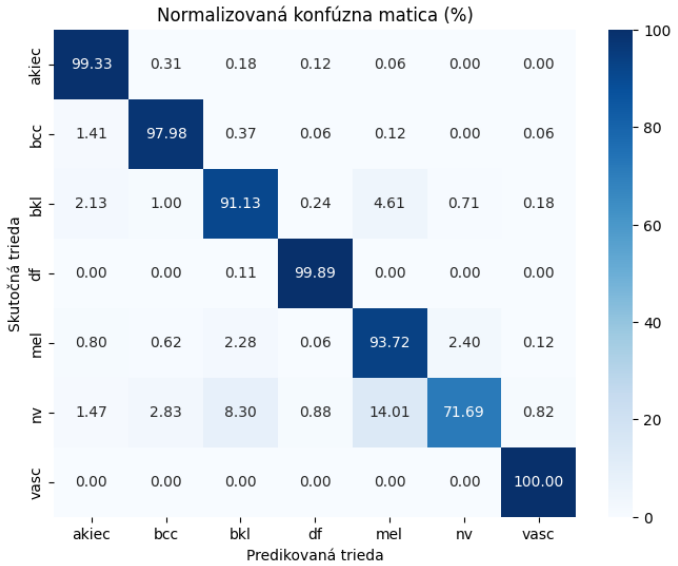
	Precision	Recall	F1-skóre
akiec	0,96	0,99	0,97
bcc	0,92	0,97	0,95
bkl	0,82	0,85	0,83
df	0,98	1,00	0,99
mel	0,89	0,87	0,88
nv	0,93	0,89	0,91
vasc	0,98	1,00	0,99
accuracy			0,92

0,96, recall 0,99 a F1-skóre 0,97, čo naznačuje vysokú presnosť aj schopnosť správne identifikovať túto kategóriu. Podobne dobré výsledky má `bcc` (0,92/0,97/0,95).

Trieda `bkl` má nižšie hodnoty (0,82/0,85/0,83), čo naznačuje vyšší počet nesprávnych klasifikácií. Naopak, `df` (0,98/1,00/0,99) a `vasc` (0,98/1,00/0,99) dosiahli takmer presné výsledky.

Pri `mel` je F1-skóre 0,88, čo ukazuje mierne nižšiu citlivosť. Trieda `nv` dosiahla F1-skóre 0,91, čo poukazuje na dobrú, ale nie dokonalú výkonnosť.

Celková presnosť modelu dosahuje 0,92, čo potvrdzuje efektivitu `CNN` pri klasifikácii kožných lézií s využitím resamplingu. Na diagonále obrázku č. 4 vidíme správne klasifikované



Obr. 4: Konfúzna matica modelu konvolučnej neurónovej siete s využitím metódy SMOTE

prípady. Napríklad trieda `df` bola správne predikovaná v 99,89 % prípadov, čo naznačuje výbornú schopnosť modelu rozpoznať túto kategóriu. Trieda `vasc` dosiahla podobný výsledok so 100 % správne klasifikovanými prípadmi bez chybných predikcií.

Niektoré triedy sa modelu rozlišujú ťažšie. Trieda `bkl` bola nesprávne klasifikovaná v 4,61 % prípadov ako `mel`, v 2,13 % ako `akiec` a v 1 % ako `bcc`. Trieda `nv` bola zamieňaná s `bkl` v 8,3 % prípadov a s `mel` v 14,01 %, čo naznačuje vizuálnu podobnosť medzi týmito kategóriami.

Tab. 2: Klasifikačný report modelu `CNN` s využitím metódy SMOTE

	Precision	Recall	F1-skóre
akiec	0,98	0,99	0,98
bcc	0,97	0,98	0,97
bkl	0,90	0,92	0,91
df	0,99	1,00	0,99
mel	0,87	0,91	0,89
nv	0,89	0,81	0,85
vasc	0,99	1,00	1,00
accuracy			0,94

V tabuľke 2 vidíme, že model dosiahol zlepšenie presnosti pri viacerých triedach. Triedy `akiec` a `bcc` vykazujú vysokú presnosť (F1-skóre 0,98 a 0,97). Trieda `bkl` zaznamenala výrazné zlepšenie (F1-skóre z 0,83 na 0,91), zatiaľ čo `df` a `vasc` si udržali takmer dokonalé hodnoty. Pri triede `mel` sa mierne zvýšil recall (0,91), no precision mierne klesla. Trieda `nv` naopak zaznamenala pokles (F1-skóre z 0,91 na 0,85), čo naznačuje problémy s presnou identifikáciou. Celková presnosť modelu predstavuje 0,94.

2.2 DenseNet

V tejto sekcii sa venujeme architektúre DenseNet, ktorá patrí medzi pokročilé konvolučné neurónové siete. DenseNet bol navrhnutý ako efektívnejšia alternatíva k tradičným `CNN` modelom, pričom jeho hlavnou výhodou je husté prepojenie vrstiev. Každá vrstva v DenseNet je spojená so všetkými predchádzajúcimi vrstvami, čo umožňuje lepší tok informácií a gradientov počas trénovania. DenseNet dosahuje vzájomné prepojenie medzi ľubovoľnými vrstvami a využíva režim preskočenia spojenia, ktorý prenáša informácie z plytkých vrstiev priamo do hlbokých vrstiev. Výsledkom je lepšia konvergencia modelu, zmiernenie problémov s degradáciou siete a redukcia problému zanikania gradientu, ktorý sa zvyčajne objavuje v hlbokých konvolučných sieťach. [3]. Problém zanikania gradientu vzniká tak, že sa gradient exponenciálne zmenšuje, keď sa pridáva viac vrstiev, čo spôsobuje, že učenie siete sa zastaví [4]. My sme využili architektúru DenseNet 121.

2.2.1 Architektúra modelu DenseNet 121

DenseNet 121 obsahuje 121 vrstiev rozdelených do štyroch hustých blokov (*Dense Blocks*), medzi ktorými sú prechodové vrstvy. Každý hustý blok obsahuje viacero konvolučných vrstiev, pričom každá vrstva dostáva ako vstup výstupy zo všetkých predchádzajúcich vrstiev v rámci daného bloku.

Model začína vstupnou vrstvou (*Stem Layer*), ktorá obsahuje konvolučnú vrstvu (jadro 7×7, 64 filtrov) a pooling vrstvu (jadro 3×3). Po nej nasledujú štyri husté bloky (*Dense Blocks*), pričom medzi nimi sú prechodové vrstvy s 1×1 konvolúciou a 2×2 poolingom.

- **Dense Block 1**: obsahuje 6 konvolučných vrstiev.
- **Dense Block 2**: obsahuje 12 konvolučných vrstiev.
- **Dense Block 3**: obsahuje 24 konvolučných vrstiev.
- **Dense Block 4**: obsahuje 16 konvolučných vrstiev (posledná s 1024 filtrami).

Na záver sa nachádza globálna pooling vrstva a plne prepojená výstupná vrstva s aktivačnou funkciou *softmax*, ktorá zabezpečuje finálnu klasifikáciu.

Architektúru DenseNet 121 sme implementovali v *TensorFlow* s predtrénovanými váhami z *ImageNetu*, pričom sme odstránili pôvodnú klasifikačnú hlavu, keďže model bol pôvodne trénovaný na 1000 tried, zatiaľ čo naša úloha zahŕňa 7 tried.

Nová klasifikačná hlava začína vrstvou *GlobalAveragePooling2D()*, ktorá redukuje výstup na jednorozmerný vektor, čím znižuje počet parametrov a zlepšuje generalizáciu. Následne sme pridali *Dense* vrstvu so 128 neurónmi (ReLU) a *Dropout* (0,5) na minimalizáciu pretrénovania. Výstupom je *Dense* vrstva so 7 neurónmi a aktivačnou funkciou *softmax*.

Predtrénované vrstvy sme zmrazili, čím sa zachovajú ich váhy a trénujú sa iba nové vrstvy. Model je kompilovaný s optimalizátorom *Adam* (*learning rate* 0,001) a stratovou funkciou *categorical_crossentropy*.

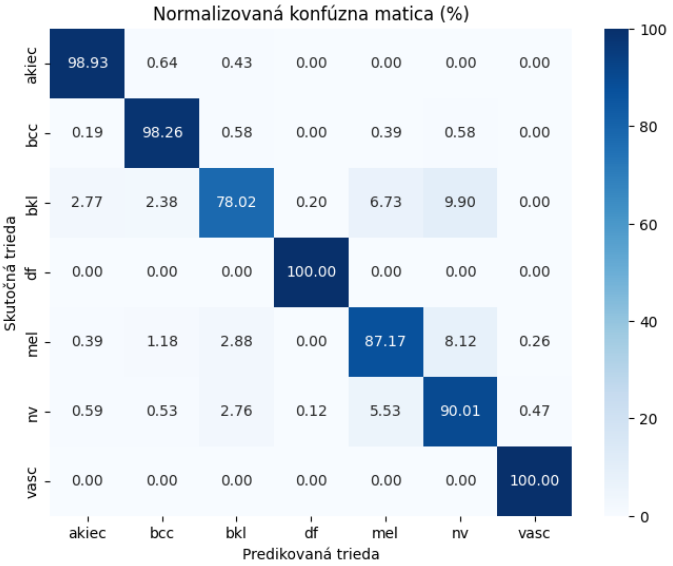
2.2.2 *Trénovanie modelu DenseNet 121*

Pri trénovaní modelu DenseNet 121 sme dáta rozdelili na tréningové (75 %) a validačné (25 %) pomocou *train_test_split() s random_state=42*. *Batch size* zostáva 16, počet epoch sme znížili na 100.

Použili sme *EarlyStopping*, ktorý monitoruje *val_loss* a zastaví trénovanie, ak sa nezlepšuje. Taktiež sme pridali *ReduceLROnPlateau* (*factor=0.2, patience=5*), ktorý pri stagnácii validačnej straty znižuje *learning rate* pre stabilnejšiu konvergenciu.

2.2.3 *Vyhodnotenie modelu DenseNet 121*

V tomto modeli sme dataset vyvažovali pomocou metódy resamplingu. Konfúzna matica modelu DenseNet 121 je zobrazená na obrázku č. 5 Model dosiahol 100 % úspešnosť pri triedach `df` a `vasc`. Chybné klasifikácie sa vyskytli najmä pri triedach `bkl, mel a nv`. Trieda `bkl` bola najčastejšie zamieňaná za `bcc` (2,38 %), `mel` (6,73 %) a `nv` (9,90 %), čo naznačuje problémy s ich odlíšením. Trieda mel bola nesprávne predikovaná ako `bkl` (2,88 %) a `nv` (8,12 %). Trieda `nv` bola zamieňaná s `bkl` (2,76 %) a `mel` (5,53 %). Oproti `CNN` modelom je pri triede `bkl` nižšia správnosť klasifikácie. V tabuľke č. 3 je zobrazený klasifikačný report modelu Densenet 121. Celková presnosť modelu dosiahla 0,92. Triedy `bdf` a `vasc` si zachovali takmer 100 % výsledky. Trieda `bcc` mierne prekonala `CNN` s resamplingom (F1-skóre 0,96) a trieda `nv` dosiahla zlepšenie oproti `CNN` so `SMOTE`. Naopak, trieda `bkl` zaznamenala pokles výkonu, čo naznačuje, že DenseNet má problém s jej rozpoznaním.



Obr. 5: Konfúzna matica modelu DenseNet 121 s využitím metódy resampling

Tab. 3: Klasifikačný report modelu DenseNet 121

	Precision	Recall	F1-skóre
akiec	0,94	0,99	0,97
bcc	0,94	0,98	0,96
bkl	0,84	0,78	0,81
df	0,99	1,00	1,00
mel	0,84	0,87	0,85
nv	0,93	0,90	0,91
vasc	0,98	1,00	0,99
accuracy			0,92

2.3 ResNet

V tejto podsekcii sa venujeme architektúre ResNet. Model ResNet (*Residual Network*) bol navrhnutý na riešenie problému degradácie presnosti pri hlbokých sieťach. Jeho hlavnou inováciou sú reziduálne spojenia (*skip connections*), ktoré umožňujú preskakovanie dvoch až troch vrstiev naraz pridaním výstupu z predchádzajúcej vrstvy k vstupu ďalšej.

Tento prístup rieši problém zanikania gradientu, pretože uľahčuje šírenie informácií z počiatočných vrstiev do hlbších vrstiev siete. Základným stavebným prvkom ResNet je reziduálny blok, obsahujúci priame spojenie (*skip connection*) medzi vstupom a výstupom bloku [4]

V našom prípade využívame architektúru ResNet 50, ktorá obsahuje 50 vrstiev.

2.3.1 *Architektúra modelu ResNet 50*

ResNet 50 obsahuje 50 vrstiev, z toho 49 konvolučných a jednu plne prepojenú vrstvu. Vstupná vrstva spracováva obrázky s rozmermi 224×224 pixelov a tromi farebnými kanálmi (`RGB`). Architektúra je rozdelená do piatich fáz, pričom každá obsahuje

viacero reziduálnych blokov.

Model začína 7×7 konvolučnou vrstvou so 64 filtrami, po ktorej nasleduje *max-pooling* (3×3) na extrakciu základných čŕt a redukciu rozmerov obrazu. Jadro architektúry tvorí sústava reziduálnych blokov, kde konvolučné vrstvy majú jadrá 1×1, 3×3 a 1×1, pričom posledná vrstva obsahuje 2048 filtrov. Priame spojenia medzi blokmi uľahčujú šírenie informácií a zabraňujú zanikaniu gradientu pri trénovaní.

Záver siete tvorí *global average pooling*, ktorý redukuje dimenzie výstupu, a následne plne prepojená vrstva s 1000 výstupmi, čo zodpovedá počtu tried v datasetoch ako ImageNet. Výstupná *softmax* funkcia generuje pravdepodobnostnú distribúciu predikovaných tried [5].

Model ResNet 50 sme implementovali pomocou knižnice *TensorFlow* a načítali z *ImageNet* s parametrom *include_top=False*, čím sme odstránili plne pripojené vrstvy. Vstupnú veľkosť obrázkov sme upravili na 64×64 pixelov (RGB) kvôli výpočtovej náročnosti.

Nový model sme zostavili pomocou funkcionálneho *API Keras*, kde ResNet 50 slúži na extrakciu čŕt. Následne sme pridali *GlobalAveragePooling2D()*, čím sme nahradili pôvodné FC vrstvy. Na zlepšenie generalizácie sme použili *Dropout(0.5)*, ktorý náhodne vypína 50 % neurónov počas tréningu.

Ďalej sme zaradili *Dense(256, activation='relu')*, ktorá spracováva extrahované črty a pomáha zachytiť zložitejšie vzory. Opäť sme aplikovali *Dropout(0.5)* na zníženie rizika pretrénovania. Výstupná vrstva *Dense(len(train_generator.class_indices), activation='softmax')* zodpovedá počtu tried v datasete a generuje pravdepodobnosti pre jednotlivé triedy.

Model sme skompilovali s *Adam optimalizátorom* (learning rate *0.0001*) pre stabilné učenie a ako funkciu straty sme použili kategóriovú krížovú entropiu, keďže ide o viactriednu klasifikáciu.

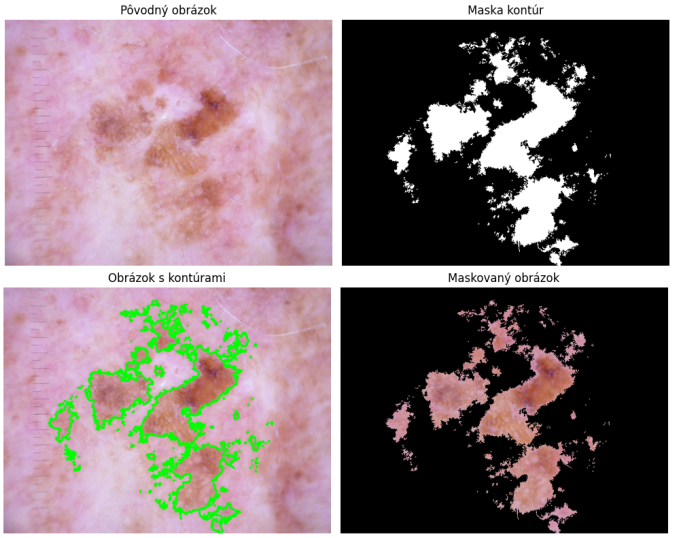
2.3.2 *Trénovanie modelu ResNet 50*

Na trénovanie sme použili *train_generator*, ktorý generuje dáta v dávkach. Model ako vstup prijíma aj hodnoty symetrií. *Early stopping* zastavil tréning na 10 epoche.

Výpis priebehu učenia je nastavený na *verbose=1*, takže sa zobrazujú podrobné informácie o trénovaní. Použili sme metódu váhovania tried cez *class_weight=class_weights_dict*.

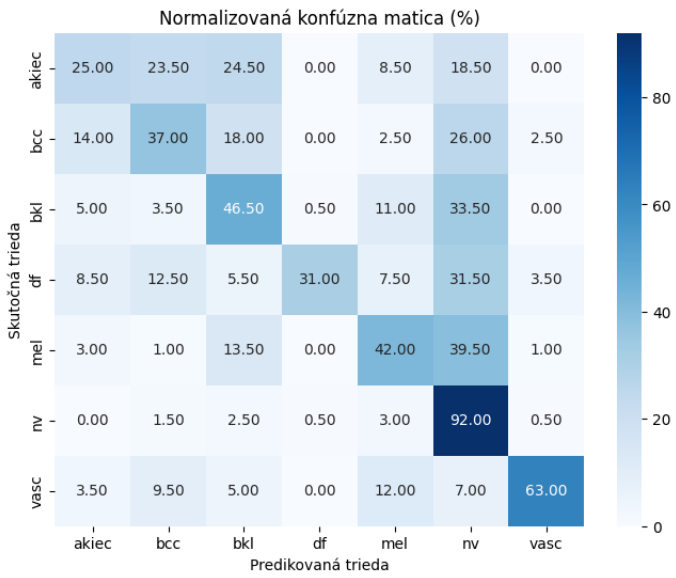
2.3.3 *Vyhodnotenie modelu ResNet 50*

V tomto modeli sme využili predspracovanie obrázkov, pri ktorom sme najskôr konvertovali obrázky do RGB farebného priestoru a následne sme pomocou prahovania vytvorili masku na základe definovaného farebného rozsahu. Potom sme detegovali kontúry a vybrali tie, ktoré sa nachádzajú v určitej vzdialenosti od stredu obrázka. Nakoniec sme z týchto kontúr vytvorili masku a aplikovali ju na pôvodný obrázok, čím sme extrahovali relevantné oblasti. Tento postup je zobrazený na obrázku č. 6 Na vyváženie datasetu používame váhovanie tried, konkrétne metódu inverzného počtu vzoriek. Pre správne vyhodnotenie konfúznej matice sme validačný dataset vyvážili pomocou metódy resampling, pričom sme pre



Obr. 6: Výsledok segmentácie obrázkov

každú triedu vybrali 200 vzoriek. Na obrázku 7 vidíme, že



Obr. 7: Konfúzna matica modelu ResNet 50 s využitím metódy váhovania tried

hoci sa diagonála zachovala, model dosiahol nižšiu presnosť v predikcii jednotlivých tried oproti ostatným modelom. Najlepšie klasifikoval triedu nv (92 %), no pri triede akiec dosiahol len 25 %, pričom ju často zamieňal s bcc, bkl a nv. V tabuľke č. 4 je zobrazený klasifikačný report modelu ResNet 50.

Celková presnosť modelu ResNet 50 dosiahla len 0,48, čo je výrazne horšie ako pri ostatných modeloch.

Triedy akiec, bcc a bkl dosiahli nízke F1-skóre (0,31 – 0,43), čo naznačuje výrazné problémy v klasifikácii. Trieda df mala vysokú precision (0,97), ale nízky recall (0,31), čo viedlo k F1-skóre 0,47. Trieda mel dosiahla F1-skóre 0,45, pričom trieda nv mala vysoký recall (0,92), ale nízku precision (0,37), čo dalo F1-skóre 0,52.

Tab. 4: Klasifikačný report modelu ResNet 50

	Precision	Recall	F1-skóre
akiec	0,42	0,25	0,31
bcc	0,41	0,37	0,39
bkl	0,40	0,47	0,43
df	0,97	0,31	0,47
mel	0,48	0,42	0,45
nv	0,37	0,92	0,52
vasc	0,89	0,63	0,74
accuracy			0,48

Najlepšie si viedla trieda `vasc` s precision 0,89 a F1-skóre 0,74.

2.4 Zhodnotenie výsledkov

`CNN` s resamplingom a `CNN` so `SMOTE` dosiahli najvyššiu presnosť – 92 % pri CNN s resamplingom a 94 % pri `CNN` so `SMOTE`. `SMOTE` prispel k lepšej klasifikácii menšinových tried, no niektoré kategórie stále vykazovali vysokú mieru chýb. Modely si dobre poradili s väčšinou kategórií, no vyskytli sa problémy pri triede `bkl`, ktorá bola často nesprávne predikovaná ako `nv` alebo `mel`.

DenseNet 121 dosiahol podobné výsledky ako CNN s resamplingom, a to presnosť na úrovni 92 %. Pri triedach `bcc a nv` bol o niečo presnejší, no pri `bkl` došlo k výraznému zhoršeniu oproti `CNN` modelom. V porovnaní s `CNN` so `SMOTE` mal nižšiu presnosť najmä pri triedach `bkl a mel`, čo naznačuje, že tento model nemusel byť dostatočne robustný na rozlíšenie niektorých jemných rozdielov medzi léziami.

ResNet 50 mal výrazne horšie výsledky s celkovou presnosťou len 48 %, čo je podstatne menej než pri ostatných modeloch. Najväčšie problémy mal pri triede `akiec`, kde správne klasifikoval len 25 % vzoriek. Model často zamieňal `akiec s bcc, bkl a nv`. Naopak, jeho najlepší výsledok bol pri triede vasc (F1-skóre 0,74), no aj tu zaostával za `CNN` a DenseNet modelmi. Nízka presnosť modelu môže byť spôsobená kombináciou viacerých faktorov, ako je prílišná hĺbka siete vzhľadom na veľkosť datasetu, nedostatočné generalizačné schopnosti pri menšinových triedach a nevhodné hyperparametre. Vzhľadom na tieto skutočnosti sa ResNet 50 neukázal ako vhodná architektúra pre túto úlohu.

`CNN` modely dosiahli najlepšie výsledky, pričom vyváženie tried pomocou `SMOTE` sa ukázalo ako efektívne. DenseNet bol konkurencieschopný, ale mal problémy pri triede `bkl`. ResNet 50 sa ukázal ako nevhodný pre túto úlohu, keďže nedokázal spoľahlivo rozlišovať medzi viacerými kategóriami.

3 Záver

Klasifikácia kožných lézií pomocou modelov hlbokého učenia predstavuje významný krok v modernizácii diagnostiky dermatologických ochorení. V tejto práci sme analyzovali výkonnosť viacerých modelov konvolučných neurónových sietí (`CNN`) pri klasifikácii lézií v datasete HAM10000. Výsledky ukázali, že pokročilé architektúry ako DenseNet a `CNN` so `SMOTE` dosahujú vysokú presnosť, pričom model ResNet 50 mal nižší výkon, najmä pri rozlíšení menej zastúpených tried.

Na základe získaných výsledkov môžeme konštatovať, že umelá inteligencia môže efektívne podporiť dermatológov pri diagnostike kožných ochorení a prispieť k rýchlejšej a presnejšej identifikácii rizikových lézií. Zároveň vzniká priestor na jej využitie v nástrojoch na samodiagnostiku, ktoré by mohli pomôcť pacientom včas rozpoznať podozrivé zmeny na koži a vyhľadať odbornú pomoc.

Napriek pokroku však stále existujú výzvy, ako je nevyváženosť datasetu, možnosť nesprávnej klasifikácie či potreba vysvetliteľnosti modelov v klinickej praxi. Budúce práce by sa mohli zamerať na ďalšie metódy zlepšenia klasifikácie, ako je integrácia viacerých modalít dát či nasadenie modelov na mobilné zariadenia.

Celkový výskum v tejto oblasti naznačuje, že kombinácia umelej inteligencie a dermatologickej diagnostiky má potenciál výrazne zlepšiť starostlivosť o pacientov a včasnú detekciu kožných ochorení.

Poďakovanie

Ďakujeme za podporu grantom Kultúrnej a edukačnej grantovej agentúry MŠVVaM SR - KEGA 010STU-4/2023 a KEGA 021STU-4/2024.

Literatúra

[1] ISIC Challenge 2018, ISIC Challenge, 2018, `https://challenge.isic-archive.com/landing/2018/`

[2] Ghorbani, Ramin and Ghousi, Rouzbeh, 2020, Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques, IEEE Access, 10.1109/ACCESS.2020.2986809

[3] Zhou, Tao and Lu, HuiLing and Zheng, Xiaomin and Qiu, Shi and Liu, YunCan, 2022, Dense Convolutional Network and Its Application in Medical Image Analysis, BioMed Research International, 10.1155/2022/2384830

[4] Agrawal, S. A. ., Rewaskar, V. D. ., Agrawal, R. A. ., Chaudhari, S. S. ., Patil, Y. ., & Agrawal, N. S., 2023, A Comprehensive Review of ResNet-50 Based Approaches, International journal of inteligent system and application in engineering https://ijisae.org/index.php/IJISAE/article/view/3452

[5] He, Kaiming and Zhang, Xiangyu and Ren, Shaoqing and Sun, Jian, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Deep Residual Learning for Image Recognition, 2016, 10.1109/CVPR.2016.90

AI chatbot a RPA - inteligentná automatizácia

Andrea Sigetiová

Ústav automobilovej mechatroniky, FEI STU v Bratislave

xsigetiova@stuba.sk

V dnešnej dobe plnej digitalizácie a technologického pokroku zohrávajú inteligentné automatizačné systémy čoraz väčší význam. V tomto článku sa budeme venovať synergickému prepojeniu technológií Robotickej automatizácie procesov RPA (Robotic Process Automation) a AI chatbotov, výsledkom čoho môžu byť napríklad tzv. virtuálni agenti. Tento článok poskytuje prehľad o novodobých trendoch a perspektívach využitia inteligentnej automatizácie v rôznych odvetviach.

1 Úvod

V súčasnej dobe sa kladie vysoký dôraz na digitalizáciu a automatizáciu s cieľom zefektívniť prácu. V tom zohráva veľkú úlohu technológia Robotickej automatizácie procesov RPA (Robotic Process Automation), ktorá umožňuje automatizovať opakujúce sa úlohy v rôznych oblastiach – od administratívy až po komplexnejšie podnikové procesy, pričom tieto úlohy sú zväčša časovo náročné a je nutné spracovať veľké objemy dát. V porovnaní s tradičnou automatizáciou, ktorá často vyžaduje zásah do infraštruktúry alebo hardvéru, RPA pracuje výhradne v softvérovom prostredí a využíva existujúce systémy bez potreby ich meniť.

Spolu s RPA sa výrazne posúva aj oblasť umelej inteligencie, najmä v oblasti Veľkých jazykových modelov LLM (Large Language Models), ktoré tvoria základ moderných chatbotov, pričom tie sú potom schopné porozumieť prirodzenému jazyku a reagovať na otázky používateľov, čo zvyšuje ich využiteľnosť aj v komplexných systémoch.

Integrácia RPA a LLM modelov predstavuje inovatívny prístup, vďaka ktorému by sa automatizovanie pracovných procesov stalo oveľa efektívnejšie so zároveň menším úsilím.

2 Technológie

V tejto časti sa zameriame na analýzu základných technológií, ktorými sú Robotická automatizácia procesov RPA (Robotic Process Automation) a Veľké jazykové modely LLM (Large Language Model), pričom tieto technológie patria v dnešnej dobe k hlavným nástrojom v oblasti automatizácie procesov.

2.1 RPA

Robotická automatizácia procesov je definovaná ako aplikácia špecifických technológií, ktoré sú založené na softvéri a algoritmoch, pričom cieľom je uľahčiť ľuďom prácu automatizovaním opakujúcich sa úloh. RPA využíva na komunikáciu s inými informačnými systémami existujúce grafické používateľské rozhrania. Túto automatizáciu vykonáva softvérový robot „bot“. [1]

RPA je často nesprávne interpretovaná ako umelá inteligencia AI (Artificial intelligence), hoci sú odlišné. Jedným z najväčších rozdielov je to, že RPA sa zameriava na procesy, zatiaľ čo AI je zameraná na dáta. Zjednodušene povedané, RPA replikuje úlohy, ktoré sú navrhnuté a riadené človekom, avšak úlohou AI je simulovať ľudskú inteligenciu. Vďaka týmto odlišnostiam sa RPA a AI dobre vzájomne dopĺňajú – AI pomáha zvládať komplexnejšie úlohy, pričom RPA pomáha rýchlejšie aplikovať nadobudnuté poznatky tým, že miesto manuálnej implementácie sa tento proces zautomatizuje. [2]

2.1.1 Výhody a nevýhody RPA

RPA sa do popredia dostala z dôvodu množstva výhod, ktorými sú najmä zvýšenie produktivity a efektivity a zároveň zníženie chybovosti. Vďaka tejto technológii je možné automatizovať veľké množstvá dát a odbremeniť tak pracovníkov, keďže by inak boli tieto úlohy vykonávané manuálne. [3][4]

Napriek mnohým výhodám tejto technológie je dôležité zdôrazniť aj jej nevýhody, ktoré sú najmä tam, kde je potrebná flexibilita a spracovanie neštruktúrovaných dát. [3][4]

2.1.2 Využitie RPA

RPA je možné použiť v akomkoľvek pracovnom odvetví, no existuje niekoľko, ktoré sa spoliehajú na digitalizáciu a automatizáciu viac, než ostatné. Medzi takéto odvetvia patrí napríklad bankové a finančné služby, poisťovníctvo, maloobchod, zdravotná starostlivosť a iné. [2]

V praxi sa táto technológia uplatňuje napríklad pri spracovaní údajov o zákazníkoch, správe dokumentov, spracovaní objednávok, fakturácii, spracovaní spätnej väzby od klientov, administratívnych činnostiach, na správu liekov na predpis, na detekciu podvodov, na boj proti praniu špinavých peňazí a iné. [2]

2.1.3 RPA softvér

Robotická automatizácia procesov sa využíva v rôznych odvetviach, pričom existujú platformy, ktoré dominujú na trhu s RPA softvérovými riešeniami.

V nasledujúcej tabuľke (Tab. 1.) je možné vidieť jednoduché porovnanie RPA softvérových riešení, pričom tam patria nielen popredné softvérové riešenia, ale aj menej známe.

Tab. 1. Porovnanie RPA softvérov [5][6][7][8]

Nástroj	Náklady	Jednoduchosť	Škálovateľnosť	Integrácia
UiPath	Stredné až vysoké	Áno	Vysoká	Vysoká
Automation Anywhere	Vysoké	Nie	Vysoká	Vysoká
Blue Prism	Vysoké	Áno	Vysoká	Vysoká
Microsoft Power Automate	Nízke	Áno	Stredná	Vysoká
Pega	Vysoké	Áno	Vysoká	Vysoká
Appian RPA	Stredné až vysoké	Áno	Vysoká	Vysoká
Robocorp	Žiadne	Nie	Vysoká	Stredná
WorkFusion	Stredné až vysoké	Áno	Vysoká	Vysoká
Robot Framework	Žiadne	Nie	Vysoká	Stredná

2.2 LLM

K jedným z popredných prístupov rozvoja komunikačných schopností strojov patrí jazykové modelovanie LM (Language Modeling), umožňujúce strojom lepšie porozumieť a generovať prirodzený text. Vo všeobecnosti sa LM snaží pomocou modelovania generatívnej pravdepodobnosti slovných sekvencií predikovať pravdepodobnosť nasledujúcich tokenov (token sa chápe ako základná jednotka textu, pričom tokenom môže byť v tomto prípade nielen slovo, ale aj jeho časť, prípadne samotný znak). [9]

2.2.1 Vývojové etapy LM

LM sa zvyčajne rozdeľuje do štyroch hlavných vývojových etáp: štatistické jazykové modely, neurálne jazykové modely, predtrénované jazykové modely a veľké jazykové modely. [9]

Štatistické jazykové modely sú modely, ktoré používajú pravdepodobnostné metódy na predikovanie pravdepodobnosti sekvencie slov v jazyku na základe kontextu a predchádzajúcich dát. Tieto modely sú závislé na veľkom množstve trénovacích dát a majú obmedzenú pamäť, čo spôsobuje problém pri zachytení celého kontextu. [9][10]

Neurálne jazykové modely využívajú neurónové siete na porozumenie a generovanie prirodzeného jazyka, pričom sa učia vzťahy medzi slovami a vďaka tomu vedia pracovať s kontextom. Tieto modely sú v porovnaní so štatistickými jazykovými modelmi komplexnejšie, pomalšie a vyžadujú viac energie na trénovanie. [9][11]

Predtrénované jazykové modely sú modely, kde sa neurónové siete najskôr trénujú na veľkých objemoch textových dát s cieľom získať reprezentácie slov závislé od kontextu. Následne sú dolaďované pre špecifické úlohy. Z toho dôvodu je nutné rozoznať, či natrénovaný model obsahuje relevantné informácie a je vhodný pre konkrétnu úlohu. [9][12]

Veľké jazykové modely založené na predtrénovaných jazykových modeloch, no s mnohonásobne väčším počtom parametrov a vrstiev neurónových sietí, z ktorých každá obsahuje parametre, ktoré je možné dolaďovať počas trénovania. Vo všeobecnosti veľké jazykové modely sú navrhnuté tak, aby porozumeli a vedeli pracovať s textom a inými formami obsahu podobne ako človek. Medzi schopnosti veľkých jazykových modelov patrí pochopenie kontextu, formulovanie relevantnej odpovede na otázky, zhrnutie obsahu textu, preklad do jazykov iných ako angličtina, generovanie kódu v rôznych programovacích jazykov a iné. Z toho dôvodu sú vhodné na použitie v moderných chatbotoch a virtuálnych asistentoch. Keďže tieto modely sa trénujú na takom obrovskom množstve neštruktúrovaných dát, môže dochádzať k tzv. halucináciám, čo sú napr. nesprávne odpovede, zaujatosť a nenávistné odpovede. Taktiež pri tvorbe veľkých jazykových modelov sú potrebné veľké prvotné investície, obrovské množstvá údajov, na ktorých sa model trénuje, technickú odbornosť a dostatočný výpočtový výkon, ktoré je nutné na vývoj a správu veľkých jazykových modelov. [9][13][14]

2.2.2 Chatbot

Chatbot je definovaný ako softvérová aplikácia, prípadne webové rozhranie, ktoré slúžia na napodobnenie ľudskej konverzácie prostredníctvom interakcií pomocou textu alebo hlasu. [15][16][17][18]

Prvotné chatboty boli v podstate interaktívne programy, ktoré odpovedali na časté otázky používateľov, pričom mali preddefinované časté otázky a ich odpovede. Neboli schopné interpretovať prirodzený jazyk a neboli schopné pochopiť komplexné otázky, a tým pádom ani neboli schopné na ne odpovedať. [19]

S postupom času sa chatboty stali schopné porozumieť prirodzenému jazyku a umožnili interakciu s používateľom v konverzačnom štýle a to vďaka algoritmom a komplexnejšiemu programovaniu, ktoré boli založené na pravidlách a boli schopné spracovať prirodzený jazyk. Takýto nový typ chatbota bol schopný rozpoznať kontext a využíval strojové učenie na neustále zlepšovanie svojej schopnosti správne spracovávať požiadavky vďaka čoraz väčšej interakcie s ľudským jazykom. [19]

Dnešné moderné AI chatboty využívajú pochopenie prirodzeného jazyka NLU (Natural language undestanding) na porozumenie aj nejednoznačného vstupu od používateľa, od preklepov, gramatických chýb, s prekladom až po nepresne formované otázky. Pokročilé AI nástroje vyhodnotia požiadavku od používateľa a na základe toho vytvoria odpoveď. Používajú strojové učenie a hĺbkové učenie, čo im pomáha neustále sa zdokonaľovať. Chatbot si každou ďalšou interakciou rozširuje databázu otázok a odpovedí a tým sa stáva presnejším a efektívnejším. Ako už bolo spomenuté, LLM sú základom moderných chatbotov, pričom chatbot v tomto prípade predstavuje rozhranie pre LLM. Vďaka rýchlemu vývoju LLM sú dnešné chatboty oveľa inteligentnejšie a majú aj širšie využitie v praxi. [19]

2.3 Synergia medzi LLM a RPA

Synergia RPA a LLM prináša nový rozmer automatizácie, a tým otvára nové možnosti podnikových procesov orientovaných na vysokú efektivitu a flexibilitu. RPA je primárne orientovaná na vykonávanie opakujúcich sa úloh so štruktúrovanými dátami a definovanými pravidlami ako napríklad vypĺňanie formulárov, odosielanie e-mailov a pod. Na rozdiel od toho LLM sa zameriava na prácu s neštruktúrovanými údajmi a schopnosti porozumieť prirodzenému jazyku. To otvára nové možnosti pre komplexnejšie procesy, na ktoré RPA zvyčajne nestačí. [20][21]

2.3.1 Kľúčové výhody

RPA boty integrované s LLM taktiež dokážu:

- Komunikovať so zákazníkmi plynulým a prirodzeným štýlom. Dokážu poskytnúť personalizované odpovede na otázky a taktiež efektívne riešiť aj zložitejšie úlohy. Vďaka tomu sa skracuje čas čakania a zlepšuje sa kvalita zákazníckej podpory. [20][21]
- Dôkladne analyzovať dáta a generovať pochopiteľné reporty, čo šetrí čas a zdroje organizáciám a zároveň zabezpečuje efektívnejšie a rýchlejšie rozhodovanie na základe kľúčových informácií. [20][21]
- Extrahovať podstatné údaje, segregovať dokumenty a zároveň vytvárať správy a prehľady na základe ich obsahu, čo urýchľuje proces a znižuje riziko chýb spôsobených ručným zadávaním. [20][21]
- Vykonať preklad medzi rôznymi jazykmi, čo zabezpečuje efektívnu interakciu so zákazníkmi, partnermi a zamestnancami po celom svete. [20][21]
- Sa stále zdokonaľovať a prispôsobovať novým informáciám a zlepšovať svoje výkony. Taktiež vedia monitorovať trendy v odvetví, pohotovo reagovať na zmeny v reguláciách a dynamické potreby zákazníkov, vďaka čomu je možné udržať vysokú kvalitu služieb a dynamicky sa adaptovať. [20][21]
- Flexibilne vytvárať pracovné procesy alebo skripty podľa požiadaviek v prirodzenom jazyku. Taktiež pomocou LLM je možné dokáže upraviť požiadavku na kód prípadne konkrétny pracovný proces. [20][21]

2.3.2 Virtuálni agenti

Virtuálni agenti predstavujú pokročilú formu chatbotov, ktorí sú kombináciou troch hlavných technológií – RPA, technika spracovania prirodzeného jazyka NLP (Natural language processing) a inteligentné vyhľadávanie. Títo virtuálni agenti dokážu nielen efektívne komunikovať s používateľmi, ale aj vykonávať používateľom zadané úlohy a automatizovať procesy. [22]

Výrazy chatbot, AI chatbot a virtuálny agent sú často zamieňané. Akýkoľvek systém simulujúci ľudskú konverzáciu je nazývaný chatbot. AI chatbot je chatbot, ktorý využíva rôzne AI technológie, od strojového učenia, po spracovanie prirodzeného jazyka NLP a porozumenie prirodzeného jazyka NLU, ktorú presne interpretujú otázku používateľa a spájajú ku špecifickej téme. Virtuálni agenti sú ďalším stupňom vývoja AI chatbotov, ktoré spájajú rôzne AI technológie s RPA v jednom rozhraní. [22]

Jedným z príkladov virtuálneho agenta je riešenie nazývané Prompt2Task, ktorý imituje ľudské správanie a to tak, že interpretuje požiadavku používateľa, spracováva externé informácie na účely generovania RPA a vykonáva úkony na smartfónoch. [23]

Prompt2Task slúži ako inteligentný navigačný asistent, ktorý dynamicky generuje RPA procesy na základe pokynov od používateľa. Tento systém sa musí neustále prispôsobovať častým aktualizáciám aplikácií a s tým spojenými dynamickými zmenami v rozhraní. Taktiež časté sú nejasnosti v požiadavkách od používateľov. Na prekonanie týchto problémov tento systém využíva viac agentov, pričom každý z nich pokrýva rozličné úlohy. Hlavnou časťou systému je tzv. dynamická báza znalostí, ktorá sa neustále aktualizuje, rozširuje a prispôsobuje svoje informácie na základe nových vstupov, skúseností a interakcií. Prompt2Task má troch primárnych agentov – Analytický agent, Agent vyhľadávania a Agent mobilnej sémantiky. Analytický agent analyzuje a spracováva vstup od používateľa. Agent vyhľadávania vyhľadáva a následne využíva online návody na automatizáciu. Agent mobilnej sémantiky, ktorý analyzuje mobilné GUI s cieľom nájdenia čo najlepšej zhody so vstupom od používateľa. Všetkých spomínaných agentov zastrešuje Hodnotiaci agent, ktorého úlohou je posudzovať správnosť rozhodovania ostatných agentov a zlepšovať ich presnosť a spoľahlivosť. [23]

3 Príprava riešenia pri tvorbe virtuálneho agenta

V tejto časti sa zameriame na kľúčové kroky potrebné na implementáciu riešenia. Pred samotnou implementáciou je potrebné si ujasniť, akej oblasti sa bude virtuálny agent venovať. Následne je nutné zabezpečiť potrebné technológie, identifikovať a implementovať potrebné definované úlohy a napokon pripraviť mechanizmus na dynamické generovanie RPA kódu na základe vstupu od používateľa.

3.1 Príprava komponentov a technológií

Navrhované riešenie pozostáva z viacerých prepojených komponentov a technológií. Na implementáciu riešenia sú potrebné tri hlavné technologické oblasti, ktorými sú chatbot, RPA systém a komunikačné rozhranie medzi nimi.

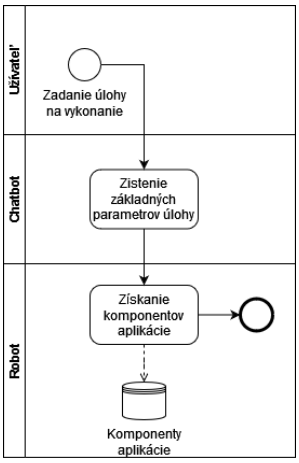
Chatbot je zodpovedný za spracovanie prirodzeného jazyka, interpretáciu požiadaviek a interakciu s používateľom. Na to je potrebné využiť veľké jazykové modely, ktoré, ako bolo spomínané, umožňujú porozumieť a generovať text v prirodzenom jazyku. Zároveň je teda nutné si vybrať vhodný spôsob vedenia konverzácie, či už použitím chatbota, QA modulu otázok a odpovedí (Question and Answer module) alebo iného riešenia, pričom my sme si vybrali QA modul z dôvodu jednoduchosti integrácie v rámci navrhovaného riešenia.

RPA systém je kľúčový pri automatizovaní úloh, preto je nutné si zvoliť vhodný RPA softvér, pričom my sme si zvolili Robot Framework kvôli jednoduchosti interakcie s inými technológiami. RPA bot následne vykoná buď preddefinované úlohy, alebo dynamicky generované automatizačné skripty na základe požiadavky používateľa.

Aby vedel chatbot komunikovať s RPA robotom je nutné si určiť komunikačné rozhranie. Na to je možné využiť napríklad REST API, WebSockets či iné vhodné komunikačné rozhranie. Komunikačné rozhranie musí byť obojsmerné, čiže nielen schopné odosielať požiadavky na RPA robota, ale ich aj vedieť od RPA robota prijať a spracovať.

3.2 Získanie komponentov z aplikácie

Táto časť sa zameriava na automatizované získavanie definovaných komponentov z aplikácie. To umožní identifikovať štruktúru, čo je dôležité pri navigácií. Cieľom tejto časti je zistiť prepojenia medzi jednotlivými oknami/stránkami a na základe toho vytvoriť hierarchickú štruktúru. Na Obr. 1 je možné vidieť diagram zobrazujúci možné riešenie tejto časti.

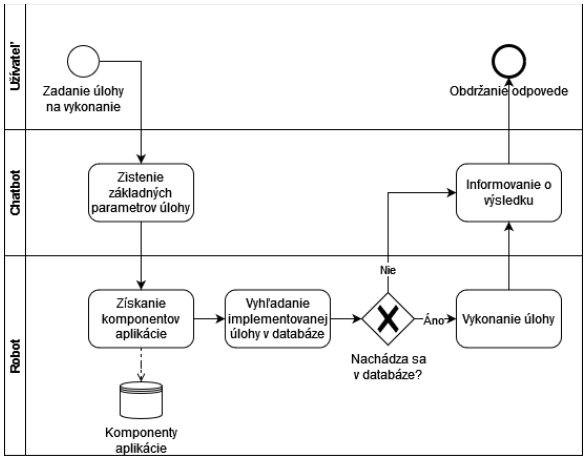


Obr. 1 Získavanie komponentov z aplikácie

Ako je znázornené na Obr. 1, získavanie potrebných elementov je realizovateľné pomocou RPA. V prípade, že cieľom je získať hierarchiu prepojení, RPA robot prehliada napr. webové stránky, extrahuje relevantné informácie a ukladá ich do štruktúrovaného formátu, ktorým môže byť napr. Excel. V ďalšom kroku sa získané údaje posielajú AI na trénovanie. AI využíva získané dáta na vytvorenie modelu, ktorý rozpozná vzory a významové vzťahy medzi jednotlivými stránkami.

3.3 Implementácia fixne definovaných úloh

V tejto časti sa zameriame na vytvorenie systému, ktorý umožňuje na základe interakcie používateľa s chatbotom vykonávať preddefinované úlohy (RPA procesy). Na to je potrebné vytvoriť si niekoľko relevantných úloh, pričom tieto úlohy musia byť jasne definované a dobre otestované. Chatbot pri tomto procese využíva LLM, pričom sa extrahuje zámer používateľa spolu s dôležitými parametrami. V prípade existujúcej preddefinovanej úlohy sa posiela požiadavka na vykonanie RPA robotom. Na Obr. 2 je možné vidieť diagram zobrazujúci možné riešenie tejto časti.

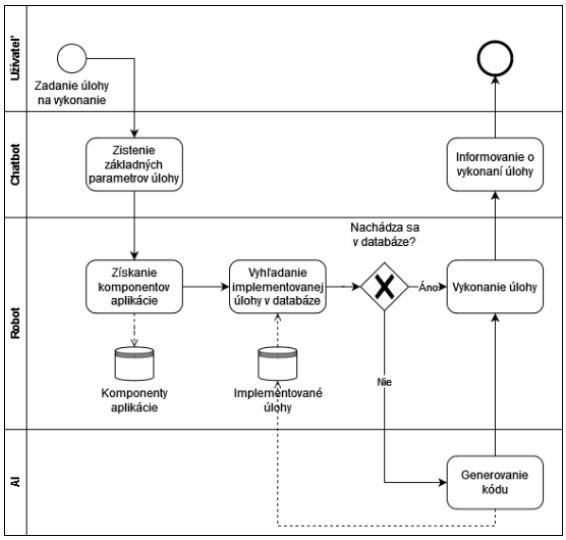


Obr. 2 Diagram znázorňujúci systém umožňujúci vykonávať preddefinované procesy na základe interakcie s chatbotom

3.4 Dynamické generovanie RPA kódu

Táto časť je zameraná na dynamické vytváranie RPA kódu na základe interakcie s používateľom a získaných dát z aplikácie.

Po tom, čo používateľ zadá požiadavku, chatbot požiadavku spracuje pomocou NLU a NLP. Následne sa zisťuje existencia už naimplementovaného rovnakého riešenia. V prípade, že existuje, robot ho vykoná. V prípade, že ešte neexistuje, systém pristupuje k dynamickému generovaniu RPA kódu. Po vygenerovaní je kód pripravený na nasadenie a vykonanie. Tento prístup umožňuje flexibilné prispôsobenie automatizačných úloh požiadavkám používateľov bez potreby hlbokej technickej expertízy. Na Obr. 3 je znázornený zjednodušený diagram procesu, ktorý spracuje vstup od používateľa, v prípade potreby vygeneruje RPA kód a následne ho vykoná.



Obr. 3 Generovanie RPA kódu na základe požiadavky používateľa a jeho následné vykonanie

4 Záver

Využitie synergie medzi RPA a LLM predstavuje dôležitý posun k inteligentnej automatizácií pracovných procesov. Spojením schopnosti porozumieť prirodzenému jazyku, čo zastrešuje chatbot, a schopnosti vykonávať úlohy, čo je úlohou RPA, vzniká riešenie, ktoré umožňuje intuitívne vytváranie a ovládanie automatizovaných procesov bez nutnosti programátorských znalostí.

Teoretický základ riešenia naznačuje, že synergia medzi RPA a LLM môže výrazne zlepšiť interakciu medzi človekom a automatizačnými nástrojmi. Doterajšia analýza dostupných riešení a príprava architektúry poskytli pevný základ pre ďalší vývoj. Budúci vývoj tejto práce by potenciálne mohol predstavovať implementáciu a testovanie riešenia jednotlivých častí. Cieľom tejto práce bolo definovať koncepčný rámec pre budúce riešenia v oblasti inteligentnej automatizácie.

Poďakovanie

Rada by som sa poďakovala Ing. Frederik Valocký, PhD. a doc. Ing. Peter Drahoš, PhD. za cenné rady a odborné usmernenia pri riešení tejto problematiky.

Literatúra

[1] DI CICCIO, C. A KOL. Business Process Management: Blockchain and Central and Eastern Europe Forum, Springer Cham, Edition 1, 2019, 344 s. ISBN 978-3-030-30428-7.
[2] RPA [online] [cit. 2025-03-28]. Dostupné na: <https://www.ibm.com/topics/rpa>
[3] PRIMER, A. Introduction to robotic process automation. Institute for Robotic Process Automation, 2015, 1-35 s.
[4] Systémy RPA: Výhody a nevýhody automatizace účetních procesů [online] [cit. 2025-03-28]. Dostupné na: <https://czechtechnology.cz/samosprava/systemy-rpa-vyhody-a-nevyhody-automatizace-ucetnich-procesu/>
[5] RPA tools comparisons [online] [cit. 2025-03-28]. Dostupné na: <https://www.besanttechnologies.com/rpa-tools-comparisons>
[6] KeyMark: RPA tools comparisons [online] [cit. 2025-03-28]. Dostupné na: <https://www.keymarkinc.com/rpa-tools-comparison/>
[7] BAWEJA, D. A Comparative Analysis of Automation Anywhere, UiPath, and BluePrism, IEEE, 2023, 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 2023, pp. 1715-1718, ISBN:979-8-3503-9926-4.
[8] KHAN, S. COMPARATIVE ANALYSIS OF RPA TOOLS-UIPATH, AUTOMATION ANYWHERE AND BLUEPRISM, 2020, vol. 8, issue. 11. 1-6 s. ISSN: 2321-8363 [online] [cit. 2025-03-28]. Dostupné na: <https://www.ijcsma.com/articles/comparative-analysis-of-rpa-toolsuipath-automation-anywhere-and-blueprism.pdf>
[9] ZHAO, W. X. A KOL. A Survey of Large Language Models, 2023, 140 s. [online] [cit. 2025-03-28]. Dostupné na: <https://arxiv.org/pdf/2303.18223>
[10] [NLP] Statistical Language Model [online] [cit. 2025-03-28]. Dostupné na: <https://medium.com/@redbeet1007/nlp-statistical-language-model-bca51c21d934>
[11] Jurafsky D., Martin J. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models, 3rd edition, 2025, chapter 7 Neural Networks [online] [cit. 2025-03-28]. Dostupné na: <https://web.stanford.edu/~jurafsky/slp3/7.pdf>
[12] Alomari, A., Idris N., Sabri A. Q. Md, Alsmadi I. Deep reinforcement and transfer learning for abstractive text summarization: A review, Computer Speech & Language, 2022, vol. 71 ISSN: 0885-2308 [online] [cit. 2025-03-28]. Dostupné na: <https://www.sciencedirect.com/science/article/pii/S0885230821000796>
[13] Large Language Models [online] [cit. 2025-03-28]. Dostupné na: <https://www.ibm.com/topics/large-language-models>
[14] Large Language Models Explained [online] [cit. 2025-03-28]. Dostupné na: <https://www.nvidia.com/en-eu/glossary/large-language-models/>
[15] chatbot [online] [cit. 2025-03-28]. Dostupné na: <https://www.techtarget.com/searchcustomerexperience/definition/chatbot>
[16] CALDARINI, G. – MCGARRY, K. A – JAF, S. Literature Survey of Recent Advances in Chatbots. Information, 2022, vol 13, no. 41. [online] [cit. 2025-03-28]. Dostupné na: <https://www.mdpi.com/2078-2489/13/1/41>
[17] ADAMOPOULOU, E. – MOUSSIADES, L. Chatbots: History, technology, and applications, vol. 2, 2020, ISSN 2666-8270 [online] [cit. 2025-03-28]. Dostupné na: <https://www.sciencedirect.com/science/article/pii/S2666827020300062>
[18] ARAI, K. – BHATIA, R. – KAPOOR, S. Proceedings of the Future Technologies Conference (FTC) 2018: Volume 2 (Advances in Intelligent Systems and Computing, 881), 1. vyd. Springer, 2019, 1198 s. ISBN 978-3030026820.
[19] What is a chatbot? [online] [cit. 2025-03-28]. Dostupné na: <https://www.ibm.com/topics/chatbots>
[20] LLM Powered Robotic Process Automation: Unleashing the Industry Transformation Revolution [online] [cit. 2025-03-28]. Dostupné na: <https://www.tangentia.com/llm-powered-robotic-process-automation-unleashing-the-industry-transformation-revolution/>
[21] CAO, J. A Study on Deploying Large Language Models as Agents, Massachusetts Institute of Technology, 2024, 59 s. [online] [cit. 2025-03-28]. Dostupné na: <https://dspace.mit.edu/bitstream/handle/1721.1/157177/cao-jiannan-sdm-sm-2024-thesis.pdf?sequence=1&isAllowed=y>
[22] What is a virtual agent? [online] [cit. 2025-03-28]. Dostupné na: <https://www.ibm.com/think/topics/virtual-agent>
[23] HUANG, T. A KOL. PromptRPA: Generating Robotic Process Automation on Smartphones from Textual Prompts, 2024, 34 s. [online] [cit. 2025-03-28]. Dostupné na: <https://arxiv.org/pdf/2404.02475>

BabyGuard: Pokojný spánok dieťaťa vďaka moderným technológiám

Bc. Matej Bašista[1]

[1]Ústav automobilovej mechatroniky, FEI STU v Bratislave

xbasistam@stuba.sk

Abstrakt - Cieľom tejto práce bolo vytvoriť mobilnú cross-platformovú aplikáciu umožňujúcu rodičom sledovať spánkové vzorce ich detí. Aplikácia bola postavená na platforme Flutter a využíva databázu Firebase na ukladanie údajov. Okrem softvérovej časti sa venovala aj hardvérovému riešeniu, ktoré umožňuje meranie spánkových vzorcov detí pomocou senzorov. Výsledkom je plne funkčná aplikácia, ktorá poskytuje rodičom nástroje na efektívne monitorovanie zdravia ich detí.

1 Úvod

Mobilné aplikácie sa stali neoddeliteľnou súčasťou našich životov, najmä v oblasti zdravia. Pre rodičov malých detí existuje rastúca potreba aplikácií, ktoré pomáhajú monitorovať ich spánok a zdravie.

V súčasnej dobe, kedy sa rodičia čoraz viac spoliehajú na technológie, je dôležité poskytnúť im nástroje, ktoré im pomôžu efektívne monitorovať zdravie ich detí. Aplikácia BabyGuard vznikla ako odpoveď na túto potrebu a ponúka rodičom komplexný nástroj na sledovanie spánkových vzorcov ich detí.

Tento článok popisuje vývoj a implementáciu aplikácie BabyGuard, ktorá je postavená na moderných technológiách, ako sú Flutter a Firebase, a zahŕňa aj hardvérové riešenie na monitorovanie spánku.

2 Aplikácia BabyGuard

Aplikácia BabyGuard využíva moderné technológie na monitorovanie spánkových vzorcov detí a zabezpečuje rodičom komplexný nástroj na sledovanie zdravia ich detí počas spánku.

Celý systém je postavený na robustnej komunikácii medzi mobilným zariadením rodiča a rôznymi IoT zariadeniami, ktoré sú zodpovedné za zber a spracovanie zdravotných údajov. Tieto zariadenia zahŕňajú senzory, mikrokontroléry a výkonné spracovacie jednotky, ktoré v reálnom čase snímajú vitálne parametre dieťaťa, ako je tep a saturácia kyslíka v krvi.

Cieľom tejto práce bolo vytvoriť mobilnú aplikáciu, ktorá by rodičom umožnila monitorovať spánkové vzorce ich detí v reálnom čase a poskytovala im dôležité informácie o ich zdravotnom stave. Tento cieľ bol dosiahnutý kombináciou moderných technológií, ako sú Flutter, Firebase a IoT zariadenia.

2.1 Súčasný stav a problém

Aplikácie na monitorovanie spánku bábätka sú obzvlášť dôležité pre predčasne narodené deti, ktorých mozog ešte dozrieva. Sledovanie pulzu a hladiny kyslíka počas spánku pomáha odhaliť zdravotné problémy, ako je apnoe alebo nepravidelný srdcový rytmus, a poskytuje včasné varovné signály pre rodičov. Nepretržité monitorovanie zvyšuje bezpečnosť a môže zabrániť vážnym stavom, ako je syndróm náhleho úmrtia dojčiat (SIDS), ktorý je hlavnou príčinou úmrtia dojčiat do 1 roka. [1]

Na trhu existuje niekoľko aplikácií na monitorovanie spánku detí, avšak väčšina z nich sa zameriava len na základné funkcie, ako je zaznamenávanie dĺžky spánku. BabyGuard sa odlišuje tým, že ponúka komplexné riešenie vrátane monitorovania tepu, saturácie kyslíka a upozornení na abnormálne hodnoty spolu s odporúčanou lekárskou pomocou v okolí, čo z neho robí unikátny nástroj pre rodičov.

2.2 Benefity aplikácie BabyGuard

Aplikácia BabyGuard prináša množstvo výhod pre rodičov, ktorí chcú mať lepší prehľad o zdravotnom stave svojich detí. Medzi hlavné benefity aplikácie patria:

- **Upozornenia na abnormálne hodnoty spánku:** Aplikácia automaticky informuje rodičov o neobvyklých alebo alarmujúcich hodnotách spánku, ktoré umožňuje identifikovať možné problémy a prijímať primerané opatrenia.
- **Rýchly prístup k odborníkom:** V prípade zaznamenania abnormálnych hodnôt aplikácia umožňuje rýchle vyhľadanie najbližších lekárov a navigáciu k nim, čím zaisťuje okamžitú pomoc a podporu.
- **Osobné účty a nastavenia:** Používatelia majú možnosť vytvorenia osobného účtu a prístupu k ich osobným údajom a nastaveniam, čo zabezpečuje individuálnu personalizáciu a lepšiu kontrolu.
- **Evidencia hmotnosti detí:** Aplikácia umožňuje sledovať vývoj hmotnosti dieťaťa a identifikovať akékoľvek zmeny, čo je dôležité pre monitorovanie rastu a zdravia.
- **Možnosť prispôsobenia profilu:** Používatelia môžu meniť svoj profilový obrázok a prispôsobiť si svoj profil podľa vlastných preferencií, čo zlepšuje používateľský zážitok a spokojnosť.

- **Voľba svetlého alebo tmavého režimu:** Aplikácia umožňuje voľbu svetlého alebo tmavého režimu, vďaka čomu sa zvyšuje čitateľnosť a ochranu očí v rôznych svetelných podmienkach.
- **Možnosť pridávania viacerých detí:** Rodičia môžu sledovať spánkové návyky a zvyky viacerých detí z jedného účtu, čo zjednodušuje a zlepšuje monitorovanie spánkových vzorcov všetkých detí rodiny.

2.3 Cross-platformová technológia

Aplikácia BabyGuard je vyvinutá pomocou technológie Flutter, ktorá umožňuje vytvárať výkonné aplikácie pre viacero platforiem, vrátane Android a iOS. Flutter je moderný framework, ktorý poskytuje vývojárom nástroje na tvorbu nielen natívnych aplikácií, ale aj aplikácií s vysokým výkonom, ktoré sú schopné pracovať na rôznych zariadeniach bez nutnosti písať samostatný kód pre každú platformu.

Jednou z hlavných výhod Flutteru je jeho flexibilita a rýchlosť. Tento framework umožňuje vývoj aplikácií, ktoré sú nielen vizuálne atraktívne, ale aj veľmi rýchle, čo je nevyhnutné pre aplikácie, ktoré sa zaoberajú citlivými dátami, ako je monitorovanie zdravia. Flutter tiež poskytuje prístup k natívnym funkciám zariadení, čím umožňuje využiť všetky možnosti daných platforiem bez kompromisov v oblasti výkonu. [2]

2.4 Cloudová databáza

Aplikácia BabyGuard využíva cloudovú databázu Firebase, ktorá bola zvolená pre svoju rýchlosť, spoľahlivosť a jednoduchú integráciu s mobilnými a IoT zariadeniami. Firebase poskytuje robustné riešenie na ukladanie a správu údajov o spánkových vzorcoch detí, čo umožňuje okamžitý prenos a synchronizáciu údajov medzi zariadeniami. Táto funkcionalita zabezpečuje plynulú komunikáciu a efektívne monitorovanie spánku detí v reálnom čase, čo je kľúčové pre poskytovanie aktuálnych a presných informácií rodičom. [3]

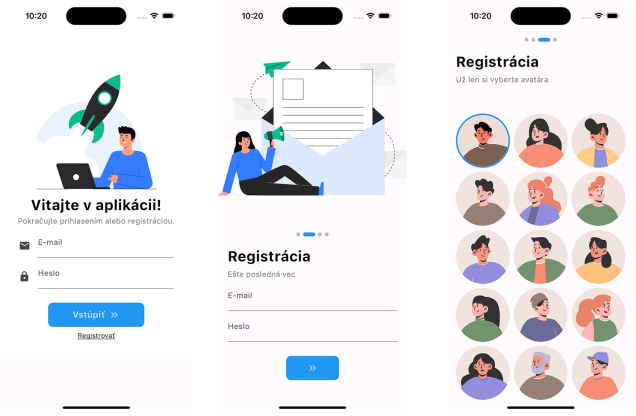
2.5 Grafické používateľské rozhranie

Grafické používateľské rozhranie je navrhnuté tak, aby bolo jednoduché a intuitívne pre používateľov všetkých vekových kategórií. Aplikácia obsahuje intuitívne a prehľadné rozloženie, ktoré zabezpečuje rýchly prístup ku všetkým dôležitým informáciám a funkciám.

Údaje z meraní sú zobrazené v podobe grafu, ktorý poskytuje prehľad o spánkových vzorcoch a zvykoch dieťaťa. Okrem toho aplikácia obsahuje aj informácie o zdravotných výhodách spánku a tipy na zlepšenie kvality spánku detí. Tieto informácie prispievajú k lepšiemu porozumeniu spánkových potrieb detí a zabezpečia, že rodičia budú mať dostatočné znalosti na to, aby mohli efektívne riadiť a monitorovať spánkové vzorce svojich detí.

2.5.1 Rozhranie a možnosti pre neprihláseného používateľa

Neprihlásený používateľ má možnosť zaregistrovať sa alebo prihlásiť do aplikácie, čím získa prístup k jej plným funkciám.

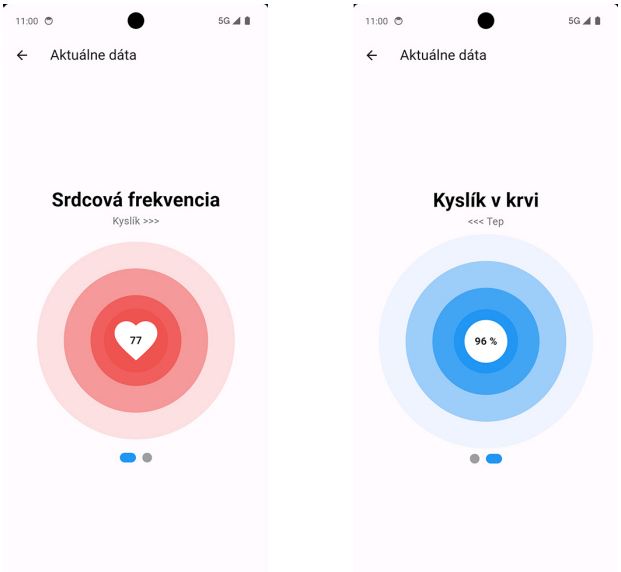


Obr. 1: Grafické rozhranie pre neprihláseného používateľa

Registrácia je jednoduchá a intuitívna, pričom si používateľ môže vybrať profilový obrázok, čo prispieva k personalizácii jeho účtu a zlepšuje celkový zážitok z používania aplikácie. Tento krok umožňuje používateľovi vytvoriť si vlastnú identitu v rámci aplikácie a zároveň ju spríjemňuje. Po úspešnom prihlásení je používateľ presmerovaný na hlavnú obrazovku, ktorá slúži ako centrálny bod pre interakciu s aplikáciou.

2.5.2 Zobrazenie dát v reálnom čase

Aplikácia BabyGuard ponúka zobrazenie údajov v reálnom čase, ktoré umožňujú rodičom sledovať spánkové vzorce ich detí počas spánku. Tento okamžitý prístup k informáciám pomáha monitorovať stav dieťaťa a poskytuje podrobné údaje o jeho spánku. V prípade abnormalít aplikácia generuje upozornenia vo forme notifikácií, ktoré obsahujú informácie o stave dieťaťa a odporúčania na ďalšie kroky, čo zvyšuje bezpečnosť a pohodlie rodičov. Ak používateľ povolí lokalizačné služby, aplikácia dokáže nájsť najbližšiu lekársku pomoc, čo je užitočné v núdzových situáciách.

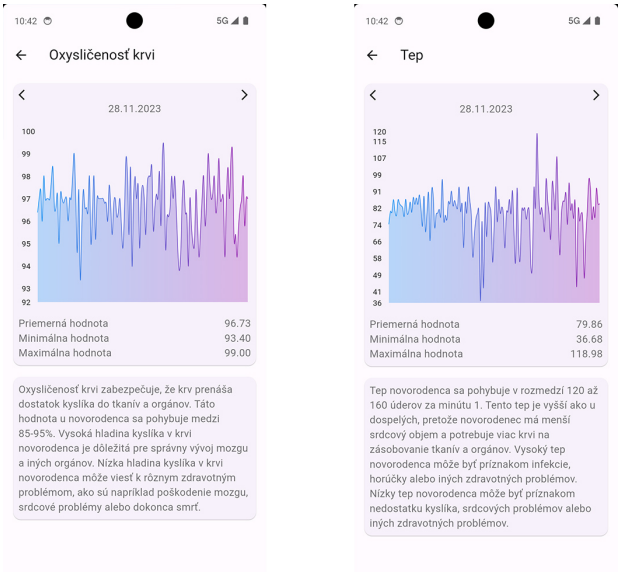


Obr. 2: Zobrazenie dát v reálnom čase

2.5.3 *Zobrazenie histórie spánku*

Aplikácia poskytuje rodičom komplexný nástroj na monitorovanie spánkových vzorcov ich detí. Prostredníctvom prehľadných grafov a detailných štatistík môžu rodičia sledovať kľúčové údaje, ako sú minimálna, maximálna a priemerná dĺžka spánku ich dieťaťa. Okrem toho aplikácia poskytuje informácie o zdravých spánkových hodnotách, ktoré pomáhajú identifikovať potenciálne oblasti na zlepšenie spánkového režimu.

Týmto spôsobom aplikácia nielen umožňuje lepšie porozumieť spánkovým návykom dieťaťa, ale aj prispieva k monitorovaniu jeho celkového zdravia a vývoja. Tento nástroj je pre rodičov veľmi užitočný, pretože im poskytuje cenné informácie, ktoré môžu využiť na zabezpečenie optimálnych podmienok pre zdravý a kvalitný spánok ich detí.



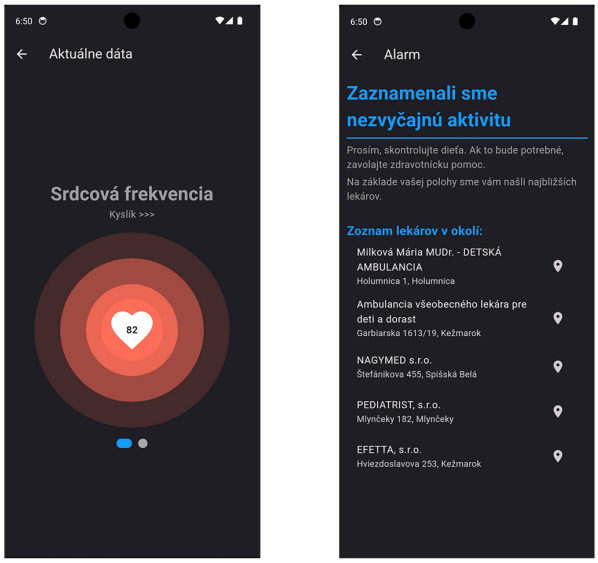
Obr. 3: Zobrazenie histórie spánku

2.5.4 *Možnosť pridávania viacerých detí*

Aplikácia poskytuje rodičom možnosť pridávať deti do svojho používateľského profilu a vytvárať pre každé dieťa osobný profil. Rodičia môžu sledovať vývoj spánkových návykov jednotlivých detí a porovnávať ich medzi sebou, čím získajú lepší prehľad o tom, ako sa spánkové návyky ich detí vyvíjajú. Tento prístup pomáha rodičom lepšie porozumieť špecifickým potrebám každého dieťaťa a prispieva k efektívnejšiemu riadeniu ich spánkového režimu.

2.5.5 *Vyhľadanie pomoci v okolí*

Aplikácia poskytuje používateľom praktickú funkciu, ktorá im umožňuje vyhľadávať pomoc v okolí a získať podrobné informácie o lekároch, nemocniciach a ďalších zdravotníckych zariadeniach nachádzajúcich sa v ich blízkosti. Táto funkcia je obzvlášť užitočná v situáciách, keď je potrebná rýchla a efektívna pomoc, ako sú núdzové prípady alebo náhle zdravotné problémy. Používatelia môžu rýchlo a jednoducho nájsť najbližšie zdravotnícke zariadenia s možnosťou navigácie

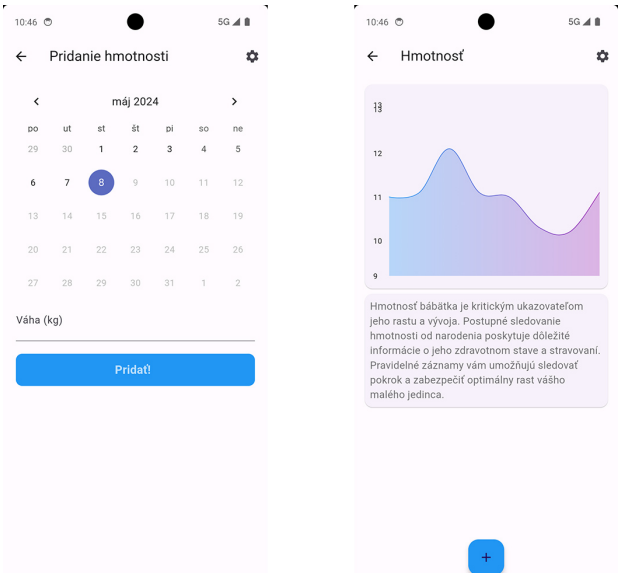


Obr. 4: Zobrazenie dát v reálnom čase s alarmom

k ním, čo im poskytuje cenné informácie o dostupných službách a možnostiach v ich okolí.

2.5.6 *Evidencia hmotnosti detí*

BabyGuard umožňuje rodičom monitorovať vývoj hmotnosti ich detí v priebehu času, čo je dôležité pre sledovanie rastu a zdravia. Používatelia môžu zaznamenávať hmotnosť v rôznych obdobiach, pričom údaje sú vizualizované v grafe, ktorý zobrazuje trendy a zmeny. Táto funkcia je predvolene zapnutá a zobrazuje sa na domovskej obrazovke, ale používatelia ju môžu vypnúť v nastaveniach, ak si to želajú. Ak je táto funkcia vypnutá, graf priebehu hmotnosti dieťaťa sa nezobrazí na domovskej obrazovke, ale používatelia stále môžu využívať túto funkciu v rámci aplikácie, ak ju potrebujú.

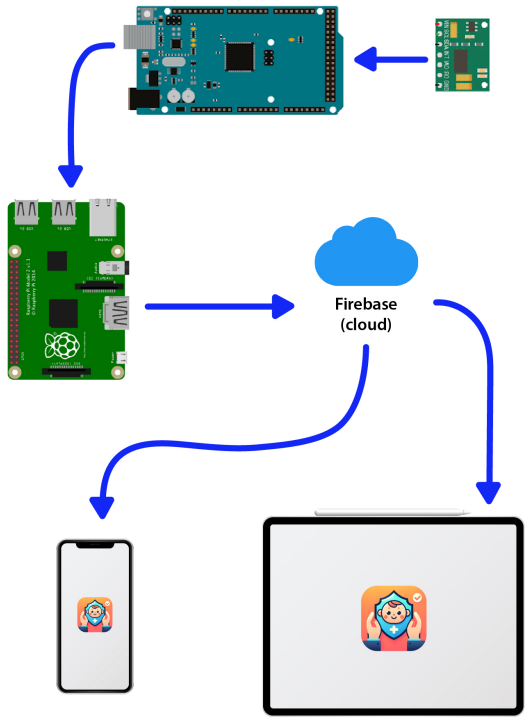


Obr. 5: Evidencia hmotnosti detí

2.6 Komunikácia a monitorovanie

Základom aplikácie BabyGuard je efektívna a stabilná komunikácia medzi mobilným zariadením rodiča a IoT zariadeniami, ktoré tvoria komplexný monitorovací systém. Tento systém pozostáva z troch hlavných komponentov: senzora, mikrokontroléra Arduino a zariadenia Raspberry Pi. Senzor, ktorý je upevnený na ruke alebo nohe dieťaťa, zbiera údaje o jeho pulze a kyslíku v krvi. Tieto údaje sú následne prenášané do mikrokontroléra Arduino, ktorý ich spracováva a pripravuje na ďalšiu komunikáciu.

Raspberry Pi, ktorý slúži ako centrálny komunikačný uzol, zabezpečuje prenos týchto údajov do cloudovej databázy Firebase. Táto databáza umožňuje ukladanie a správu údajov v reálnom čase, čo poskytuje rodičom okamžitý prístup k informáciám o spánkových vzorcoch a zdravotnom stave ich dieťaťa. Rodičia môžu tieto údaje sledovať prostredníctvom mobilnej aplikácie, ktorá zobrazuje prehľadné grafy a štatistiky, ako aj upozornenia v prípade akýchkoľvek abnormalít.



Obr. 6: Schéma systému BabyGuard

2.7 Meranie a posielanie hlásení

Meranie spánkových vzorcov je vykonávané senzorom MAX30100, ktorý je špeciálne navrhnutý na monitorovanie tepu a hladiny kyslíka v krvi. Tento senzor je pripojený k mikrokontroléru Arduino, ktorý slúži na spracovanie a riadenie senzorových údajov.

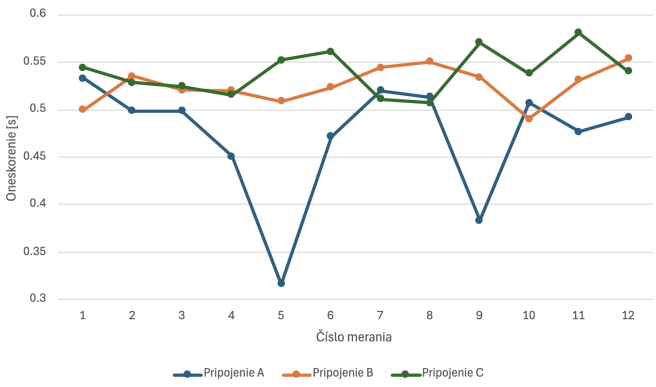
- **MAX30100** meria tep a saturáciu kyslíka v krvi (SpO2) pomocou optických senzorov, ktoré snímajú odraz svetla v tkanivách a na základe toho určujú hodnoty.
- **Arduino** je zodpovedné za čítanie týchto hodnôt zo senzora a ich odoslanie na Raspberry Pi, ktoré sa stará o ďalšie spracovanie.

Aplikácia BabyGuard je navrhnutá tak, aby v prípade zistenia abnormálnych hodnôt (napr. nízka hladina kyslíka, nízky alebo vysoký tep) automaticky generovala upozornenie, ktoré sa okamžite pošle rodičovi prostredníctvom alarmu v mobilnom zariadení spolu s kontaktom na lekársku pomoc v okolí.

2.8 Rýchlosť a oneskorenie

Na testovanie latencie a oneskorenia bola použitá kombinácia mikrokontrolérov Arduino, Raspberry Pi a mobilnej aplikácie Flutter. Testovali sme rýchlosť prenosu údajov zo senzorov do Raspberry Pi a následne do aplikácie cez cloudovú databázu Firebase.

Testy ukázali, že aplikácia vykazuje veľmi nízke oneskorenie, čo je kľúčové pre použitie v reálnom čase. Väčšina meraní sa pohybovala v rozsahu do pol sekundy, čo zaručuje, že rodičia dostanú aktuálne informácie o zdravotnom stave svojich detí bez významného oneskorenia.



Obr. 7: Graf zobrazujúci latenciu pre jednotlivé pripojenia

Otestované boli pripojenia s rôznymi parametrami s cieľom overiť, ako ovplyvňuje rýchlosť internetu latenciu aplikácie:

Tab. 1: Tabuľka s parametrami pripojení

Pripojenie	Nahrávanie	Sťahovanie	Latencia
A	35 Mbps	6 Mbps	8 ms
B	4 Mbps	14 Mbps	44 ms
C	1 Mbps	10 Mbps	46 ms

2.9 Testovanie

Počas vývoja aplikácie BabyGuard boli rodičia aktívne zapojení do testovania, čo nám umožnilo získať cennú spätnú väzbu a overiť funkčnosť aplikácie v reálnych podmienkach. Testovanie prebiehalo v niekoľkých fázach, pričom rodičia mali možnosť vyjadriť sa k rôznym aspektom aplikácie, ako je

používateľské rozhranie, funkčnosť a spoľahlivosť. Väčšina rodičov ocenila najmä jednoduchosť a intuitívnosť aplikácie, ktorá im umožnila rýchlo a efektívne monitorovať zdravotné parametre ich detí, ako sú tep, saturácia kyslíka a dĺžka spánku.

2.10 Budúca práca

V budúcnosti plánujeme rozšíriť funkcionalitu aplikácie BabyGuard o ďalšie užitočné nástroje a služby, ktoré by rodičom poskytli ešte viac možností na monitorovanie zdravia ich detí. Medzi hlavné oblasti, na ktoré sa chceme zamerať, patrí:

- **Integrácia s ďalšími senzormi:** Plánujeme integrovať ďalšie senzory, ktoré by umožnili monitorovať ďalšie zdravotné parametre detí, ako je teplota tela, dýchanie.
- **Rozšírenie o kameru:** Chceme pridať možnosť sledovania spánku detí pomocou kamery, čo by rodičom umožnilo sledovať ich deti v reálnom čase.
- **Integrácia s ďalšími platformami:** Plánujeme rozšíriť podporu aplikácie na ďalšie platformy, ako sú Windows, MacOS, čo by umožnilo rodičom prístup k dôležitým informáciám o zdraví ich detí z rôznych zariadení.
- **Detská pestúnka:** V pláne je taktiež pridanie možnosti sledovania dieťaťa a upozornenie rodičov v prípade, že sa dieťa zobudí.
- **Analýza a predikcia pomocou AI:** Pri dostatku dát plánujeme využiť umelú inteligenciu na analýzu a predikciu potenciálnych nebezpečných situácií počas spánku.

3 Záver

Aplikácia BabyGuard je plne funkčným nástrojom, ktorý rodičom umožňuje efektívne monitorovať spánkové vzorce ich detí. Vďaka kombinácii pokročilých technológií, ako je Flutter a Firebase, spolu s hardvérovými riešeniami na meranie spánkových údajov, poskytuje aplikácia rodičom cenné informácie o zdraví ich detí. Tento systém nielen zvyšuje bezpečnosť, ale aj pohodlie rodičov, ktorí majú prístup k dôležitým zdravotným údajom v reálnom čase.

Poďakovanie

Na tomto mieste by som rád vyjadril svoju vďaku Ing. Lukášovi Beňovi, PhD., za jeho odborné vedenie a cenné rady počas celého procesu prípravy tejto práce. Jeho podpora a expertíza mi pomohli prekonať mnohé výzvy a umožnili mi dôkladne rozvinúť koncept aplikácie BabyGuard.

Rovnako by som chcel poďakovať všetkým odborníkom a vývojárom, ktorých práca a príspevky do oblasti cross-platformového vývoja mobilných aplikácií ma inšpirovali a obohatili tento projekt.

Literatúra

[1] Augusto Sola, Sergio G. Golombek Safe oxygen saturation targeting and monitoring in preterm infants: can we. *Acta Paediatrica*, 2014 DOI: 10.1111/apa.12692

[2] S. Sharma, S. Khare, V. Unival and S. Verma Hybrid Development in Flutter and its Widgets *2022 International Conference on Cyber Resilience (ICCR), Dubai, United Arab Emirates, 2022* DOI: 10.1109/ICCR56254.2022.9995973

[3] U. A. Madaminov and M. R. Allaberganova Firebase Database Usage and Application Technology in Modern Mobile Applications *2023 IEEE XVI International Scientific and Technical Conference Actual Problems of Electronic Instrument Engineering (APEIE), Novosibirsk, Russian Federation, 2023* DOI: 10.1109/APEIE59731.2023.10347828

Sekcia 4

Elektronika a fotonika, Elektrotechnika, Elektroenergetika

Projektovanie VN sietí a zariadení v BIM a analýza siete prostredníctvom integrácie programu ETAP s Revitom

Patrik Hakala[1], doc. Ing. Peter Janiga, PhD.[1],

[1] Ústav elektroenergetiky a aplikovanej elektrotechniky, FEI STU v Bratislave

xhakalap@stuba.sk

Abstrakt – Táto práca sa zaoberá návrhom a analýzou 22 kV vzdušnej distribučnej siete s využitím moderných digitálnych nástrojov. Cieľom je demonštrovať praktické využitie Building Information Modeling (BIM) pri navrhovaní elektrickej infraštruktúry v prostredí Autodesk Revit a následnej simulácie v analytickom softvéri ETAP. Modelovaná sieť má dĺžku 10 km a obsahuje všetky základné prvky vrátane stožiarov, vodičov, transformátorov a ochranných zariadení. Táto práca predstavuje úvodnú fázu širšieho výskumu, ktorému sa plánujem venovať podrobnejšie v diplomovej práci. V nej sa zameriam na rozšírenie návrhu o digitálne dvojča, prepojenie s GIS, ako aj na podrobnejšiu analýzu prevádzky siete v reálnom čase.

1 Úvod

Spoľahlivá distribúcia elektrickej energie je jedným zo základných pilierov moderného hospodárstva a komfortu. Návrh distribučných sietí však už dnes nie je len o výpočtoch prúdov a dimenzovaní káblov, ale čoraz viac sa presúva do digitálneho priestoru. V tejto súvislosti získava na význame koncept Building Information Modeling (BIM), ktorý prepája technické, priestorové a prevádzkové dáta do jedného jednotného systému.

Cieľom tejto práce je vytvoriť koncepčný návrh 22 kV vzdušnej siete v prostredí BIM a vykonať základné analýzy jej funkčnosti pomocou simulačného softvéru ETAP. Zvolený príklad má dĺžku 10 km a je navrhnutý ako funkčný model, ktorý môže slúžiť ako základ pre ďalšie rozširovanie a optimalizáciu.

Vzhľadom na rozsah a technickú náročnosť problematiky je táto práca chápaná ako úvodná etapa k pripravovanej diplomovej práci. V nej plánujem rozpracovať detailnejšie možnosti automatizácie návrhu, prepojenie s reálnym územným plánom pomocou GIS, ako aj vývoj a testovanie digitálneho dvojčaťa siete s možnosťou monitorovania a prediktívnej údržby.

2 Teoretický základ

2.1 BIM v elektroinžinierstve

Building Information Modeling (BIM) predstavuje komplexný proces tvorby, správy a zdieľania digitálnych informácií o stavebnom diele počas celého jeho životného cyklu – od návrhu, cez výstavbu až po prevádzku a údržbu. Hoci sa BIM spočiatku využíval najmä v oblasti architektúry a statiky, v posledných rokoch sa jeho potenciál ukazuje aj v technických profesiách, predovšetkým v oblasti elektrotechniky.

BIM umožňuje inžinierom a projektantom vytvoriť presný 3D model elektroinštalácií, ktorý neobsahuje iba geometrické dáta, ale aj technické a funkčné informácie. Každý prvok – či už ide o rozvádzač, kábel, vývod alebo stĺp – má v modeli priradené parametre ako dimenzovanie, napätie, prúd, výrobca či trieda ochrany. Vďaka tomu je možné automaticky generovať výkazy materiálu, kontrolovať kolízie s inými profesiami, sledovať stavebné fázy a znižovať riziko chýb pri realizácii.

Zavedenie BIM do oblasti elektroenergetických sietí, najmä distribučných, otvára nové možnosti. V prípade návrhu vonkajších vedení možno využiť topografické podklady, GIS dáta a integrovať návrh s environmentálnymi či legislatívnymi obmedzeniami. Výsledný model môže slúžiť ako základ pre digitálne dvojča siete, ktoré umožňuje online sledovanie parametrov, prediktívnu údržbu a efektívne plánovanie modernizácie.

2.2 Vzdušné distribučné siete napätia 22kV

Distribučné siete stredného napätia (VN – vysoké napätie) tvoria dôležitú spojnicu medzi transformáciou z prenosovej sústavy a koncovým užívateľom. V stredoeurópskom priestore sa najčastejšie používa napäťová hladina 22 kV, ktorá predstavuje optimálny kompromis medzi prenosovou kapacitou a technickými požiadavkami na izoláciu a bezpečnosť.

Vzdušné vedenia 22 kV sa bežne používajú v menej urbanizovaných oblastiach, kde nie je ekonomicky výhodné budovať podzemné káblové trasy. Ich výhodou je nižšia

investičná náročnosť, jednoduchšia lokalizácia porúch a rýchlejšia realizácia. Na druhej strane sú však viac vystavené vplyvom počasia (vietor, námraza, búrky) a vyžadujú dôslednú údržbu vegetácie v ochrannom pásme.

Základné komponenty 22 kV vzdušnej siete:

- **Stožiare:** Najčastejšie železobetónové (typ J6), výška 9–12 m, rozstup cca 100 m.
- **Vodiče:** Typ AlFe (hliník-oceľ), zvyčajne prierez 35 mm² až 70 mm², nesené na izolátoroch.
- **Izolátory:** Keramické alebo kompozitné, navrhnuté na vonkajšie prostredie a napätie 24 kV.
- **Transformačné stanice:** Obvykle 22/0.4 kV, výkon 250 – 1000 kVA, pripojené priamo k vedení.
- **Ochranné prvky:** Odpojovače, poistky, ističe, prepäťové ochrany, uzemnenie.

Pri návrhu siete treba zohľadniť nielen mechanické a elektrické parametre, ale aj požiadavky normy **STN EN 50341-1** (vzdušné vedenia nad 1 kV), bezpečnostné vzdialenosti, ochranu pred skratom a úbytok napätia.

2.3 Použité softvéry – Revit a ETAP

Autodesk Revit je BIM platforma, ktorá umožňuje modelovanie infraštruktúry v 3D priestore s priradenými technickými dátami. Využil som ho na vytvorenie geometrického modelu siete, rozmiestnenie prvkov a výkazy materiálov.

ETAP (Electrical Transient Analyzer Program) je analytický softvér, ktorý umožňuje simuláciu elektrických javov – od základnej load flow analýzy až po zložité dynamické simulácie porúch. V tejto práci slúžil najmä na overenie návrhu siete, výpočet napäťových poklesov, skratových prúdov a koordináciu ochrany.

V diplomovej práci plánujem tieto nástroje integrovať s ďalšími platformami – najmä **GIS (geografické informačné systémy)** a **SCADA** – za účelom vytvorenia plnohodnotného digitálneho dvojčaťa siete.

2.4 Praktické riešenie projektu

Modelovanie siete prebiehalo v prostredí **Autodesk Revit**, kde bola navrhnutá 10 km dlhá vzdušná trasa pozostávajúca zo 100 stožiarov. Vzdialenosť medzi stožiarmi bola štandardizovaná na 100 metrov. Všetky komponenty siete boli vložené ako parametrické objekty – tzv. „Revit Family“, vrátane stožiarov, výložníkov, vodičov a ochranných prvkov.

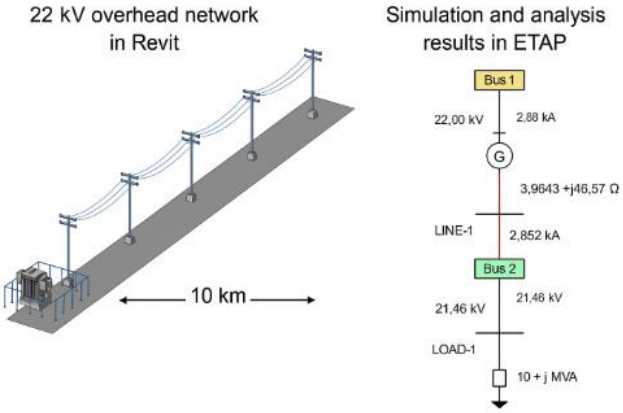
Sieť bola rozdelená na úseky podľa typického vedenia VN: hlavná vetva, odbočka, koncové body. **Transformačné stanice** boli umiestnené na začiatku a v polovici siete, každá s výkonom 630 kVA (22/0.4 kV). V modeli boli vytvorené aj „conduit“ trasy pre vodiče a vedenia, ktoré slúžia na vizuálnu aj technickú reprezentáciu trasy fázy A, B, C.

Na efektívne rozmiestnenie stožiarov som využil skriptovací nástroj **Dynamo**, kde boli vstupné súradnice čerpané z CSV súboru. Tento prístup sa ukázal ako veľmi efektívny a do budúcnosti ho plánujem rozšíriť o **automatické generovanie stožiarov podľa digitálneho výškového modelu terénu**, čomu sa chcem podrobne venovať v mojej diplomovej práci.

Vstupné parametre návrhu:

- Napätie siete: 22 kV
- Počet stožiarov: 100
- Dĺžka vedenia: 10 000 m
- Typ stožiarov: železobetónové J6
- Vodiče: AlFe 35 mm²
- Transformátory: 2x 630 kVA
- Ochrana: odpájače každé 2 km, prepäťové ochrany, uzemnenie

Z modelu bolo možné automaticky získať **výkaz materiálu (BoM)**, ktorý obsahoval zoznam všetkých komponentov, ich rozmery a množstvá. Tento výkaz predstavuje základ pre rozpočet stavby a verejné obstarávanie, pričom jeho export do formátu Excel alebo IFC je otázkou niekoľkých kliknutí.



Obr.1 Vizualizácia siete s návrhom v Revite a simuláciou v ETAPe

Po vytvorení modelu v Revite bol projekt exportovaný do formátu CSV a následne naimportovaný do simulačného prostredia **ETAP**. Vytvoril som jednopólovú schému siete, ktorá kopíruje fyzický návrh z BIM modelu. Následne boli pridané zaťažovacie body, trafostanice, ochranné zariadenia a nastavené elektrické parametre každého úseku.

V ETAP som vykonal tieto analýzy:

- **Load Flow**: Pri maximálnom zaťažení boli napäťové poklesy v prijateľných hodnotách. Na konci siete bol pokles 4,8 %, čo je v súlade s normou STN EN 50160.
- **Skratová analýza**: Trojfázové skraty boli simulované na viacerých miestach – najvyšší skratový prúd dosiahol 6,2 kA. Dimenzovanie vodičov aj ochrán bolo dostatočné.

- **Koordinácia ochrany**: Odpájače boli správne nastavené tak, aby vypínali poruchové úseky bez ovplyvnenia zvyšku siete.
- **Analýza spoľahlivosti**: Model bol použitý na výpočet ukazovateľov SAIDI a SAIFI, ktoré zodpovedali štandardom distribučných sietí.

Získané výsledky potvrdili správnosť návrhu a zároveň poskytli **konkrétne odporúčania na zlepšenie návrhu siete** – napríklad zvýšenie prierezu vodičov v posledných úsekoch pre zaistenie výkonovej rezervy.

V rámci diplomovej práce plánujem túto časť značne rozšíriť – najmä o **dynamické simulácie (napr. opakované spínanie, výpadky, self-healing siete)** a o **prepojenie simulácií s BIM modelom**, čím vznikne základ pre digitálne dvojča reálnej siete.

2.5 Presnosť a výhody BIM modelu

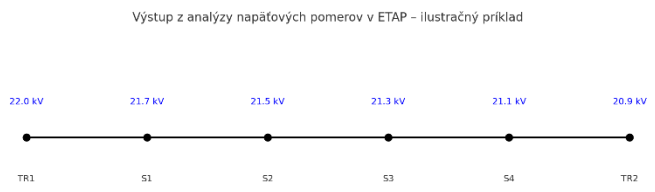
V prostredí Autodesk Revit bolo možné vytvoriť 3D model, ktorý odráža reálnu geometriu siete, umožňuje sledovanie trasy vedenia, výšok stožiarov, rozostupov aj detailov ako sú izolátory či odpájače. Prínosom bola možnosť priradiť každému prvku **parametre**, ktoré sa automaticky premietli do výkazov a výstupných súborov.

Už v tejto fáze bolo jasné, že model dokáže podporovať nielen návrh, ale aj budúcu **prevádzku a údržbu**. Takto navrhnutý systém je pripravený na rozšírenie o ďalšie vrstvy dát – napríklad informácie o inštalácii, výmene, životnosti komponentov a servisných zásahoch.

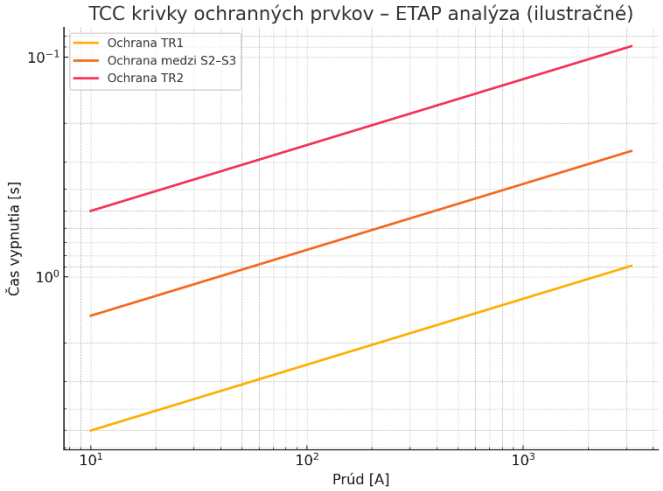
2.6 Analytické výstupy

Analýzy v ETAP-e potvrdili správnosť návrhu. Skratové prúdy, úbytky napätia aj zaťaženie transformátorov boli v súlade s platnými normami a technickými požiadavkami. Zvlášť užitočná bola analýza selektivity ochranných prvkov, ktorá pomohla optimalizovať rozmiestnenie odpájačov.

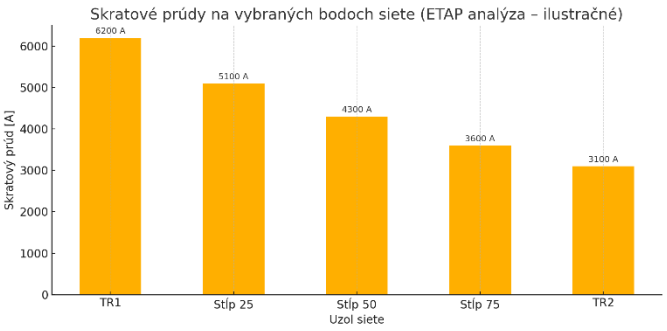
Výsledky ukázali, že správne nastavený model poskytuje dostatočnú výkonnostnú rezervu a dokáže reagovať aj na výpadky či poruchy. V ďalšej fáze projektu (v diplomovej práci) sa plánujem zamerať na **automatizáciu procesov nastavovania ochrán**, ako aj na **dynamické simulácie reálnych prevádzkových stavov**.



Obr.2 Ilustračný výstup z analýzy v ETAPe – napäťové pomery pozdĺž siete



Obr.3 Ilustračný výstup z analýzy v ETAPe –TCC krivky (selektivita ochrán)



Obr.4 Ilustračný výstup z analýzy v ETAPe – Skratové prúdy

2.7 Význam pre ďalší rozvoj

Už táto úvodná práca ukázala, že prepojenie návrhových a analytických nástrojov môže viesť k vytvoreniu jednotného digitálneho modelu siete. Ten môže byť v budúcnosti doplnený o:

- Live dáta zo senzorov
- Geografické vrstvy z GIS
- Prevádzkové informácie z SCADA

V diplomovej práci sa plánujem podrobnejšie venovať práve týmto aspektom – teda vytvoreniu **digitálneho dvojčaťa siete**, ktoré bude možné nielen analyzovať, ale aj priebežne aktualizovať na základe údajov z terénu.

3 Záver

O Cieľom tejto práce bolo vytvoriť koncept návrhu a simulácie 22 kV vzdušnej distribučnej siete s využitím nástrojov BIM a elektrotechnickej analýzy. Na konkrétnom príklade 10 km dlhého vedenia bolo ukázané, ako možno v prostredí Autodesk

Revit modelovať prvky siete, rozmiestniť ich v priestore, priradiť im technické parametre a pripraviť celý projekt na výpočtovú simuláciu.

Pomocou softvéru ETAP bola následne sieť analyzovaná z hľadiska elektrických veličín, prevádzkovej spoľahlivosti a ochrany. Výsledky potvrdili, že navrhnuté vedenie spĺňa požiadavky na dimenzovanie, úbytok napätia, skratové prúdy aj ochranu pred poruchami. Integrácia BIM a výpočtových nástrojov sa ukázala ako praktický spôsob, ako zjednotiť návrh, dokumentáciu aj analytickú časť projektu.

Táto práca bola zámerne spracovaná ako **úvodná fáza hlbšieho výskumu**, ktorý plánujem rozvíjať v **mojej diplomovej práci**. V nej sa budem podrobne venovať:

- Integrácii modelu s **geografickým informačným systémom (GIS)**,
- Automatizovanému návrhu siete pomocou **Dynamo a parametrického generovania** komponentov,
- Prepojeniu BIM modelu s reálnymi dátami zo siete cez **SCADA alebo IoT senzory**,
- Návrhu a testovaniu **digitálneho dvojčaťa elektrickej distribučnej siete**,
- A tiež detailnej **simulácii prevádzky siete počas porúch a rekonfigurácií.**

Záverom možno konštatovať, že moderné prístupy ako BIM a pokročilá elektrotechnická simulácia otvárajú nové možnosti v návrhu a prevádzke elektrických sietí. Ich kombinácia zvyšuje efektivitu, znižuje chybovosť a zároveň vytvára základ pre digitálnu transformáciu v oblasti elektroenergetiky.

Literatúra

[1] Autodesk. *Revit for Electrical Design*. Dostupné online: https://www.autodesk.com.
[2] ETAP. *Electrical Power System Analysis & Operation Software*. Dostupné online: https://etap.com.
[3] Górny, A., & Węgrzyński, W. “Application of BIM in Electrical Engineering.” IEEE Transactions on Industrial Applications, 2021.
[4] Ďurica, R. *Distribučné siete v praxi*. STU Bratislava, 2019.
[5] IEC 60909 – *Short-Circuit Currents in Three-Phase Systems*. International Electrotechnical Commission.
[6] IEC 60287 – *Electric Cables – Current-Carrying Capacity*. IEC, Geneva.

Analýza výsledkov cyklického výkonového testu MOS tranzistora podporená 3D elektrotepelnými simuláciami

Patrik Suhaj, Aleš Chvála.

[1] Ústav elektroniky a fotoniky, Fakulta elektrotechniky a informatiky, Slovenská technická univerzita v Bratislave

xsuhaj@stuba.sk

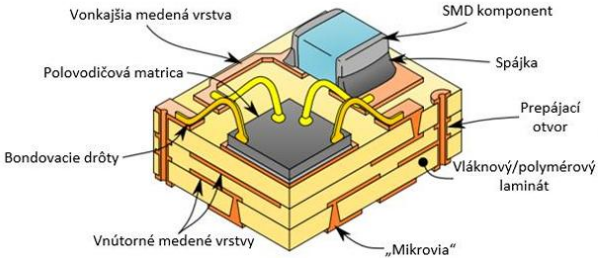
Abstrakt – Táto práca sa zaoberá analýzou výsledkov cyklického výkonového testu výkonového MOS tranzistora s využitím 3D elektrotepelných simulácií. Cieľom je hodnotenie spoľahlivosti a odolnosti súčiastky voči cyklickému namáhaniu. Testovanie bolo realizované na výkonovom tranzistore v puzdre PowerFLAT 5x6, ktorý bol osadený na dosku plošného spoja (DPS) pomocou spájky. Počas experimentu bol tranzistor podrobený cyklickému zahrievaniu a chladeniu, pričom sa monitorovali teplotné a elektrické charakteristiky. Optická analýza odhalila degradáciu spájky, ktorá bola potvrdená aj 3D tepelnými a elektrickými simuláciami v programe COMSOL Multiphysics. Výsledky poukazujú na vplyv tepelného cyklovania na mechanickú stabilitu kontaktov a poskytujú cenné informácie pre optimalizáciu spoľahlivosti výkonových MOS tranzistorov.

1 Úvod

Výkonový MOS tranzistor je trojelektródové zariadenie, obsahuje elektródy source (S), drain (D), gate (G). Je to unipolárny tranzistor čo znamená, že je závislý len od jedného typu nosiča náboja. V závislosti od typu nosiča náboja sa výkonový MOS tranzistor delí na NMOS, kde sú majoritné nosiče náboja elektróny a na PMOS kde sú majoritnými nosičmi náboja naopak diery.[1]

Pre výkonové MOS tranzistory je špecifický vertikálny dizajn, kde sú drain a source na opačných stranách polovodičovej dosky. Takáto konfigurácia zlepšuje schopnosť tranzistora zvládať vyššie napätia a prúdy bez rizika poškodenia. Medzi typy výkonových MOS tranzistorov s vertikálnou štruktúrou patrí V-MOSFET, VD-MOSFET/D-MOSFET a U-MOSFET. [2]

Výkonové MOS tranzistory ale aj iné súčiastky môžeme osadiť na DPS rôznymi spôsobmi. Príkladom je osadenie súčiastky na dosku plošného spoja (DPS), ktorá je na dosku pripevnená spájkou. Ide o technológiu povrchovej montáže (SMT), ktorá je zobrazená na obrázku nižšie (Obr. 1). Pri tejto technike sa komponenty umiestňujú priamo na povrch dosky pričom sa spájkujú len kontaktné plôšky na súčiastkach. Prepájacie otvory slúžia len ako vodivé spojenia medzi vrchnou spodnou stranou DPS. Malé rozmery povrchovo montovaných súčiastok (SMD) umožňujú vysokú hustotu osadenia komponentov. [3,4]



Obr. 1. Technológia povrchovej montáže (SMT) [4]

Na zistenie spoľahlivosti a odolnosti výkonových polovodičových zariadení vrátane výkonových MOS tranzistorov sa používa cyklický výkonový test. Počas testu je testované zariadenie (DUT) cyklicky namáhané až pokým nedosiahne jedno z dvoch kritérií zlyhania. Prvé kritérium je zvýšenie napätia v zapnutom stave DUT tesne pred vypnutím záťažového prúdu o 5 % a druhé kritérium je zvýšenie tepelného odporu R_{th} o 20 %. [5]

Keďže priame meranie teploty PN-prechodu je náročné, používa sa nepriama metóda. Na odhad teploty PN-prechodu sa využíva tepelno-citlivý elektrický parameter (TSEP), konkrétne metóda V_{SD}(T). Táto metóda meria úbytok napätia medzi terminálmi source a drain MOS tranzistora, keď cez diódu prechádza malý konštantný prúd. Odhadovaná teplota sa nazýva virtuálna teplota prechodu. T_{vj}.[5]

Na overenie správneho fungovania súčiastky sa vo väčšine používajú simulácie, ktoré nahrádzajú dlhotrvajúce ručné merania, ktoré môžu byť ovplyvnené aj inými rušivými faktormi. Simulácie elektrických alebo tepelných vlastností možno vykonávať v programe COMSOL Multiphysic. Pomocou simulácii je taktiež možné simulovať vzniknuté poruchy počas cyklického výkonového testu.

Elektrická simulácia slúži na simuláciu elektrických parametrov, ako sú prúd, napätie alebo odpor, pričom využíva model „Electric Currents“. Jej presnosť je však podmienená zanedbateľnými indukčnými účinkami. Základom tejto simulácie je elementárna rovnica Ohmového zákona

$$J = \sigma \cdot E \tag{1}$$

kde J je hustota elektrického prúdu, σ je merná elektrická vodivosť a E je intenzita elektrického poľa. Tento model obsahuje niekoľko predefinovaných parametrov ako napríklad „Current Conservation“, v ktorom sa definujú vodivé časti

modelu alebo „Electric Insulation“, v ktorom sa naopak definujú izolované oblasti. Tento simulačný model ponúka aj iné parametre, ktoré si v prípade potreby možno pridať. [8]

Tepelná simulácia je jedna z najdôležitejších simulácií, pretože teplota ovplyvňuje množstvo iných parametrov súčiastky. Využíva sa model „Heat Transfer in Solids“, ktorý je aktívny vo všetkých častiach súčiastky. Simulácia vychádza z tepelnej rovnice

$$\rho C_{\mathrm{p}} \frac{\partial T}{\partial t} + \rho C_{\mathrm{p}} u \cdot \nabla T \cdot \nabla q = Q + Q_{\mathrm{ted}} \tag{2}$$

kde ρ je hustota látky (kg/m^3), C_{P} je merná tepelná kapacita pri konštantnom tlaku (J/kg·K), $\frac{\partial T}{\partial t}$ je parciálna derivácia absolútnej teploty podľa času, člen $u \cdot \nabla T$ hovorí o prenose tepla prúdením pričom u je vektor rýchlosti translačného pohybu, člen ∇q predstavuje divergenciu tepelného toku, Q je parameter, ktorý zohľadňuje primárny zdroj tepla (W/m^3) a Q_{ted} je termoelatrické tlmenie, ktoré zohľadňuje termoelastické účinky v pevných látkach (W/m^3).

Člen q z rovnice (2), ktorý predstavuje tepelný tok vedením (W/m^2) možno rozpísať podľa nasledujúceho vzťahu

$$q = -k \cdot \nabla T \tag{3}$$

kde k je tepelná vodivosť materiálu (W/J·K) a ∇T je gradient teploty (K/m).

Tepelná simulácia takisto obsahuje niekoľko preddefinovaných parametrov ako napríklad „Thermal Insulation“, ktorý definuje okrajové podmienky. Aj tento model ponúka viaceré parametre ktoré možno pridať, napríklad „Heat Source“, teda tepelný zdroj, ktorý definuje zdroj tepla, v prípade MOS tranzistorov môže ísť o čip. [9]

2 Experimentálna časť

Táto časť je zamenená na analýzu výsledkov cyklického výkonového testu MOS tranzistora osadeného na DPS podporenú 3D elektrotepelnými simuláciami.

2.1 Analyzovaný výkonový MOS tranzistor

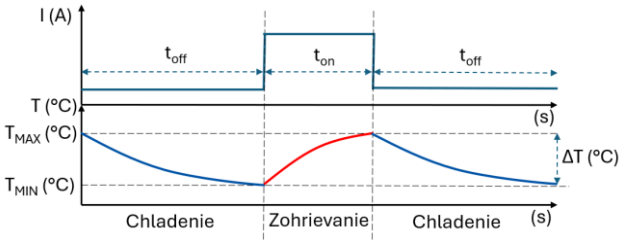
Analyzovaný výkonový MOS tranzistor je v puzdre PowerFLAT 5x6 a je osadený na doske plošného spoja (DPS). Tranzistor je k DPS pripevnený spájkou. Štandardná štruktúra kontaktov puzdra PowerFLAT 5x6 pozostáva zo štyroch terminálov kontaktu drain, troch terminálov kontaktu source a jedného terminálu kontaktu gate (Obr. 2).



Obr. 2. Výkonový MOS tranzistor v puzdre PowerFLAT 5x6 osadený na DPS

2.2 Cyklický výkonový test

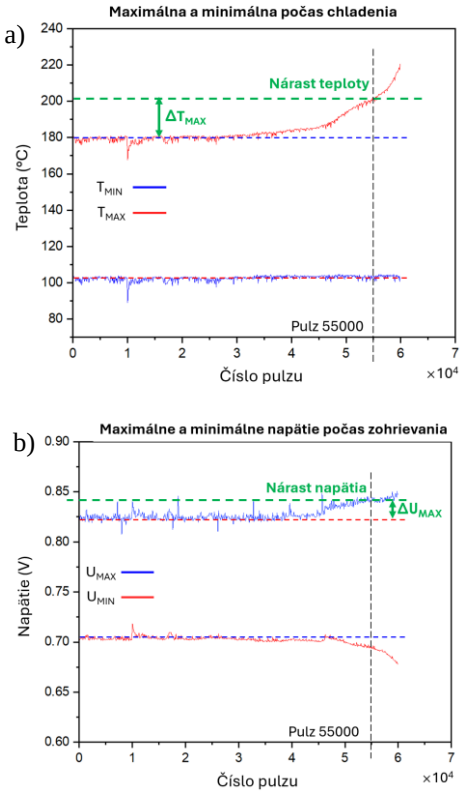
Cyklický výkonový test pozostával z periodicky opakujúcich sa impulzov. Časový priebeh cyklického výkonového testu je zobrazený na obrázku (Obr. 3). Jeden cyklus trval 11 sekúnd, pričom jednu sekundu sa tranzistor zohrieval (t_{on}) pri prúde 10 A a desať sekúnd sa tranzistor chladil (t_{off}). Počas zohrievania sa tranzistor zohrial na teplotu 180 °C a následne sa počas chladenia ochladil na teplotu 105 °C.



Obr. 3. Časový priebeh cyklického výkonového testu zobrazujúci čas zohrievania a čas chladenia

2.3 Analýza charakteristík výkonového cyklického testu

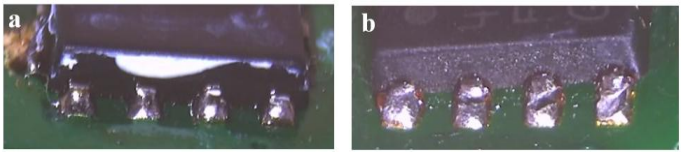
Pri analýze charakteristík výkonového cyklického testu tranzistora na obrázkoch nižšie (Obr. 4) sme odhalili poruchu, ktorá sa začala prejavovať okolo pulzu 55000, kedy sme zaznamenali nárast teploty približne o 20 °C (ΔT_{MAX}) a rovnako nárast napätia približne o 0,02 V (ΔU_{MAX}).



Obr. 4. Charakteristiky z cyklického výkonového testu: a) maximálna a minimálna teplota pri chladení, b) maximálne a minimálne napätie pri zohrievaní

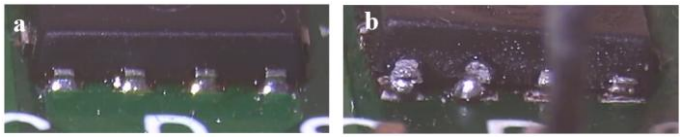
2.4 Optická analýza vzorky

Pri optickej analýze vzorky sme zistili, že spájka, ktorou je tranzistor pripevnený k DPS je značne poškodená a vykazuje známky degradácie. Na obrázkoch nižšie je porovnanie stavu spájky pred testom a po teste. Na obrázku (Obr. 5) je zobrazený stav spájky na kontaktoch source a gate pred testom. Na obrázku vedľa je zobrazený stav spájky na kontaktoch source a gate po teste kde je viditeľná značná degradácia spájky.



Obr. 5. Stav spájky na kontaktoch source a gate: a) pred testom, b) po teste

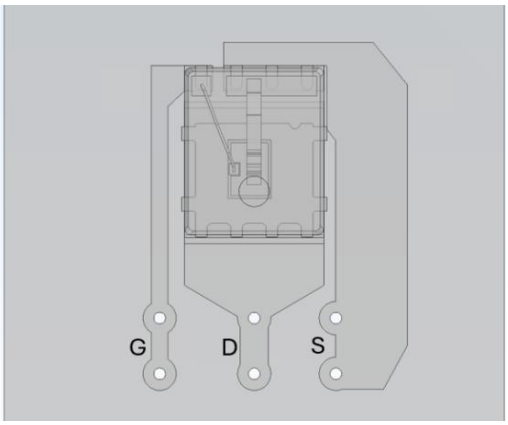
Obrázok nižšie (Obr. 6) zobrazuje stav spájky na kontaktoch drain pred testom a po teste a aj v tomto prípade je viditeľná značná degradácia spájky na všetkých kontaktoch. Degradácia spájky mohla viesť k uvoľneniu kontaktov.



Obr. 6. Stav spájky na kontaktoch drain: a) pred testom, b) po teste

2.5 Simulácie v programe COMSOL Multiphysic

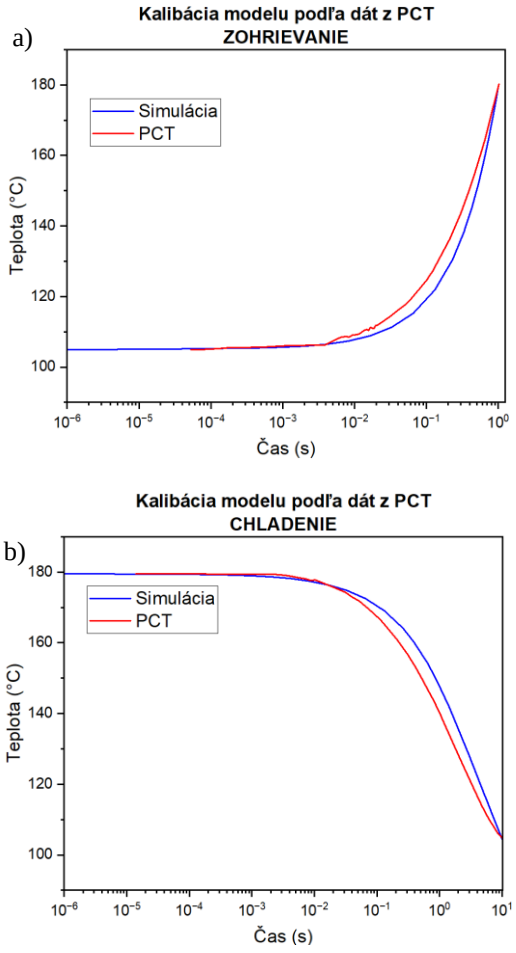
Zistenú chybu sme overovali prostredníctvom simulačného programu COMSOL Multiphysic. Ako prvé sme vytvorili 3D model, ktorý zodpovedá reálnej vzorke. Výsledný 3D model je zobrazený na obrázku nižšie (Obr. 7). Mriežku 3D modelu sme zvolili tak aby boli výsledky dostatočne presné a zároveň aby nebol čas simulácie príliš dlhý.



Obr. 7. 3D model výkonového tranzistora osadeného na DPS

2.5.1 Kalibrácia 3D modelu výkonového MOS tranzistora osadeného na DPS

Následne sme pomocou tepelnej a elektrickej simulácie kalibrovali 3D model podľa dát z cyklického výkonového testu (PCT) tak, aby sme dostali čo najpresnejšie výsledky. V prípade tepelnej simulácie sme na kalibráciu použili parameter „heat flux“, ktorý udáva množstvo tepelnej energie, ktoré prechádza cez jednotkovú plochu za jednotku času. Pre kalibráciu elektrickej simulácie sme využili materiálový parameter kremíka, z ktorého pozostáva čip výkonového tranzistora. Konkrétne išlo o materiálový parameter elektrická vodivosť σ (S/m). Výsledky kalibrácie sú zobrazená na grafe nižšie (Obr. 8).



Obr. 8. Výsledok kalibrácie podľa dát z PCT: a) korelácia simulačnej krivky z krivkou merania pri zohrievaní (t_{on}), b) korelácia simulačnej krivky z krivkou merania pri chladení (t_{off})

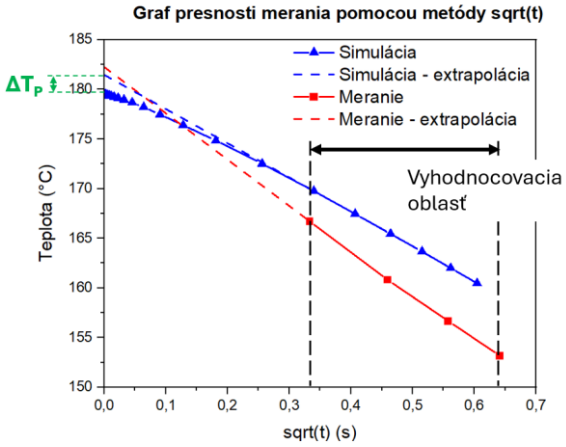
2.5.2 Vyhodnotenie presnosti merania

Na vyhodnotenie presnosti merania sme využili metódu odmocniny času

$$\Delta T(t) = \frac{P}{A} \cdot \frac{2}{\sqrt{\pi c \rho \lambda}} \cdot \sqrt{t} \qquad (4)$$

kde P (W) predstavuje výkonovú stratu, A (mm²) je aktívna plocha čipu (mm²), c (J/(kg·K)) je merná tepelná kapacita, ρ (kg/m³) je hustota, λ W/(m·K) je tepelná vodivosť a t (s) je čas [9]. Táto metóda je vhodná pretože zohľadňuje fyziku šírenia tepla po vypnutí tranzistora.

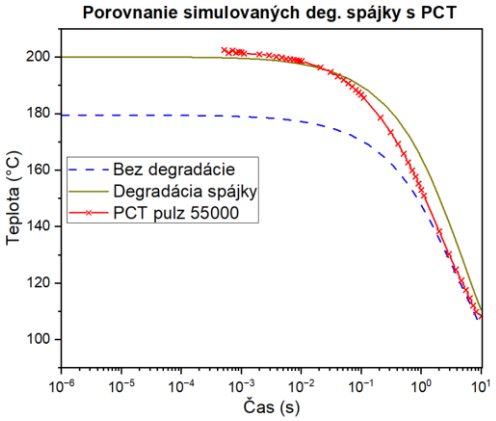
Po extrápolácii bodov krivky merania a simulácie vo vyhodnocovacej oblasti sme zistili, že odchýlka počiatočných teplôt ΔT_P je približne 1 %. Výsledok vyhodnotenia presnosti merania je zobrazený na grafe nižšie (Obr. 9).



Obr. 9. Výsledok vyhodnotenia presnosti merania pomocou metódy odmocniny z času

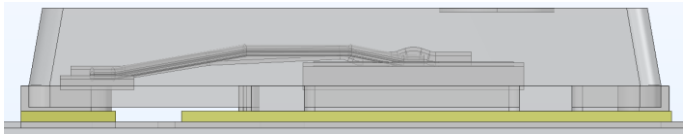
2.5.3 Simulácia degradácie spájky pomocou tepelnej simulácie

V simulácii degradácie spájky sme menili jej tepelnú impedanciu v materiálových parametroch. Pre simuláciu ktorá zodpovedá analyzovanej poruche kedy teplota stúpla o 20 °C, sme museli zmenšiť tepelnú impedanciu spájky približne 100 krát. To značí výraznú degradáciu spájky čo potvrdzuje aj stav spájky po teste. Degradáciu spájky tak isto potvrdzuje aj tvar krivky simulácie na grafe (Obr. 10), kde sa krivka priblížila ku krivke, ktorá zodpovedá poruche zistenej pri cyklickom výkonovom teste.



Obr. 10. Porovnanie simulovanej degradácie spájky s meraním

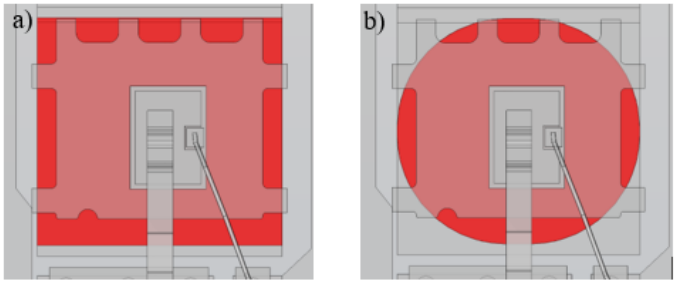
Nasledujúci obrázok (Obr. 11) zobrazuje polohu spájky na 3D modely, ktorou je výkonový MOS tranzistor pripevnený na DPS.



Obr. 11. Poloha spájky v 3D modely

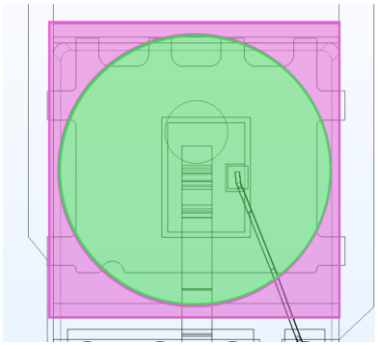
2.5.4 Simulácia pomocou elektrickej simulácie

Pri tejto simulácii sme sa zamerali na simuláciu rozdielu napätia ΔU_{MAX} = 0,02 V, ktorý sme zistili pri analýze. Ako prvé sme menili parameter elektrickej vodivosti spájky v materiálových parametroch, daný rozdiel sme dostali ak sme elektrickú vodivosť znížili približne 250 krát, čo tak isto značí že na dosiahnutie takého rozdielu muselo dôjsť k výraznému zhoršeniu elektrických vlastností spájky, alebo inú formu poškodenia ako utrhnutie nožičiek či iné mechanické poškodenie. Preto sme sa ďalej zamerali na simulácie, kde sme simulovali rôzne formy poškodenia spájky. V prvom prípade sme zaoblením upravili geometriu spájky tak aby sme simulovali jej delamináciu na rohoch (Obr. 12). Pri tejto simulácii sme však dosiahli rozdiel napätia ΔU_{MAX} približne 1×10^{-5} V.



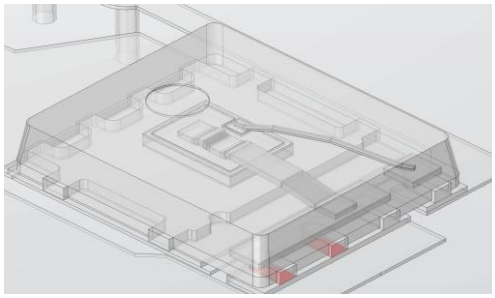
Obr. 12. Delaminácia spájky simulovaná zaoblením rohov – a) bez zaoblenia, b) so zaoblením

Rovnaký rozdiel napätia ako v prípade simulácie delaminácie spájky sme dosiali aj pri simulácii vzduchovej dutinky v spájke. (Obr. 13).



Obr. 13. Simulácia vzduchovej dutinky (zelená) v spájke (ružová)

Ako poslednú elektrickú simuláciu sme robili simuláciu odpojenia dvoch kontaktov source (Obr. 14), ktoré mohlo byť tak isto príčinou stúpnutia napätia ΔU_{MAX} pri cyklickom výkonovom teste. Pri tejto simulácii bol však rozdiel napätia ΔU_{MAX} približne 3×10^{-4} V.



Obr. 14. Simulácia dvoch odpojených kontaktov source (červená)

3 Záver

Táto práca analyzovala vplyv cyklického výkonového testu na výkonový MOS tranzistor v puzdre PowerFLAT 5x6 osadený na DPS pomocou spájky. Experimenty preukázali, že dlhodobé cyklické namáhanie vedie k postupnej degradácii spájky, čo spôsobuje nárast teploty a zvýšenie elektrického odporu kontaktov. Optická analýza vzorky po teste potvrdila degradáciu spájky. Naše výsledky boli overené 3D elektrotepelnými simuláciami v programe COMSOL Multiphysics, ktoré preukázali, že porucha vznikla z dôvodu výrazného zhoršenia tepelných a elektrických vlastností spájky. Simulované výsledky vykazovali vysokú koreláciu s experimentálnymi meraniami, pričom presnosť vyhodnotenia bola potvrdená metódou odmocniny času. Tepelné simulácie potvrdzujú poruchu spájky, ktorá bola viditeľná aj pri optickej analýze. Elektrické simulácie zúžili kruh možných porúch, ktoré mohli byť príčinou stúpnutia napätia pri PCT a budú predmetom ďalšieho výskumu. Zistenia tejto práce sú dôležité pre vývoj spoľahlivých výkonových elektronických súčiastok, pričom poukazujú na potrebu optimalizácie spájkovacích materiálov a montážnych technológií na zvýšenie odolnosti MOS tranzistorov voči tepelnému namáhaniu.

Poďakovanie

Táto práca bola podporená projektom FastLane. Projekt je podporovaný iniciatívou Chips Joint Undertaking (JU) a jej členmi, vrátane dodatočného financovania zo strany Rakúska, Francúzska, Nemecka, Rumunska a Slovenska, na základe grantovej dohody č. 101139788. Táto práca bola taktiež podporená grantom VEGA 1/0543/23, financovaným Ministerstvom školstva, výskumu, vývoja a mládeže Slovenskej republiky. Táto publikácia bola čiastočne financovaná z prostriedkov EÚ NextGenerationEU prostredníctvom Plánu obnovy a odolnosti pre Slovensko v rámci projektu č. 09I01 03V04-00032.

Literatúra

[1] *MOSFET – Structure, working principle, symbol and applications* [online]. [cit. 2025-03-24]. Dostupné z: https://www.electricalclassroom.com/mosfet-symbol-working-structure/

[2] BALIGA, B. Jayant. Fundamentals of power semiconductor devices. New York: Springer, 2008. ISBN 978-0-387-47313-0.

[3] MOHD NAN, Noor Mariamadzliza a spol. *The Application of PCB, Mounted-Components and Solder Paste in Surface Mount Technology Assembly*, National Conference on Electronic Design (2005) pp. 145-151. [cit. 2025-03-24].

[4] BUTTAY, Cyril, Christian MARTIN a Florent MORNEL. *Application of the PCB Embedding Technology in Power Electronics – State of the Art and Proposed Development*, 3D-PEIM, 2018 pp.1-10 doi: 10.1109/3DPEIM.2018.8525236 [cit. 2025-03-24].

[5] GÖTHNER, Fredrik a spol. *Challenges of SiC MOSFET Power Cycling Methodology* EPE'18 ECCE Europe 2018, pp. P.1-P.8. [cit. 2025-03-24].

[7] *AC/DC Module: User's Guide* [online]. [cit. 2025-03-24]. Dostupné z: https://doc.comsol.com/5.4/doc/com.comsol.help.acdc/ACDCModuleUsersGuide.pdf

[8] *Heat Transfer Module: User's Guide* [online] [cit. 2025-03-24]. Dostupné z: https://doc.comsol.com/5.4/doc/com.comsol.help.heat/HeatTransferModuleUsersGuide.pdf

[9] DENG, Erping. *Measurement Error Caused by the Square Root t Method Applied to IGBT Devices during Power Cycling Test*, 2020 32nd ISPSD, Vienna, Austria, 2020, pp. 545-548, doi: 10.1109/ISPSD46842.2020.9170083.

Návrh nízkopríkonového rekonfigurovateľného FIR filtra v HDL jazyku

Bc. Richard Bagín[1], Ing. Lukáš Nagy, PhD.[1]

[1]Oddelenie návrhu a testovania integrovaných obvodov, Ústav elektroniky a fotoniky, FEI STU v Bratislave

xbagin@stuba.sk

Abstrakt – Článok približuje proces návrhu a testovania rekonfigurovateľného FIR filtra optimalizovaného na nízku spotrebu v nanometrových CMOS technológiách. Pri jeho syntéze možno nastaviť maximálny počet koeficientov filtra a číselnú reprezentáciu dát. Z pohľadu užívateľa sú rekonfigurovateľné hodnoty koeficientov a ich aktívny počet, teda rád filtra. Všetky pamäťové elementy filtra môžu byť preusporiadané do 1-bitového skenovacieho reťazca (scan chain) pre výrobný test a možno ho využiť aj na ladenie. Pre jednoduchú interakciu s filtrom bola vytvorená MATLAB GUI aplikácia, ktorá umožňuje nastaviť koeficienty, spracovávať vzorky, sken a zobrazenie pamäťových elementov.

1 Úvod

Digitálne FIR (finite impulse response) filtre majú široké uplatnenie a sú základným stavebným blokom spracovania signálov v mnohých oblastiach každodenného života. Tiež majú zaručenú stabilitu ako aj možnosť lineárnej fázovej charakteristiky. Rekonfigurovateľný filter má koeficienty nastaviteľné, a tak rovnakým hardvérom možno realizovať filtre s rôznymi charakteristikami. Pokiaľ je rekonfigurácia riadená ďalším algoritmom, jedná sa o adaptívne spracovanie signálov.

Neustály pokrok v multimediálnych mobilných aplikáciách vyžaduje nízky príkon a vysoký výkon VLSI (very large scale integration) systémov DSP (digital signal processing), kde hlavnou časťou spotreby energie, na ktorú je potrebné myslieť pri návrhu, je taktovanie [1]. Na tieto obvody napájané z batérií či zberačov energie sa tak kladú čoraz vyššie nároky na znižovanie ich spotreby a maximalizáciu ich výpočtového výkonu, ktoré sú však protichodné. Preto je žiadané prispôsobiť obvod už nie len funkčným požiadavkám, ale aj špecifickej výrobnej technológii.

2 Číselná reprezentácia dát

Dôležitým výstupom tohto výskumu je aj stanovenie prípadov, kedy a s akou presnosťou je vhodnejšie použiť číselnú reprezentáciu s pohyblivou rádovou čiarkou (floating-point založenú na princípoch IEEE 754) voči číselnej reprezentácii s pevnou rádovou čiarkou (Q-formát) na základe porovnania chyby výstupu spôsobenej kvantovaním, maximálnej hodinovej frekvencie, odhadovanej spotreby energie a plochy na čipe v jednotlivých konfiguráciách pre konkrétnu technológiu.

2.1 Pevná rádová čiarka (Q-formát / fixed-point)

Q-formát využíva niekoľko najmenej významových bitov na zaznamenanie zlomkových hodnôt. Spôsob označovania je „Q$i.f$“. Takto zaznamenané číslo možno vyjadriť ako

$$n = -b_{i+f}2^i + \sum_{k=0}^{i+f-1} b_k 2^{k-f} \tag{1}$$

kde $n \in \mathbb{Q}$ je zaznamenané číslo, i je počet celočíselných bitov, f je počet zlomkových bitov, b_{i+f} je znamienkový bit a b_k sú potom jednotlivé celočíselné a zlomkové bity. Znamienkový bit nie je započítaný do počtu celočíselných bitov. Celkový počet bitov tak je $N = 1 + i + f$. Netreba zabúdať, že $i, f \in \mathbb{Z}$, teda počet celočíselných či zlomkových bitov môže byť aj záporný. Rozsah zaznamenateľných čísel je od -2^i do $2^i - 2^{-f}$. Pre lepšie pochopenie je v tabuľke 1 uvedených niekoľko príkladov.

Tab. 1: Príklady číselnej reprezentácie v rôznych Q-formátoch

	Bity	Min.	Max.	Rozlíšenie
Q2.5	8	$-2^2 = -4$	$2^2 - 2^{-5} = 3{,}96875$	0,03125
Q5.-2	4	$-2^5 = -32$	$2^5 - 2^{-(-2)} = 28$	4
Q-2.5	4	$-2^{-2} = -0{,}25$	$2^{-2} - 2^{-5} = 0{,}21875$	0,03125

2.2 Pohyblivá rádová čiarka (floating-point)

Formát s pohyblivou rádovou čiarkou možno použiť na zaznamenanie hodnôt v rozsahu väčšom než 2^{N+b_0}, kde N je počet bitov a b_0 je hodnota nultého bitu. Existujú viaceré štandardy, no najpoužívanejším je štandard IEEE 754, ktorý definuje aj najbežnejšie používané dátové typy, 32-bitový *float* (single-precision) a 64-bitový *double* (double-precision). Racionálne čísla sú zaznamenané v tomto formáte ako

$$n = \begin{cases} (-1)^{S_b} \cdot \left(1 + M_b \cdot 2^{-W_M}\right) \cdot 2^{E_b - B} & , \ E_b \neq 0 \\ (-1)^{S_b} \cdot \left(0 + M_b \cdot 2^{-W_M}\right) \cdot 2^{(E_b+1) - B} & , \ E_b = 0 \end{cases} \tag{2}$$

kde $n \in \mathbb{Q}$ je zaznamenané číslo, S_b je znamienkový bit, M_b je celočíselná hodnota uložená v bitoch mantisy, W_M je počet bitov mantisy, E_b je celočíselná hodnota uložená v bitoch exponentu a $B; B \in \mathbb{Z}; B = 2^{W_E - 1} - 1$ je hodnota posunu exponentu (bias), kde W_E je počet bitov exponentu.

Štandard definuje aj zaznamenanie špeciálnych výrazov ako nekonečno, limita blížiaca sa k nule zľava, limita blížiaca sa k nule sprava (zároveň 0), či prázdna množina (*NaN*). Prehľad prípadov reprezentácie zaznamenaných bitov je v tabuľke 2.

Tab. 2: Interpretácia bitov podľa normy IEEE 754

Podmienka	Hodnota
$E_b \notin \{0;\ 2^{W_E}-1\}$	$(-1)^{S_b}\left(1+2^{-W_M}M_b\right)2^{E_b-B}$
$E_b=0 \quad \wedge \quad M_b\neq 0$	$(-1)^{S_b}\left(0+2^{-W_M}M_b\right)2^{1-B}$
$S_b=0 \wedge E_b=0 \quad \wedge \quad M_b=0$	$0;\ 0^+$
$S_b=1 \wedge E_b=0 \quad \wedge \quad M_b=0$	0^-
$S_b=0 \wedge E_b=2^{W_E}-1 \wedge M_b=0$	$+\infty$
$S_b=1 \wedge E_b=2^{W_E}-1 \wedge M_b=0$	$-\infty$
$E_b=2^{W_E}-1 \wedge M_b\neq 0$	NaN

Pri konverzii čísel n kde $n \in \mathbb{R}$; $n = I + F$; $I \in \mathbb{Z}$; $F \in \mathbb{R}$; $|F| < 1$ s väčšou presnosťou na čísla s menšou presnosťou (napríklad na celé čísla), je dôležité dbať na správne zaokrúhľovanie, pričom existuje niekoľko typov zaokrúhľovania:

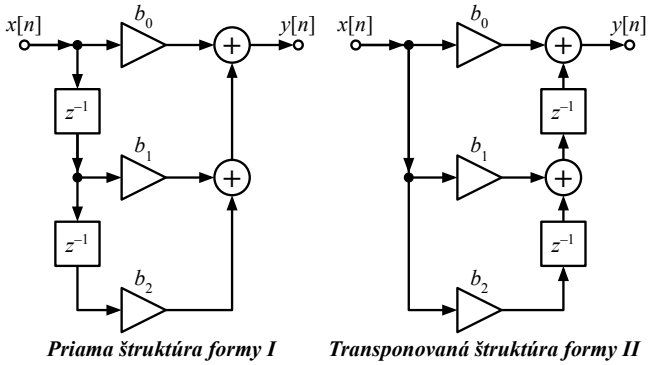
- *ceil*$_{\text{MATLAB}}$ / *towards positive infinity*$_{\text{IEEE 754}}$
  - k väčšiemu číslu
- *floor*$_{\text{MATLAB}}$ / *towards negative infinity*$_{\text{IEEE 754}}$
  - k menšiemu číslu
- *fix*$_{\text{MATLAB}}$ / *towards zero*$_{\text{IEEE 754}}$
  - k nule, resp. k číslu s menšou absolútnou hodnotou
- *convergent*$_{\text{MATLAB}}$ / *towards nearest, tie to even*$_{\text{IEEE 754}}$
  - k najbližšiemu číslu a ak $F = 0,5$, tak k najbližšiemu párnemu číslu
- *nearest*$_{\text{MATLAB}}$
  - k najbližšiemu číslu a ak $F = 0,5$, tak k najbližšiemu väčšiemu číslu
- *round*$_{\text{MATLAB}}$ / *towards nearest, tie away from zero*$_{\text{IEEE 754}}$
  - k najbližšiemu číslu a ak $F = 0,5$, tak k najbližšiemu číslu s väčšou absolútnou hodnotou

Počítače a programovacie jazyky vo všeobecnosti využívajú zaokrúhľovanie typu *towards nearest, tie to even*. To je východzie aj pre grafické karty, no tie zväčša možno nakonfigurovať na ktorýkoľvek typ zaokrúhľovania definovaný v IEEE 754 [3].

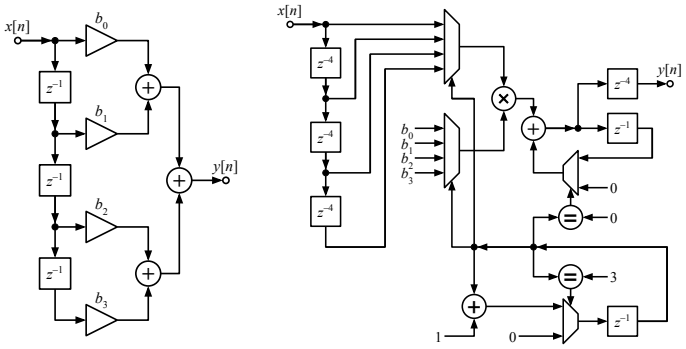
3 Možnosti štruktúry filtra

Podľa spôsobu výpočtu a obsahu pamäťových elementov (PE) možno FIR filter realizovať buď priamou štruktúrou vo forme I, kedy si filter pamätá hodnoty vzoriek, alebo transponovanou štruktúrou vo forme II, kedy si filter pamätá hodnoty medzivýsledkov, resp. súčinov. Tieto štruktúry, pre jednoduchosť iba FIR filtra 2. rádu, sú na obrázku 1.

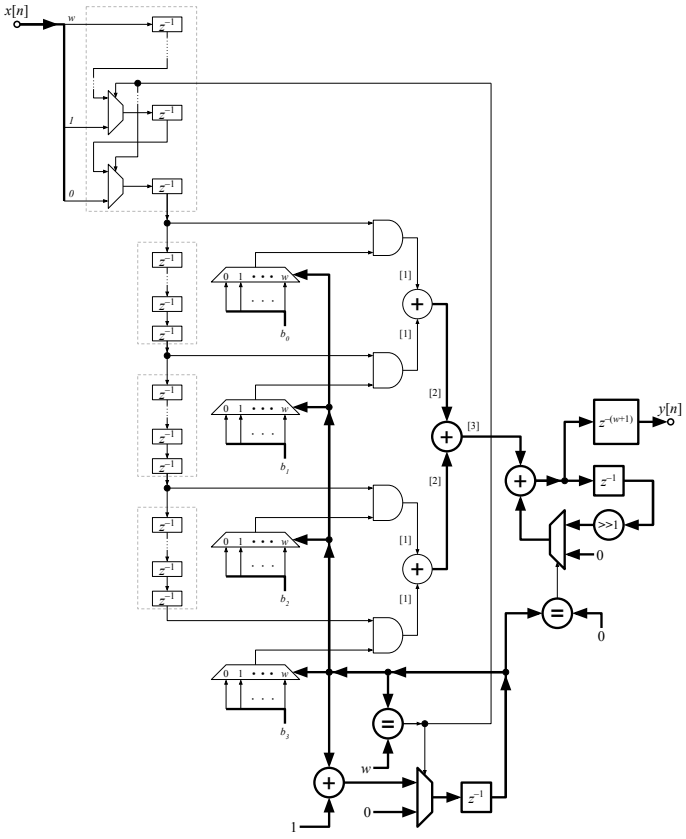
Keďže sčítať alebo vynásobiť naraz i len 2 čísla je náročné na plochu na čipe, podľa zapojenia sumátorov a násobičov môžu byť digitálne filtre realizované napríklad 1) s paralelným násobením a kaskádovým sčítaním, 2) so sériovým násobením a sčítaním alebo 3) s násobením bit po bite. Najvyššiu rýchlosť dosahuje spôsob 1) a najmenšiu plochu spôsob 3), no je vhodný len pre pevnú rádovú čiarku. Tieto zapojenia sú na obrázkoch 2 a 3 pre FIR filter 3. rádu so štruktúrou priamej formy I.



Obr. 1: Prikladná štruktúra FIR filtra 2. rádu



Obr. 2: Násobenie vo FIR filtri 3. rádu pre prípad 1) paralelne vľavo a 2) sériovo vpravo



Obr. 3: Násobenie vo FIR filtri 3. rádu pre prípad 3) bit po bite

Navrhovaný FIR filter je realizovaný v priamej štruktúre formy I so sériovým násobením a sčítaním 1 násobičom-akumulátorom. Využíva hradlovanie hodinového signálu po blokoch riadené hlavným stavovým automatom a izoláciu operandov pre násobič-akumulátor. Násobenie bit-po-bite (na obrázku 3) je pre dáta v Q-formáte najvhodnejšie. Samozrejme, podľa pomeru bitovej šírky čísel voči počtu násobení. No pre dáta s pohyblivou rádovou čiarkou je tento spôsob značne nevýhodný, preto bolo zvolené sériové násobenie a sčítanie ako vhodná alternatíva pre oba prípady reprezentácie uchovaných dát. Transponovaná štruktúra formy II by znížila „clock to output“ oneskorenie, no pri Q-formáte by mali medzivýsledky premenlivú, postupne zvyšujúcu sa bitovú šírku, čo by viedlo buď k zložitejšej implementácii, alebo naopak k plytvaniu PE pri zanedbaní tohto faktu, a tým aj k zbytočnému zvýšeniu spotreby. Priama štruktúra formy I tak predstavovala vhodnú alternatívu pre oba prípady reprezentácie uchovaných číselných dát.

4 Optimalizácia pre nízku spotrebu

V digitálnych CMOS obvodoch sú 3 hlavné druhy výkonových strát, a to dynamické straty, straty spôsobené skratovým prúdom a statické straty spôsobené zvodovým prúdom.

$$P = P_{dyn} + P_{short} + P_{stat} \tag{3}$$

Na každý typ strát majú vplyv rôzne parametre. Tie najvýraznejšie vplývajúce na spotrebu energie sú uvedené v rovnici 4

$$P = \left(\alpha C_{ld} f_{sw} {V_{dd}}^2\right) + (V_{dd} I_{sc}) + (V_{dd} I_{lk}) \tag{4}$$

kde α je faktor pravdepodobnosti nastania prechodu spôsobujúceho dynamickú výkonovú spotrebu, C_{ld} je elektrická kapacita záťaže, f_{sw} je frekvencia hodín, V_{dd} je napájacie napätie, I_{sc} je skratový prúd tečúci cez CMOS štruktúru, keď sú oba, NMOS a PMOS tranzistory aktívne a I_{lk} je zvodový prúd.

Hlavnou výhodou digitálnych CMOS obvodov je zanedbateľná statická spotreba. Dominantné sú tak dynamické straty, čomu napovedá aj trend neustáleho zvyšovania frekvencie hodín. Vo väčšine dobre navrhnutých CMOS obvodov tvoria dynamické straty 90 % celkových strát, čo napovedá existencii 3 rozmerov voľnosti pre nízkopríkonový návrh: 1) napätie, 2) elektrická kapacita a 3) aktivita. Ak zníženie výpočtového výkonu nie je akceptovateľné, optimalizácia spotreby zahŕňa minimalizáciu predošlých 3 faktorov. Pri HDL (hardware description language) návrhu je možné zamerať sa práve len na aktivitu dát. Tú je možné obmedziť metódami ako izolácia operandov, hradlovanie hodín [4], prúdové spracovanie operácií (pipelining) [5] či použitím viacbitových preklápacích obvodov [6].

5 Požadované parametre

Predložený FIR filter je navrhovaný ako univerzálny blok s plánovaným použitím v IO (integrovaných obvodoch) vyvíjaných v rámci projektov riešených na Ústave elektroniky a fotoniky FEI STU. Tieto projekty sú zamerané na IO pre senzorické aplikácie a IO napájané zo zberačov energie. Prvou aplikáciou bude predspracovanie dát pre blok hľadania bodu maximálneho výkonu (MPPT) zberača energie [2]. Keďže okrem porovnania číselných reprezentácií bude filter následne zakomponovaný do rôznych ASIC návrhov, rozhranie musí odpovedať rozhraniu už na čipe prítomných digitálnych alebo zmiešaných blokov. Z toho vyplývajú požadované parametre:

- Vstupné a výstupné dáta: 8-bitové celočíselné, iba kladné
- Rád filtra ≥ 50
- Vnútorná reprezentácia čísel:
  - Q-formát: Q1.8 (10 bitov) a Q1.14 (16 bitov)
  - Floating-point: binary16 / half-precision (16 bitov)
- Min. frekvencia spracovania vzoriek: 100 kHz
- Max. hodinová frekvencia: 100 MHz
- Asynchrónny reset pamäťových elementov na log. 0
- Filter nevyužíva prúdové spracovanie operácií
- Testovací režim využívajúci 1-bitový skenovací reťazec
- Maximálna plocha na čipe[(1)]: 200 µm × 200 µm
- Max. odhadovaná spotreba energie[(1)]:
  - statická: 0,1 mW[(2)] / 0,8 mW[(3)]
  - dynamická: 0,4 mW[(2)] / 1,0 mW[(3)]

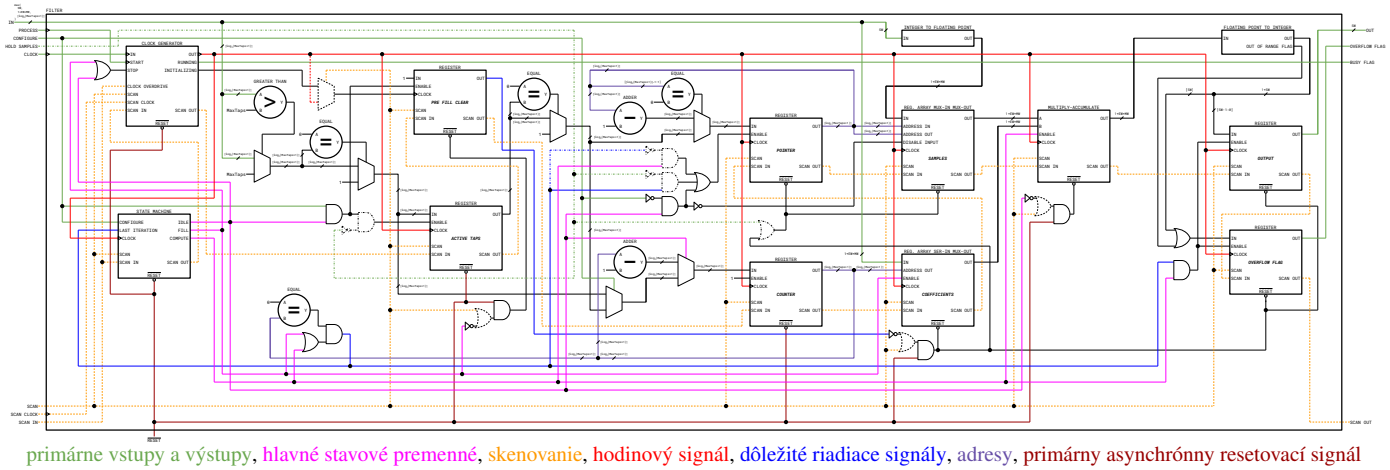
[(1)] v technológii UMC 65 nm
[(2)] typicky: TT tranzistory, V_{dd} = 1,2 V a T = 27 °C
[(3)] najhorší prípad: FF tranzistory, V_{dd} = 1,32 V a T = 125 °C

6 Opis navrhnutého FIR filtra

Navrhnutý digitálny filter je synchrónny sekvenčný logický obvod schopný implementovať akýkoľvek typ filtra s konečnou impulzovou odozvou, vďaka rekonfigurovateľným hodnotám koeficientov. Vlastnosti filtra tak možno meniť, napríklad ich prispôsobovať podľa výstupných hodnôt. Modul FIR filtra vo floating-point aritmetike zložený z podblokov je na obrázku 4. Popis jeho vstupných a výstupných portov je v tabuľke 3.

Tab. 3: Popis vstupov a výstupov SystemVerilog modulu rekonfigurovateľného FIR filtra

Port	Popis
not_reset	Asynchrónny reset
clock	Hlavný vždy bežiaci vstupný hodinový signál
process_in	Nábežnou hranou spúšťa spracovanie vzorky alebo ukladania hodnoty koeficientu či počtu aktívnych koeficientov
configure	Pri vysokej úrovni tohto vstupu po nábežnej hrane hodinového vstupu *process_in* filter prechádza do procedúry konfigurácie koeficientov, pri nízkej úrovni tohto vstupu po nábežnej hrane *process_in* filter spracováva hodnotu na vstupe *in* ako vzorku
hold_samples	Pri konfigurácii koeficientov zachováva predošlé vzorky a udržuje predošlý počet aktívnych koeficientov
in	Vstupná dátová zbernica s hodnotou vzorky (celé č.), koeficientu (podľa číselnej reprezentácie) alebo počtu koeficientov (celé č.)
scan_chain	Rekonfiguruje vnútorne prepojenia bitov do skenovacieho reťazca
scan_chain_clock	Hodinový signál pre skenovanie
scan_chain_in	Dátový skenovací vstup
scan_chain_out	Dátový skenovací výstup
out	Výstupná dátová zbernica s hodnotou výstupnej vzorky
overflow_flag	Indikuje pretečenie alebo podtečenie
busy_flag	Indikuje priebeh spracovávania vzorky alebo ukladania počtu a hodnôt koeficientov



primárne vstupy a výstupy, hlavné stavové premenné, skenovanie, hodinový signál, dôležité riadiace signály, adresy, primárny asynchrónny resetovací signál

Obr. 4: Vnútorné zapojenie rekonfigurovateľného FIR filtra využívajúceho reprezentáciu čísel s pohyblivou rádovou čiarkou

Parametrami hierarchicky najvyššieho modulu v HDL kóde pri syntéze možno zvoliť bitovú šírku vzoriek a maximálny počet koeficientov. Koeficienty sú štandardne v Q-formáte, kedy možno parametrami zvoliť bitovú šírku celočíselnej a zlomkovej časti koeficientov. Avšak zadefinovaním makra *IMPLEMENT_IN_FLOATING_POINT_ARITHMETICS* je použitá číselná reprezentácia s pohyblivou rádovou čiarkou (floating-point), kedy parametrami možno zvoliť bitovú šírku exponentu a mantisy. Ďalej makrom *IMPLEMENT_SCAN_CHAIN* možno vytvoriť vo filtri 1-bitový skenovací reťazec PE. Jeho zadefinovaním filtru automaticky pribúdajú vstupy a výstup pre sken. Zákazník alebo aplikácia následne dokáže nakonfigurovať rád filtra, resp. počet aktívnych koeficientov a ich hodnoty.
Poznámky:

- Filter potrebuje dobežnú hranu hodín a (*aktívne koeficienty* + 1) impulzov na vstupe *CLOCK* na spracovanie vzorky.
- Výstupné vzorky sú pri číselnej reprezentácii v Q-formáte zaokrúhlené nadol ($floor_{\text{MATLAB}}$, $towards\ negative\ infinity_{\text{IEEE 754}}$). Pri reprezentácii s pohyblivou rádovou čiarkou sú zaokrúhlené k nule (fix_{MATLAB}, $towards\ zero_{\text{IEEE 754}}$).

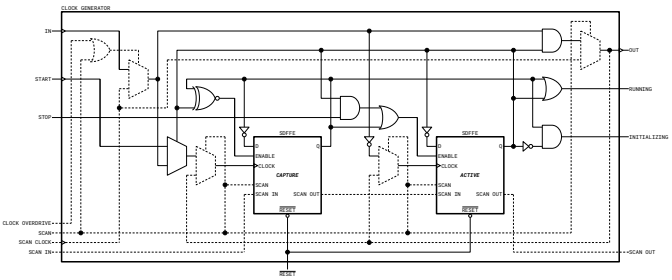
6.1 Vybrané navrhnuté podbloky

Značná časť vývoja bola zameraná na analýzu floating-point aritmetiky pre jej komplexnosť a potrebu reprodukcie pre ASIC návrh z dôvodu neprítomnosti podpory v HDL syntéze.

Bloky **AdderFloatingPoint** a **MultiplierFloatingPoint** sčítavajú / násobia čísla s pohyblivou rádovou čiarkou podľa normy IEEE 754 (binary-*n*). Avšak podporujú voliteľnú bitovú šírku operandov, danú parametrami počtu bitov mantisy a exponentu. Výsledok je zaokrúhlený zaokrúhľovaním typu *towards nearest, tie to even*.

IntegerToFloatingPoint a **FloatingPointToInteger** prevádzajú čísla medzi celočíselnou znamienkovou reprezentáciou s parametrizovanou bitovou šírkou a reprezentáciou s pohyblivou rádovou čiarkou podľa IEEE 754, no pre ľubovoľnú bitovú šírku exponentu a mantisy danú parametrami. Celé čísla väčšie ako tie zaznamenateľné vo floating-point formáte, sú prevedené na $\pm\infty$, naopak hodnota *NaN* na 0. Čísla sú zaokrúhľované k nule.

Blok **ClockGenerator** bezpečne hradluje hodinový signál zo vstupu *IN* na výstup *OUT* pomocou riadiacich vstupov *START* a *STOP*. Jeho zapojenie je na obrázku 5.



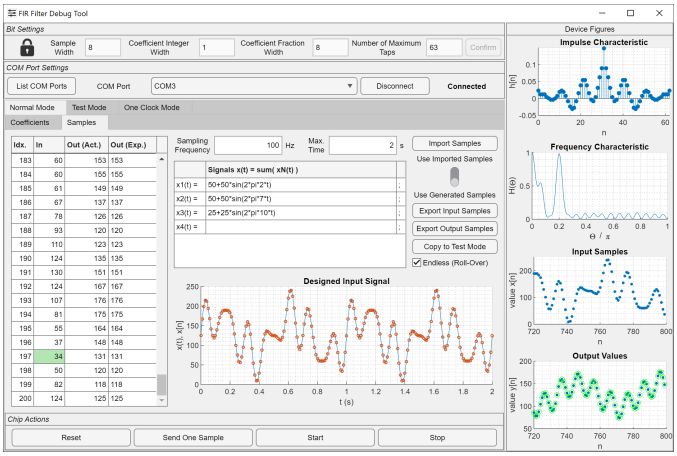
Obr. 5: Vnútorné zapojenie generátora hodinového signálu

7 Testovanie

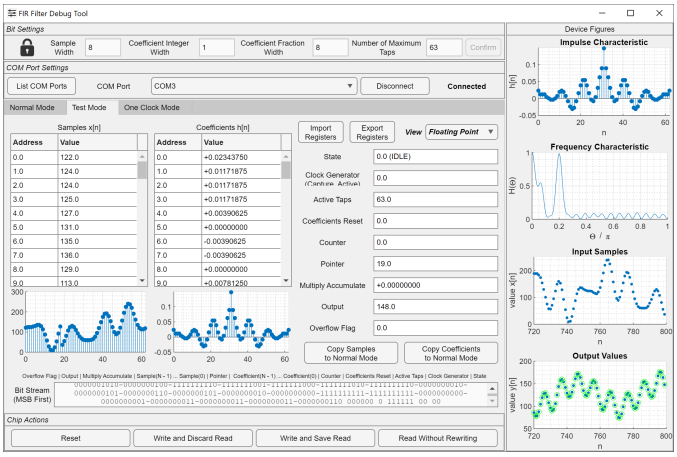
Návrh bol verifikovaný po blokoch a následne ako celok pri rôznych konfiguráciách. Pre vytvorenie testovacieho systému boli využité rôzne softvérové prostriedky. Prezentované výsledky boli dosiahnuté nasledovne. FIR filter bol navrhnutý a verifikovaný v HDL jazyku SystemVerilog. Následne bol zosyntetizovaný a implementovaný do FPGA Altera Cyclone II EP2C5T144C8. Na generovanie vstupných dát, charakteristík filtrov a spracovanie výsledkov meraní bol použitý softvér MATLAB. Koeficienty filtrov boli vypočítané MATLAB funkciou *firpm(...)*. Tá využíva metódu rovnomerného zvlnenia (equiripple). Takto navrhnuté FIR filtre majú vždy najnižší rád. Metóda využíva algoritmus Parks-McClellan, ktorý pre nájdenie vhodnej aproximácie využíva algoritmus výmeny Remez. Koeficienty, vzorky a skenovacie dáta sú medzi vyvinutou počítačovou MATLAB aplikáciou a FIR filtrom prenášané cez USB PHY mikrokontroléra STM32F103C8T6, ktorého firmvér je napísaný v C/C++. Vývody mikrokontroléra nastavujú vstupy a čítajú výstupy FPGA. MATLAB aplikácia zároveň predikuje aj správnu odozvu filtra a slúži aj na ladenie a ukladanie dát. Odhad spotreby energie konfigurácií FIR filtra vychádza zo simulácií.

7.1 MATLAB aplikácia

Aplikácia dokáže importovať a exportovať vzorky, koeficienty a stav filtra (sken) do súboru. Nahrať koeficienty do filtra, poslať a prijať vzorky a skenovací reťazec. Vzorky možno navrhnúť aj ručne pomocou rovníc. V okne testovacieho režimu vidno stav všetkých bitov filtra, kde je možné nastaviť zobrazenie v binárnych či hexadecimálnych hodnotách alebo v reálnych číslach. Takisto možno manuálne nastaviť vstupy a sledovať priebeh výpočtu aj po každom hodinovom impulze. Ako ukážka sú uvedené len 2 zo 4 okien aplikácie na obrázkoch 6 a 7.



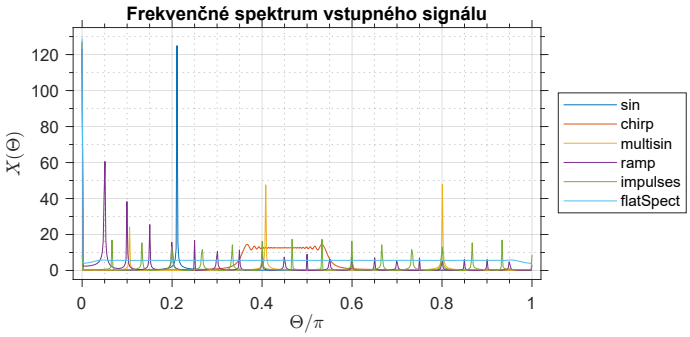
Obr. 6: Obrazovka MATLAB aplikácie pre návrh, spracovanie a ukladanie vstupných a výstupných vzoriek



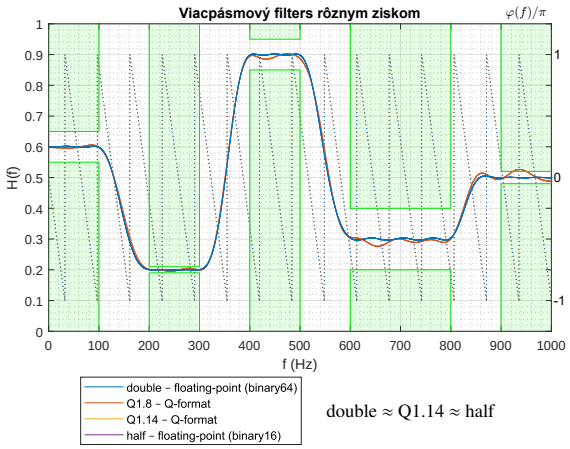
Obr. 7: Obrazovka MATLAB aplikácie pre testovací sken režim

7.2 Merania

Keďže uviesť predpisy vstupných signálov a frekvenčné charakteristiky filtrov by bolo príliš rozsiahle, sú uvedené aspoň ich frekvenčné spektrá z 1024 vzoriek (na obrázku 8), 1 charakteristika (na obrázku 9) a zoznam meraných typov FIR filtrov s rádmi v „{}“. Merané boli 2 dolnopriepustné {4, 60}, 2 hornopriepustné {8, 61}, 2 pásmovozádržné {44, 48}, 2 pásmovopriepustné {44, 55}, 2 hrebeňové {27, 10}, 2 výrezové (notch) {62, 62}, 3 viacpásmové {61, 62, 62} filtre, diferenciátor {1}, oneskorenie {49} a Hilbertov transformátor {62}.



Obr. 8: Spektrum vstupných signálov ($x[n] \in \langle 0; 255 \rangle$)



Obr. 9: Modulová a fázová frekvenčná charakteristika viacpásmového FIR filtra 62. rádu

7.3 Prediktory pre verifikáciu

Na verifikáciu správnosti funkcie aritmetických blokov a celkového FIR filtra boli vyvinuté SystemVerilog prediktory ich korektného správania. Vývoj prediktorov bol náročný, keďže napríklad každá floating-point operácia musela byť bitovo upravená, aby odpovedala nastavenej bitovej šírke mantisy a exponentu. Ďalej, aj samotné poradie sčítaní medzivýsledkov podľa normy IEEE 754 produkuje spravidla rozdiel v navrhnutom FIR filtri, ktorý kvôli minimalizácii plochy zjednodušením hardvéru sčítava súčiny od posledného koeficientu k nultému.

7.4 Výsledky

Pre záverečné testy boli zvolené nasledujúce číselné formáty:

- Half-precision (16 bitov, $M_W = 5$, $E_W = 10$) – najmenšia normou etablovaná bitová šírka
- Q1.8 (10 bitov) – odhadovaná bitová šírka prvého využitia
- Q1.14 (16 bitov) – rovnaká bitová šírka k floating-point konfigurácii

Odchýlka výstupu bola vypočítaná zo všetkých meraní súhrnne. Tabuľka 6 udáva rozdiel výstupných hodnôt meraného FIR filtra voči výpočtu s ideálnymi koeficientmi (double-precision) pri 8-bitovom neznamienkovom kvantovaní vstupu a výstupu s takým zaokrúhľovaním, ktoré je pre danú aritmetiku použité vo FIR filtri. Percentuálna hodnota je vypočítaná voči maximálnej hodnote výstupu, teda v použitej konfigurácii voči 255.

Tab. 4: Dosiahnuté parametre FIR filtra pre ASIC v technológii UMC 65 nm pri 63 koeficientoch a so vstupom *hold_samples*

Formát	Parameter						
	$f_{CLK_{MAX}}$ [1] (MHz)	Plocha[2] (μm^2)	Registre	Dynamická spotreba P_{dyn} (μW)		Statická spotreba P_{stat} (μW)	
				Typická[3]	Najhorší prípad[4]	Typická[3]	Najhorší prípad[4]
Q1.8	94,94	13 976	1 190	52,347	63,030	1,4518	360,51
Q1.14	83,79	18 569	1 574	85,708	101,47	1,7678	497,92
half-precision	31,28	22 498	1 812	295,68	351,26	2,2029	597,48
$Q1.8_{scan}$	74,77	21 099	1 191	53,603	69,255	1,4399	497,16
$Q1.14_{scan}$	66,05	27 341	1 575	76,497	96,466	1,9961	667,18
$half\text{-}prec._{scan}$	28,60	37 788	2 064	311,72	374,64	2,7165	911,02

[1] najhorší prípad, reg→reg [2] pri 70 % využití plochy
[3] TT tranzistory; V_{dd} = 1,2 V; T = 27 °C [4] FF tranzistory; V_{dd} = 1,32 V; T = 125 °C

V tabuľke 4 si možno v posledných dvoch stĺpcoch všimnúť nárast v statickej spotrebe až o dva rády. Toto značné zvýšenie je spôsobené exponenciálnou závislosťou statickej spotreby od pracovnej teploty.

Tab. 5: Dosiahnuté parametre FIR filtra realizovaného v FPGA Altera Cyclone II EP2C5T144C8 pri maximálnom počte koeficientov 63 a so vstupom *hold_samples*

Formát	Parameter					
	$f_{CLK_{MAX}}$ (MHz)	Registre	LUT	P_{dyn}[1] (mW)	P_{stat}[1] (mW)	$P_{I/O}$[2] (mW)
Q1.8	51,50	1 191	945	3,50	18,02	14,28
Q1.14	47,85	1 575	1 212	4,80	18,02	15,00
half-precision	11,68	1 813	2 267	43,85	18,12	15,00
$Q1.8_{sken}$	50,02	1 191	2 144	4,36	18,02	14,78
$Q1.14_{sken}$	48,71	1 575	2 795	5,69	18,03	15,49
$half\text{-}prec._{sken}$	11,18	2 064	4 445	41,39	18,11	15,49

[1] spotreba jadra [2] spotreba vstupno-výstupných obvodov

Tab. 6: Chyba výstupu FIR filtra s rôznou číselnou reprezentáciou a kvantovaním

Formát	Maximálny absolútny rozdiel výstupu (číslice / %)	Priemerný absolútny rozdiel výstupu (číslice / %)
Q1.8	5 / 1,9608	0,5844 / 0,2292
Q1.14	1 / 0,3922	0,0205 / 0,0080
half-precision	1 / 0,3922	0,0441 / 0,0173

8 Záver

Pre spomínanú prvú plánovanú aplikáciu predspracovania dát pre MPPT bola napríklad zvolená konfigurácia: 8-bitové vzorky, max. 63 koeficientov vo formáte Q1.8, nakoľko zistená presnosť výpočtu je dostatočná a žiadaná je viac minimalizácia spotreby energie a plochy na čipe.

Keďže návrh bol orientovaný na konkrétnu aplikáciu v špecifickej technológii, s inými návrhmi FIR filtrov určených pre iné technológie či aplikácie, nebol v tomto štádiu výskumu porovnávaný. Možno konštatovať, že pre podľa požiadaviek navrhnutý FIR filter pri rovnakej bitovej šírke koeficientov obvodovo zložitejšia číselná reprezentácia dát s pohyblivou rádovou čiarkou (floating-point) nemá jednoznačné výhody oproti číselnej reprezentácii s pevnou rádovou čiarkou (Q-formát). Podľa tabuliek uvedených vo výsledkoch (v časti 7.4) možno však podľa váhy jednotlivých požiadaviek každej aplikácie určiť najvhodnejšiu alternatívu.

Literatúra

[1] V. K. Solanki, M. K. Hoang, Z. Lu and P. K. Pattnaik. Intelligent Computing in Engineering: Select Proceedings of RICE 2019. *Advances in Intelligent Systems and Computing, vol. 1125*, DOI: 10.1007/978-981-15-2780-7, ISBN: 978-981-15-2779-1, Springer, 2020

[2] A. Hudec, L. Nagy, M. Kovac, L. Kohutka and V. Stopjakova. Maximum Power Point Tracking Circuit for an Energy Harvester in 130 nm CMOS Technology. *2020 International Conference on Applied Electronics (AE)*, DOI: 10.23919/AE49394.2020.9232903, 1-4, IEEE, 2020

[3] Floating Point and IEEE 754 Compliance for Nvidia GPUs. [online], [cit. 2025-03-10], `https://docs.nvidia.com/cuda/floating-point/index.html#rounding-modes`

[4] M. Pristach. Návrh optimalizovaných architektur digitálních filtrů pro nízkopříkonové integrované obvody. *Dizertačná práca*, 71, VUT Brno, 2015

[5] S. Goyal, K. G. Singh, and R. M. Mehra. Design of Low Power FIR Filter *Global Journal of Advanced Engineering Technologies*, 3:1–5, 2014

[6] L. Touil, A. Hamdi, I. Gassoumi and A. Mtibaa. Design of Low-Power Structural FIR Filter Using Data-Driven Clock Gating and Multibit Flip-Flops *Journal of Electrical and Computer Engineering*, 2020:1–9, 2020

Návrh µLaborátoria satelitu pre meranie viskozity mazív

Bc. Martin Brázda

[1] Ústav vedy a výskumu, FEI STU v Bratislave
[2] Názov inštitúcie/pracoviska autora

zodpovedny.xbrazda@stuba.sk

Abstrakt – Tento článok sa zaoberá konceptom návrhu konštrukcie mikrolaboratória pre meranie viskozity mazív v satelite veľkosti 1U. Článok obsahuje prehľad hlavných metód, ktoré sa využívajú na meranie viskozity mazív a kvapalín v prostredí mikrogravitácie. Vysvetľuje na akom princípe sú jednotlivé metódy založené, a popisuje ich výhody a nevýhody. V ďalšej časti sa práca podrobnejšie zaoberá aplikáciou týchto metód prostredníctvom navrhnutého modelu, ako aj ich princíp fungovania v danom modeli. V poslednej časti sa práca zaoberá zhrnutím vlastnosti a parametrov jednotlivých časti modelov, navrhnutím prístupov merania, ako aj celkovými parametre meracieho laboratória.

1 Úvod

Moderné satelitné misie čoraz častejšie vyžadujú palubné laboratória pre získavanie nových poznatkov z rôznych oblastí vedy v podmienkach mikrogravitácie. Návrh takýchto laboratórií však prináša mnohé výzvy najmä pri menších satelitov, ako najmä obmedzený priestor – všetky komponenty musia byť navrhnuté tak, aby sa zmestili do veľmi malého objemu, energetické limity – satelity majú prísne obmedzenia na spotrebu elektrickej energie, čo ovplyvňuje výber senzorov a meracích metód a hmotnostné obmedzenia – každý gram hmotnosti je kritický, preto musí byť celý systém optimalizovaný pre minimálnu hmotnosť.

V rámci tejto práce sa zameriavame na návrh kompaktného laboratória na meranie viskozity mazív v prostredí mikrogravitácie. Práca vychádza z analýzy existujúcich riešení a navrhuje inovácie na zvýšenie efektivity a spoľahlivosti merania, pre budúce misie.

2 Metódy merania viskozity

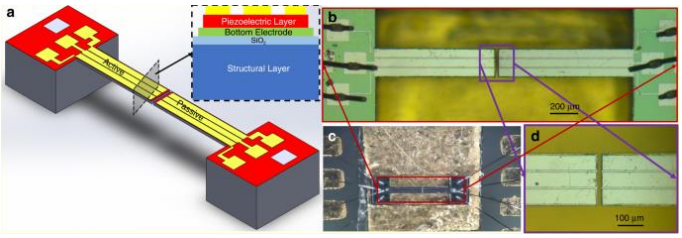
Meranie viskozity mazív je dôležité pri budúcom vývoji mazív a kvapalných systémov používaných v extrémnych podmienkach, aké sa vyskytujú vo vesmíre. Existuje niekoľko metód merania viskozity, pričom každá z nich má svoje výhody a obmedzenia.

2.1 Tip-Coupled Two-Cantilever (TCTC)

Táto metóda je založená na princípe dvoch piezoelektrických mikrokantileveroch (typ súčiastky MEMS), ktoré sú ponorené v meranej látke. Pričom jeden kantilever je aktívny a druhý pasívny. Aktívny kantilever funguje ako aktívny oscilátor a vibruje pri nastavenej frekvencii, pasívny kantilever je umiestnený v definovanej vzdialenosti, a pohybuje sa v dôsledku prenosu energie cez kvapalinu. Prenos energie z aktívnej konzoly na pasívnu konzolu sa uskutočňuje prostredníctvom šmykových vĺn generovaných aktívnou konzolou. Tok šmykových vĺn vytvorených kmitajúcou doskou sa riadi charakteristickou dĺžkou δ.

$$\delta = \sqrt{\frac{2\eta}{\omega}} \quad (1)$$

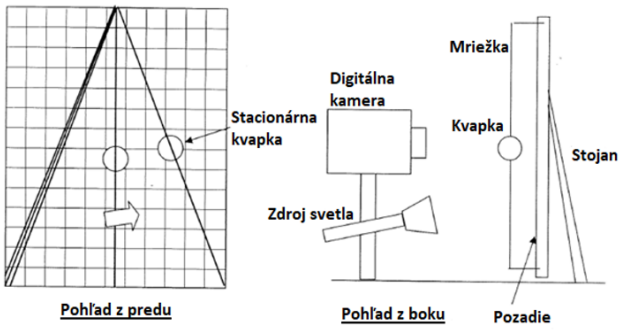
kde ω je frekvencia kmitov a η je kinematická viskozita kvapaliny. Hĺbka prieniku šmykových vĺn má rastúci trend s kinematickou viskozitou, a klesajúci trend s frekvenciou. Keďže medzera medzi konzolami je pre vopred vyrobený systém vždy pevná, môže existovať horná hranica frekvencie a dolná hranica viskozity [1]. Výsledný ukazovateľ viskozity sa tak určuje priamo, ako pomer amplitúd aktívneho a pasívneho kantileveru. Výhodou tejto metódy je priame a kompaktné meranie, možnosť merania pri rôznych frekvenciách a nízka spotreba energie. Nevýhodou tejto metódy je vysoká citlivosť na šum pri nízkych frekvenciách a technologická náročnosť výroby.



Obr. 1. (a) Schematické znázornenie konštrukcie viskozimetra TCTC. (b) Optický obrázok viskozimetra TCTC. (c) Optický obrázok toho istého zariadenia na doske plošných spojov. (d) Zväčšený optický obrázok systému, na ktorom je vidieť medzeru medzi dvoma kantilevermi [1]

2.2 Metóda koalescencie kvapiek

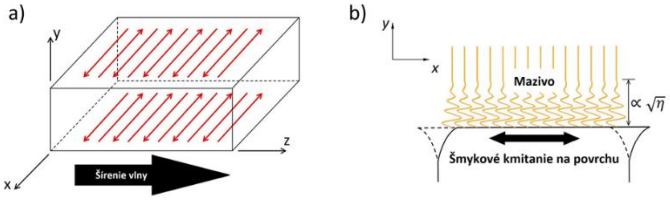
Metóda koalescencie kvapiek je založená na základe rýchlosti spájanie dvoch kvapiek kvapaliny (meranej látky). Kvapky sú vystavené povrchovému napätiu, ktoré spôsobuje ich spájanie, pričom výsledná viskozita je určená rýchlosťou relaxácie tvaru kvapiek. Experimentálne údaje získané optickými senzormi sa následne využívajú na výpočet dynamickej viskozity kvapaliny. Výhodou tejto metódy je rýchlosť merania a jej umiestnenie kvapalín v priestore mikrogravitácie, čo pomáha eliminovať vplyv heterogénnej nukleácie spôsobenej nádobou. Nevýhodou tejto metódy je nutnosť použitia presného zariadenia na sledovanie pohybu kvapiek a analýza veľkého objemu dát.



Obr. 2. Meracie pracovisko koalescencie kvapiek [2]

2.3 Akustická metóda SH-SAW (Shear-Horizontal Surface Acoustic Wave)

Táto metóda využíva šírenie šmykových horizontálnych akustických vĺn na povrchu senzora na meranie viskozity kvapaliny. Generované vlny interagujú s kvapalinou, pričom spôsobujú zmenu fázového posunu a amplitúdy signálu. Viskozita sa následne určuje z výpočtov fázového posunu a amplitúdy odrazenej akustickej vlny. Výhodou tejto metódy je možnosť merania v reálnom čase. Nevýhodou tejto metódy je citlivosť na hrúbku povrchovej vrstvy a materiálové vlastnosti testovanej látky.



Obr. 3. (a) Šírenie SH-SAW a pohyb častíc (červenou farbou), (b) Strata energie SH-SAW na rozhraní pevná látka - kvapalina v dôsledku viskozity kvapaliny [3]

2.4 Metóda rozdielového tlakové kapilárne viskozimetra DPV

Metóda DPV (Differential Pressure Viscometers) je založená na princípe merania odporu kvapaliny v podobe času potrebného na prietok maziva cez presne definovaný úzky kapilárny otvor. Mazivo prechádza cez kapilárny otvor pod vplyvom vonkajšieho napätia alebo za pomoci piestu, ktorý je poháňaný krokovým motorom. Pri použití krokového motora je potrebné merať dva hlavné parametre, a to čas potrebný na prechod látky cez kapiláru a elektrický prúd potrebný na pohon motora. Na základe týchto údajov sa následne vypočíta dynamická viskozita látky. Výhodou tejto metódy je flexibilita pre široký rozsah viskozít látok, presnosť merania vďaka jemnému nastaveniu tlaku a rýchlosti pretláčania. Nevýhodou tejto metódy je komplexnosť systému a energetická náročnosť meraní.



Obr. 4. Tlakový kapilárny viskozimeter [4]

3 Výber metód merania viskozity a návrh meracích jednotiek

Táto časť sa zameriava na návrh mikro-laboratória určeného na meranie rôznych typov mazív s odlišnými hodnotami viskozity. Hlavným cieľom je overiť presnosť a spoľahlivosť dvoch vybraných meracích metód pre budúce misie. Prvou zvolenou metódou je Tip-Coupled Two-Cantilever (TCTC), ktorá sa vyznačuje nízkou energetickou spotrebou a kompaktnými rozmermi, čo je výhodné pre aplikácie v obmedzených priestorových podmienkach. Táto metóda však využíva miniatúrne komponenty, ktoré môžu byť náchylné na poškodenie vplyvom otrasov počas prepravy a vynesenia na obežnú dráhu.

Z tohto dôvodu bola implementovaná aj druhá metóda, ktorá kombinuje akustickú metódu SH-SAW a vibračnú metódu, známu z laboratórnych viskozimetrov. Tento prístup umožňuje dosiahnuť robustnejšiu konštrukciu s vyššou mechanickou odolnosťou, čím sa minimalizuje riziko zlyhania experimentu počas misie.

3.1 Meracia jednotka TCTC

3.1.1 Konštrukcia a technické parametre

Meracia jednotka založená na metóde TCTC (Obr. 5) pozostáva z nami navrhnutého monolitického hliníkového bloku s vytvorenými štyrmi oddelenými komorami. Každá komora je dimenzovaná na objem maziva 230 mm³. Blok je umiestnený na silikónovej podložke, ktorá zabezpečuje jeho odizolovanie od sekundárnej dosky plošného spoja (DPS1) a zároveň bráni úniku maziva. Na vnútornom okraji čipu každej komory je vytvorený jeden integrovaný teplotný senzor.

V strede čipu sú umiestnené dva integrované piezoelektrické kantilevery s vrstvou zirkontitanátu olova (PZT), ktorá slúži ako aktívny prvok. Konzoly sú vyrobené na 25 µm hrubej kremíkovej doštičky (SOI), pričom ich rozmery sú šírka 200 µm, dĺžka 800 µm a vzájomná vzdialenosť koncov konzol je 30 µm.

3.1.2 Princíp merania a spracovanie dát

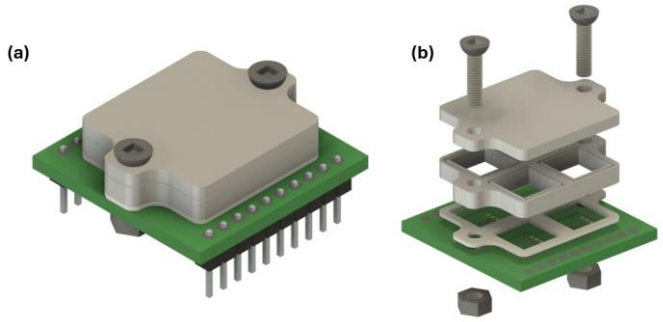
Pri aplikácii striedavého elektrického poľa na aktívnu konzolu začne táto konzola vibrovať s frekvenciou aplikované striedavého napätia, čím sa uvedie do pohybu okolitá kvapalina. Kvapalina následne pôsobí časovo premenlivou silou na

pasívnu konzolu, čím sa prenáša vibračná energia, ktorá uvedie pasívnu konzolu do kmitavého pohybu. Rozdiel amplitúdy kmitania pasívnej konzoly a aktívnej konzoly je priamo závislý od vlastností kvapaliny, pričom dominantným faktorom je jej viskozita.

Na vrchnej časti komory sa nachádza uzáver s drážkou vytvorenou pre možnosť aplikovania tesniacej látky, pričom celý systém je mechanicky uzatvorený pomocou skrutiek. Na spodnej časti DPS1 sú umiestnené zosilňovače signálu, pričom zosilnený signál sa následne cez piny konektorovej lišty privádza do mikrokontroléra (MCU) MSP430FR5994 [7], ktorý obsahuje 12 a 16-bitové ADC prevodníky a pamäť typu FRAM odolnú voči elektromagnetickému rušeniu. Dáta sú následne uchovávané v pamäti, kým nie sú prenesené do palubného počítača (OBC) pomocou dátovej zbernice I^2C.

3.1.3 Prevádzkové parametre a spotreba energie

Merania sa vykonávajú postupne v jednotlivých komorách, pričom jedno meranie trvá 15 sekúnd a interval medzi meraniami je 5 sekúnd. Merania sa opakujú každých 60 minút alebo na základe príkazu OBC. Maximálna spotreba energie jednotky by v prípade poruchy a fungovania celého systému nemala nepresiahnuť 50mW, pričom spotreba za bežných podmienok by mala byť 20mW. Celková hmotnosť modulu je do 30 gramov.



Obr. 5. (a) Model TCTC časti meracieho laboratória, (b) Rozložený pohľad TCTC časti meracieho laboratória

3.2 Meracia jednotka SH-SAW a vibračná metóda

3.2.1 Konštrukcia a technické parametre

Druhá časť mikro-laboratória (Obr. 6) využíva kombináciu metódy SH-SAW a vibračnej metódy. Túto jednotka sme navrhli z monolitického hliníkového bloku s vytvorenými 12 oddelenými komorami, pričom každá komora je navrhnutá na objem 340 mm^3 maziva. Monolitický hliníkový blok je obklopený dvomi silikónovými podložky, ktoré ho oddeľujú zo spodnej časti od hlavnej DPS, a z hornej časti od sekundárnej dosky plošného spoja (DPS2). Silikónová podložka zároveň zabezpečuje utesnenie systému pred únikom maziva. Sekundárna DPS2 zároveň slúži ako uzáver komôr, pričom je mechanicky uzatvárateľná pomocou skrutiek.

Na spodnej strane sekundárnej DPS2 sú prispájkované MEMS reproduktory (Adap UT-P2023 [6]), ktoré využívajú piezoelektrický jav na generovanie vibrácií. Na vrchnej strane hlavnej DPS je umiestnené pole MEMS mikrofónov, ktoré detegujú vibrácie spôsobené interakciou akustických vĺn s médiom.

3.2.2 Princíp merania a spracovanie dát

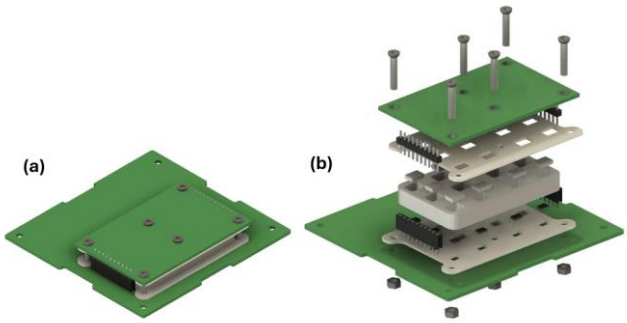
Pri aplikácii striedavého elektrického signálu do MEMS reproduktorov sa vytvára mechanická vibrácia, ktorá sa šíri cez médium. Viskozita sa následne vypočíta na základe fázového posunu a amplitúdy zaznamenaných vĺn vo forme vibrácií. Signály sú následne zosilnené a prevedené na digitálny signál prostredníctvom 12 a 16-bitových ADC prevodníkov v MCU.

3.2.3 Prevádzkové parametre a spotreba energie

Aby sa zabránilo interferencii medzi meraniami, merania sa vykonávajú postupne. Meranie každej komory trvá 10 sekúnd, pričom interval medzi meraniami je 5 sekúnd. Celý merací cyklus sa opakuje každých 60 minút alebo na základe príkazu OBC.

Z 12 komôr je 9 určených na meranie troch rôznych mazív, pričom každé mazivo sa testuje v troch samostatných komorách. Zvyšné tri komory slúžia ako referenčné vzorky s mazivom so známou viskozitou a stabilnými vlastnosťami, odolné voči zmenám. Ak je to potrebné, tieto komory môžu byť využité na ďalšie merania, avšak za podmienky, že aspoň jedna z komôr si zachová svoju referenčnú funkciu.

Maximálna spotreba energie jednotky by v prípade poruchy a fungovania celého systému nemala nepresiahnuť 750mW, pričom spotreba za bežných podmienok by mala byť 100mW. Celková hmotnosť modulu by nemala presiahnuť hodnotu 150 gramov.



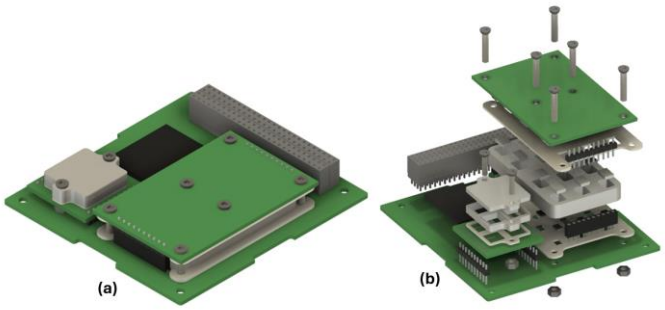
Obr. 6. (a) Model jednotka SH-SAW a vibračnej časti meracieho laboratória, (b) Rozložený pohľad jednotka SH-SAW a vibračnej časti meracieho laboratória

3.3 Návrh a konštrukcia mikro-laboratória

Mikro-laboratórium na meranie viskozity mazív (Obr.7) bolo kompletne navrhnuté a zostrojené nami s využitím modelovacieho softvéru Autodesk Fusion 360. Pri návrhu sme zohľadnili nielen technické požiadavky jednotlivých meracích jednotiek, ale aj priestorové obmedzenia a mechanickú odolnosť celého zariadenia. Hlavným konštrukčným materiálom pre komory oboch meracích jednotiek je hliník typu 6061, ktorý bol vybraný pre svoju nízku hmotnosť, vysokú pevnosť a dobrú tepelnú vodivosť.

Tab. 1. Tabuľka parametrov mikro-laboratória

Parametre	**Meracia jednotka TCTC**	**Meracia jednotka SH-SAW**	**Celého zariadenia**
Hmotnosť	30g	150g	220g
Bežná spotreba	20mW	100mW	20 – 120mW
Poruchová spotreba	50mW	750mW	50 – 800mW
Materiál komôr	Hliník 6061	Hliník 6061	
Materiál skrutiek	Hliník 6061	Hliník 6061	
Meracie komory	4	9 až 12	16
Referenčné komory	0	1 až 3	3



Obr. 7. (a) Model meracieho laboratória, (b) Rozložený pohľad meracieho laboratória

4 Záver

Navrhnuté a skonštruované mikro-laboratórium predstavuje kompaktné a efektívne riešenie na meranie viskozity rôznych typov mazív v podmienkach mikrogravitácie. Vďaka integrácii dvoch odlišných meracích metód – TCTC a kombinácie SH-SAW s vibračnou metódou – sme dosiahli vyššiu presnosť a spoľahlivosť meraní, pričom každá metóda ponúka špecifické výhody v oblasti mechanickej odolnosti, energetickej spotreby a kompaktnosti.

Celková konštrukcia laboratória bola navrhnutá v programe Autodesk Fusion 360 s dôrazom na minimalizáciu hmotnosti a energetickej náročnosti, pričom použitý hliník typu 6061 zabezpečuje dostatočnú pevnosť a odolnosť proti mechanickému poškodeniu. Modulárny návrh umožňuje jednoduchú integráciu do existujúcich systémov satelitu typu 1U, nízku spotrebu energie, nízku hmotnosť a zaručuje efektívne fungovanie aj v podmienkach s obmedzenými zdrojmi. Realizáciou tohto návrhu sme vytvorili základ pre budúce testovanie a optimalizáciu mikro-laboratória s cieľom jeho nasadenia v reálnom satelite.

Poďakovanie

Týmto spôsobom by som sa chcel poďakovať svojmu konzultantovi pri vytváraní tejto práce Ing. Michalovi Dzurišovi, ako aj všetkým, ktorým sa akýmkoľvek spôsobom pričinili k vzniku tejto práce.

Literatúra

[1] TIWARI, Sudhanshu, Ajay DANGI a Rudra PRATAP, 2023. A tip-coupled, two-cantilever, non-resonant microsystem for direct measurement of liquid viscosity. Microsystems & Nanoengineering [online]. 2023, roč. 9, č. 1, s. 1–9. ISSN 2055-7434. Dostupné na: doi:10.1038/s41378-023-00483-6

[2] GODFREY, Brian Michael, 2011. EXPERIMENTAL RESULTS FOR VISCOSITY MEASUREMENTS PERFORMED ON THE INTERNATIONAL SPACE STATION USING DROP COALESCENCE IN MICROGRAVITY. no date. Dostupné na: https://trace.tennessee.edu/cgi/viewcontent.cgi?article=2135&context=utk_gradthes

[3] TYREAS, G., R.S. DWYER-JOYCE, R.S. NOTAY a D.J. DURHAM, 2022. Measuring lubricant viscosity at a surface and in a bearing film using shear-horizontal surface acoustic waves. Proceedings of the Institution of Mechanical Engineers, Part J [online]. 2022, roč. 236, č. 8, s. 1511–1530. ISSN 1350-6501. Dostupné na: doi:10.1177/13506501221092867

[4] anon., no date. How to measure viscosity | Anton Paar Wiki. Anton Paar [online] [cit. 20.3.2025]. Dostupné na: https://wiki.anton-paar.com/en/how-to-measure-viscosity/

[5] IZUMO, Naoto, D DIVISION a D CO, no date. Physical Quantity Measured by a Vibration Viscometer Subtitle: The JCSS Standardization of Viscosity. no date. Dostupné na: https://www.aandd.jp/support/materials/23rd_sensing_forum_sv.pdf

[6] anon., no date. Ganymede Series MEMS speakers. USound [online]. [cit. 20.3.2025]. Dostupné na: https://usound.com/product/ganymede/

[7] anon., no date. MSP-EXP430FR5994 Development kit | TI.com [online] [cit. 20.3.2025]. Dostupné na: https://www.ti.com/tool/MSP-EXP430FR5994#description

Simulácie pre optimalizáciu batériového systému v elektrickom autobuse MHD

Bc. Boris Bajla[1]

[1]Ústav elektroniky a fotoniky, Fakulta elektrotechniky a informatiky STU v Bratislave

xbajla@stuba.sk

Práca sa zameriava na návrh a simuláciu batériového systému pre elektrický autobus MHD na linke 39 v Bratislave. Pomocou reálnych meraní trasy, rýchlosti a počtu cestujúcich bol vytvorený model v prostredí MATLAB Simulink. Simulácia zohľadňuje topografiu, hmotnosť a dynamiku jazdy. Výsledky ukazujú vplyv prevýšenia a zaťaženia na spotrebu energie, ako aj prínos regeneratívneho brzdenia. Navrhnutý systém umožňuje absolvovať niekoľko obojsmerných jázd bez nutnosti priebežného dobíjania.

1 Úvod

Budúcnosť hromadnej dopravy je významne prepojená s rozvojom a zavádzaním batériových elektrických autobusov, ktoré sa čoraz viac považujú za životaschopnú a udržateľnú alternatívu k tradičným autobusom poháňaným fosílnymi palivami. Nasadenie elektrických autobusov bude čoskoro kľúčovým prvkom pri transformácii mobility smerom od fosílnych palív, keďže obnoviteľné zdroje energie primárne produkujú elektrickú energiu, elektromotory predstavujú perspektívne riešenie pre ekologickú a udržateľnú dopravu. [10] Elektrické autobusy ponúkajú výhody, ako sú nulové výfukové emisie, čo je rozhodujúce pre zlepšenie kvality ovzdušia v mestských oblastiach. [1] Ich potenciál znížiť emisie CO_2 by mohol dosiahnuť až 75 % v závislosti od spôsobu výroby elektrickej energie. [12] Elektrické autobusy poskytujú vysokú účinnosť, nižšiu úroveň hluku a lepšie zrýchlenie v porovnaní s tradičnými dieselovými autobusmi. [1] Môžu tiež obsahovať funkcie, ako je rekuperácia brzdnej energie a ukladanie energie v superkapacitoroch. [10]

Nórsko patrí medzi svetových lídrov v oblasti elektrických vozidiel a má ambiciózny cieľ dosiahnuť úplne bezemisný autobusový park do konca roku 2025. [1] Oslo, hlavné mesto krajiny, je príkladom tohto pokroku. Do decembra 2023 prešlo takmer celé mesto na batériové autobusy, až na niekoľko výnimiek spôsobených špecifickými infraštruktúrnymi obmedzeniami. [5] Odráža to širší globálny trend elektrifikácie verejnej dopravy.

Prebiehajúci vývoj v oblasti technológií batérií sa zameriava na zlepšenie kapacity, zvýšenie bezpečnosti, zníženie výrobných a prevádzkových nákladov. Úspešná integrácia elektrických autobusov do existujúcich sietí verejnej dopravy si vyžaduje dôkladné zváženie výberu trasy, modelu elektrického autobusu a stratégie nabíjania vzhľadom na dostupnú alebo novú infraštruktúru. [13] Hlavnou výzvou sú vysoké počiatočné náklady na batériové systémy, ktoré musia spĺňať požiadavky miestnych liniek a zároveň minimalizovať finančné riziko pre dopravcov. [11]

Znalosť aktuálnej a predpokladanej hodnoty stavu nabitia batérie umožňuje dopravným podnikom efektívnejšie riadenie rozmiestnenia autobusov a harmonogramov nabíjania, čím sa zabezpečí, že nabíjanie prebieha v optimálnom čase, čo môže znížiť celkové náklady na energiu. [14]

Simulácie pre presný odhad spotreby energie umožňujú prevádzkovateľom zabezpečiť, že autobus spoľahlivo dokončí trasu bez rizika vybitia batérie. [4] Kapacita batérie sa časom znižuje v dôsledku starnutia a degradačných procesov, čo môže mať vplyv na výkon a vyvolávať bezpečnostné obavy. [20] Presným predpovedaním stavu nabitia batérie a pochopením jeho vplyvu na stav batérie (napr. vyhýbanie sa hlbokému vybíjaniu) možno vyvinúť stratégie na predĺženie životnosti batérie, čím sa znížia náklady spojené s jej výmenou. [16] Simulačné modely na báze reálnych meraní sa používajú pre odhad degradácie batérie na základe zmien stavu nabitia počas jazdy, nabíjania na zastávkach alebo nabíjania na centrálnom termináli. [19] Na širšej úrovni sú simulácie dôležité pri modernizácii miest, keďže dokážu predpovedať strategicky najvýhodnejšie lokality pre nabíjaciu infraštruktúru. Tým umožňujú optimalizovať jej rozmiestnenie a využitie, čo je nevyhnutné pre efektívne nasadenie elektrifikovanej dopravy. [4]

2 Prehľad publikovaných výsledkov

Určenie odporúčanej kapacity batériového systému pre elektrický autobus je zložité a závisí od viacerých faktorov, vrátane zamýšľaného použitia autobusu (mestský, regionálny), denného počtu kilometrov, topografie trás a dostupnej nabíjacej infraštruktúry. [1] Napríklad medzimestské autobusy môžu vyžadovať väčšie batérie (320-680 kWh) kvôli dlhším vzdialenostiam. [2] Mestské autobusy v jednej štúdii potrebovali na pokrytie svojich denných potrieb 640 kWh batériu na nočné nabíjanie. [2] Štúdia autobusových liniek toto tvrdenie potvrdila, keď ukázala, že minimálna kapacita batérie sa pohybuje od 18 kWh do viac ako 70 kWh v závislosti od spotreby energie na trase. Pre osemhodinový prevádzkový deň v najhorších podmienkach môže byť pre 18 m dlhý autobus

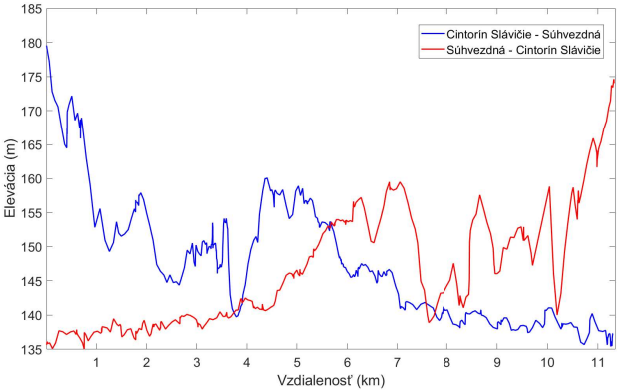
potrebná využiteľná kapacita batérie väčšia ako 400 kWh, čo si potenciálne vyžaduje inštalovanú kapacitu približne 625 kWh. [17] Dostupnosť a typ nabíjacej infraštruktúry významne ovplyvňujú potrebnú kapacitu batérie. [2]

- Nočné nabíjanie v sklade: Ide o stratégiu, pri ktorej sa autobusy s väčšími batériami nabíjajú v depe počas hodín mimo prevádzky, zvyčajne pomocou pomalých nabíjačiek. Autobusy tak môžu začať deň s plným nabitím a pokryť svoje denné trasy. [2] Spoliehanie sa výlučne na veľké batérie a nočné nabíjanie však môže viesť k ťažším vozidlám a nemusí byť optimálne pre všetky typy trás alebo prevádzkové požiadavky. [15]
- Príležitostné nabíjanie (rýchle nabíjanie): Mnohé veľké mestá zavádzajú alebo skúmajú príležitostné nabíjanie, pri ktorom sa autobusy s menšími batériami počas prestávok rýchlo nabíjajú vysokým výkonom na strategických miestach na svojich trasách, ako sú autobusové zastávky alebo konečné stanice. [2] To umožňuje znížiť potrebnú kapacitu batérie a predĺžiť prevádzkový dosah bez potreby veľmi dlhého nabíjania.
- Bleskové nabíjanie, ako sa realizuje v Ženeve so systémom TOSA, umožňuje veľmi rýchle nabíjanie (napr. 15-20 sekúnd pri výkone 600 kW) na určených zastávkach, čo umožňuje používanie menších a ľahších batérií. Tým sa eliminuje potreba nadzemného vedenia na celej trase. [6]
- Hybridné stratégie nabíjania: Veľké mestá čoraz častejšie využívajú hybridný prístup, ktorý kombinuje nočné nabíjanie v skladoch s príležitostným nabíjaním na trase. To umožňuje dosiahnuť rovnováhu medzi dostatočným počiatočným nabitím na každodennú prevádzku a jeho doplnením o rýchle nabíjanie počas dňa na udržanie vyššieho stavu nabitia a využitie menšieho počtu batérií. [19] V meste Eindhoven sa používa kombinácia rýchlonabíjačiek (450 kW) na strategických miestach a nízkovýkonných nabíjačiek (30 kW) v depe na nočné nabíjanie. [13]

3 Výber linky MHD

Pre účely našej práce sme sa rozhodli použiť autobusovú linku 39 v Bratislave. Táto linka premáva medzi zastávkami Súhvezdná a Cintorín Slávičie údolie a prechádza okolo našej školy, čo ju robí študentom veľmi blízkou. Dĺžka trasy linky 39 je približne 11.2 km a zahŕňa 21 zastávok. Medzi významné zastávky patria Zochova, Kollárovo námestie, Račianske mýto a Trnavské mýto, ktoré sú dôležitými prestupnými bodmi v rámci mestskej hromadnej dopravy.

Linka 39 je náročná z viacerých hľadísk. Autobusy sú často v ranných hodinách plné, čo predstavuje zvýšenú záťaž pre použitie batériovo napájaného elektrického autobusu. Trasa linky prechádza cez celé mesto, kde môže byť doprava náročná, a autobusy sa často musia zastaviť v kolónach. Navyše, elevácia trasy nie je rovnomerná - smerom na zastávku Cintorín Slávičie údolie musí autobus prekonať stúpanie. Tieto faktory robia linku 39 ideálnym kandidátom pre štúdium implementácie elektrických autobusov v náročných mestských podmienkach.



Obr. 1: Výškový profil trasy

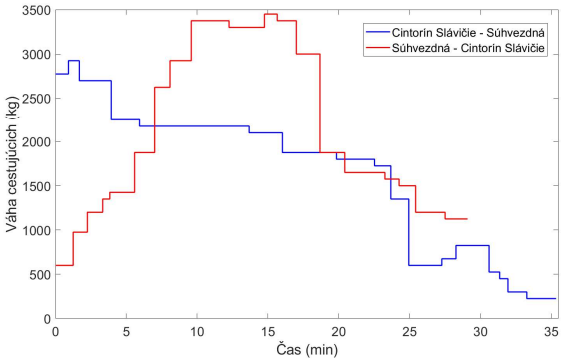
4 Meranie reálnych dát

Pre meranie dát linky 39 sme použili smartfón iPhone 13 s aplikáciou GeoLogger. [7] Táto aplikácia umožňuje zaznamenávať GPS súradnice a elevéciu trasy v reálnom čase. Na základe GPS súradníc a času je potom možné vypočítať rýchlosť pohybu meraného autobusu. Okrem toho sme manuálne počítali počet nastupujúcich a vystupujúcich cestujúcich na jednotlivých zastávkach, spolu s časom ich nástupu a výstupu, čo je možné pozorovať na grafe Obrázok 2. Presnosť tohto merania odhadujeme približne na ±3 osoby. Pri výpočtoch sme predpokladali, že priemerná hmotnosť jedného cestujúceho je 75 kg. Získané údaje – rýchlosť autobusu, výškový profil trasy a časovo sa meniaca hmotnosť cestujúcich – vstupujú do našej simulácie ako faktory ovplyvňujúce energetickú spotrebu a výkon modelovaného elektrického autobusu.

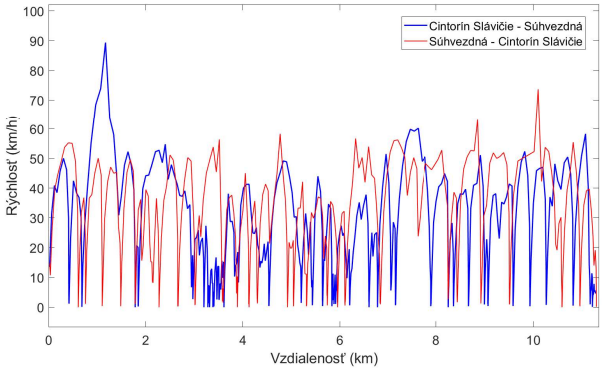
Meranie trasy v smere Cintorín Slávičie – Súhvezdná sa uskutočnilo 20. 3. 2025 o 08:45:22 CET. Autobus bol stredne zaplnený cestujúcimi a premávka v meste bola veľmi hustá a spomaľovala jeho pohyb, čo je možné pozorovať grafe Obrázok 3. Napriek tomu počas cesty nenastali žiadne výrazné komplikácie. Meranie trasy v opačnom smere, Súhvezdná – Cintorín Slávičie, prebehlo v ten istý deň o 09:24:43 CET. Aj v tomto prípade bol autobus stredne zaplnený, avšak premávka bola výrazne redšia, čo sa prejavilo na kratšom čase jazdy – cesta trvala o 6 minút menej. Porovnanie výškových profilov v grafe Obrázok 1 ukázalo, že smer Súhvezdná – Cintorín Slávičie vykazuje väčšie prevýšenie a teda predstavuje elevačne náročnejší úsek.

5 Model autobusu

Na autobusovej linke č. 39 v Bratislave aktuálne premávajú dva typy kĺbových autobusov. Starším z dvojice je SOR NB 18. Tento kĺbový autobus má dĺžku 18750 mm a je poháňaný šesťvalcovým naftovým motorom. Jeho maximálna kapacita je 150 cestujúcich + 1 vodič, pričom celková prípustná hmotnosť vozidla je 27000 kg a pohotovostná hmotnosť predstavuje 16830



Obr. 2: Váha cestujúcich v meranom autobuse



Obr. 3: Rýchlosť meraného autobusu

kg. [9] Druhou, modernejšou alternatívou je Otokar Kent C 18,75. Tento model taktiež disponuje dĺžkou 18 750 mm a pohonom zabezpečeným šesťvalcovým motorom. Autobus má celkovú kapacitu 155 cestujúcich + 1 vodič, celkovú hmotnosť 28000 kg a pohotovostnú hmotnosť 17450 kg. [8] Tieto vozidlá tvoria referenčný základ pre porovnanie s elektrickým pohonom v rámci našej navrhovanej simulácie.

V prípade linky č. 39 nie je možné nasadenie trolejbusov z dôvodu neexistujúcej trolejovej infraštruktúry na celej trase. Vhodnou náhradou za vozidlá so spaľovacím pohonom tak môže byť plne elektrický autobus napájaný z batériového systému. Jedným z perspektívnych kandidátov je kĺbový autobus Volvo 7900 Electric Articulated. Tento model má dĺžku 18660 mm a maximálnu celkovú hmotnosť 30000 kg.[3] Pohonnú sústavu tvoria dva elektrické motory s výkonom 2×200 kW, pracujúce s dvojstupňovou prevodovkou. Energiu zabezpečuje batériový systém s pracovným napätím 600V, využívajúci lítium-iónovú technológiu typu NCA (nikel–kobalt–hliník) s celkovou kapacitou až 564kWh.[3]

V rámci našej práce sme v softvérovom prostredí MATLAB Simulink vytvorili zjednodušený model elektrického autobusu, inšpirovaný technickými parametrami moderných batériových elektrických vozidiel pre mestskú hromadnú dopravu. Model simuluje kompletnú pohonnú jednotku elektrického vozidla, pričom simulácia zohľadňuje parametre v Tabuľka 1. Hodnoty ako napríklad valivý odpor kolies, straty v diferenciále a prevodovke sú bezstratové, pre zjednodušenie modelu.

Pohon zabezpečuje jednosmerný elektrický motor s výkonom 400 kW a napätím 600 V. V rámci modelu sa neuvažuje s tepelnými stratami. Riadiaci algoritmus využíva pulzno-šírkovú moduláciu (PWM) v kombinácii s H-mostíkovým meničom, ktorý zároveň umožňuje aj regeneratívne brzdenie, teda spätné získavanie energie počas spomaľovania. Riadiaca jednotka interpretuje vstupné signály a reguluje výkon dodávaný do motora na základe požadovanej jazdnej dynamiky. Model predpokladá ideálne správanie vodiča bez neželaných akcelerácií a zbytočného brzdenia.

6 Návrh batériového balíka

Napájanie systému zabezpečuje batériový balík, ktorý dodáva elektrickú energiu do riadiacej jednotky a následne do motora. Spotreba energie, rovnako ako jazdná dynamika, sú v simulácii vyhodnocované v závislosti od aktuálneho zaťaženia, sklonu trasy a rýchlosti jazdy. Ako základný stavebný prvok batériového systému bola zvolená batéria U27-36XP od spoločnosti Valence Technology, založená na $LiFePO_4$ (lítium-železo-fosfát) chémií. Táto batéria poskytuje nominálne napätie 38,4 V a energetickú kapacitu 1920 Wh pri kompaktných rozmeroch. [18] Batériový systém bol modelovaný v prostredí MATLAB, pričom jeho návrh vychádzal z kapacitných a výkonových požiadaviek simulovaného autobusu nasledovne.

Napätie jedného článku:

$$U_{\text{cell}} = 38.4\,\text{V}$$

Celkové systémové napätie:

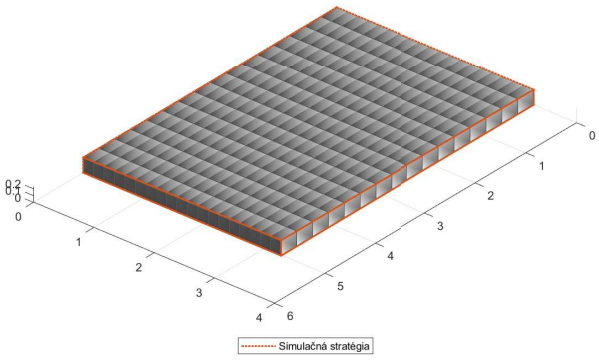
$$U_{\text{total}} = 16 \times 38.4 = 614.4\,\text{V}$$

Energia jedného článku:

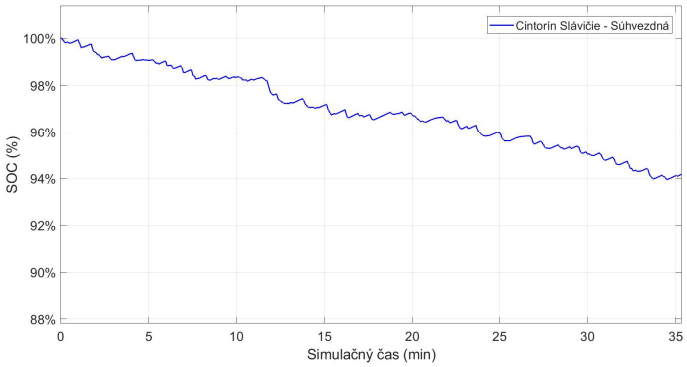
$$E_{\text{cell}} = 1.92\,\text{kW h}$$

Tab. 1: Parametre modelu elektrického autobusu

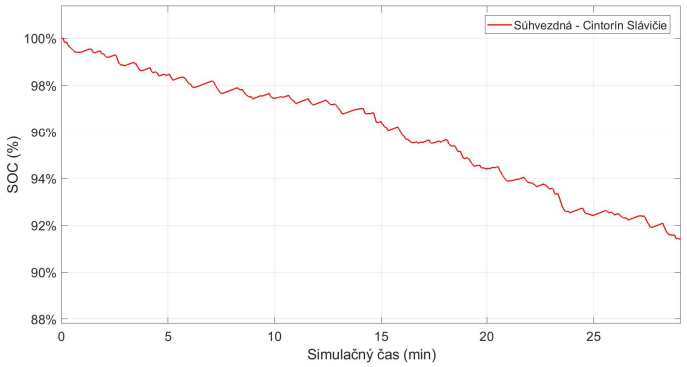
Parameter	Hodnota
Hmotnosť vozidla	19000 kg [3]
Vzdialenosť ťažiska ku prednej náprave	5945 mm [3]
Vzdialenosť ťažiska ku zadnej náprave	6.605 mm [3]
Vzdialenosť ťažiska od zeme	1000 mm
Gravitačné zrýchlenie	9.81 m/s^2
Čelná plocha vozidla	8.415 m^2 [3]
Koeficient aerodynamického odporu	0.7
Hustota vzduchu	1.18 kg/m^3
Valivý polomer kolies	0.5 m
Sklz pri maximálnej adhéznej sile	10 %



Obr. 4: Model batériového systému



(a) Stav nabitia batérie pre trasu Cintorín Slávičie - Súhvezdná



(b) Stav nabitia batérie pre trasu Súhvezdná - Cintorín Slávičie

Obr. 5: Stav nabitia batérie pre obidve trasy

Počet článkov:

$$N = 16 \times 19 = 304$$

Celková kapacita batériového systému:

$$E_{\text{total}} = 304 \times 1.92 = 583.68\,\text{kW h}$$

7 Výsledky

Na základe výsledkov simulácie bola analyzovaná zmena stavu nabitia batérie počas obojsmernej jazdy na trase linky 39 medzi zastávkami Cintorín Slávičie údolie a Súhvezdná. Výsledky sú znázornené na grafoch Obrázok 5. Pri smere Cintorín Slávičie údolie - Súhvezdná bol zaznamenaný pokles stavu nabitia batérie z 100 % na približne 94 % počas približne 35 minút jazdy. Tento pokles je spôsobený predovšetkým vysokou intenzitou dopravy, čo malo za následok časté zastavovanie a rozbiehanie, ktoré je energeticky náročné. Výškový profil tejto trasy je negatívny, pričom v takýchto prípadoch zohráva dôležitú úlohu regeneratívne brzdenie, ktoré prispieva k čiastočnému zníženiu spotreby.

Pri smere Súhvezdná - Cintorín Slávičie údolie poklesol stav nabitia až na hodnotu približne 91 %, a to aj napriek nižšej premávke a kratšiemu času jazdy. Tento jav je spôsobený najmä zvýšeným prevýšením trasy v tomto smere a zároveň vyšším počtom cestujúcich. Vyššia hmotnosť a stúpanie sa prejavujú výraznejším odberom energie z batérie.

Z uvedených výsledkov je zrejmé, že topografia trasy elevácia má výrazný vplyv na rýchlosť vybíjania batérie, dokonca viac než dopravná situácia. Regeneratívne brzdenie pritom zohráva kľúčovú úlohu pri predlžovaní dojazdu, najmä v hustej mestskej premávke s častým brzdením alebo na trase so záporným výškovým profilom.

Na základe výsledkov simulácie možno predpokladať, že elektrický autobus s navrhnutou batériovou konfiguráciou by bol schopný absolvovať približne štyri obojsmerné cykly na linke 39, kým by bolo potrebné opätovné nabitie batérie. Tento fakt poukazuje na obmedzený denný dojazd pri použití vysoko kapacitnej batérie, čo môže byť nevýhodou z hľadiska hmotnosti a nákladov. Ako alternatíva sa javí možnosť využitia menšej batérie v kombinácii s priebežným nabíjaním počas prestojov na konečných zastávkach, prípadne aj na vybraných medzizastávkach, čo by výrazne znížilo nároky na hmotnosť a náklad batériového systému.

Tab. 2: Parametre batériového balíka

Parameter	Hodnota
Počet batérii zapojených sériovo	16
Počet batérii zapojených paralelne	19
Celková kapacita batériového systému	583.68 kWh
Celkové napätie batériového systému	614.4 V
Objem	3.706 m^3
Hmotnosť	5684.8 kg

8 Záver

V tejto práci sme sa zamerali na simuláciu a optimalizáciu batériového systému pre elektrický autobus nasadený na linke 39 v Bratislave. Na základe reálnych meraní trasy, rýchlosti jazdy, elevácie a počtu cestujúcich bol vytvorený model elektrického autobusu a batériového systému v prostredí MATLAB Simulink. Výsledky simulácií potvrdili, že topografia trasy a hmotnosť vozidla vrátane cestujúcich majú zásadný vplyv na spotrebu energie. Zároveň sa ukázalo, že regeneratívne brzdenie môže významne prispieť k efektívnejšiemu využitiu energie, najmä v podmienkach mestskej prevádzky s častým zastavovaním.

Navrhnutý batériový systém s kapacitou 583 kWh umožňuje absolvovať približne štyri obojsmerné cykly bez potreby dobíjania. Tento prístup je však spojený s vyššou hmotnosťou a nákladmi na batériu. Alternatívne riešenie, založené na menšej kapacite v kombinácii s priebežným nabíjaním počas prestojov, sa ukazuje ako perspektívnejšia a flexibilnejšia možnosť pre efektívne nasadenie elektrických autobusov v mestskom prostredí. Práca potvrdzuje význam modelovania a simulácií ako nástroja pre návrh udržateľnej a ekonomicky efektívnej verejnej dopravy.

Poďakovanie

Rád by som poďakoval doc. Ing. Miroslavovi Mikoláškovi, PhD., za odborné rady a usmernenie pri realizácii tejto práce.

Literatúra

[1] Linda Ager-Wick Ellingsen, Rebecca Jayne Thorne, Julia Wind, Erik Figenbaum, Mia Romare, and Anders Nordelöf. Life cycle assessment of battery electric buses. *Transportation Research Part D: Transport and Environment*, 112:103498, 2022.

[2] Hussein Basma, Charbel Mansour, Marc Haddad, Maroun Nemer, and Pascal Stabat. Energy consumption and battery sizing for different types of electric bus service. *Energy*, 239:122454, 2022.

[3] Volvo Buses. Volvo 7900 electric articulated - data sheet, 2024. Accessed: 2025-03-17.

[4] Changyin Dong, Zhuozhi Xiong, Ni Li, Xinlian Yu, Mingzhang Liang, Chu Zhang, Ye Li, and Hao Wang. A real-time prediction framework for energy consumption of electric buses using integrated machine learning algorithms. *Transportation Research Part E: Logistics and Transportation Review*, 194:103884, 2025.

[5] EV Transport Climate Pathway (EVTCP). Norway - electric vehicle transport climate pathway, 2025. Accessed: 2025-03-14.

[6] Transports Publics Genevois. Tosa: Flash-charging battery buses for urban public transport in geneva, 2018. Accessed: 2025-03-17.

[7] GeoLogger - GPS Speed Tracker. GeoLogger - GPS Speed Tracker. `https://apps.apple.com/us/app/geologger-gps-speed-tracker/id1256293008`, 2025. Accessed: 2025-03-22.

[8] imhd.sk. Otokar kent c 18.75, n.d. Accessed: 2025-03-24.

[9] imhd.sk. Sor nb 18 city, n.d. Accessed: 2025-03-24.

[10] Reinhart Kühne. Electric buses – an energy efficient urban transportation means. *Energy*, 35(12):4510–4513, 2010. The 3rd International Conference on Sustainable Energy and Environmental Protection, SEEP 2009.

[11] Moataz Mohamed, Mark Ferguson, and Pavlos Kanaroglou. What hinders adoption of the electric bus in canadian transit? perspectives of transit providers. *Transportation Research Part D: Transport and Environment*, 64:134–149, 2018. The contribution of electric vehicles to environmental challenges in transport. WCTRS conference in summer.

[12] G.J. Offer, D. Howey, M. Contestabile, R. Clague, and N.P. Brandon. Comparative analysis of battery electric, hydrogen fuel cell and hybrid vehicles in a future sustainable road transport system. *Energy Policy*, 38(1):24–29, 2010.

[13] Gregor Papa, Marina Santo Zarnik, and Vida Vukašinović. Electric-bus routes in hilly urban areas: Overview and challenges. *Renewable and Sustainable Energy Reviews*, 165:112555, 2022.

[14] Harikishan Perugu, Sonya Collier, Yi Tan, Seungju Yoon, and Jorn Herner. Characterization of battery electric transit bus energy consumption by temporal and speed variation. *Energy*, 263:125914, 2023.

[15] Matthias Rogge, Sebastian Wollny, and Dirk Uwe Sauer. Fast charging battery buses for the electrification of urban public transport—a feasibility study focusing on charging infrastructure and energy storage requirements. *Energies*, 8(5):4587–4606, 2015.

[16] Ali Shehabeldeen, Ahmed Foda, and Moataz Mohamed. A multi-stage optimization of battery electric bus transit with battery degradation. *Energy*, 299:131359, 2024.

[17] Philipp Sinhuber, Werner Rohlfs, and Dirk Uwe Sauer. Study on power and energy demand for sizing the energy storage systems for electrified local public transport buses. pages 315–320, 01 2012.

[18] Valence. U27-36xp. `https://www.valence.com/stackable-modules-12-819v/u27-36xp`. Accessed: 2025-03-24.

[19] Yongxing Wang, Feixiong Liao, Jun Bi, and Chaoru Lu. Optimal battery electric bus system planning considering heterogeneous vehicles, opportunity charging, and battery degradation. *Renewable Energy*, 237:121596, 2024.

[20] Yifei Xu and Hengzhao Yang. A battery capacity trajectory prediction framework with mileage correction for electric buses. *Journal of Energy Storage*, 110:115301, 2025.

Zlepšenie elektrochemických vlastností elektród na báze MoS_2 prostredníctvom chemickej aktivácie

Marcel Husár[1], Peter Ondrejka[1], Michaela Sojková[2], Patrik Novák[3], Miroslav Mikolášek[1]

[1] Ústav elektroniky a fotoniky, Slovenská technická univerzita, Ilkovičova 3, 841 04 Bratislava, Slovensko
[2] Elektrotechnický ústav SAV, Dúbravská cesta 9, 84104 Bratislava, Slovensko
[3] Ústav jadrového a fyzikálneho inžinierstva, Slovenská technická univerzita v Bratislave, Ilkovičova 3, 841 04 Bratislava, Slovensko

xhusarm1@stuba.sk

Abstrakt – V tejto práci sme porovnávali dve metódy prípravy MoS_2 elektród pre superkondenzátory s cieľom zvýšiť špecifickú kapacitu a cyklickú stabilitu. Zatiaľ čo priama hydrotermálna syntéza dosiahla kapacitu 10,8 mF/cm², trojfázový postup zahŕňajúci chemickú aktiváciu a oxidáciu viedol k výrazne vyššej kapacite 71,3 mF/cm² a výbornej stabilite po 100 000 cykloch. Výsledky potvrdzujú prínos povrchovej úpravy substrátu pre optimalizáciu vlastností MoS_2 elektród.

1 Úvod

Superkondenzátory sa vďaka vysokej výkonovej hustote a výbornej cyklickej stabilite považujú za perspektívne zariadenia pre moderné systémy pre uskladnenie energie [1]. Napriek týmto výhodám však superkondenzátory stále čelia obmedzeniam v oblasti energetickej hustoty, čo motivuje výskum zameraný na vývoj nových a účinnejších elektródových materiálov. Mimoriadnu pozornosť pritom získali dvojrozmerné (2D) materiály, ktoré vynikajú vysokým špecifickým povrchom, laditeľnou elektronickou štruktúrou a mechanickou stabilitou [2,3].

Molybdén disulfid (MoS_2) patrí medzi najviac skúmané dvojrozmerné (2D) materiály. Jeho štruktúra je vrstvená, podobná grafitu, pričom jednotlivé vrstvy sú viazané slabými van der Waalsovými silami. Táto vrstvená štruktúra umožňuje iónom elektrolytu prenikať a interkalovať medzi vrstvy, čím sa zvyšuje počet dostupných aktívnych miest pre uskladnenie náboja. MoS_2 tak ukladá náboj kombináciou dvoch mechanizmov. Prvým je elektrická dvojvrstva kde ukladanie náboja prebieha nefaradaickým procesom na rozhraní elektróda-elektrolyt, formovaním elektrickej dvojvrstvy. Druhým je ukladanie náboja prostredníctvom pseudokapacity, teda faradaickými procesmi spojenými s rýchlymi a reverzibilnými redoxnými reakciami, počas ktorých dochádza k interkalácii iónov medzi vrstvy MoS_2. Kombináciou týchto mechanizmov môže MoS_2 dosiahnuť vyššiu energetickú hustotu v porovnaní s konvenčnými uhlíkovými materiálmi, ktoré využívajú prevažne uskladnenie energie v podobe elektrickej dvojvrstvy. V dôsledku toho sa MoS_2 vyznačuje aj vyššou špecifickou kapacitou [2,4].

Pre efektívne využitie týchto vlastností je kľúčové zabezpečiť kontrolovanú morfológiu elektrochemickej vrstvy. Rôzne postupy prípravy môžu významne ovplyvniť elektrochemické vlastnosti materiálu, najmä v súvislosti s aktívnou povrchovou plochou, pórovitosťou a mierou interkalácie [2].

V našej práci sme sa zamerali na optimalizáciu prípravy elektród na báze MoS_2 pre superkondenzátory. Porovnávali sme dve metódy prípravy. Prvá metóda zahŕňala priamu hydrotermálnu syntézu. Druhá, zložitejšia metóda zahŕňala chemickú aktiváciu povrchu molybdénového substrátu pomocou roztoku hydroxidu draselného (KOH), nasledovanú tepelnou oxidáciou pred samotnou hydrotermálnou syntézou.

Morfologická charakterizácia bola vykonaná pomocou skenovacej elektrónovej mikroskopie (SEM). Elektrochemické vlastnosti sme hodnotili na základe meraní špecifickej kapacity a cyklickej stability pomocou galvanostatického nabíjania a vybíjania. Získané výsledky ukazujú, že chemická aktivácia a oxidácia substrátu majú výrazný vplyv na elektrochemické správanie MoS_2 vrstiev pri ich použití v superkondenzátoroch.

2 Experimentálna časť a výsledky

2.1 Použité materiály

Molybdénová fólia (99,9 %, hrúbka 0,1 mm), tiomočovina (CH_4N_2S, ≥95 %), hydroxid draselný (KOH, ≥95 %) a izopropylalkohol (IPA, ≥99,5 %) boli zakúpené od spoločnosti Sigma-Aldrich. Počas celého experimentu bola použitá deionizovaná voda. Všetky chemikálie boli analytickej čistoty a použité bez ďalšej úpravy.

2.2 Príprava elektród

Molybdénovú fóliu sme nastrihali na substráty s rozmermi 2,5 x 1 cm. Substráty sme pred samotnou syntézou čistili v izopropylalkohole počas 10 minút a v 1M roztoku kyseliny chlorovodíkovej po dobu 20 minút, čím sme odstránili povrchovú mastnotu a prirodzene vzniknuté oxidy. Po opláchnutí deionizovanou vodou sme substráty nechali vyschnúť a použili na následnú syntézu MoS_2 vrstiev.

Elektródy sme pripravovali dvoma odlišnými metódami. Prvá metóda, označená ako HTS, spočívala v priamej hydrotermálnej syntéze. Pripravili sme roztok tiomočoviny s koncentráciou 50 g/l rozpustený v deionizovanej vode. Molybdénové substráty sme následne umiestnili do 50 ml teflónových autokláv naplnených pripraveným roztokom. Autoklávy boli utesnené a vložené do laboratórnej pece, kde prebiehala syntéza pri teplote 180 °C počas 18 hodín (Obr. 1).

Druhá metóda, označená ako KOH+OXI+HTS, pozostávala z troch krokov. Po úvodnom čistení bol substrát ponorený do 1 M roztoku hydroxidu draselného pri izbovej teplote. Tento krok viedol k čiastočnému leptaniu a povrchovej aktivácii kovu, čím sa zvýšila reaktivita povrchu. Po dôkladnom oplachu deionizovanou vodou nasledovala tepelná oxidácia pri 500 °C počas 1 hodiny, čo viedlo k tvorbe povrchových oxidov typu MoO_3. Následne bola realizovaná hydrotermálna syntéza identickým spôsobom ako pri prvej metóde, teda pri teplote 180 °C počas 18 hodín.

Po ukončení syntézy sme všetky elektródy opláchli deionizovanou vodou a nechali vysušiť na vzduchu. Výsledné elektródy tvorili molybdénové substráty, s tenkou vrstvou MoS_2 na povrchu.

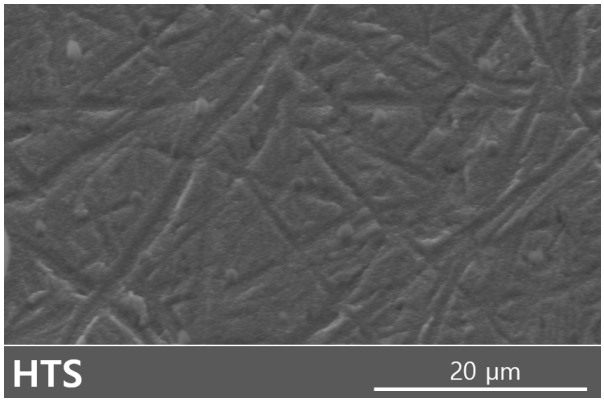


Obr. 1 - Zariadenia použité pri hydrotermálnej syntéze MoS_2 vrstiev: a) teflónové autoklávy a b) laboratórna pec.

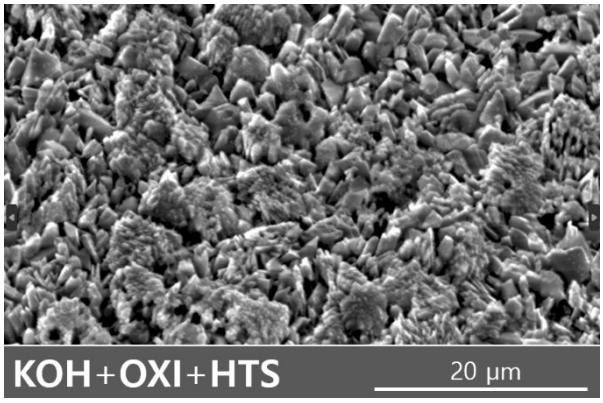
2.3 Morfologická charakterizácia

Na posúdenie vplyvu jednotlivých syntetických prístupov na výslednú morfológiu vrstiev sme použili SEM.

Povrch elektródy získanej priamou syntézou bol homogénny, bez výraznej štruktúry, s hladkým a kompaktným charakterom a minimálnym výskytom pórovitosti (Obr. 2). Naopak, povrch elektródy pripravenej pomocou trojkrokovej metódy vykazoval značnú mikro- a nanoštruktúru s nerovnomernou topografiou a vysokou pórovitosťou (Obr. 3). Takáto morfológia je z hľadiska elektrochemickej aktivity priaznivá, pretože poskytuje väčšie množstvo aktívnych miest pre interakciu s elektrolytom.



Obr. 2 - SEM snímka povrchu HTS elektródy pripravenej priamou hydrotermálnou syntézou.

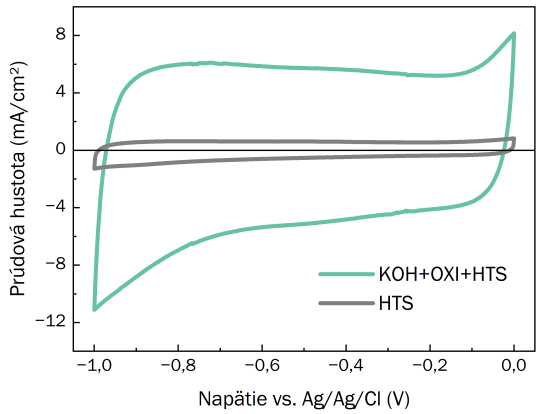


Obr. 3 - SEM snímka povrchu KOH+OXI+HTS elektródy pripravenej postupom s chemickou aktiváciou povrchu.

2.4 Elektrochemická charakterizácia

Elektrochemické vlastnosti pripravených elektród sme vyhodnotili pomocou cyklickej voltametrie (CV), galvanostatického nabíjania a vybíjania (GCD) a elektrochemickej impedančnej spektroskopie (EIS). Merania sme realizovali v trojelektródovom zapojení s použitím potenciostatu MultiEmStat4 značky PalmSens. Pracovnú elektródu tvorila pripravená MoS_2 elektróda, pričom sme na jej povrch aplikovali epoxidovú vrstvu, aby sme zabezpečili konštantnú a presne definovanú aktívnu plochu počas merania. Elektrolyt tvoril 1 M vodný roztok Na_2SO_4. Všetky merania prebiehali v napäťovom rozsahu od –1,0 V do 0,0 V, ktoré zodpovedá stabilnému pracovnému oknu MoS_2 elektród.

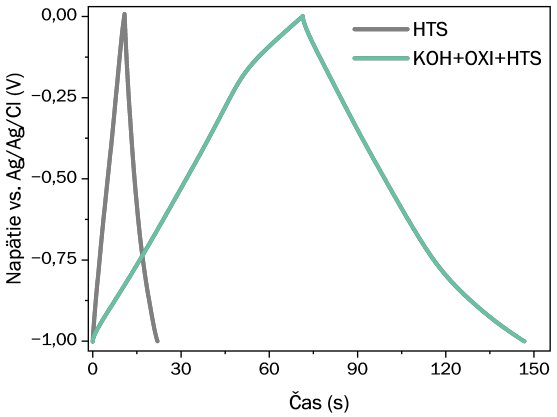
CV krivky (Obr. 4) odhalili výrazné rozdiely v elektrochemickom správaní oboch elektród. Merania boli realizované pri skenovacej rýchlosti 100 mV/s. Elektróda označená ako KOH+OXI+HTS vykazovala vyššiu intenzitu prúdovej odozvy, čo poukazuje na zvýšenú elektrochemickú aktivitu v porovnaní s HTS elektródou.



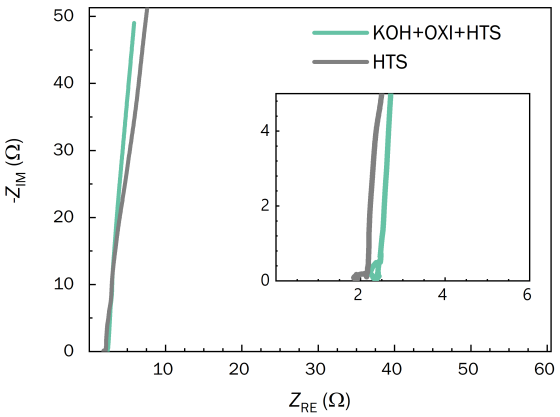
Obr. 4 - CV krivky pripravených elektród pri skenovacej rýchlosti 100 mV/s.

GCD merania sme realizovali pri prúdovej hustote 1 mA/cm^2. HTS elektróda vykazovala krátky vybíjací profil (Obr. 5), čo zodpovedá nízkej špecifickej kapacite 10,8 mF/cm^2. Naopak, elektróda KOH+OXI+HTS dosiahla výrazne dlhší vybíjací profil a zodpovedá špecifickej kapacite 71,3 mF/cm^2, čo predstavuje viac než šesťnásobné zlepšenie. Tento nárast kapacity pripisujeme zväčšenej elektrochemicky aktívnej povrchovej ploche a lepšej dostupnosti redoxne aktívnych miest. Tie vznikli v dôsledku chemickej aktivácie a následnej tepelnej oxidácie substrátu pred hydrotermálnou syntézou.

EIS merania (Obr. 6) ukázali, že elektróda KOH+OXI+HTS vykazuje vyšší sériový odpor v porovnaní s elektródou HTS. Tento nárast odporu možno pripísať komplexnejšej a poréznejšej mikroštruktúre aktívnej vrstvy, ktorá môže znižovať elektrickú vodivosť materiálu. Napriek tomu elektróda vykazuje výrazne vyššiu špecifickú kapacitu, čo poukazuje na dominantný vplyv zväčšenej aktívnej povrchovej plochy a väčšieho počtu elektrochemicky aktívnych miest. Z výsledkov vyplýva, že chemická aktivácia povrchu má zásadnejší vplyv na elektrochemický výkon ako hodnota sériového odporu.



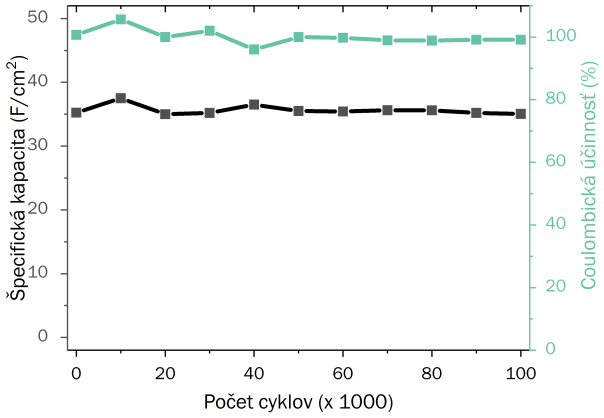
Obr. 5 - GCD krivky pripravených elektród pri prúdovej hustote 1 mA/cm^2.



Obr. 6 - EIS krivky pripravených elektród.

2.5 Cyklická stabilita

Okrem zvýšenej kapacity sme pri elektróde KOH+OXI+HTS zaznamenali aj vynikajúcu cyklickú stabilitu. GCD meranie so 100 000 cyklami potvrdilo, že elektróda si udržala 100 % svojej počiatočnej kapacity bez známok degradácie. Coulombická účinnosť sa udržiavala stabilne nad úrovňou 95 %, čo poukazuje na vysokú mieru reverzibility elektrochemických procesov prebiehajúcich v aktívnej vrstve elektródy (Obr. 7).



Obr. 7 - Závislosť špecifickej kapacity a coulombickej účinnosti od počtu cyklov pre elektródu KOH+OXI+HTS.

3 Záver

Na základe uskutočnených experimentov sme preukázali, že chemická aktivácia povrchu molybdénového substrátu pred hydrotermálnou syntézou má výrazný vplyv na výsledné štrukturálne a elektrochemické vlastnosti MoS_2 vrstiev. Porovnaním priamej hydrotermálnej syntézy so syntézou zahrňujúcou chemickú aktiváciu sme zistili, že úprava substrátu vedie k tvorbe výrazne členitej, poréznej vrstvy s rozvinutými mikro- a nanoštruktúrami.

Elektróda KOH+OXI+HTS, pripravená metódou zahŕňajúcou chemickú aktiváciu substrátu, viedla k viac než šesťnásobnému zvýšeniu špecifickej kapacity (71,3 mF/cm²) v porovnaní s priamou hydrotermálnou syntézou (10,8 mF/cm²).

Tento výsledok podčiarkuje význam aktívnej povrchovej plochy a dostupnosti redoxne aktívnych miest pre efektívne

ukladanie náboja. Okrem toho si elektróda KOH+OXI+HTS udržala 100 % svojej pôvodnej kapacity aj po 100 000 cykloch.

Tieto výsledky potvrdzujú, že MoS_2 má vysoký aplikačný potenciál pre vývoj superkondenzátorových elektród s výbornými vlastnosťami a že vhodná úprava substrátu zohráva kľúčovú úlohu v optimalizácii jeho funkčných parametrov.

Poďakovanie

Táto práca bola podporená výzvou pre doktorandov a mladých výskumných pracovníkov STU na naštartovanie výskumnej kariéry (Grant 23-03-03-B). Výzva pre doktorandov a mladých výskumníkov bola podporená z výzvy 09I03-03-V05, financovanej zo zdrojov POO SR a EU NextGenerationEU prostredníctvom Plánu obnovy a odolnosti Slovenska v rámci projektu č. 2DSuperCap - 09I05-03-V02-00037.

Literatúra

[1] PATEL, A. et al. Review on recent advancements in the role of electrolytes and electrode materials on supercapacitor performances. In *Discover Nano* [online]. 2024. Vol. 19, no. 1, s. 188. [cit. 2025-03-22]. . Dostupné na internete: <https://doi.org/10.1186/s11671-024-04053-1>

[2] GUPTA, S.P. et al. Molybdenum chalcogenides for supercapacitor applications: A critical review. In *Journal of Energy Storage* [online]. 2023. Vol. 73, s. 109065. [cit. 2025-02-18]. . Dostupné na internete: <https://www.sciencedirect.com/science/article/pii/S2352152X23024635>

[3] ÖZTÜRK, O. - GÜR, E. Layered Transition Metal Sulfides for Supercapacitor Applications. In *ChemElectroChem* [online]. 2024. Vol. 11, no. 11, s. e202300575. [cit. 2025-03-22]. . Dostupné na internete: <https://onlinelibrary.wiley.com/doi/abs/10.1002/celc.202300575>

[4] KARAMI, K. et al. Synthesis and characterization of MoS2-COOH /Co composite as electrode material for Asymmetric supercapacitors (SCs) and devices. In *Materials Science in Semiconductor Processing* [online]. 2024. Vol. 184, s. 108841. [cit. 2025-03-22]. . Dostupné na internete: <https://www.sciencedirect.com/science/article/pii/S1369800124007376>

Memristor Technology for Neural Network Hardware Synapses

Natalia Kamila Calvo[1,2], Boris Hudec[2], Peter Ballo[1]

[1]Institute of Nuclear and Physical Engineering, FEI STU in Bratislava
[2]Institute of Electrical Engineering SAS, Bratislava

xcalvo@stuba.sk

Abstract – This paper addresses the growing demand for energy-efficient, high-performance computing by exploring memristors as promising components for neuromorphic systems. We investigated the impact of the dielectric layer thickness and the top electrode material on the performance of analog-type memristors. Samples with varying HfO_2 and TiO_2 thickness were analysed, revealing that increasing the dielectric layer thickness leads to higher breakdown voltages and reduced current levels. Memristors with different top electrodes were compared in terms of I-V characteristics, breakdown voltages, and endurance. Among the tested materials, Ni exhibited the most stable memory window and favourable electrical properties, making it the optimal choice for this memristor configuration. The findings provide insight into the optimalization of memristor characteristics through material selection.

1 Introduction

Modern technologies, societal demands, and the increasing complexity of computational challenges are driving the development of modern computing. One of the greatest challenges are how to efficiently process and analyse enormous amounts of data while advancing the development of low-power autonomous decision-making systems. We aim to create intelligent systems with brain-like functionality. Because of that, artificial intelligence and neuromorphic computing are two trending topics regarding computing development. [1]

Artificial intelligence (AI) creates algorithms that are able to solve complex problems and simulates aspects of human intelligence, the purpose of neuromorphic computing is to build systems that replicate the operational principles of the brain. [2]

Artificial intelligence has advanced at an unprecedented pace in the past few years and has caused an exponential increase in demand for computing power, while at the same time Moore's law has been significantly slowing down. Moreover, most current AI systems use computer designs that keep memory and processing separate. These designs have to constantly move information back and forth between the storage and the processing unit, creating a so called von Neuman bottleneck, which significantly impacts the efficiency and performance. It requires a lot of energy, consumes time and causes a slowdown that has become more noticeable when processing big data sets. [1] [2]

The hardware and electronic components we have relied on for the last decades are approaching limitations difficult to overcome. The electronics we have been using are no longer sufficient to meet the growing demands of advanced AI and complex computational tasks.

1.1 The Impact of Memristors on the Electronic Industry

Moore's Law and the von Neumann bottleneck present two significant challenges for the future of computing development. The physical limits of transistor miniaturization and the inefficiency of separated processing and memory units indicate that we need new paradigms to continue increasing computational power and efficiency. One of the most promising solutions is the use of memristors – devices capable of emulating brain synapses.

The memristor is considered the fourth fundamental device in electrical circuits alongside the resistor, capacitor and inductor. Its existence was theoretically predicted by Leon Ong Chua in 1971, and in 2008, a team at Hewlett-Packard Labs created the first memristor. [3]

Memristors link the magnetic flux and electric charge. They are two terminal devices that can adapt their resistance based on the voltage applied to their terminals. After turning the power supply off, the memristor "remembers" past signals by retaining the resistance value that depends on the integral of the past electric current waveform.

These properties make memristors ideal for a range of applications. Memristors are crucial for neuromorphic computing systems which mimic the perception and processing mechanisms of the human brain using bio-inspired electronic designs. In this type of architecture memristors can be used as a processing unit that emulates synapses between two neurons. It incorporates a set of operations and weights that enables the simulation of neural signal propagation. [4]

Memristors also hold immense potential in fields such as in-memory computation, biologically inspired computing or single chip memory and storage integration. The use of memristors in non-volatile storage is particularly promising. [4]

Memristors are unique devices that can help us overcome problems and limitations of traditional silicon-based computing. Their nanoscale size may allow for the continued advancement in computing power as transistors reach their limits. At the same time, they have the potential to mitigate the von Neumann

bottleneck by combining memory and processing functions in a single device.

2 Types of Memristors

Memristors are a promising technology. They are capable of surpassing technological barriers while enhancing neural networks in artificial intelligence. They are typically composed of a metal – insulator – metal (MIM) stack, representing the bottom electrodes – dielectric layer – top electrodes, respectively. [5] [6]

Memristive devices have different switching mechanisms that depend on the material composition of the MIM stack. The switching mechanism governs the transition from high resistance state to low resistance state and vice versa. Memristors can be classified into many groups: conductive filament memristors, electronic effect memristors, phase change memory, interface-type memristors, etc. The resistive switching behaviour includes digital and analog types. [5]

Digital memristors show an abrupt change between the high resistant state (HRS) and low resistant state (LRS). The resistive ratio between them can reach several magnitudes making them ideal to store binary information. Digital type memristors typically require a forming process. An example of digital memristors are filamentary type memristive devices that change their resistive state with the formation and rupture of filaments. [6]

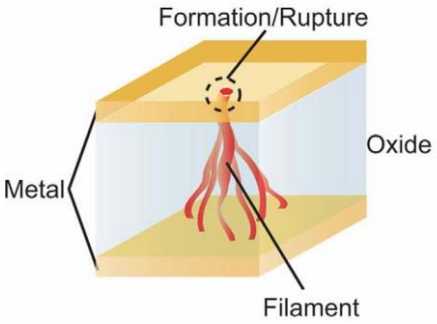


Figure 1. Filamentary type memristor [7]

Analog-type memristors, exhibit switching behaviour in the as-fabricated device without requiring a forming process. The I-V curve of the analog memristor shows a hysteresis loop under cyclic voltage sweep. They show a reproducible and gradual multi-level resistive switching under consecutive electrical stimulus. That is because the resistance state change is driven by the electrochemical migration of oxygen vacancies or charge carriers in the dielectric layer. The migration takes place across the entire area of the dielectric layer, creating an area dependence. These properties make analog-type memristors ideal for neuromorphic applications. [6]

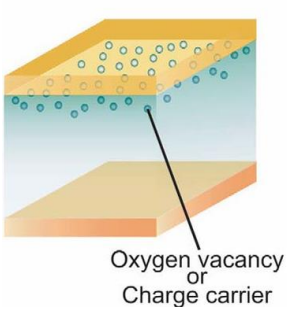


Figure 2. Analog-type memristor [7]

3 Experimental Part

The aim of this work was to investigate how different parameters of the dielectric layer and the electrodes affect the resistance states of analog-type memristive devices. We measured samples with different top electrodes and samples with varying thickness of the dielectric layer.

The bottom electrodes were deposited by physical vapour deposition (PVD) of TiN on SiO_2/Si substrate. The dielectric films consisted of a TiO_2 layer and HfO_2 layer, which were both grown with atomic layer deposition (ALD). After that a lithography process on the sample was executed. Next, the top electrodes were deposited on the sample by an evaporation method which was followed by a lift off technique. Thanks to the lithography process, we were able to fabricate around 500 devices (memristors) on a 1x1 cm^2 sample. The devices of circular (diameter *d*) and/or square (side *a*) shape varied in dimension with *d* and *a* = 200, 100, 80, 65, 50, and 40 µm.

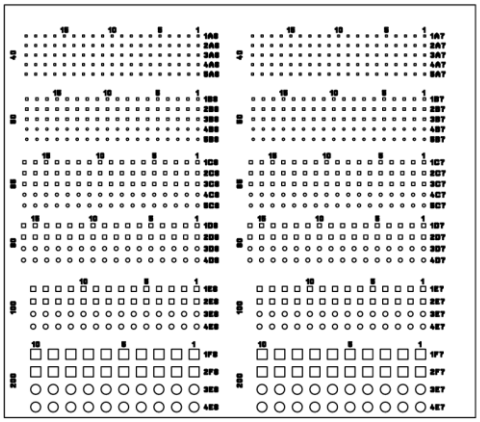


Figure 3. Schematic of a sample with many devices

The resistance measurements were conducted with a source-measure unit: Keithley 4200 SCS in the source-voltage operation mode. To measure devices of µm-size we used two needles, one for the top electrode which was biased and the other for the bottom electrode which was grounded. By applying voltage and measuring the current we obtained the I-V characteristics of the measured memristors.

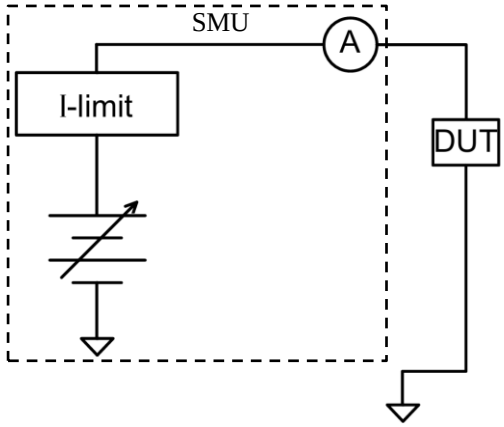


Figure 4. Measurement schematic

To change the resistance state from a high resistance state to a low resistance state (*set process*) we applied a positive voltage. To return to the high resistance state (*reset process*) we applied a negative voltage. The applied voltage caused the trapping of electrons affecting the resistance of the film.

3.1 Impact of the Dielectric Layer Thickness

The memristor structure employed in this part of the experiment was Ni/HfO_2/TiO_2/TiN. To study the influence of the dielectric layer we measured 4 samples with varying thickness of HfO_2 and 2 samples with varying thickness of TiO_2. The thickness depended on the amount of ALD cycles. The ALD technology allows a very precise thickness tuning of the deposited layers and the process is highly repeatable.

The HfO_2 layer was deposited with 35, 40, 45 and 50 cycles, which corresponded to a thickness of 2.8 nm, 3.2 nm, 3.6 nm and 4 nm, respectively. The results of the I-V characteristic measurements for samples with different thickness are in figure 6. The thickness difference of 1.2 nm led to more than 1 order of magnitude difference in the measured current.

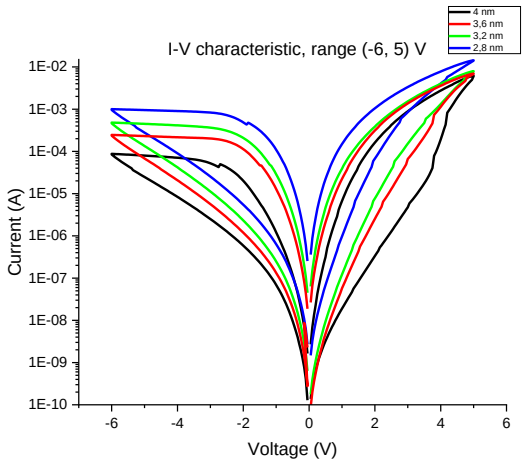


Figure 6. I-V characteristics for samples with varying dielectric layer thickness (HfO_2)

Next, we measured the breakdown voltage. Figure 7 shows resulting breakdown voltages in both polarities for respective HfO_2 thickness. As expected, the samples with a thicker HfO_2 layer had a higher breakdown voltage value.

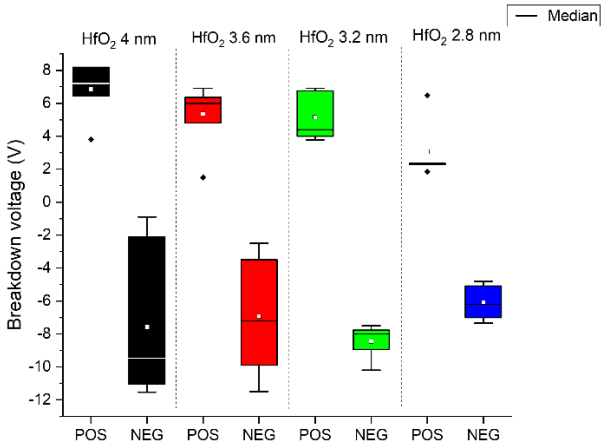


Figure 7. Breakdown voltages of samples with varying thickness of HfO_2 layer

For the measurements of the thickness influence of the TiO_2 layer we used two samples, one with 7 nm of TiO_2 and the other one with 10 nm of TiO_2. Both had a 3,2 nm HfO_2 layer. The results of the I-V characteristic measurements are in figure 8. In the sample with a thinner TiO_2 layer the current value was around 1.5 order of magnitude higher.

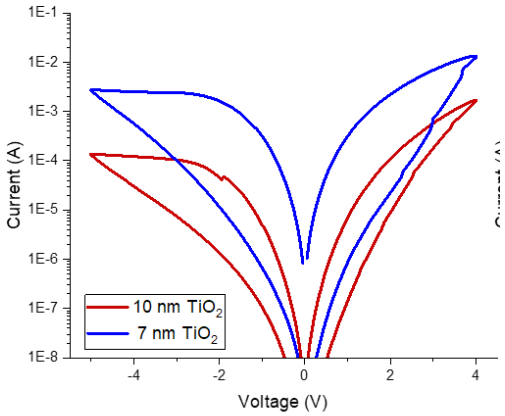


Figure 8. I-V characteristic for samples with varying dielectric layer thickness (TiO_2)

3.2 Impact of the Material of the Top Electrode

To study the impact of the top electrode material we fabricated 4 samples with different top electrodes: Ni, Al, Pt, and TiN, which differ in their work function and oxygen affinity. The comparison of the I-V characteristics is in Figure

9. We also conducted measurements of the breakdown voltage in both polarities (see Figure 10).

Another important parameter of memristive devices is their endurance. To measure the endurance, we used 5 ms pulse measurements in a sequence *set*, *read*, *reset*, *read* with applied voltage of 4/-1/-5/-1 V, respectively. The results of the *read* operations are depicted in Figures 11(a), 11(b), 11(c) and 11(d).

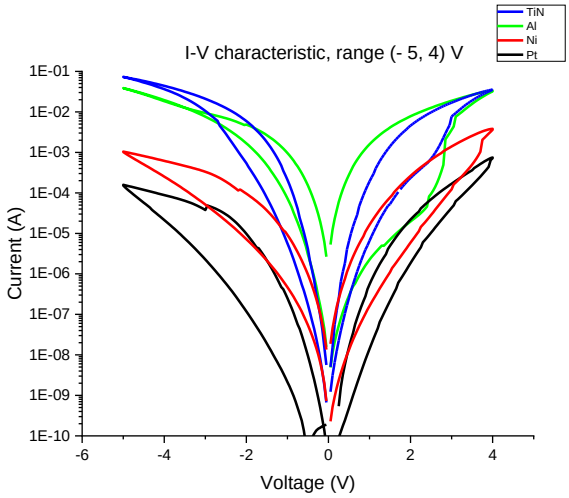


Figure 9. I-V characteristics of samples with different top electrodes

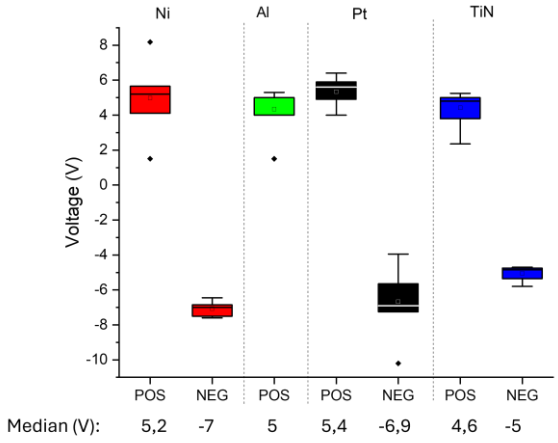


Figure 10. Breakdown voltages of samples with different top electrodes

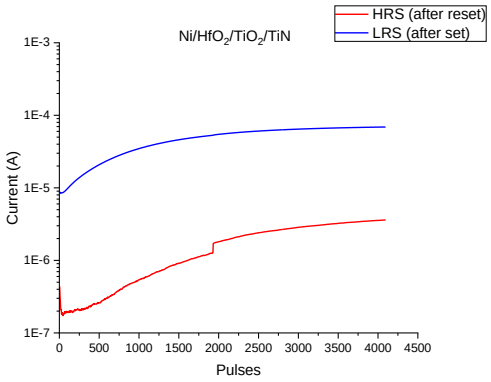


Figure 11(a). Cycling measurements for sample with Ni top electrode

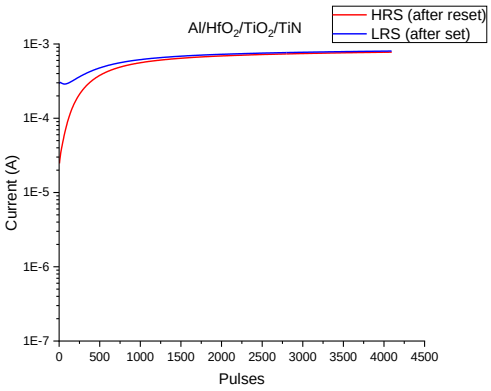


Figure 11(b). Cycling measurements for sample with Al top electrode

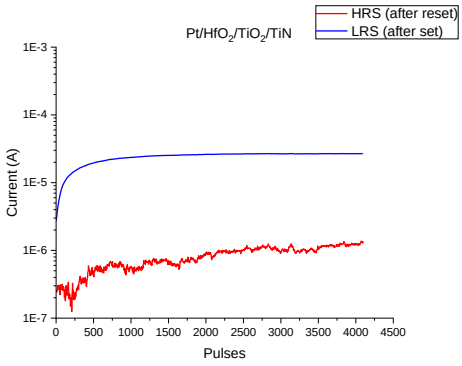


Figure 11(c). Cycling measurements for sample with Pt top electrode

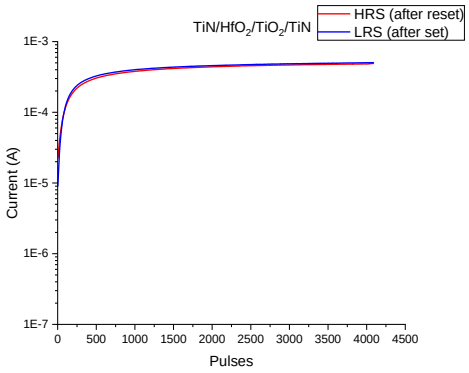


Figure 11(d). Cycling measurements for sample with TiN top electrode

Memristors with Ni top electrode (*high work function, medium-low oxygen affinity)* had slightly higher current values compared to those with Pt top electrodes, due to Ni's work function of 5,15 eV. Among the four types of memristors, these had the largest memory windows, which remained relatively stable during cycling. The average breakdown voltages were 5,2 V/ -7,0 V which are favourable values. Based on these and other previous results, we identified the Ni top electrode as ideal for this type of memristor and used it in subsequent experiments.

Memristors with an Al top electrode (*low work function, high oxygen affinity*) showed high hysteresis in the positive polarity but could not be effectively reset with negative voltage. During cycling it became particularly evident that with each additional cycle the HRS shifted significantly higher, quickly closing the memory window. This was likely caused by the gradual oxidation of the top electrode interface, which is undesirable. The average breakdown voltage was 5 V in the positive direction, but we could not measure it in the negative direction due to extremely high currents exceeding the instrument limit (100 mA). As they did not exhibit desirable behaviour, we did not investigate them further.

Memristors with a Pt top electrode (*high work function, low oxygen affinity*) displayed good characteristics with sufficient hysteresis even at low voltages. These memristors had the lowest currents, attributable to Pt's high work function (5,65 eV). This resulted in a "noisy" HRS characteristic during cycling, as they exhibited sub-microampere values, which is not entirely desirable. The breakdown voltages were 5,4 V/ -6,9 V which are considered good values.

Memristors with a TiN top electrode (*low work function, low oxygen affinity*) exhibited weak memory windows, during cycling this window became unmeasurable. The average breakdown voltages 4,6 V/ -5 V were low and the currents flowing through this memristors were relatively high.

4 Conclusion

The experimental results demonstrated that both dielectric layer thickness and top electrode material significantly influence the performance of analog-type memristors:

- Increasing the thickness of the HfO_2 layer led to higher breakdown voltages and reduced current values. Increasing the thickness of the TiO_2 layer also led to reduced current values, confirming that the dielectric layer can be tailored to optimize device behaviour.
- The choice of top electrode material also played a crucial role: Ni electrodes showed the most promising performance, offering stable memory windows, suitable breakdown voltages and consistent endurance during cycling.

This study contributes to the understanding of material selection in memristor design.

Acknowledgments

This work was supported by the project Memristor technology R&D for industry, ITMS code 09I05-03-V01-00003. I would also like to thank my colleagues Bc. Matej Horský, Mgr. Fridrich Egyenes, PhD., and my supervisor Ing. Boris Hudec, PhD. for their patience, guidance and valuable insights.

Literature

[1] MEHONIC, A. et. al. 2020. Memristors—From In-Memory Computing, Deep Learning Acceleration, and Spiking Neural Networks to the Future of Neuromorphic and Bio-Inspired Computing. In *Advanced Intelligent Systems*. 2020, vol. 2, 2000085.

[2] GILL, S. S. et al. 2024. Modern computing: Vision and challenges. In *Telematics and Informatics Reports, Elsevier*. 2024, vol 13, 100116.

[3] CHUA, L. 1971. Memristor-The missing circuit element. In *IEEE Transactions on Circuit Theory*. 1971, vol. 18(5), 507-519.

[4] VERA-TASAMA, A. et al. 2019. Memristors: A perspective and impact on the electronic industry. In *2019: Latin American Electron Devices Conference (LAEDC), IEEE*, Armenia, Colombia, 2019, pp. 1-4.

[5] CHEN, W. et al. 2023. Essential Characteristics of Memristors for Neuromorphic Computing. In *Advanced Materials.* 2023, vol 9, 2200833.

[6] TIOTTO, T. F. et al. 2022. A Compact Model of Interface-Type Memristors Linking Physical and Device Properties. In *arXiv preprint arXiv:2210.01455*. (2022).

[7] SAWA, A. 2008. Resistive switching in transition metal oxides. In *Materials Today*. 2008, vol 11(6), 1369-7021.

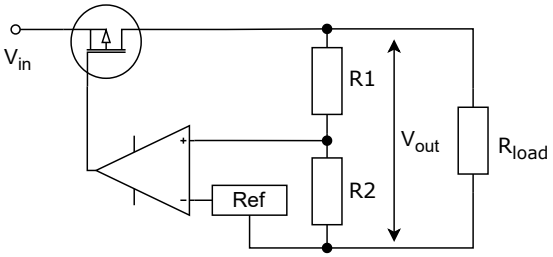
Napájací zdroj s ultra nízkym šumom a ultra nízkym únikovým prúdom

Bc. Viktor Štovčík

Ústav elektrotechniky, Fakulta elektrotechniky a informatiky, Slovenská technická univerzita v Bratislave
Ilkovičova 3, Bratislava, Slovensko

xstovcik@stuba.sk

Článok sa venuje návrhu, realizácii a testovaniu prototypu ultra nízko šumového napájacieho zdroja s ultra nízkym únikovým prúdom. Prístroj je navrhnutý primárne pre testovanie veľmi citlivých obvodov v elektrickej metrológii, ale vďaka modulárnemu dizajnu a batériovému napájaniu sa dá použiť aj v podmienkach, kde nie je možnosť využiť klasický laboratórny zdroj. Je použitá precízna elektronika, stabilná referencia a prevodníky s vysokým rozlíšením, takže nájde uplatnenie aj ako prenosný kalibrátor pre jednoduchšie aplikácie.



Obr. 1: Univerzálna schéma lineárnych stabilizátorov

1 Úvod

V procese testovania, vývoja a merania precíznej a citlivej elektroniky, ako sú napríklad vysoko kvalitné termostatované oscilátory, ultra stabilné napäťové referencie či presné analógové obvody pre meraciu techniku, je kladený vysoký dôraz na parametre napájacích zdrojov, ktoré do veľkej miery ovplyvňujú celkové výsledky merania a parametre skúmaného obvodu. Medzi nežiadúce faktory, ktoré je v prípade napájacích zdrojov potrebné eliminovať v čo najväčšom rozsahu, patrí šum a únikový prúd.

Bežne používané laboratórne zdroje sú síce cenovo dostupné a možno ich využiť na širokú škálu aplikácií, ale kvalita výstupného napätia nie je dostatočná pre náročné metrologické aplikácie. Vykazujú vysoký únikový prúd, ktorý môže ovplyvniť presnosť meraní a produkujú výrazné zvlnenie a šum na výstupe. Z tohto dôvodu sme sa rozhodli navrhnúť a zrealizovať napájací zdroj, ktorý bude možné používať pri meraniach citlivej elektroniky.

Na trhu je dostupných niekoľko typov napájacích zdrojov s veľmi kvalitným výstupom. Často sa využívajú pre napájanie laserových diód, prevodníkov s vysokým rozlíšením, napäťovo riadených oscilátorov či fázových závesov [1].

Zdroje rady UltraPS pre napájanie laserových diód so stabilným výstupným napätím od firmy Aerodiode dosahujú výstupné zvlnenie napätia 5 μV_{rms} a šumovú hustotu 500 $pV/\sqrt{Hz}$ pri všetkých hodnotách napätia. Výstupné prúdy sa pohybujú v intervale od 1 do 8 A v závislosti od modelu a hodnoty výstupného napätia [2].

Výstupný šum batériového zdroja SIM928 od firmy Stanford Research Systems dosahuje hodnoty 10 μV_{rms} pri šírke pásma 1 kHz. Výstupné napätie je však nastaviteľné v rozmedzí ±20 V a výstupný prúd dosahuje hodnôt ±10 mA [3].

Zdroje LP5393 a LP5394 od japonskej firmy NF Corporation dosahujú výstupné zvlnenie typicky 10 μV_{rms} nezávisle od zaťaženia so šírkou pásma 10 Hz - 20 MHz. LP5393 disponuje nastaviteľným výstupným napätím v rozmedzí ±12 - ±15 V a LP5394 v rozmedzí 0 - ±15 V. Oba modely dokážu do záťaže dodať ±100 mA [4].

2 Šumové vlastnosti lineárnych stabilizátorov

Lineárne stabilizátory vo všeobecnosti pozostávajú z regulačného prvku, chybového zosilňovača, referencie a snímača výstupného napätia. Typické zapojenie blokov je na obrázku 1. V prípade lineárnych stabilizátorov je regulačný prvok tvorený bipolárnymi alebo unipolárnymi tranzistormi. Hlavnými zdrojmi šumu v lineárnych stabilizátoroch napätia sú referencia, rozdielový zosilňovač. [5]

V tabuľke 1 je možné vidieť porovnanie šumových vlastností uvedených stabilizátorov v daných frekvenčných pásmach.

Prvý vybraný stabilizátor je z rady LM78XX, keďže v nenáročných aplikáciách sú zrejme najviac využívané. Zo starej generácie precíznych stabilizátorov možno spomenúť μA723 disponujúci nastaviteľným výstupným napätím. Kategória LDO stabilizátorov je zastúpená modernejším stabilizátorom LT3080, ktorý tiež disponuje nastaviteľným výstupným napätím. Zo skupiny nízkošumových stabilizátorov bol vybraný LT3045 s vy-

nikajúcimi šumovými parametrami a nastaviteľným výstupným napätím.

Ako výkonový člen zdroju bol zvolený precízny buffer LT3040, ktorý je primárne určený pre použitie v kombinácii s ADC alebo napäťovými referenciami. Topologicky je odlišný od štandardných stabilizátorov spomenutých vyššie. Okrem nízkeho šumu uvedeného v tabuľke disponuje veľmi nízkou hodnotou 1/f šumu, konkrétne 0.6 μV_{pk-pk}. Vďaka absencii vnútornej referencie šumový príspevok chybového zosilňovača buffera predstavuje 4 $nV/\sqrt{Hz}$, ktorý je v mnohých prípadoch prevážený šumom vstupného napätia. Výhodou v prípade potreby vyššieho výstupného prúdu alebo ešte nižšej hodnoty výstupného šumu je zaradenie viacerých bufferov paralelne.

Tab. 1: Šumové vlastnosti lineárnych stabilizátorov napätia (*výrobca neuvádza či ide o pk-pk alebo rms hodnotu)

Typ	Šírka pásma	Výstupné šumové napätie
LM78XX [6]	10Hz-100kHz	40 μV*
UA723 [7]	50Hz-10kHz	20 μV*
LT3080 [8]	10Hz-100kHz	40 μV_{rms}
LT3045 [9]	10Hz-100kHz	0.8 μV_{rms}
LT3040 [10]	10Hz-100kHz	1.2 μV_{rms}

3 Únikový prúd a jeho minimalizácia

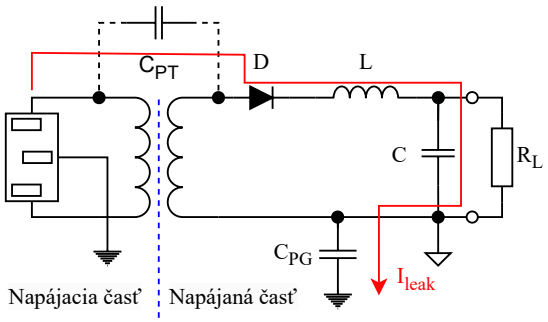
Únikový prúd vzniká v dôsledku parazitnej kapacitnej väzby medzi napájacími a napájanými obvodmi prístroja, ktoré sú znázornené na obrázku 2. V závislosti od kvality spracovania zariadenia sa rušenie môže naviazať na napájané obvody ako súhlasný mód, ktoré následne plávajú na naviazanom striedavom potenciáli. Po pripojení obvodu ku zdroju sa pre súhlasný mód vytvorí impedancia, ktorou začne zo zdroja tiecť únikový prúd [11].

Pre minimalizáciu únikového prúdu bola výkonová časť oddelená od zvyšku obvodu pomocou planárneho transformátora. Problémom pri planárnych transformátoroch býva vysoká kapacitná väzba medzi primárnou a sekundárnou stranou. Medzi primárnu a sekundárnu stranu boli pridané tieniace polygóny, ktoré je možné pripojiť na potenciál primárnej alebo sekundárnej zeme, čo zníži hodnotu kapacitnej väzby.

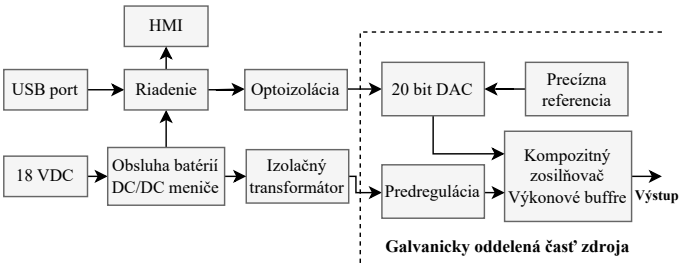
4 Technické parametre a návrh zdroja

Požadované technické parametre zariadenia sú nasledovné:

- Rozsah výstupného napätia: 0 - 15 V
- Rozsah výstupného prúdu: 0 - 500 mA, optimálne 1 A
- Modulárny dizajn
- Režim konštantného napätia (CV) a konštantného prúdu (CC)
- Napájanie zo siete pomocou adaptéra
- Napájanie z batérií
- Lokálne ovládanie pomocou užívateľského rozhrania
- Diaľkové ovládanie pomocou SCPI príkazov



Obr. 2: Smer toku únikového prúdu cez parazitné kapacity



Obr. 3: Bloková schéma zdroja

- Návrh optimalizovať pre dosiahnutie najnižšieho dostupného šumu výstupného napätia

5 Popis blokov zdroja

Základná štruktúra prototypu ultra nízko šumového zdroja s ultra nízkym únikovým prúdom je rozdelená na galvanicky neoddelenú a galvanicky oddelenú časť, ktoré je možné vidieť v blokovej schéme na obrázku 3. Galvanicky oddelená časť obsahuje bloky, ktoré sa podieľajú na generovaní výstupného napätia, či zabezpečujú meranie požadovaných parametrov. Galvanicky neoddelená časť obsahuje bloky, ktoré zabezpečujú napájanie pre jednotlivé časti zariadenia, spracovanie nameraných údajov či požiadavky ohľadom zmeny parametrov, ako je výstupné napätie a prúd.

Jednotlivé moduly budú mať rozmer 4U a bude ich možné používať samostatne alebo v rámci šasi. Šasi bude slúžiť ako nabíjacia stanica a v prípade potreby batériového napájania si moduly budú zdieľať napájacie napätie na DC zbernici. Šasi bude schopné riadiť viacero modulov naraz pomocou zabudovaného mikroprocesora a dynamického prideľovania adresy modulom na základe pozície.

5.1 Napájacia časť

Zdroj je možné napájať zo siete 18 Voltovým AC/DC adaptérom, alebo Li-ion batériovými článkami, ktoré sú súčasťou zariadenia. Zvolené boli batériové články typu Li-ion, vďaka dostupnosti množstva jednoduchých 'battery management' obvodov. Hlavným dôvodom implementácie batériového napájania

bola eliminácia vedeného rušenia zo sieťového adaptéra. Ďalším dôležitým kritériom boli šumové vlastnosti batériového zdroja, ktoré sú výrazne nižšie oproti bežne používaným sieťovým zdrojom (najmä spínaným zdrojom).

Napájacia časť bola navrhnutá tak, aby sa v prípade zapojenia zdroja do siete využíval adaptér ako hlavný zdroj energie a zároveň nabíjal batérie, ktoré v tomto prípade nedodávajú zdroju energiu. Pri odpojení adaptéru zdroj prechádza do režimu, kedy je napájaný batériami. V prípade poklesu napätia v batériách pod zadefinovanú hodnotu sa batérie odpoja, aby nedošlo ku ich poškodeniu a zdroj sa vypne. Pri vybití batérií, ktoré je nastavené na približne 15 V trvá plné nabitie približne 2:45 - 3 hodiny.

5.2 DC/DC meniče

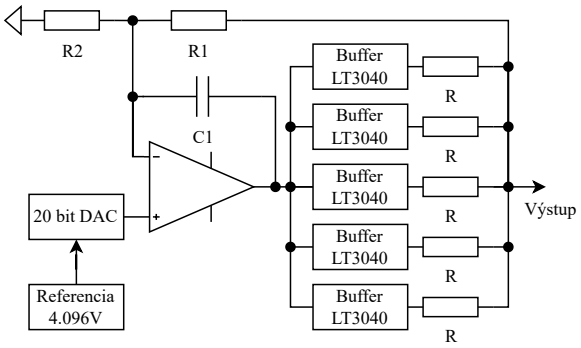
V galvanicky neoddelenej časti je zaradený znižujúci DC/DC menič, ktorý transformuje vstupné napätie na +5 V, potrebné pre napájanie mikroprocesora, prevodníka USB - UART, optoizolátorov, posuvných registrov slúžiacich na ovládanie relé ako aj samotné relé, podsvietenie tlačidiel na prednom paneli a LCD displej. Pri návrhu bol odhadovaný maximálny odber 500 mA vrátane rezervy. Vybraný bol menič s maximálnym výstupným prúdom 2 A, čo pre danú aplikáciu predstavuje nadmernú výkonovú rezervu.

Pre napájanie galvanicky oddelenej časti bolo nutné použiť DC/DC menič a transformátor. Po prieskume trhu bolo zistené, že nízko šumové meniče s PWM reguláciou výstupného napätia, integrovanými mosfetmi a minimálnym počtom externých súčiastok majú spínače dimenzované nanajvýš na 1 A. Zámerom použitia zvoleného meniča bolo, aby pri spodnej napäťovej hranici batérií predstavujúcej 15 V menič pracoval s maximálnou striedou - 50% a so zvyšujúcim sa vstupným napätím sa striedа znižovala. Aj napriek vhodne vypočítaným hodnotám nastavovacích prvkov táto funkcionalita nefungovala. Vyskúšaných bolo niekoľko kombinácií nastavovacích prvkov, avšak bez úspechu. Keďže obvodové prvky nebolo možné meniť, tak napájacie napätie muselo byť znížené z pôvodných 24 V na 18 V.

Oscilácie na nábežných a dobežných hranách spínaného napätia sa prejavili aj na galvanicky oddelenej strane, preto bol ku každému spínaču zaradený RC snubber. Aj napriek zaužívaným postupom výpočtov výsledné hodnoty nebolo možné použiť. Pri použití vypočítaných hodnôt sa oscilácie značne zmenšili, ale menič sa opakovane prehrieval a nefungoval správne. Po zvolení nižších hodnôt kondenzátora sa oscilácie zväčšili, ale menič už pracoval.

Sekundárne výkonové vinutia napájajú buffre vo výkonovom stupni. Pre minimalizáciu stratového výkonu na výkonových bufferoch je každé vinutie privedené cez relé na usmerňovač v závislosti od nastavovaného napätia. Pomocou relé je následne možné pre požadovanú hodnotu napätia nastaviť adekvátne napájacie napätie. Sekundárne napájacie vinutia slúžia na napájanie ostatných obvodov v galvanicky oddelenej časti, ako sú ADC, DAC, nízko šumová referencia, posuvné registre slúžiace na zmenu prúdového obmedzenia a operačných zosilňovačov.

Transformátor je uzavretý EILP jadrom zo štandardného feritového materiálu pre transformátory N87 pri použití spínacej frekvencie meniča v rozmedzí od 25 kHz do 500 kHz, keďže menič pracuje so spínacou frekvenciou 300 kHz [12].



Obr. 4: Bloková schéma výkonového stupňa

5.3 Výkonová časť

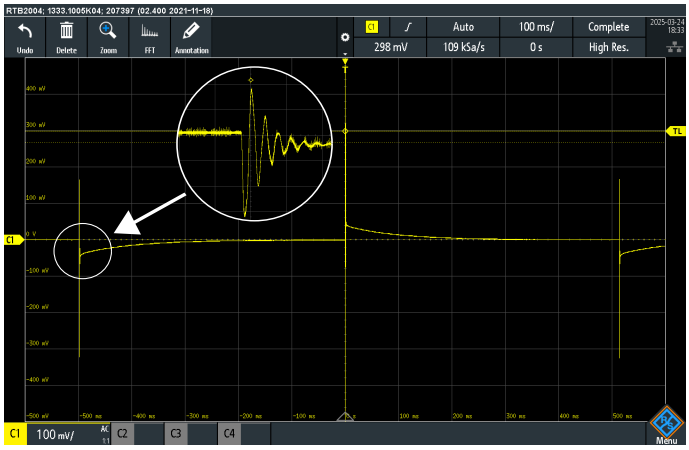
Na obrázku 4 je možné vidieť základnú topológiu výkonového stupňa. Výstupné napätie je generované 20-bitovým DAC, ktorý využíva nízkošumovú referenciu. Generované napätie z DAC je následne privedené do kompozitného zosilňovača, ktorý ho upraví do podoby výstupného napätia.

Keďže DAC nedokáže generovať výstupné napätie v plnom rozsahu a požadovaný prúd, bolo potrebné použiť buffre, ktoré zabezpečia dostatočný prúd pre záťaž. Po zaradení bufferov však vzniká problém s presnosťou a zosilnením signálu. Na vyriešenie tohto problému bolo zvolené zapojenie kompozitného zosilňovača, ktoré umožňuje dosiahnuť kombináciu vysokej presnosti, požadovaného rozsahu výstupného napätia a dostatočného výstupného prúdu.

Kompozitný zosilňovač je tvorený nízko šumovým a nízko offsetovým operačným zosilňovačom a piatimi paralelne zapojenými ultra nízko šumovými buffermi. Zapojenie piatich bufferov paralelne bolo realizované za účelom zvýšenia výstupného prúdu modulu a dodatočnej minimalizácie šumu. Zosilnenie kompozitného zosilňovača je nastavené rezistormi R1 a R2. V blokovej schéme sú prezentované jedným symbolom, avšak v realite sú realizované ako sériovo-paralelne siete rezistorov rovnakých hodnôt pre kompenzáciu driftu. Kondenzátor C1 slúži na stabilizáciu kompozitného zosilňovača, bez ktorého by celý stupeň osciloval.

Nastaviteľné prúdové obmedzenie je realizované prepínaním rezistorov pri jednotlivých bufferoch. Keďže spojité ovládanie obmedzenia v tejto aplikácii nebolo potrebné, prúdové obmedzenie je nastaviteľné v intervaloch 250 mA, 500 mA, 750 mA, 1 A.

Na výstup je zaradený diferenciálny dolnopriepustný filter 4. rádu pre minimalizáciu vysokofrekvenčného rušenia pochádzajúceho zo spínaného transformátora.



Obr. 5: Odozva zdroja s impulznou záťažou. $R_1 = 33.8\ \Omega$, $R_2 = 199\ \Omega$, $t_1 = t_2 = 500$ ms. Detail: $R_1 = 33.8\ \Omega$, $R_2 = 199\ \Omega$, $t_1 = t_2 = 100$ ms, 500 μs/div, 100 mV/div.

5.4 Riadiaca časť

Úlohou riadiacej časti je zabezpečenie nasledujúcich procesov:

- komunikácia s nadradeným systémom - PC, šasi
- obsluhovať prepínanie napäťových odbočiek pomocou relé zabezpečujúcich predreguláciu výkonových členov
- obsluhovať prepínanie stavu výstupného relé
- zasielanie kódu nastaveného napätia do DAC
- nastavenie rozsahu prúdového obmedzenia
- spracovanie nameraných dát z ADC
- meranie a spracovanie dát ohľadom nabíjania batérií
- stav nabitia batérií
- výpis nastaveného napätia a prúdového obmedzenie alebo meraného výstupného napätia a prúdu na LCD displej

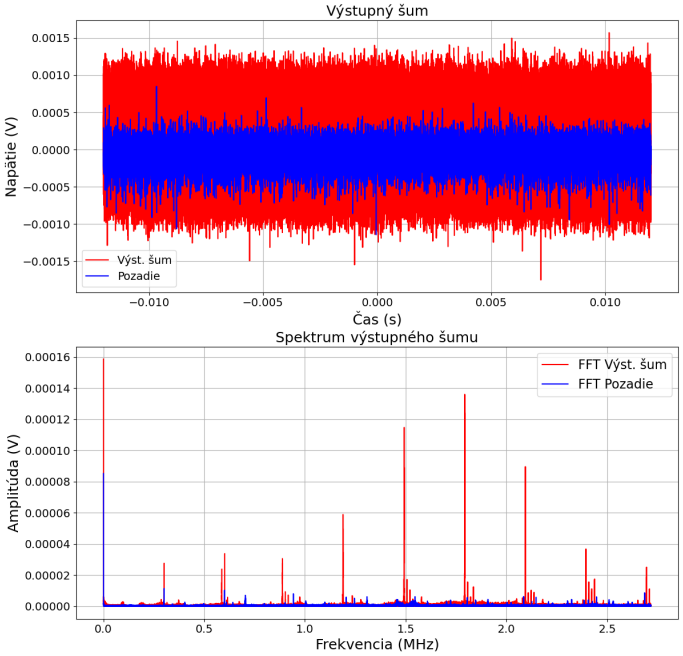
Ako hlavná riadiaca jednotka bol zvolený mikroprocesor ATmega 2560 na základe dostatočného počtu GPIO pinov a prítomnosti viacerých hardvérových sériových liniek.

Komunikácia medzi prevodníkmi a posuvnými registrami v rámci modulu je realizovaná prostredníctvom SPI. Rozhranie medzi galvanicky izolovanou a galvanicky neizolovanou časťou je vytvorené pomocou optoizolátorov. Sériová linka 0 je využívaná na programovanie a komunikáciu cez USB konektor na zadnej strane modulu a sériová linka 2 je vyvedená na backplane konektor a jej úlohou je komunikácia so šasi.

6 Dosiahnuté parametre

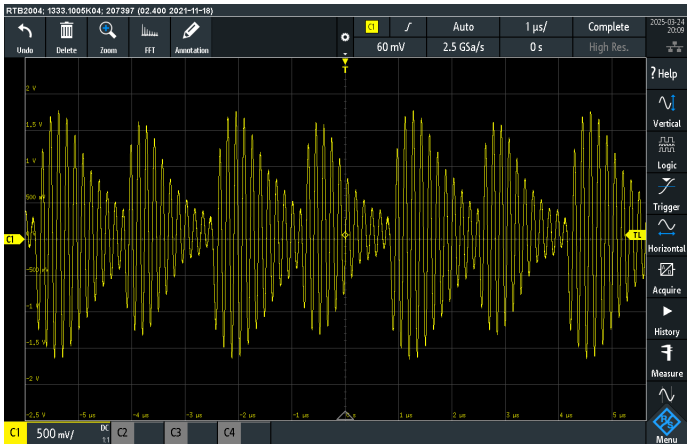
Pri meraní impulznej odozvy bola použitá elektronická záťaž v impulznom režime. Pri meraní impulznej odozvy zdroj pracoval v režime konštantného napätia. Na obrázku 5 sú zreteľne viditeľné napäťové skoky pri jednotlivých impulzoch. Ďalej sú viditeľné detailným priebehom prítomných prekmitov. Čas ustálenia je ovplyvnený stabilizačným kondenzátorom C1 v kompozitnom zosilňovači (obrázok 4) a výstupným filtrom.

Nasledujúce meranie pozostávalo z merania šumu pri maximálnej záťaži, ktoré je možné vidieť na obrázku 6.



Obr. 6: Tvar šumového pozadia a výstupného šumu v časovej a frekvenčnej oblasti

Na zdroji bolo nastavených 15 V s prúdovým obmedzením 1 A. Na výstup bola pripojená rezistívna záťaž s odporom 9.8 Ω. Prvým krokom bolo zmeranie šumového pozadia (modrý priebeh) - osciloskop a záťaž pripojené ku zdroju s vypnutým výstupom. Nasledovalo meranie výstupného šumu - zapnutý výstup s pripojenou záťažou a osciloskopom (červený priebeh). Pri meraní výstupného šumu bol zdroj pracujúci v režime konštantného prúdu. Po vypočítaní FFT je vidieť frekvenčné zložky prítomného vysokofrekvenčného rušenia. Určená RMS hodnota výstupného šumu pri maximálnom výkone je 437.353 μV_{rms}.



Obr. 7: Únikový prúd pri napájaní z batérií. 500 nA/div, 1μs/div.

Meranie únikového prúdu bolo realizované pri niekoľkých napájacích zdrojoch: sieťový adaptér, napájanie z laboratórneho zdroja a batérií. Ako je zrejmé z obrázku 7, pri batériovom napájaní tvar únikového prúdu obsahuje vysokofrekvenčné

zložky spôsobené DC/DC meničom. Určená hodnota únikového prúdu pri batériovom napájaní je 3.2 μA_{pk-pk}.

7 Záver

Bolo zhotovené zariadenie s rozsahom jednosmerného výstupného napätia 0 - 15 V a výstupným prúdom 0 - 1 A. Zariadenie je možné napájať viacerými spôsobmi (AC/DC adaptér, laboratórny zdroj, zabudované batérie), má modulárny dizajn, podporuje lokálne a vzdialené ovládanie.

Únikový prúd pri napájaní z batérie dosahuje hodnotu 3.2 μA_{pk-pk} a výstupný šum pri maximálnom výkone 437.353 μV_{rms}.

Kritickou časťou, ktorá bude vyžadovať úpravu, je DC/DC menič pri transformátore. Dôvodom je generácia vysokofrekvenčného rušenia prítomného v galvanicky oddelenej časti, ktoré sa nepriaznivo prejavuje na výstupnom napätí aj na únikovom prúde.

Z používateľského hľadiska by bolo vhodné upraviť napájaciu sekciu, aby pracovala na napájacom napätí 24 V namiesto 18 V, keďže 24 V adaptéry sú omnoho bežnejšie. Zároveň voľba kvalitného napájacieho adaptéra pozitívne ovplyvní parametre výstupného napätia a únikového prúdu pri napájaní zo siete.

Literatúra

[1] C. Felton, "Achieving Ultra-Low Noise Power," Analog Devices, Technical Article, Apr. 2018. [Online]. Available: https://www.analog.com/en/resources/technical-articles/achieving-ultra-low-noise-power.html. [Accessed: 16-Mar-2025].

[2] Aerodiode, "PN A298 Ultra-Low Noise Linear Power Supply," Datasheet, 2024. [Online]. Available: https://www.aerodiode.com/wp-content/uploads/2024/11/PN_A298-Ultra-low-noise-linear-power-supply.pdf. [Accessed: 16-Mar-2025].

[3] Stanford Research Systems, "SIM928 Isolated Voltage Source," Datasheet, [Online]. Available: https://www.thinksrs.com/downloads/pdfs/catalog/SIM928c.pdf. [Accessed: 16-Mar-2025].

[4] NF Corporation, "LP Series - Low Noise DC Power Supply," Datasheet, [Online]. Available: https://www.nfcorp.co.jp/english/pro/mi/loc/dc_power_supply/lp/dl/en_spec_lp_series.pdf. [Accessed: 16-Mar-2025].

[5] B. Shaffer, "Power supply sequencing," Texas Instruments, Application Report SLYT201, Jul. 2004. [Online]. Available: https://www.ti.com/lit/an/slyt201/slyt201.pdf. [Accessed: 16-Mar-2025].

[6] Texas Instruments, "LM340, LM78XX Series 3-Terminal Positive Regulators," Datasheet, Jan. 2018. [Online]. Available: https://www.ti.com/lit/ds/symlink/lm340.pdf. [Accessed: 16-Mar-2025].

[7] Texas Instruments, "UA723 Precision Voltage Regulator," Datasheet, Jan. 2018. [Online]. Available: https://www.ti.com/lit/ds/symlink/ua723.pdf. [Accessed: 16-Mar-2025].

[8] Analog Devices Inc., "LT3080 - 1.1A Single Resistor Low Dropout Regulator," Datasheet, Apr. 2009. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/3080fc.pdf. [Accessed: 16-Mar-2025].

[9] Analog Devices Inc., "LT3045 - Ultralow Noise, Ultrahigh PSRR RF Linear Regulator," Datasheet, Aug. 2016. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/lt3045.pdf. [Accessed: 16-Mar-2025].

[10] Analog Devices Inc., "LT3040 - Ultralow Noise, Ultrahigh PSRR RF Linear Regulator," Datasheet, Jan. 2016. [Online]. Available: https://www.analog.com/media/en/technical-documentation/data-sheets/lt3040.pdf. [Accessed: 16-Mar-2025].

[11] D. Hrušovský, "Etalón napätia 10 V," Diplomová práca, 2024. [Online]. Available: https://opac.crzp.sk/?fn=detailBiblioFormChildMIH7N&sid=AA0D9432EF6DE826CF161B4F11D3&seo=CRZP-detail-kniha. [Accessed: 18-Mar-2025].

[12] TDK Electronics, "Ferrite materials,"TDK Electronics. [Online]. Available: https://www.tdk-electronics.tdk.com/en/529404/products/product-catalog/ferrites-and-accessories/ferrite-materials. [Accessed: Mar. 16, 2025].

[13] Texas Instruments, "Designing for Low-Noise Operation in High-Speed Data Converters," Application Report SNVA559C, Nov. 2014. [Online]. Available: https://www.ti.com/lit/an/snva559c/snva559c.pdf. [Accessed: 16-Mar-2025].

Sekcia 5

Jadrové a fyzikálne inžinierstvo

Feasibility of incorporating a DT neutron generator in a recently constructed laboratory

Matej Kleštinec[1], Branislav Vrban[1]

[1] Slovak University of Technology in Bratislava, Faculty of Electrical Engineering and Information Technology, Institute of Nuclear and Physical Engineering, Ilkovičova 3, 841 04 Bratislava, Slovakia

xklestinec@stuba.sk

Abstract – This paper explores the potential implementation of a DT neutron generator at the newly constructed laboratory in the B block of FEI STU. Utilizing a developed simulation model, the study investigates the radiation risks to the environment surrounding the controlled area in the context of the DT neutron generator, which is characterized by an isotropic distribution of high energy neutrons.

1 Introduction

Neutron generators are among the most prevalent sources of neutrons, serving as a substitute for research reactors. The recently constructed laboratory at the B block of the FEI STU will be equipped with a DD neutron generator and an X-ray tube, thereby expanding the research capabilities of the faculty. The implementation of an additional neutron generator employing the DT fusion reaction would further expand the laboratory's research and experimental capabilities by allowing for the utilization of neutrons with different energies, specifically 2.5 MeV for the DD neutron generator and 14 MeV for the DT neutron generator. However, the use of higher energy neutrons introduces a radiation risk due to their increased penetrability, which has the potential to elevate the radiation risk to workers and the public.

To investigate the radiation situation in and around the laboratory, a detailed 3D model was created in the SCALE 6.2 simulation package developed and maintained at Oak Ridge National Laboratories (ORNL) [1]. The necessity of such an investigation stems from the need to ensure the safety of laboratory personnel and the general public, who may move freely in the laboratory's vicinity.

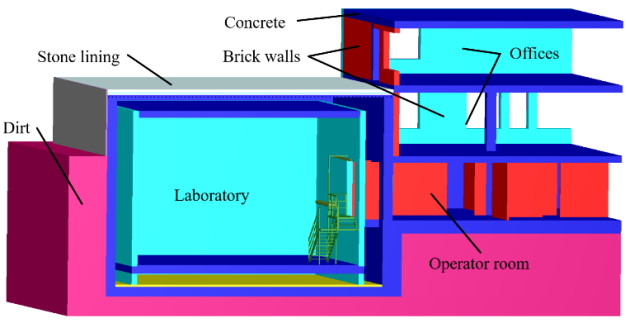


FIGURE. 1. Lateral view of the laboratory model with description of adjacent structures

A lateral perspective of the laboratory is illustrated in Fig. 1, accompanied by a detailed description of the neighboring structures. The laboratory's design employs a cube-in-a-cube configuration, offering inherent natural radiation shielding due to the substantial dimensions of the room, coupled with its partial subterranean positioning.

During the operation of the DT neutron generator, the generator will be at least partially shielded. Water is the optimal shielding material due to its neutron shielding properties, its availability, its low cost and its versatility. There are stackable water containers available that would enable us to shield the generator according to our needs. However, the utilization of water-based shielding could potentially pose a radiation concern in the form of tritium generation and accumulation.

2 Methods and techniques

2.1 Radiation doses

In the previous studies [2], the operator's room has been identified as the location with the highest radiological risks. To mitigate this risk, the room has been equipped with additional shielding barriers in the form of a 23 cm thick polyethylene moveable gate with a 3 mm lead lining. This shielding is sufficient for the case of a DD neutron generator, however, additional new simulations are necessary for the case of DT generated high-energy neutrons. The new investigation of the radiation situation also considered the generation of secondary gamma radiation, which is produced when neutrons pass through the structural and shielding materials of the laboratory. The new simulations were performed under the assumption of an isotropic neutron flux density distribution and a uniform neutron energy of 14 MeV. The employed simulation program is of a stochastic nature, operating based on the Monte Carlo method. The simulations utilized the FW-CADIS (Forward Weighted Consistent Adjoint Driven Importance Sampling) variance reduction method [3].

The DT neutron generator that was examined is the MP 320 Neutron Generator from Thermo Scientific [4]. This particular neutron generator produces neutrons through a DT fusion reaction, as outlined in equation (1), with neutron energy of 14 MeV, and a yield of 1E8 n/s. The accelerating voltage is 90 kV, and the operating current equals to 60 μA.

$$ {}_{1}^{2}D + {}_{1}^{3}T \rightarrow {}_{2}^{4}He + {}_{0}^{1}n + 17{,}586\ MeV \qquad (1) $$

The investigation is focused on the case of isotropic neutron flux density distribution and operation of just one neutron generator. In the simulations, the braking γ radiation produced by the generator operation was not considered, as previous simulations have demonstrated that its contribution to the total radiation load is negligible when compared to the contribution from the neutrons.

We assessed the ambient dose equivalent in office areas located on the east side of the laboratory, in parks situated on the north and south sides, in garages on the west side, and on the roof of the laboratory. The detectors were always positioned 120 cm above the floor and 50 cm from the exterior walls of the laboratory. In order to conservatively assess the radiation risks, the calculations were conducted without any additional shielding of the neutron generator.

Fig. 2 and Fig. 3 illustrate the locations of the individual detectors employed in the investigation of the radiation situation for the offices located to the east of the laboratory.

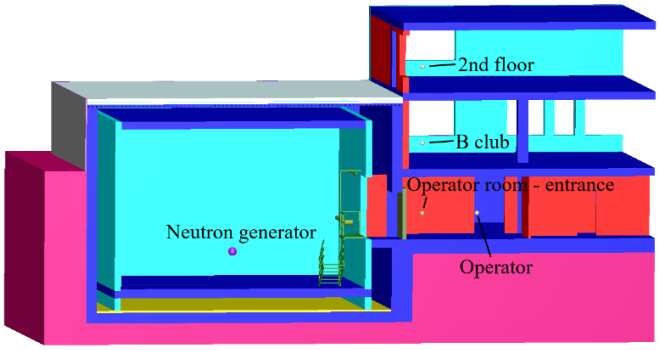


FIGURE. 2. Lateral view of the laboratory illustrating the detector positions for the east direction

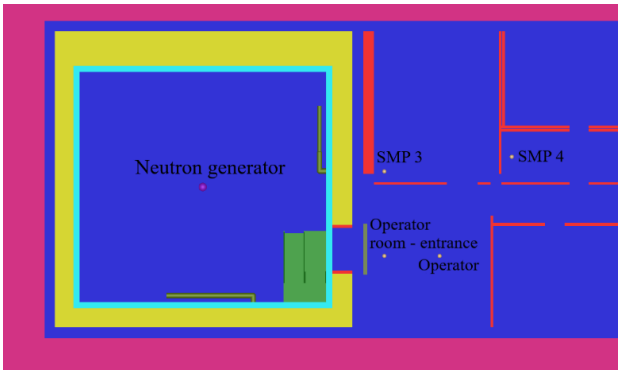


FIGURE. 3. Laboratory floor plan illustrating the detector positions for the east direction

2.2 Neutron activation

The utilization of a neutron generator also poses a radiation risk in the form of potential activation of the components of the neutron generator, the surrounding air, and adjacent construction and shielding materials. In light of these concerns, a thorough radiation risk analysis was conducted to assess the potential for neutron activation. The primary radiation concern pertains to the potential formation of tritium from neutron activation within the assumed shielding material. Existing literature from IAEA-related publications indicates that the emission of tritium from bare metal hydride is negligible, therefore we can exclude the construction materials of the neutron generator from the analysis. The focal point of the presented analysis is the potential for the activation of shielding materials, which are primarily water-based.

The analysis was conducted under the assumption that the neutron flux density is constant at the distance of 0.5 m from the source, with a value of 3.18E7 n/s/m^2. Two approaches were taken into consideration: one where all produced neutrons are in the fast spectrum with an energy of 14 MeV and the second where all produced neutrons are imaginary slowed down into the thermal spectrum with energy of 0.0253 eV.

The primary concern is related to the production of tritium in the proposed water-based shielding closely surrounding the neutron generator due to the potentially harmful effects of tritium on living organisms [5]. This assertion has been substantiated through studies conducted on mice that were exposed to regulated doses of tritiated water over an extended period. These studies have indicated that the ingestion of water containing elevated levels of tritium may potentially elevate the risk of cancer development. The study found that mice ingesting tritiated water with activity below 1.4E8 Bq/l, equal to 3.6 mGy/day, differed only little from control mice. However, increasing the dose of tritium resulted in a higher probability of carcinogenesis.

Tritium can enter the human body in several ways: inhalation, absorption through the skin, and ingestion. Tritium is likely to be produced by a neutron generator shielded by water, which poses a risk to personnel and students of the faculty. Therefore, it is imperative to accurately analyze the potential production of tritium via neutron interactions to mitigate the associated risks.

3 Results

3.1 Radiation doses

The results for neutron ambient dose rate equivalents presented in this paper have a relative uncertainty below 15 % for all the displayed values.

3.1.1 East direction – offices + operator room

The results for the east direction are presented in Table 1. The highest ambient dose equivalent rate is found behind the entrance to the laboratory. In order to be in compliance with our set limit of equivalent dose of 6 mSv/year allowing the simpler operational procedures, the maximum yearly operation time needs to be at maximum 1500 hours. In areas accessible by the public and faculty employees, the ambient dose equivalent rate is less than 50 nSv/hour, equivalent to approximately half of the natural radiation dose, and thus poses little to no radiation risk.

Table 1. Results for the east direction

Position	Ambient Dose – Neutron [μSv/h]	Ambient Dose – Photons [μSv/h]
Operator room – entrance	3.6	0.1

B club	0.0	0.0
2nd floor	0.0	0.0
Operator	0.7	0.0
SMP 3	0.1	0.0
SMP 4	0.0	0.0

3.1.2 South direction – park

The results for the south direction are presented in Table 2. The calculated ambient dose equivalent rates for the southern park are below the natural radiation dose at less than 40 nSv/hour and thus poses little to no radiation risk to the public.

Table 2. Results for the south direction

Position	Ambient Dose – Neutron [nSv/h]	Ambient Dose – Photons [nSv/h]
-750	4.37	0.12
-600	13.04	0.40
-450	24.10	0.61
-300	29.18	0.89
-150	34.85	1.70
0	35.59	1.56
150	33.53	1.34
300	28.23	1.05
450	21.26	0.61
600	12.55	0.46
750	3.78	0.08

3.1.3 West direction – garages

The results for the west direction are presented in Table 3. The calculated ambient dose equivalent rates are well below the natural radiation dose, which is primarily ensured by a 2 m thick layer of dirt between the laboratory wall and the adjacent garages which serves as natural shielding. Thus, the operation of the neutron generator poses little to no radiation risk.

Table 3. Results for the west direction

Position	Ambient Dose – Neutron [nSv/h]	Ambient Dose – Photons [nSv/h]
Southwest – high	0.08	0.00
Southwest – middle	0.01	0.00
Southwest – low	0.00	0.00
West – high	0.30	0.08
West – middle	0.05	0.00
West – low	0.03	0.00
Northwest – high	0.10	0.00
Northwest – middle	0.01	0.00
Northwest – low	0.01	0.00

3.1.4 North direction – park

The results for the north direction are presented in Table 4. The calculated ambient dose equivalent rate for the northern park are below the natural radiation dose and thus poses little to no radiation risk to the general public.

Table 4. Results for the north direction

Position	Ambient Dose – Neutron [nSv/h]	Ambient Dose – Photons [nSv/h]
-750	5.08	1.15
-600	22.02	57.75
-450	23.97	18.93
-300	32.95	1.99
-150	36.87	1.67
0	37.71	1.76
150	38.67	2.51
300	30.37	1.12
450	23.80	0.87
600	13.63	1.33
750	3.69	0.29

3.1.5 Roof

The results for the roof of the laboratory are presented in Table 5. The calculated ambient dose equivalent rate is elevated right above the neutron generator, however, considering an operational time of 1500 hours, we would not exceed the permissible limit for the general public of 1 mSv/year. Additionally, the roof is inaccessible to the public and will be monitored by security cameras to ensure that no one is present during neutron generator operation. The area with the highest calculated ambient dose is located directly above the neutron generator. Nevertheless, as access to the roof is limited, it does not pose a significant radiation risk.

Table 5. Results for the roof

Position	Ambient Dose – Neutron [nSv/h]	Ambient Dose – Photons [nSv/h]
Northwest	32.02	5.85
North	189.47	19.80
Northeast	35.68	3.95

West	115.50	36.27
Middle	480.12	52.16
East	138.27	39.35
Southwest	28.18	3.28
South	114.03	11.19
Southeast	34.99	3.25

3.2 Neutron activation

In order to assess the risk of potential neutron activation, it was assumed that all neutrons possessed a single energy, either 14 MeV or 0.0253 eV. Additionally, it was postulated that the neutron flux density would be uniform, with a value of 3.18E7 n/s/m^2. The analyzed volume was 10 m^3 of demineralized water. The possible neutron interactions are illustrated in Table 6, accompanied by the corresponding neutron cross sections for thermal and fast neutrons. The neutron cross sections used in this analysis were obtained from the NEA JANIS database [6] and the ENDF/B-VIII.0 [6] and JEFF-3.3 [6] nuclear data libraries.

Table 6. Possible neutron interactions

Interaction	Thermal neutron cross section [m^2]	Fast neutron cross section [m^2]
^{1}H (n, γ) ^{2}H	3.33E-29	2.96E-33
^{2}H (n, 2n) ^{1}H	0	1.66E-29
^{2}H (n, γ) ^{3}H	5.06E-32	9.50E-34
^{16}O (n, γ) ^{17}O	1.70E-32	2.83E-33
^{16}O (n, p) ^{16}N	0	4.21E-30
^{16}O (n, d) ^{15}N	0	1.69E-30
^{16}O (n, α) ^{13}C	0	1.42E-29
^{16}O (n, 2α) ^{9}Be	0	6.05E-42
^{17}O (n, γ) ^{18}O	3.83E-31	2.08E-32
^{17}O (n, p) ^{17}N	0	4.02E-30
^{17}O (n, d) ^{16}N	0	1.83E-30
^{17}O (n, t) ^{15}N	0	5.16E-31
^{17}O (n, α) ^{14}C	2.35E-29	2.62E-29
^{18}O (n, n + α) ^{14}C	0	1.93E-30
^{18}O (n, γ) ^{19}O	1.54E-32	1.10E-31
^{18}O (n, p) ^{18}N	0	1.44E-32
^{18}O (n, α) 16C	0	2.15E-30

It is possible to calculate the exact number of specific neutron interactions that occur during any given operational time of the neutron generator, given the exact volume, isotopic composition, and neutron cross sections. For this analysis, the assumed conservative operational time is 2000 continuous hours. The final activity of unstable isotopes was calculated via a known neutron activation equation (2), where N represents the number of target atoms, σ is the neutron cross section, Φ is the neutron flux, t_a is the activation time, and t_w is the time between the end of activation and the start of the count, and λ is the decay constant for unstable isotopes.

$$A = N\sigma\phi\left(1 - e^{-\lambda t_a}\right)e^{-\lambda t_w} \quad (2)$$

Utilizing equation (2), we calculated that for thermal neutron interactions, the total generated activity is 30.1 Bq/m^3, or 6.6 Bq/m^3 after 60 s waiting time a for fast neutron interactions, the total generated activity is 6610 Bq/m^3, or 50.3 Bq/m^3 after 60 s waiting time.

However, this method is only applicable to unstable isotopes and does not provide a definitive quantity of tritium atoms produced during the operation of the neutron generator. To comprehensively account for tritium production, it is necessary to calculate the precise amount of tritium atoms generated within a specified operational time. The first approach was to conservatively account for all possible tritium creation by disregarding the time dependency of the parent isotopes by only accounting for the initial natural isotope amount in our given volume as well as disregarding the diminishment over time due to radioactive decay of tritium and diminishment due to other tritium-neutron interactions. To this end, we formulated equation (3), in which N is the reaction product quantity, N_p is the parent isotope quantity at the start, σ is the neutron cross section, Φ is the neutron flux rate, and t_o is the operational time of the neutron generator.

$$N = N_p\sigma\phi t_o \quad (3)$$

Utilizing equation (3) we can calculate the final number of tritium atoms created during a specified operational time of the neutron generator. The results for an operational time of 2000 hours are presented in Table 7. It should be noted that interactions which do not lead to the creation of tritium were excluded.

Table 7. Total number of tritium atoms created during 2000 hour operation of the neutron generator

Interaction	Number of interactions for thermal neutron interaction	Number of interactions for fast neutron interaction
Build-up from ^{1}H ^{2}H (n, γ) ^{3}H	2.92E-02	4.87E-08
^{2}H (n, γ) ^{3}H	2.21E+08	4.14E+06
^{17}O (n, t) ^{15}N	0	1.41E+10

If we postulate that the generated neutrons have a single energy, the total amount of tritium created is typically less than 1E10 atoms, even with our conservative approach, with the primary contributor being ^{17}O via fast neutron interaction. The total amount of tritium generated by thermal neutrons is 2.21E8 atoms which corresponds to 2.21E-15 g. The total amount of tritium generated by fast neutrons is 1.41E10 atoms which corresponds to 1.41E-13 g. A consideration of the total number of tritium produced in the two proposed scenarios results in an

activity of 2.51 Bq/m^3 for fast neutron interaction and 0.04 Bq/m^3 for thermal neutron interaction.

It is important to note that the threshold for ^{17}O (n,t) ^{15}N reaction is 8.38 MeV neutron energy. Given the inherent moderation properties of water, it is more probable that this interaction occurs only in the region closest to the neutron generator. It is likely that the neutrons are moderated below the required energy for this reaction before a significant amount of tritium is generated. However, this interaction contributes to the creation of tritium and must be more closely evaluated before the implementation of water shielding of the neutron generator.

4 Conclusion

The first investigations of the radiation risks involved in the operation of a DT neutron generator revealed that faculty employees and the general public do not face any significant radiation risk associated with the operation of the generator. However, it was determined that the maximum possible yearly operational time should be 1500 hours in order to comply with the set limit of 6 mSv/year in all investigated locations.

The laboratory will not operate both the DD and the DT neutron generators simultaneously, so the total yearly operational time is likely to be lower than the proposed limit.

Furthermore, an activation analysis of potential water-based radiation shielding was conducted, which revealed even in our proposed conservative scenario that the potential tritium risk associated with such shielding is negligible, both in terms of generated activity and the amount of tritium generated.

Additionally, there is a potential for attenuation of the neutron flux due to the proposed shielding, which requires further investigation, as it could potentially increase the yearly operational time of the neutron generator by lowering the ambient doses in the operator room.

Literature

[1] READEN, B. T. and JESSEE, M. A. 2018. SCALE Code System, Version 6.2.3. Oak Ridge, Tennesse: Oak Ridge National Laboratory. ORNL/TM-2005/39, USA, 2764 pages, 2018.

[2] VRBAN, B. et al. 2025. Assessing radiation exposure and room effect in the new lab with a DD neutron generator at STU. In *The European Physical Journal Special Topics*. ISSN 1951-6401, 2025.

[3] WAGNER, JOHN C. et al. 2007. Forward-Weighted CADIS Method for Global Variance Reduction. In *Transactions of the American Nuclear Society*. ISSN 0003-018X, 2007, vol. 97, no. 1, p. 630-633.

[4] THERMO FISHER SCIENTIFIC. 2007. Thermo Scientific MP 320. available online at https://assets.thermofisher.com/TFS-Assets/CAD/Specification-Sheets/D10497~.pdf.

[5] HIDEKI, M. et al. 2021. Health effects triggered by tritium: how do we get public understanding based on scientifically supported evidence?. in *Journal of Radiation Research*. DOI https://doi.org/10.1093/jrr/rrab029, 2021, vol. 62, no. 4, p. 557-563.

[6] NEA JANIS. available online at https://www.oecd-nea.org/jcms/pl_39910/janis.

Simulácia radiačnej situácie v oblasti jadrovej elektrárne Mochovce v prípade ťažkej havárie

Brigita Kovácsová

Ústav jadrového a fyzikálneho inžinierstva, FEI STU v Bratislave

xkovacsovab@stuba.sk

Abstrakt – Táto práca sa zaoberá simuláciou radiačnej situácie v okolí jadrovej elektrárne Mochovce v prípade hypotetickej ťažkej havárie s únikom rádioaktívnych látok do atmosféry. Na základe meteorologických údajov boli analyzované trajektórie šírenia kontaminantov, dávkové charakteristiky a účinnosť ochranných opatrení, ako sú ukrytie v budovách, či jódová profylaxia. Výsledky poukazujú na význam včasnej reakcie, najmä pre citlivé skupiny obyvateľstva a zdôrazňujú potrebu funkčných havarijných plánov a efektívnej komunikácie. Simulácia slúži ako podpora pri plánovaní civilnej ochrany a riadení radiačných núdzových situácií.

1 Úvod

Bezpečnosť jadrových elektrární (JE) patrí medzi najprísnejšie a najregulovanejšie oblasti technickej infraštruktúry modernej spoločnosti. Aj napriek sofistikovaným systémom riadenia, preventívnym opatreniam a prísnym legislatívnym rámcom nemožno úplne vylúčiť možnosť výskytu havarijných udalostí. V takýchto prípadoch je nevyhnutné disponovať účinnými nástrojmi na predikciu a manažment potenciálnych dôsledkov. Práve simulácie radiačných situácií zohrávajú kľúčovú úlohu pri hodnotení rizík, určovaní potrebných ochranných opatrení a optimalizácii reakčných protokolov.

Táto práca sa zaoberá modelovaním šírenia rádioaktívnych látok v okolí jadrovej elektrárne Mochovce v hypotetickom scenári ťažkej havárie s únikom kontaminantov do atmosféry. Cieľom simulácie je vyhodnotiť dopady takejto udalosti na okolitú populáciu a životné prostredie, identifikovať najviac ohrozené oblasti a navrhnúť primerané ochranné stratégie. Simulácia vychádza z reálnych meteorologických údajov, pričom zohľadňuje rôzne faktory ako smer a rýchlosť vetra, typ a množstvo uniknutých izotopov, časové rozloženie dávky a účinnosť ochranných opatrení, ako je pobyt v interiéri či jódová profylaxia.

Osobitná pozornosť je venovaná hodnoteniu dávok prijatých obyvateľstvom prostredníctvom inhalácie a vonkajšej expozície, pričom analýza reflektuje rozdiely medzi vekovými skupinami. Výsledky sú interpretované z pohľadu odporúčaných limitných hodnôt a praktických opatrení, ktoré je vhodné prijať na minimalizáciu rizík.

Významnou súčasťou tejto štúdie je aj analýza efektívnosti havarijnej pripravenosti, funkčnosti varovných systémov a výzvy spojené s ich implementáciou v praxi. Opierajúc sa o historické príklady z minulých jadrových havárií na Slovensku aj vo svete, práca zdôrazňuje potrebu transparentnej a rýchlej komunikácie, dôslednej edukácie verejnosti a pravidelného testovania reakčných mechanizmov. Simulácia je preto nielen technickým modelom, ale aj nástrojom na podporu rozhodovania v rámci civilnej ochrany.

Cieľom práce je ukázať, že aj v prípade vážnych scenárov existuje priestor pre účinnú reakciu, ak sú k dispozícii presné údaje a efektívne plány. Zároveň sa kladie dôraz na zrozumiteľnú interpretáciu dát, aby mohli slúžiť ako podklad pre ďalšie výskumy, rozhodovanie orgánov verejnej správy či edukáciu obyvateľstva.

2 Jadrová bezpečnosť a spoľahlivosť

Jadrová bezpečnosť (JB) je stav a schopnosť jadrového zariadenia a jeho personálu zabrániť vzniku havárií a katastrof na týchto zariadeniach, účinne zabrániť nedovolenému úniku rádioaktívnych materiálov alebo ionizujúceho žiarenia do životného prostredia, a tak ohroziť zdravie a život ľudí. Hlavnými prostriedkom jadrovej bezpečnosti jadrových zariadení je zabrániť nekontrolovanému vzniku štiepnych reťazových reakcií, zabezpečiť dochladzovanie jadrových materiálov a kontrolovať tesnosť kontajnmentu.

JB vyžaduje neustálu snahu o dokonalosť. Kritériá na spoľahlivosť a bezpečnosť jadrových elektrární sú veľmi prísne a sú priebežne hodnotené individuálne pre jednotlivé JE národnými aj medzinárodnými dozornými úradmi, aby bolo možné minimalizovať pravdepodobnosť výskytu neočakávaných udalostí. Účinnosť tohto úsilia závisí od znalosti a pochopenia cieľov a zásad jadrovej bezpečnosti a ich vzájomného pôsobenia, ako aj od ich vhodného praktického uplatňovania [1].

Zákona 541 z 9. septembra 2004 o mierovom využívaní jadrovej energie (atómový zákon) a o zmene a doplnení niektorých zákonov definuje jadrovú bezpečnosť ako povinnosť prevádzkovateľa, teda držiteľa povolenia.

Držiteľ povolenia na jadrové zariadenie nesie plnú zodpovednosť za jeho jadrovú bezpečnosť. Zároveň musí pravidelne vyhodnocovať úroveň bezpečnosti zariadenia a priebežne informovať príslušný úrad o plnení požiadaviek súvisiacich s jadrovou bezpečnosťou. Držiteľ povolenia je tiež zodpovedný za zavádzanie preventívnych opatrení proti vzniku havárií.

3 Havarijná pripravenosť a havarijné plány

Jadrové havárie predstavujú mimoriadne vážne a komplexné riziká tak pre ľudské zdravie, ako aj pre životné prostredie. Dôsledky takýchto udalostí môžu byť citeľné aj dlhodobo, čoho príkladom sú následky katastrof v Černobyle alebo vo Fukušime. Preto zohráva havarijné plánovanie a pripravenosť na núdzové situácie kľúčovú úlohu pri zabezpečovaní jadrovej bezpečnosti. Dobre vypracované havarijné plány majú za cieľ zabezpečiť rýchle a koordinované konanie na ochranu obyvateľstva [3]. Vďaka dobre pripraveným havarijným plánom a kultúre bezpečnosti sa ukázalo po Fukušime, že jadrová havária sa dá zvládnuť omnoho lepšie a bez ohrozenia životov v porovnaní s Černobyľskou haváriou.

Základným cieľom havarijných plánov pre jadrové mimoriadne situácie je predchádzanie škodám, prípadne minimalizácia už vzniknutých škôd. Tieto plány obsahujú podrobné scenáre pre rôzne závažné udalosti, spolu s okamžitými opatreniami a definovanými úlohami.

V prípade jadrovej havárie je mimoriadne dôležitá rýchla evakuácia obyvateľstva, keďže intenzita žiarenia sa v čase mení a okamžitá reakcia môže zachrániť životy. Primerané informovanie o mimoriadnych situáciách, napríklad prostredníctvom SMS správ, hlásení v rádiu a televízii, ako aj cez sociálne médiá, môže znížiť pravdepodobnosť vzniku paniky a napomáha koordinovanému úniku obyvateľstva.

Efektívne nasadenie záchranných síl je takisto nevyhnutné – koordinované konanie hasičov, záchranárov a policajtov umožňuje lokalizovať zdroj žiarenia, zastaviť únik rádioaktívnych látok a minimalizovať radiačnú záťaž osôb.[4]

Ochrana verejného zdravia je hlavným cieľom každého núdzového plánu. V prípade jadrových havárií nejde iba o zabránenie fyzickým poraneniam, ale aj o prevenciu vdýchnutia rádioaktívnych látok a ich absorpcie cez pokožku.

Vďaka dobre fungujúcim systémom núdzovej pripravenosti sa znižuje počet ochorení a úmrtí spôsobených žiarením, obzvlášť medzi najohrozenejšími skupinami, ako sú deti, seniori a chronicky chorí pacienti. Navyše vopred pripravené protokoly umožňujú rýchle podávanie jódových tabliet, ktoré znižujú vplyv rádioaktívneho jódu na štítnu žľazu. [3]

Prevencia paniky a zachovanie spoločenskej stability úzko súvisí s dôveryhodnou a rýchlou komunikáciou. Spoľahlivé a jasné informácie znižujú šírenie špekulácií a vymyslených či nepravdivých správ, čím napomáhajú plynulému fungovaniu spoločnosti.

Pri jadrovej havárii rozhoduje každá minúta, preto je rýchla koordinácia udalostí a vybudovanie vhodného logistického zázemia mimoriadne dôležité. Núdzové plány musia obsahovať presne rozdelené úlohy medzi členmi záchranných zložiek, ako aj spôsob a komunikačné kanály na miestnej, regionálnej a národnej úrovni.[4]

Medzi logistické výzvy patrí:

- dostupnosť ochranných prostriedkov a meracích zariadení,
- nasadenie mobilných dekontaminačných jednotiek,
- zabezpečenie zásob pitnej vody a potravín pre obyvateľov v uzavretých zónach.

Jedným z najdôležitejších krokov pri zvládaní jadrových havárií je rýchle a dôveryhodné informovanie obyvateľstva. Únik rádioaktívnych látok je neviditeľný a bez zápachu, a preto je jeho priama detekcia nemožná. Reakcia ľudí na oficiálne výstrahy je preto kľúčová.

Prostriedky varovania zahŕňajú [6]:

- Sirény okamžitý výstražný signál, ktorý ľudí upozorňuje na potrebu sledovať podrobnejšie informácie.
- Hromadné priame varovania cez mobilné siete pre osoby nachádzajúce sa v ohrozených oblastiach.
- Podrobné informovanie prostredníctvom miestnych médií o nevyhnutných krokoch, napríklad či majú ľudia ostať doma, začať evakuáciu, užiť jódové tablety atď.
- Webové stránky samospráv a štátnych orgánov, ktoré poskytujú neustále aktualizované informácie, kontakty, mapy a pokyny.

V určitých prípadoch, napríklad pri úniku rádioaktívneho prachu alebo plynu do ovzdušia, môže byť evakuácia nebezpečnejšia ako zotrvanie na mieste. V takýchto situáciách úrady vyzývajú obyvateľstvo, aby ostalo vo vnútri budov, pričom je nevyhnutné dodržať špecifický protokol.

Hlavné opatrenia:

- Hermetické uzatvorenie okien a dverí, ktoré zabraňuje prenikaniu vonkajšieho vzduchu.
- Vypnutie ventilačných systémov, klimatizačných zariadení a ventilátorov.
- Presun do uzatvorenej miestnosti. Ideálne je zvoliť vnútornú miestnosť bez okien, ktorá minimalizuje možnosť kontaktu s vonkajším prostredím.

Ak rozsah alebo lokalizácia rádioaktívneho úniku vyžaduje evakuáciu, tá sa stáva nevyhnutnou. Takéto operácie môžu byť úspešné len vtedy, ak prebiehajú podľa vopred pripraveného a podrobného plánu.

Typy evakuácie:

- Preventívna evakuácia uskutočňuje sa v prípade predvídateľného rizika, napríklad po zemetrasení, keď existuje riziko pre bezpečnosť jadrového reaktora.
- Núdzová evakuácia vykonáva sa okamžite po udalosti, ktorá už nastala, napríklad po výbuchu reaktora.

4 Sprostredkovanie informácii obyvateľom

Krízová komunikácia nie je len doplnkovým prvkom núdzového plánovania, ale jedným z jeho kľúčových pilierov. Počas jadrovej havárie, keď je nebezpečenstvo neviditeľné, rýchlo sa šíri a môže mať dlhodobé následky, môžu byť včasné, jasné a cielené informácie životne dôležité.

Cieľom efektívnej komunikácie je zabezpečiť, aby obyvateľstvo jasne pochopilo aktuálnu situáciu, presne vnímalo hrozby, ktoré im hrozia, dostalo jasné a ľahko vykonateľné pokyny, a aby sa zároveň zvýšil ich pocit bezpečia, čím sa minimalizuje pravdepodobnosť paniky a chaosu.

Rôznorodosť komunikačných kanálov umožňuje rôznym skupinám obyvateľstva, nezávisle od veku, technologických znalostí a miesta bydliska, prístup k relevantným informáciám. Dobre fungujúci systém pozostáva z viacerých vzájomne sa dopĺňajúcich prvkov.

Jeden z prvých varovných systémov pri jadrových haváriách sú sirény, ktoré sa môžu okamžite aktivovať a poskytujú zvukovú výstrahu na veľkej ploche. Sú obzvlášť užitočné v oblastiach s obmedzeným pokrytím mobilnej siete.

Technológia Cell Broadcast umožňuje zasielať varovné správy priamo na mobilné zariadenia ľudí nachádzajúcich sa v ohrozených oblastiach bez potreby internetového pripojenia. Tento systém nevyžaduje mobilné dátové prenosy a funguje aj pri extrémnom zaťažení siete. Výstraha okamžite zasiahne desaťtisíce ľudí a často obsahuje konkrétne pokyny. [6]

Klasické masmédiá sú vhodné na poskytovanie detailných informácií v reálnom čase. Počas jadrovej krízy je dôležité nielen informovať o aktuálnej situácii, ale aj komunikovať pokojným tónom a vysvetľovať udalosti. Mestské rozhlasové systémy spravované miestnymi samosprávami sú kľúčové pre menšie obce. Verejnoprávne televízie môžu spustiť špeciálne vysielania určené na nepretržité informovanie obyvateľov. Kritické informácie zahŕňajú evakuačné trasy, distribučné miesta jódových tabliet a lokalizácie dekontaminačných centier. [3]

Webové stránky a sociálne siete umožňujú aktualizovať oficiálne a aktuálne správy v priebehu niekoľkých minút. Webová stránka SHMÚ (Slovenský hydrometeorologický ústav) a jeho sociálne siete publikujú mapy zobrazujúce meteorologické riziká a šírenie rádioaktívnych látok. UJD SR (Úrad jadrového dozoru SR) slúži ako dôveryhodný zdroj pre šírenie oficiálnych informácií a technických vysvetlení týkajúcich sa jadrových udalostí. Facebookové stránky miestnych samospráv často predstavujú najrýchlejší zdroj informácií pre obyvateľstvo, najmä ak zároveň odpovedajú na otázky občanov. Kľúčová je nielen rýchlosť, ale aj dôveryhodnosť informácií, preto majú oficiálne zdroje prednosť pred správami zo sociálnych sietí.

Tlačové vyhlásenia a krízové linky - Centrálne a regionálne krízové štáby pravidelne vydávajú tlačové vyhlásenia, ktoré štruktúrovane oznamujú najdôležitejšie udalosti a opatrenia. Komunikácia polície a zložiek civilnej ochrany zohráva ústrednú úlohu. Call-centrá a krízové linky poskytujú pomoc tým, ktorí nemajú prístup k online informáciám alebo potrebujú osobnú pomoc, napríklad pri hľadaní nezvestných príbuzných či evakuácii osôb so zdravotným postihnutím.

Efektívna komunikácia je jedným zo základných pilierov krízového manažmentu. Ak komunikácia meškâ, je nepresná alebo sa nedostane ku všetkým ľuďom, môže to mať tragické následky, ako sú panika, úmrtia, šírenie dezinformácií, ba dokonca aj strata dôvery voči úradom. História nám poskytuje viacero príkladov, ktoré ukazujú, že napriek vybudovaným systémom môžu nastať chyby. Ich identifikácia a pochopenie sú kľúčové pre prevenciu takýchto situácií v budúcnosti.

Porucha varovných systémov jadrových elektrární predstavuje závažný problém, ktorý môže znamenať bezpečnostné riziko pre personál, obyvateľstvo aj životné prostredie. Niektoré prípady, keď došlo k zanedbaniu varovania alebo k neposkytnutiu adekvátnych informácií.

Medzi najčastejšie chyby patria:

- Oneskorené alebo nedostatočné varovanie obyvateľstva.
- Nesprávne pochopenie údajov alebo chybná interpretácia signálov.
- Cenzurovanie informácií alebo nedostatočná transparentnosť.

5 Cvičenie obyvateľov na havárie

Jadrové havárie patria medzi najvážnejšie potenciálne katastrofy modernej doby. Kľúčom k úspešnému zvládaniu krízových situácií je, aby celé obyvateľstvo vrátane detí, starších ľudí a osôb so špeciálnymi potrebami vedelo, ako sa správať v prípade núdze. Preto sú mimoriadne dôležité pravidelné nácviky a testovanie funkčnosti výstražných systémov.

Cieľom cvičení civilnej ochrany nie je len preveriť pripravenosť príslušných orgánov, ale aj umožniť obyvateľstvu vyskúšať si v simulovaných podmienkach, čo najbližších realite, ako reagovať v prípade jadrovej mimoriadnej situácie. Konkrétnymi cieľmi týchto cvičení sú otestovanie reakčných protokolov, ako napríklad evakuácia obyvateľstva, distribúcia jódových tabliet či rýchle uzatvorenie postihnutej oblasti. [3]

Ani najlepší plán nemá význam, ak prostriedky, ktoré by mohli zachrániť životy, nie sú spoľahlivé. Preto je pravidelné testovanie výstražných systémov – najmä sirén, mobilných varovaní a miestnych rozhlasových hlásení – absolútne nevyhnutné. Testy sirén nie sú len technickou kontrolou systému, ale zároveň slúžia ako istý druh „psychologického tréningu“ pre obyvateľstvo, aby spoznalo jednotlivé varovné signály a naučilo sa na ne reagovať bez paniky. Evakuácia škôl ukazuje, ako je možné organizovane, včas a pokojne opustiť budovu školy v prípade núdze.

Samotné centrálne plánovanie nestačí. Počas krízových situácií je prvý kontakt medzi obyvateľstvom a miestnou správou zabezpečený prostredníctvom starostov, miestnych krízových manažérov, škôl a občianskych organizácií.

Informovanie obyvateľov o cvičení:

Obyvatelia sú včas informovaní o termíne a obsahu cvičení pomocou plagátov, letákov, miestnych novín a stránok na sociálnych sieťach. Primerané informovanie pred cvičením zvyšuje účasť obyvateľov a pomáha predchádzať nedorozumeniam.

Povinnosti starostov:

- Zabezpečiť existenciu miestneho krízového plánu, ktorý detailne zahŕňa evakuáciu, informovanie a distribúciu jódových tabliet.

- Vytvorenie zoznamu evakuovaných osôb s osobitným dôrazom na starších ľudí, osoby so zdravotným postihnutím a deti.
- Vypracovanie plánu dopravy na zabezpečenie prístupu k zhromažďovacím miestam, vrátane prepravy osôb s obmedzenou mobilitou.

Zapojenie komunity:

Aktívna účasť škôl, miestnych občianskych združení, cirkevných organizácií a dobrovoľníckych skupín posilňuje komunitnú súdržnosť a zohráva kľúčovú úlohu aj v reálnych krízových situáciách – či už v oblasti šírenia informácií alebo pri poskytovaní starostlivosti evakuovaným osobám.

Jadrové havárie si vyžadujú rýchle reakcie, premyslené rozhodnutia a stabilnú komunitnú spoluprácu. Cvičenia civilnej ochrany a testovanie sirén nie sú len „administratívnou povinnosťou“, ale nástrojmi, ktoré pripravujú ľudí na neočakávané situácie. Vďaka nim sa nielen orgány, ale aj obyvatelia učia zachovať pokoj a zároveň vykonávať konkrétne potrebné kroky.[4]

Ten, kto trénuje, sa nebojí a dobre pripravená komunita nielenže prežije, ale môže byť aj vzorom pre celý svet.

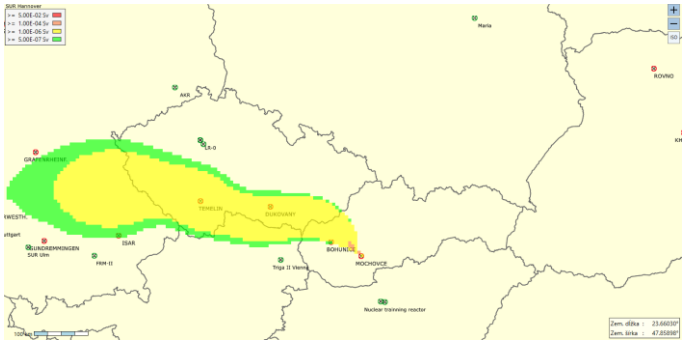
6 Simulácia situácie v EMO

V rámci havarijných plánov je nevyhnutné plánovať evakuáciu a evakuačné trasy tak, aby boli obyvatelia zasiahnutí, čo najmenej. Evakuačné trasy závisia od konkrétneho počasia a meteorologických predpovedí, keďže vietor roznáša rádioaktivitu. Pre výpočet optimálnych evakuačných tras je vytvorený program ESTE, ktorý určí evakuačné podmienky pre konkrétne obce a najvhodnejší čas evakuácie. Tento program spravuje UJD a je dostupný aj v JE.

Na FEI STU máme školskú licenciu programu ESTE, v ktorej sa dá simulovať šírenie rádioaktivity s obmedzením na rozptyl 3 rádionuklidov: ^{131}I, ^{133}Xe a ^{137}Cs. Školská verzia nedisponuje konkrétnymi údajmi pre jadrové materiály a ich výskyt v „havarovanom“ jadrovom reaktore, ale má aktuálne meteorologické mapy, ktoré spoľahlivo určujú smer šírenia rádioaktivity v rôznych výškach nad podložím.

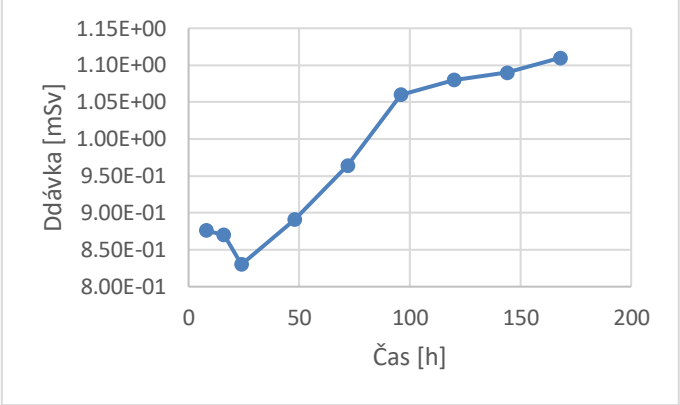
V rámci praktickej časti som simulovala priebeh a dopad ťažkej havárie, kedy v JE zlyhajú všetky bezpečnostné bariéry (pokrytie paliva, tesnosť primárneho okruhu, poškodenie hermetickej zóny) a rádioaktivita sa šíri do okolia. V rámci vyhodnocovania výsledkov v tejto práce som sa sústredila na deň 8.11.2024 a šírenie rádioaktivity do prostredia počas nasledujúcich 7 dní. Mapa šírenia rádioaktivity je zobrazená na Obr. 1.

8. novembra fúkal severozápadný vietor – smer na Trnavský kraj. V tento deň bola maximálna rýchlosť vetra 16,6 km/h. Počasie bolo bez zrážok a priemerná vlhkosť dosiahla 55% (Slovensko, Bratislavský kraj). Na druhý deň po havárii bol viditeľný rozdiel v sile vetra v porovnaní s predchádzajúcim dňom, pretože vietor už zasahoval nielen Českú republiku, ale aj Nemecko. Na Slovensku sa vietor opäť prevažne pohyboval smerom k Trnavskému kraju, pričom zasiahol sever Českej republiky, menšiu severnú časť Rakúska a malú oblasť Nemecka pri českej hranici. V tomto období panovalo v týchto štyroch krajinách podobné počasie: v Českej republike a Rakúsku bolo hmlisto, v Nemecku prevažne zamračené.



Obr. 1. Šírenia mraku v 8.11.2024.

Najväčšia dávka bola 7 hodín po nehode v obciach Slepčany a Červený Hrádok. Obe obce ležia na severe od Mochoviec. V Slepčanoch bola nameraná dávka 1,11 mSv, v Červenom Hrádku 47,8 µSv od začiatku úniku. Časový priebeh efektívnej dávky od vzniku havárie po 7 deň (168 hodín) je zobrazený na Obr. 2.

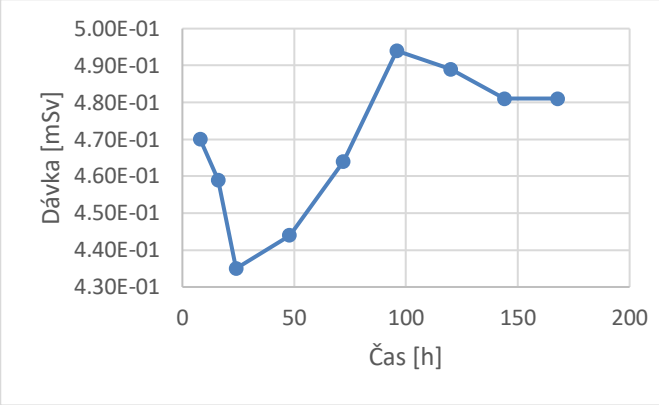


Obr. 2. Efektívna dávka od začiatku úniku.

V prvých 24 hodinách dávka mierne klesá, následne dosiahne lokálne minimum po 24 hodinách. To môže indikovať, že prvá vlna sa už čiastočne rozptýlila alebo boli priaznivé poveternostné podmienky. Medzi 24. a 96. hodinou však možno pozorovať nepretržitý a rýchly nárast. Toto poukazuje na kumulatívny efekt spôsobený trvajúcim únikom alebo ukladaním rádioaktívnych látok. V období od 96. do 168. hodiny sa rast dávky spomalí a stabilizuje sa na hodnote približne 1,09 mSv. To môže naznačovať oslabenie zdroja kontaminácie alebo zlepšenie meteorologických podmienok.

Kumulovaná dávka je nízka, jódová profylaxia nie je potrebná. Evakuácia nie je nevyhnutná len na základe kumulovanej dávky. Odporúča sa zotrvanie v interiéri počas prvých 48–72 hodín, kedy je rast najprudší. Je opodstatnené pokračovať v nepretržitom monitorovaní, pretože aj keď rast dávky spomalil, ešte sa úplne nezastavil.

Pobyt v interiéri predstavuje mimoriadne účinnú formu ochrany (viď. Obr. 3), pričom namerané hodnoty sú takmer o polovicu nižšie ako tie, ktoré sme videli na predchádzajúcom grafe (vonkajšie hodnoty). Z toho môžeme usudzovať, že evakuácia nie je potrebná, ak obyvatelia zostanú v uzavretých priestoroch. Uzatvorenie okien a vypnutie ventilácie môže ešte viac znížiť namerané hodnoty.



Obr. 3. Efektívna dávka v budove.

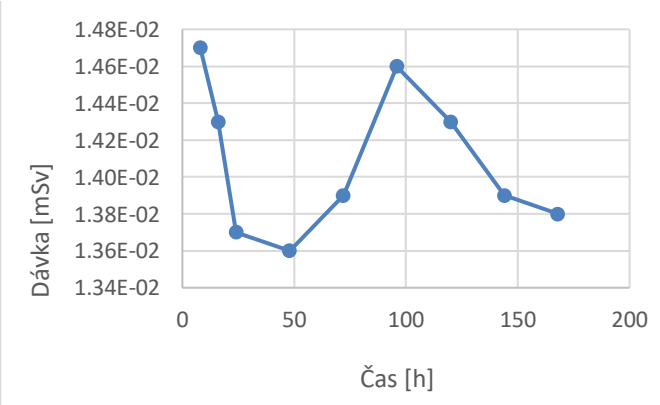
Naproti tomu, na obrázku 4 je znázornené, akú dávku dostane osoba vo vonkajšom prostredí z rádioaktívneho oblaku šíriaceho sa vzduchom.

Na začiatku môžeme pozorovať pokles, čo možno pripísať prirodzenému rozptylu rádioaktívneho oblaku a zmene smeru vetra, avšak po 48 hodinách vidíme opäť nárast. Medzi 96 a 168 hodinami je už pozorovateľný kontinuálny pokles. Atmosférická aktivita klesá, oblak sa pravdepodobne rozptýlil, a tým spôsobuje nižšiu expozíciu.

Tieto hodnoty samy osebe nevyžadujú ochranné opatrenia, avšak odporúča sa vyhnúť dlhodobému pobytu vonku počas prvého týždňa. Ochrana vo vnútri budov významne znižuje túto dávku.

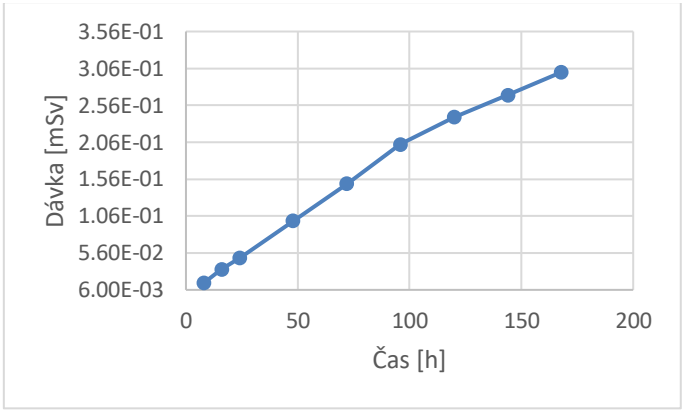
Dávka pochádzajúca z oblaku je nízka a nemá kumulatívny charakter. Dá sa dobre eliminovať pobytom v interiéri. Riziko predstavuje len krátkodobo, najmä v prípade ďalšieho úniku rádioaktívnych látok.

Na prvý pohľad sa môže zdať zvláštne, že dávka v budove je vyššia ako dávka pochádzajúca z mraku. Dávka v budove je totiž súčtom viacerých zložiek. Okrem dávky z mraku zahŕňa aj dávku pochádzajúcu z depozitov rádioaktívnych látok a vnútornú dávku spôsobenú inhaláciou kontaminovaného vzduchu. Preto je celková dávka v budove vyššia ako samotná dávka z mraku. Malá hodnota dávky z mraku je spôsobená tým, že zobrazuje len krátkodobú vonkajšiu expozíciu počas prechodu mraku, ostatné dávky zahŕňajú dlhodobé kumulatívne účinky.



Obr. 4. Dávka z mraku od začiatku úniku.

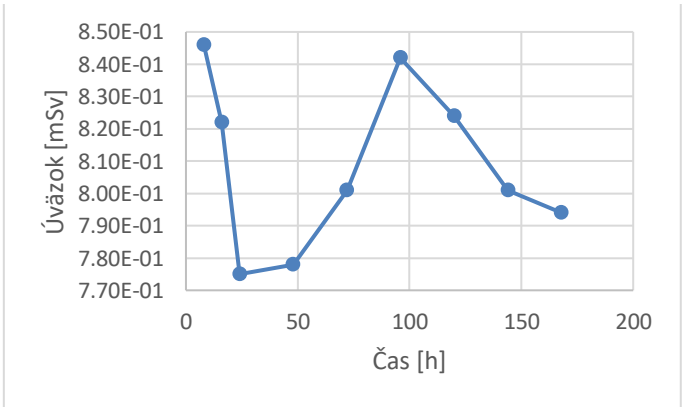
Obrázok č. 5 zobrazuje dávku pochádzajúcu z externého žiarenia rádioaktívnych látok, ktoré sa usadili na povrchu pôdy alebo na iných povrchoch. Dávka lineárne a stabilne rastie s časom. Je to pravdepodobne spôsobené tým, že usadené rádioaktívne látky nemiznú, naopak, priebežne prispievajú k vonkajšej radiačnej záťaži. Dávka nestagnuje ani neklesá. V tomto prípade sa neodporúča zdržiavať dlhodobo vonku, najmä v kontaminovaných oblastiach. Odporúča sa dekontaminácia povrchov. Môže byť potrebná kontrola potravín a krmív, pokiaľ došlo k usadeniu rádioaktívnych látok na poľnohospodárskej pôde. Deti môžu byť v tomto prípade obzvlášť citlivé, preto by sa mali vyhýbať aktivitám vonku. Táto dávka sa kumuluje a môže mať dlhodobý vplyv na zdravie. Aj keď táto dávka neodôvodňuje evakuáciu, preventívne opatrenia sú rozhodne odporúčané.



Obr. 5. Dávka z depozície od začiatku úniku.

Obrázok č. 6 zobrazuje, akú dávku dostane človek vdýchnutím rádioaktívnych látok od začiatku udalosti. V prvých 24 hodinách môžeme pozorovať pokles, čo môže byť spôsobené odchodom prvotného rádioaktívneho vrcholu alebo tým, že obyvateľstvo rýchlo zostalo v interiéroch. Po 24 hodinách však dávka opäť začína stúpať, čo pravdepodobne súvisí s novým oblakom rádioaktívnych izotopov alebo nepriaznivým pohybom vzdušných más. Po 4 dňoch dávka znovu klesá a postupne sa stabilizuje. Tento trend môže potvrdzovať, že kritická fáza už prešla, no stále sú vo vzduchu prítomné aktívne izotopy. Pobyt v interiéri počas kontaminovaných období minimálne do 96 hodín výrazne znižuje túto dávku. Minimalizácia vetrania môže byť taktiež veľmi dôležitá.

Významná časť inhalačnej dávky sa vytvorí v prvých 100 hodinách, avšak podľa grafu neprekračuje hodnotu 1 mSv. Toto samo osebe neodôvodňuje samostatnú evakuáciu, môže však odôvodňovať preventívne ochranné opatrenia, najmä pre citlivé skupiny obyvateľstva.



Obr. 6. Úväzok inhaláciou od začiatku úniku.

V Tabuľke č. 1 sú uvedené hodnoty dávok, ktorým sa dá predísť. Podľa údajov poskytuje jódová profylaxia najvýraznejšie zníženie dávky, najmä u dojčiat (33 mSv), čo sa už približuje evakuačnej hranici 50 mSv. U dospelých má jódová profylaxia taktiež významný účinok, keď znižuje dávku o 14,5 mSv. Evakuácia je užitočná, ale redukuje dávku výrazne menej. Účinok ukrytia sa v interiéri (vnútorný úkryt) je miernejší, no stále lepší než neurobiť vôbec nič. Cieľom evakuácie je minimalizovať radiačnú záťaž obyvateľstva v kritickom období, s osobitným dôrazom na dojčatá a citlivé skupiny obyvateľov.

Tab. 1. Odvrátená dávka.

č.		[Sv]
1	Evakuácia pred príchodom mraku, z voľného priestoru	0,00115
2	Ukrytie-budova pred príchodom mraku, z voľného priestoru	0,000475
3	I-profylaxia, dospelý, účinok pred príchodom mraku	0,0145
4	I-profylaxia, kojenec, účinok pred príchodom mraku	0,033

Priorita opatrení je z hľadiska času okamžitá, teda ešte pred príchodom rádioaktívneho oblaku, a to na základe predpovede. Distribúcia jódovej profylaxie sa vykonáva prioritne pre dojčatá a deti, keďže podľa údajov môžeme takto zabrániť ožiareniu rádioaktívnym jódom až do 33 mSv. Podanie jódových tabliet sa odporúča aj dospelým. V dotknutých oblastiach (Slepčany a okolie) je potrebné vyzvať obyvateľov, aby zostali vnútri budov a uzatvorili ventilačné systémy.

V krátkom časovom horizonte, pred príchodom alebo počas príchodu rádioaktívneho oblaku, je evakuácia odporúčaná iba pre najcitlivejšie skupiny obyvateľstva, ako sú deti vo veku 0–5 rokov, tehotné ženy a osoby so závažnými zdravotnými problémami. Dávka, ktorej sa dá predísť (1,15 mSv), je stále nižšia ako evakuačný prah 50 mSv, ale jej zníženie predstavuje minimalizáciu rizika. Hromadná evakuácia na základe dostupných údajov nie je opodstatnená, keďže dávky nedosahujú núdzové limitné hodnoty a riziká spojené s evakuáciou, ako napríklad stres či dopravné zápchy, môžu byť vyššie než samotná radiačná dávka.

Okrem toho je potrebné jasne informovať obyvateľstvo, napríklad o účele jódových tabliet, o tom, kedy ich treba užiť atď. Obyvatelia by mali byť informovaní aj o výhodách ukrytia sa v interiéri, čím sa môže znížiť množstvo vdýchnutých rádioaktívnych izotopov. V prípade evakuácie je potrebné určiť konkrétne evakuačné trasy a časové rámce, pričom treba dbať na to, aby sa pohyb obyvateľstva vyhýbal hlavnému smeru vetra. V tomto prípade by evakuačná trasa mala byť juhozápadne alebo severovýchodne – teda kolmo na smer šírenia mraku.

7 Záver

Simulácia radiačnej situácie v okolí jadrovej elektrárne Mochovce v prípade ťažkej havárie poskytla cenné poznatky o dynamike šírenia rádioaktívnych látok a ich potenciálnom dopade na zdravie obyvateľstva a životné prostredie. Výsledky ukázali, že smer a intenzita vetra zohrávajú kľúčovú úlohu pri určovaní postihnutých oblastí, pričom aj v rámci relatívne krátkeho časového obdobia môže dôjsť k výrazným zmenám v distribúcii dávky.

Simulované údaje potvrdili význam okamžitých ochranných opatrení ukrytie v budovách, vypnutie ventilácie a včasné podanie jódovej profylaxie ktoré dokážu zásadne znížiť radiačnú záťaž. Kumulované dávky v skúmaných oblastiach síce nedosiahli hranicu, ktorá by si vyžadovala plošnú evakuáciu, no opodstatňujú preventívne odporúčania a zvýšenú pripravenosť. Samozrejme, že efektívne dávky pre reálne zastúpenie rádionuklidov v JE Mochovce by boli vyššie ako pre školskú verziu programu.

Dôležitým aspektom tejto práce bolo poukázať na význam dobre pripravených havarijných plánov, funkčnej krízovej komunikácie a pravidelného školenia obyvateľstva. Simulácie podobného typu by sa mali stať neoddeliteľnou súčasťou plánovania a hodnotenia jadrovej bezpečnosti, keďže umožňujú vopred identifikovať potenciálne riziká, optimalizovať reakčné časy a minimalizovať negatívne dopady v prípade havárie.

Literatúra

[1] V. Slugeň, M. Florek, P. Uhrík, J. Markuš, a J. Tomek, *Jadrové zariadenia, jadrová bezpečnosť*. Bratislava, Slovensko: Slovenská nukleárna spoločnosť, 2009. [Online]. Dostupné na: https://edu.snus.sk/ucebnice/Slugen-JZ.pdf

[2] SLOVENSKO. 2004. Zákon č. 541/2004 Z. z. o mierovom využívaní jadrovej energie (atómový zákon) a o zmene a doplnení niektorých zákonov. Vyhlásený 16. októbra 2004. Účinný od 1. októbra 2019. Zbierka zákonov Slovenskej republiky.

[3] WORLD HEALTH ORGANIZATION. 1999. *Guidelines for Iodine Prophylaxis following Nuclear Accidents, 1999*. [online]. Geneva: World Health Organization, 1999. [2025.03.23]. Dostupné:https://iris.who.int/bitstream/handle/10665/66143/WHO_SDE_PHE_99.6.pdf

[4] FEDERAL EMERGENCY MANAGEMENT AGENCY. 2010. *Developing and Maintaining Emergency Operations Plans: Comprehensive Preparedness Guide (CPG) 101* [online]. Washington, D.C. : Federal Emergency Management Agency, 2010. [2025.03.23]. Dostupné na: https://www.fema.gov/sites/default/files/2020-05/CPG_101_V2_30NOV2010_FINAL_508.pdf

[5] INTERNATIONAL FEDERATION OF RED CROSS AND RED CRESCENT SOCIETIES. 2018. *Public Awareness and Public Education for Disaster Risk Reduction: Action-oriented key messages for households and schools* [online]. Geneva : International Federation of Red Cross and Red Crescent Societies, 2018.[2025.03.23]Dostupné:https://www.ifrc.org/sites/default/files/PAPE-2.0-English.pdf

[6] ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. 2013. *Crisis Communication: Facing the Challenges in Nuclear* [online]. Paris : OECD Nuclear Energy Agency,2013.[2025.03.23].Dostupné:https://www.oecd.org/content/dam/oecd/en/publications/reports/2013/06/crisis-communication-facing-the-challenges_g1g3750a/9789264205055-en.pdf

Predlžovanie životnosti jadrových elektrární a využitie pozitrónovej anihilačnej spektroskopie pri testovaní starnutia materiálov

Marek Kováč

Ústav Jadrového a Fyzikálneho Inžinierstva
FEI STU

xkovacm7@stuba.sk

Abstrakt – Tento príspevok popisuje dôležitosť dlhodobej prevádzky jadrových elektrární, ktorá sa dá dosiahnuť len pri použití spoľahlivých konštrukčných materiálov schopných odolávať záťaž dlhšie ako 60 rokov. V článku popisujem testovanie materiálov namáhaných radiačným a tepelným poškodením pomocou nedeštruktívnej techniky pozitrónovej anihilácie – metódy doby života pozitrónov.

1 Úvod

V dnešnej dobe sa čoraz viac kladie dôraz na obnoviteľné zdroje energie, medzi ktoré sa dostáva do popredia aj jadrová energia s vysokým produkovaným výkonom na malej ploche. Jadrová elektráreň (JE) je komplexné zariadenie, ktoré musí spĺňať vysoké bezpečnostné normy a je neustále pod dohľadom dozorných orgánov a verejnosti. Nakoľko výstavba nového energetického zdroja prináša so sebou veľké finančné, materiálové a hlavne časové náklady, predlžovanie životnosti bezpečných a spoľahlivých jadrových elektrárni sa zdá byť ako efektívne riešenie zvyšujúcej sa globálnej spotreby energie. Pri výstavbe elektrární sa životnosť odhadovala na 30-40 rokov. Dnes vieme, že pri pravidelnej kontrole, údržbe a výmene materiálov sa môže životnosť jadrových zariadení predĺžiť na 60-80 rokov. Treba pri tom dbať na neustále zvyšovanie jadrovej bezpečnosti a najmä na bezpečnosť zdravia ľudí a prírody v okolí elektrárne.

2 Testovanie starnutia materiálov v JE

Konštrukčné materiály JE musia byť vyrobené špeciálne, a musia odolávať technicko-fyzikálnym podmienkam pri prevádzke. Sú pod záťažou ako napríklad ťah, tlak, vysoká teplota, radiačné poškodenie a korózia. V rámci nášho výskumu starnutia materiálov pozorujeme vzorky, ktoré pochádzajú z primárneho okruhu a okolia tlakovej nádoby reaktora (TNR).

Na porovnanie radiačného poškodenia, poškodenia spôsobeného koróziou a tepelným starnutím sme si zvolili vzorky z hlavného cirkulačného čerpadla, kapsule pre svedočné vzorky a komponenty úchytu príruby reaktora. Tieto vzorky boli v prevádzke JE od 24 do 32 rokov.

2.1 Vzorky

Naše vzorky popísané v Tabuľke 1 sme čerpali z projektu DELISA-LTO [1], ktoré sme prevzali narezané z laboratória nedeštruktívnych testov VUJE, a.s.. Materiály pochádzajú z rôznych jadrových elektrární v rámci Európy.

Tab. 1. Vzorky

Vzorka	Pôvod vzorky	Opis vzorky
6-c	VVER 1000 Ukrajina	Skrutka príruby TNR, 30 rokov, 305 °C
6-d	VVER 1000 Ukrajina	Matica príruby TNR, 30 rokov, 240 °C
9-a	VVER 440 Paks, Maďarsko	Kapsula svedočných vzoriek, 24 rokov, 278 °C
9-b	VVER 440 Paks, Maďarsko	Kapsula svedočných vzoriek, 32 rokov, 278 °C
10	VVER 440 Paks, Maďarsko	Hlavné cirkulačné čerpadlo (HCČ), 27 rokov, 266 °C

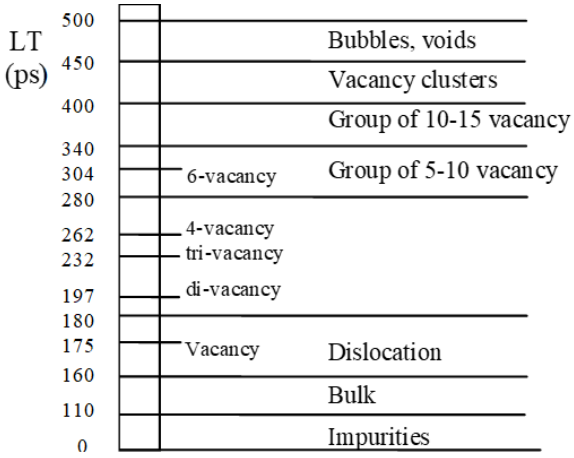
2.2 Prevádzka pri teplote 300 °C

Ako môžeme vidieť z Tab. 1, tak väčšina vzoriek je namáhaná vysokou teplotou počas dlhého obdobia. Predpokladáme, že toto dlhodobé tepelné zaťaženie má vplyv na štruktúru materiálov a na ich defekty. Dlhšie časové namáhanie teplom pôsobí na materiál priaznivým spôsobom tak, že zoceľuje štruktúru materiálu a robí ho pevnejším. Je to vplyvom vzniku precipitácii medi, alebo karbidov chrómu, niklu a mangánu [2]. Tieto precipitácie ovplyvňujú pnutie a pohyb defektov v štruktúre, materiál tak môže strácať pružnosť a stať sa krehkejším [3]. Zmenu v koncentrácii a veľkosti defektov vplyvom času a teploty namáhania môžeme pozorovať pomocou techniky pozitrónovej anihilácie. Táto zmena sa prejaví v technike zmenou hodnôt intenzity a doby života pozitrónov vo vakančných typoch defektov.

2.3 Aplikácia techniky Pozitrónovej anihilácie na výskum starnutia materiálov v prostredí JE

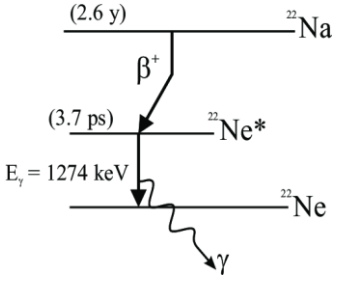
Pozitrónová anihilačná spektroskopia má vysokú citlivosť na defekty vakančného typu [4]. Medzi najrozšírenejšie techniky pozitrónovej anihilácie je meranie doby života pozitrónov (PALS). Technika PALS je založená na prestupe pozitrónov z pozitrónového zdroja do vzorky a meraní času od vzniku pozitrónu v zdroji po jeho anihiláciu s elektrónom vo vzorke. Čas, za ktorý pozitrón anihiluje závisí od veľkosti a koncentrácie defektov v materiáli.

Všeobecne, dĺžka života pozitrónov anihilovaných v menších vakanciách kovov bola experimentálne a teoreticky dokázaná, leží v rozmedzí od 150-300 ps [5]. Ku defektom väčšieho rozmeru zaraďujeme defekty, v ktorých pozitrón anihiluje v čase od 400 ps. Jednotlivé doby života pozitrónov a k nim priradené defekty sú na Obrázku 1.



Obr. 1 Defekty vakančného typu a doba života pozitrónov, ktoré v nich anihilujú [5].

Ako zdroj pozitrónov sme pri našich meraniach využili rádioaktívny sodík ^{22}Na, ktorý emituje Beta časticu – pozitrón a o 4 ps aj gama časticu s energiou 1274 keV (Obr. 2). Pozitrón anihiluje s energiou 511 keV. Z časového posunu pri záchyte týchto dvoch energii vieme určiť dobu života pozitrónu.

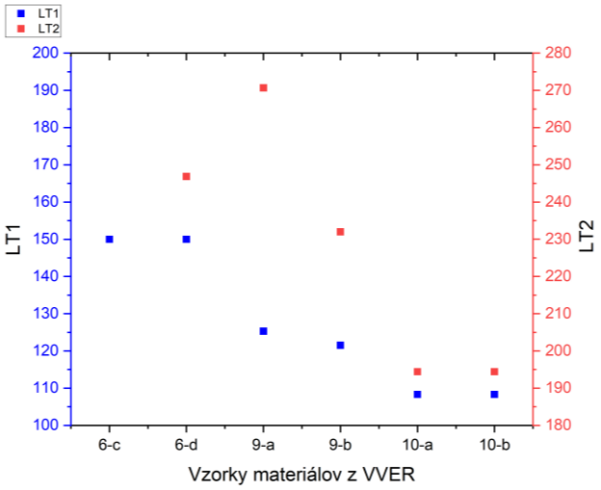


Obr. 2 Pozitrónový zdroj na FEI STU.

V rámci meraní sme použili model štandardného záchytu s 3 komponentami, hoci pre niektoré vzorky sa podarili nájsť len 2 komponenty:

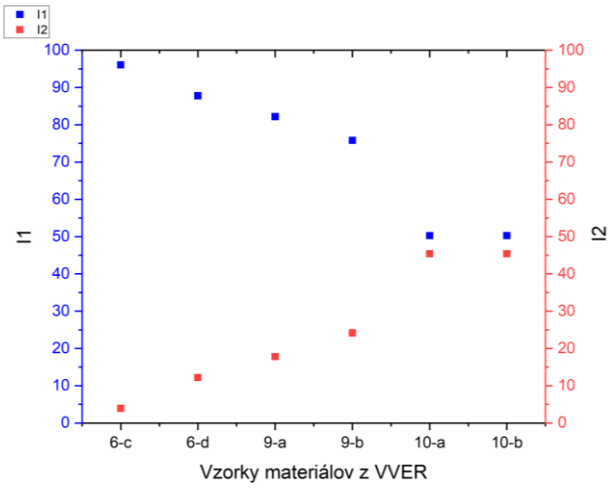
- LT1 – anihilácia pozitrónu v nepoškodenej/ resp. bezdefektnej štruktúre
- LT2 – anihilácia pozitrónu v defektoch.

Výsledky merania všetkých vzoriek sú zobrazené na Obrázku 3. Z tohto obrázku je vidieť, že niektoré vzorky (najmä tie, v ktorých sa nepodarila odseparovať vo fitovaní 3. komponenta, majú hodnotu LT1 vyššiu a teda nie je to bezdefektná štruktúra, ale štruktúra obsahujúca dislokácie s dobou života pozitrónov približne 140 ps). Vzorky a výsledky z PALS sú detailnejšie vyhodnotené a porovnané v ďalších kapitolách.



Obr. 3 Grafické znázornenie dôb života pozitrónov LT (ps) pre bezdefektnú štruktúru a dislokácie (LT1) a pre vakancie (LT2) typu mono- a di- vakancií.

Obrázok 4 zobrazuje intenzity (resp. podiel), s ktorými anihilovali jednotlivé skupiny pozitrónov. Intenzita 1 (I1) popisuje bezdefektnú štruktúru, kde anihilovala viac ako polovica pozitrónov. Intenzita 2 (I2) je proporčná ku koncentrácii defektov a mení sa so zaťažením materiálu počas prevádzky JE.



Obr. 4 Grafické znázornenie intenzity dôb života pozitrónov I (%) pre jednotlivé LT1 a LT2.

2.4 Výsledky pre meranie vzorky HCČ

Vzorka 10 pochádza z hlavného cirkulačného čerpadla, kde sa nevyskytuje priame radiačné poškodenie. Vzorka tak bola namáhaná len tepelným zaťažením a koróziou. Vzorku 10 sme si zvolili ako ukazovateľ konzistentnosti merania PALS po uplynutí časového intervalu pol roka, keďže merania DELISA-LTO vzoriek vyžadujú približne 1 rok. Vzorku z HCČ sme rozdelili na dve merania 10-a, 10-b.

Z Tabuľky 2 môžeme vidieť, že obe merania vzorky HCČ boli prakticky totožné. Mierne odchýlky merania nastávajú až na rádovo okolo 10^4 mieste, čo je zanedbateľný rozdiel v súvislosti s neistotou merania a môžeme povedať, že merania sú totožné. Tieto výsledky merania PALS sú bez iných vonkajších vplyvov na materiál a z časového hľadiska konzistentné s minimálnou odchýlkou.

Tab. 2. Doba života a intenzita vzorky 10. Neistota merania ±20%.

Vzorka	LT1[ps]	I1[%]	LT2[ps]	I2[%]
10-a	108	50,27	194	45,4
10-b	108	50,27	194	45,4

Zo spracovaných údajov môžeme vidieť, že vo vzorke sa z polovice nachádza základný materiál a defekty sú prítomné vo forme mono-vakancií.

2.5 Výsledky pre meranie vzorky kapsula svedočných vzoriek

Kapsula svedočných vzoriek je vytvorená z austenitickej ocele 08Ch18N10T s 18% chrómu a 10% niklu s vysokou korozívnou a vysokoteplotnou odolnosťou. Pri prevádzke JE sa nachádzala medzi TNR a šachtou reaktora nad aktívnou zónou, aby sledovala tepelné starnutie TNR intenzívnejšie ako radiačné starnutie. Z tohto materiálu máme 2 vzorky:

- 9-a bola v prevádzke 24 rokov,
- 9-b bola v prevádzke 32 rokov.

Výsledky z merania týchto vzoriek sú zobrazené v Tab. 3 a Obr. 3 a Obr. 4.

Tab. 3. Doba života a intenzita vzoriek 9-a, 9-b. Neistoty merania sú na úrovni ± 20%.

Vzorka	LT1[ps]	I1[%]	LT2[ps]	I2[%]
9-a	125	82,2	271	17,8
9-b	122	75,8	232	24,2

Môžeme vidieť že pri vzorke 9-b je pokles doby života pozitrónov anihilovaných vo vakanciách, z 270 ps (vzorka 9-a) na 232 ps (vzorka 9-b). Toto bol najväčší rozdiel vo vzorkách, ktorý môže naznačovať, že zvýšením doby trvania tepelného zaťaženia sa defekty zmenšili, ale ich koncentrácia sa mierne zväčšila (Viď Obr. 4). Rozdiel v strednej dobe života pozitrónov medzi vzorkou v JE po dobu 24 rokov a 32 rokov je pokles o 3 ps, čo je zanedbateľná zmena (v rámci neistoty merania), preto môžeme posúdiť, že nové defekty sa neformovali, ale zrejme sa preskupovali existujúce defekty.

2.6 Výsledky pre meranie vzorky úchytu príruby TNR

Vzorky matice a svorky uchyteného systému TNR, vzorky 6-c a 6-d, boli vystavované tlakovému a čiastočne aj radiačnému poškodeniu po dobu 30 rokov. Vzorka 6-c je skrutka a bola bližšie k reaktoru, takže predpoklad je väčšie radiačné poškodenie. Výsledky sú zobrazené v Tab. 4.

Tab. 4. Doba života a intenzita vzoriek 6-c, 6-d. Neistoty merania sú na úrovni ± 20%.

Vzorka	LT1[ps]	I1[%]	LT2[ps]	I2[%]
6-c	150	96,06	2700	3,94
6-d	150	87,80	246	12,20

Ako však môžeme z grafu (Obr. 3) vidieť, doba života LT1 má hodnotu 150 ps, čo môžeme pokladať za anihiláciu pozitrónov v dislokáciách. Intenzita týchto dôb života je v priemere 90%, čo naznačuje že v materiáli sa nenachádzajú nejaké menšie defekty typu mono-, di- alebo tri- vakancií. Pri vzorke 6-c môžeme pozorovať pri dobe života LT2 prudký nárast času anihilácie. Môžeme predpokladať, že časť týchto pozitrónov (cca 4%) anihilovali vo vzduchu pri implantácii medzi Kaptonovou fóliou a vzorkou.

Tak isto môžeme predpokladať faktor tepelného žíhania, ktorý by spôsoboval úplné vymedzenie väčších defektov. Túto teóriu by sme však mohli potvrdiť iba pri porovnaní so vzorkou z tohto istého komponentu veka reaktora, ale po rôznej dobe prevádzky.

Vzorka 6-d už vykazuje známky tri– až tetra-vakancií s intenzitou (cca 12%) ktoré by mohli byť spôsobené polohou vzorky v JE a intenzitou radiačného a tepelného poškodenia oproti vzorke 6-c.

3 Zhodnotenie výsledkov meraní

Nedeštruktívne meracie techniky nám dovoľujú predlžovať životnosť JE vďaka opakovateľnosti merania vzorky, rýchlosti vyhodnocovania a možnosti online monitoringu.

PALS je perspektívna technika na vyhodnocovanie najmenších defektov a je porovnateľná iba so zdĺhavou TEM. PALS bola využitá pri štúdovaní tepelného starnutia vzoriek z VVER reaktorov v európskom projekte DELISA-LTO.

Z meraní sme zistili, že typ, veľkosť a početnosť defektov v materiáli vysoko závisí od polohy a typu záťaže komponentu, z ktorých sme čerpali vzorky. Z výsledkov merania vzorky 9 môžeme povedať, že tepelná záťaž je prospešná pre redukovanie väčších defektov na menšie, ale s vyššou početnosťou menších defektov.

Meranie vzorky 10 nás uistilo v konzistentnosti merania pomocou PALS. Preto môžeme zaradiť túto techniku medzi spoľahlivé a presné meracie metódy.

Poďakovanie

Týmto by som chcel vyjadriť vďaku garantovi tejto práce Ing. Jane Šimeg Veterníkovej, PhD., za pomoc pri tvorbe práce, meraní vzoriek a poskytnutie potrebných dát.

Literatúra

1. Webstránka projektu DELISA-LTO. [online, cit. marec 2025] www.delisa-lto.eu

2. Slugen, V.; Brodziansky, T.; Simeg Veternikova, J.; Sojak, S.; Petriska, M.; Hinca, R.; Farkas, G. (2022). Positron Annihilation Study of RPV Steels Radiation Loaded by Hydrogen Ion Implantation. Materials, 15(20), 7091.

3. Boåsen, M., Lindgren, K., Öberg, M., & Thuvander, M. (2022). Analysis of thermal embrittlement of a low alloy steel weldment: Microstructure and mechanical properties. Engineering Failure Analysis, 131, 105833.

4. Hautojarvi, P.; Dupasquier, A.; Manninen, M.J.; Vehanen, A.; Mijnarends, P.E.; West, R.N.; Nieminen, R.M. Positrons in solids. Ed.: Hautojarvi, P., Springer-Verlag: Berlin/Heidelberg/New York, 1979, p.255, ISBN 978-3-642-81318-4.

5. Krause-Rehberg, R.; Leipner, H.S. Positrons Annihilation in semiconductors – Defect studies. Springer-Verlag: Berlin, 1997, p. 375, ISBN 3-540-64371-0.

Registrácia 14 MeV neutrónov s Si Timepix detektorom

Matej Balušík, Andrea Šagátová

Ústav jadrového a fyzikálneho inžinierstva, FEI STU v Bratislave, Ilkovičova 3, 841 04, Bratislava, Slovensko

xbalusikm@stuba.sk

Abstrakt – Tento článok sa zaoberá detekciou 14 MeV neutrónov pomocou hybridného polovodičového pixelového detektora Timepix s Si senzorom. Hlavným cieľom experimentu bolo vyhodnotiť citlivosť a účinnosť detektora pri identifikácii rýchlych neutrónov vznikajúcich z D-T reakcie. Experimentálne výsledky naznačujú, že detektor Timepix2 s kremíkovým senzorom je schopný detegovať rýchle neutróny, pričom efektívnosť konverzného materiálu bola potvrdená zvýšením detegovaných udalostí v mieste jeho aplikácie na detektor. Tento výskum poskytuje dôležité podklady pre ďalšie využitie detektorov Timepix v oblasti radiačnej detekcie vrátane medicíny a výskumu v oblasti jadrovej fyziky.

1 Úvod

Hybridné polovodičové pixelové detektory rodiny Timepix sa čoraz viac využívajú na rôzne merania, vrátane charakterizácie zloženia zmiešaných radiačných polí v zložitých radiačných prostrediach na vysokovýkonných urýchľovačoch, v rádioterapii a vo vesmírnom prostredí. Detektory Timepix umožňujú radiačné zobrazovanie, spektrometriu širokého rozsahu typov žiarenia a sledovanie jednotlivých častíc, a to vďaka presnému meraniu odovzdanej energie a energetických strát, polohy interakcie a smeru trajektórie každej jednej častice.

Skúmali sme odozvu detektora Timepix2 (Si) na rýchle neutróny s dobre definovaným spektrom. Na prípravu neutrónov sme využili jadrovú reakciu vyvolanú deutériom na tríciovom terči T(d,n)α. Signál z detektorov sme analyzovali z hľadiska energetickej a morfologickej štruktúry stôp zaregistrovaných častíc v dvojdimenzionálnom poli 256×256 pixelov detektora s cieľom klasifikácie typov časticových udalostí pre budúce určenie zloženia neznámych radiačných polí. Na základe rozlíšených tried udalostí v detektore sme určili detekčnú účinnosť registrácie neutrónov, pričom kremíkový detektor bol upravený štyrmi rôznymi typmi konverzných vrstiev na báze polyetylénu (PE) pre zefektívnenie detekcie neutrónov.

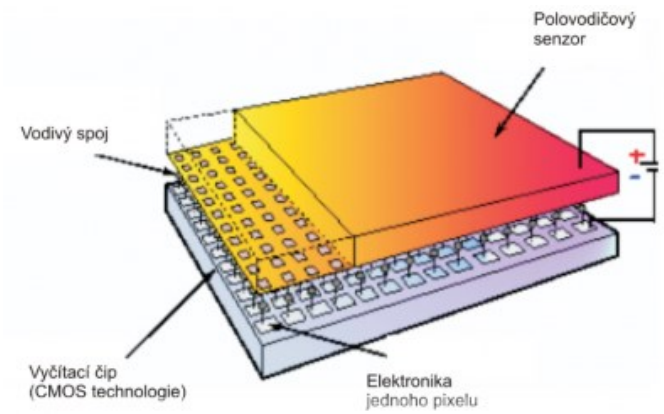
2 Teória

Pixelový detektor typu Timepix je srdcom radiačnej kamery MiniPIX, ktorá je určená na zobrazovanie a detekciu ionizujúceho žiarenia (Obr.1). Kamera sa skladá z troch základných stavebných blokov:

Senzor -sa skladá z polovodičového monolytu, na ktorom je vytvorená dvojdimenzionálna pixelová štruktúra s 256×256 pixelov s veľkosťou 55×55µm.

Vyčítavacia elektronika – určená na prenos a spracovanie dát zo senzora realizovaná ako Timepix2 CMOS čip, vyvíjaný kolaboráciou viacerých výskumných inštitútov a krajín s CERNom.

Rozhranie detektor-počítač –na ovládanie prenosu dát z detektora do počítača a ovládanie nastavení detektora realizované cez dosku plošného spoja a USB pripojenie k počítaču.



Obr.1 – Schéma detektora Timepix

Princíp činnosti

Keď ionizujúce žiarenie dopadne na senzorovú vrstvu, dochádza k **ionizácii materiálu** a vzniku párov elektrón-diera. Tieto náboje sú elektrickým poľom usmernené k príslušným elektródam, kde sa elektronicky zaznamenávajú pre každú zaregistrovanú časticu zvlášť. Timepix čip pracuje v troch základných režimoch: Time-over-Threshold (ToT) meria dĺžku času, počas ktorého je signál nad prahovou hodnotou, pričom ToT je úmerný energii prenesenej časticou do každého pixelu; Time-of-Arrival (ToA) zaznamenáva presný čas príchodu častice, čo je užitočné pre časovú koreláciu udalostí; a Counting Mode, ktorý spočíva v počítaní častíc dopadajúcich na jednotlivé pixely a využíva sa pri radiačnom zobrazovaní.

2.1 Interakcia neutrónov s látkou

Interakcie neutrónov s látkou rozdeľujeme na dva hlavné spôsoby: rozptyl, pri ktorom dochádza k zrážke neutrónu a terčového jadra a následnej zmene smeru a k prerozdeleniu kinetickej energie medzi časticou a jadrom, a absorpcia, pri ktorej môže dôjsť k nepružnému rozptylu, radiačnému záchytu,

záchytu s vyslaním častice alebo k štiepeniu jadra. Interakcia pri ktorej sa neutróny neabsorbujú je pružný rozptyl.[1]

2.1.1 Nepružný rozptyl

Nepružný rozptyl neutrónov nastáva, keď rýchly neutrón (typicky s energiou nad niekoľko stoviek keV alebo v oblasti MeV) narazí na terčové jadro, pričom odovzdá časť svojej kinetickej energie a excitáciou ho privedie na vyššiu energetickú úroveň. Na rozdiel od pružného rozptylu, pri ktorom jadro zostáva v základnom stave, nepružný rozptyl vedie k dočasne excitovanému jadru. Táto interakcia je sprostredkovaná jadrovou silou a závisí od jadrového potenciálu. Jadro pri nej získa dostatočnú energiu na prechod do excitovaného stavu. [1]

2.1.2 Pružný rozptyl

Celková kinetická energia sústavy sa zachováva, dochádza len k jej prerozdeleniu medzi časticami sústavy. Neutrón odovzdá časť svojej kinetickej energie inej častici a zmení smer pohybu (výsledkom je spomaľovanie neutrónov). Je to dominantný mechanizmus interakcie rýchlych > 0,1 MeV neutrónov s látkou a ľahkých jadier A < 20 (spôsobí ich spomaľovanie). Keďže nedochádza k vzniku zloženého jadra dochádza k rozptylu na potenciáli jadra. Táto interakcia sa dá zjednodušene vysvetliť pomocou mechanizmu zrážky dvoch dokonale tuhých gúľ.

2.1.3 Záchyt s emisiou častice

Podobne ako pri radiačnom záchyte vzniká zložené vzbudené jadro pri ktorého deexcitácií dôjde k uvoľneniu alfa častice alebo protónu. Tento jav je možný iba pri ľahkých jadrách, ktorých potenciálová bariéra na uvoľnenie kladne nabitej alfa častice alebo protónu je nízka. Napr.:

$$n + {}^{3}He \longrightarrow {}^{3}H + {}^{1}H$$

alebo

$$n + {}^{10}B \longrightarrow {}^{7}Li + {}^{4}He$$

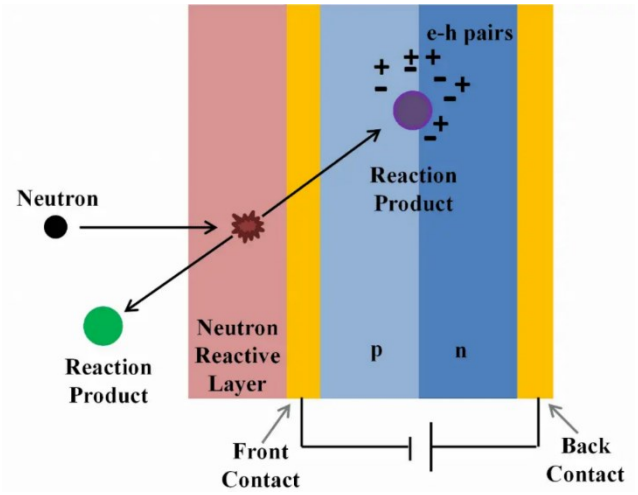
2.2 Detekcia neutrónov v Si senzore

Rýchle neutróny môžu byť detegované vďaka jadrovým reakciám, pri ktorých vznikajú nabité častice ako produkty reakcie alebo prostredníctvom konverzie na nabité častice pružným rozptylom s ľahkými jadrami ako napr. jadro vodíka.

Keď neutróny dopadajú na polovodič, reagujú s atómami kremíka (Si), pričom tieto atómy môžu byť buď rozptýlené pružným rozptylom alebo môže neutrón na nich vyvolať jadrovú reakciu. Výsledné energetické ióny sú absorbované v aktívnej oblasti Si, kde vytvárajú elektrón-dierové páry, ktorých počet je úmerný energii iónu. Pomocou vnútorného elektrického poľa sú tieto nosiče náboja zachytávané pixelmi a odvádzané čítacou elektronikou, čím vzniká prúdový impulz.[2]

Hlavné reakcie, ktoré ovplyvňujú odozvu detektora na neutróny, sú interpretované ako pružný a nepružný rozptyl ^{28}Si(n,n')^{28}Si. [3]

Ďalšie interakcie neutrónov v kremíku sú spôsobené priamymi jadrovými reakciami, ako sú (n,α) a (n,p) na izotope 28-Si, a reakciami na konverzných vrstvách detektora. Produkty týchto reakcií zahŕňajú protóny a alfa častice, ktoré následne interagujú a sú detegované priamo v kremíkovom senzore (Obr. 2). [3]



Obr.2 – Interakcia neutrónov s konverznými vrstvami na Si detektore

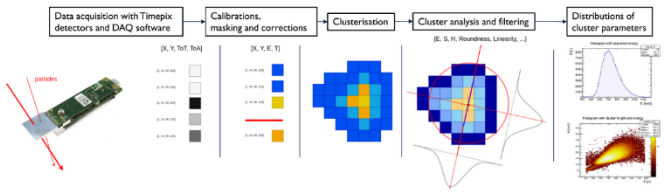
Ďalšie neutrónmi indukované interakcie zahŕňajú pružný rozptyl a neionizačné interakcie prebiehajúce priamo v kremíkovej matici senzora. Predpokladá sa, že tieto procesy prebiehajú homogénne v celom objeme senzora. (Tab.1)

Tab.1 Materiálové vlastnosti Si	Si
Hustota [g/cm^3]	2.3
Šírka zakázaného pásma [eV]	1.12
Energia potrebná na vytvorenie e dierového páru [eV]	3.6

2.3 Spracovanie dát

Data Processing Engine (DPE) resp. TraX Engine [1], je softvér vyvinutý spoločnosťou Advacam slúžiaci na spracovanie a analýzu údajov z detektorov Timepix.

DPE vykonáva kalibráciu a korekciu údajov s cieľom transformovať surové dáta z pixelov na súbory korelovaných pixelových klastrov (Obr. 3). Tento proces zahŕňa aplikáciu potrebných kalibrácií a korekcií, analýzu klastrov a vytvorenie distribúcií ich parametrov ako napríklad energetické spektrá, LET alebo informácie o morfológii klastrov (Tab.2). Okrem toho DPE umožňuje klasifikáciu častíc, identifikáciu ich typu a energie,.[4]



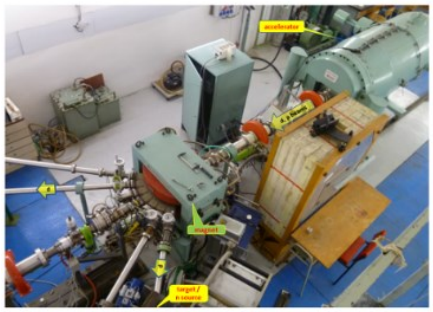
Obr. 3 – Diagram zobrazujúci jednotlivé stupne spracovania dát pomocou DPE.

3 Experiment

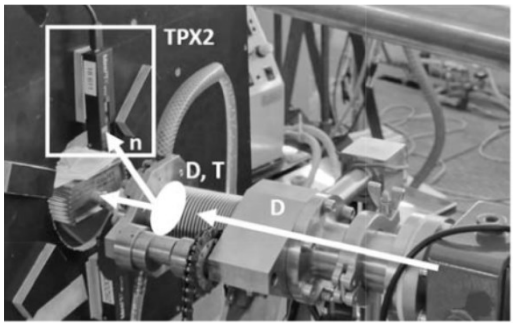
Hlavným cieľom experimentu bola detekcia rýchlych neutrónov produkovaných v DT (deuterium-trícium) reakcii v Timepix detekotre.

$$D + T \rightarrow {}^{4}_{2}He\ (3.52\ MeV) + n\ (14.06\ MeV)$$

Experimenty sme uskutočnili na urýchľovači Van de Graaff (VdG) na Ústave experimentálnej a aplikovanej fyziky ČVUT v Prahe (Obr. 4, 5), kde bola energia neutrónov dopadajúcich na detektory (pod uhlom 90°) 14.7 MeV (deuteróny 160keV) a v laboratóriách CEA Cadarache vo Francúzsku (Obr. 6), na detektory (pod uhlom 90°) s energiou neutrónov 14.05 MeV (deuteróny 45keV) na detektore. Na zvýšenie detekcie neutrónov boli jednotlivé časti Si detektora pokryté štyrmi vrstvami konvertorov (Obr.7).



Obr.4 Fotografia VdG urýchľovača



Obr.5 – Usporiadanie Timepix detektora s VdG urýchľovačom



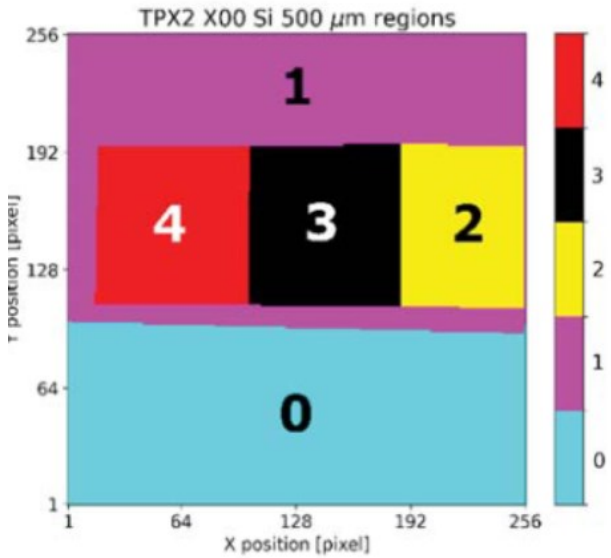
Obr.6 - Usporiadanie Timepix detektora v CEA Cadarache

Na detekciu neutrónov sme použili MiniPIX Timepix2 detektor s kremíkovým senzorom s hrúbkou 500µm so 4 rôznymi konverznými vrstvami (Obr. 7). Detekcia neutrónov prebiehala rôznymi mechanizmami. Neutróny sa rozptyľovali na jadrách polovodiča alebo s nimi interagovali prostredníctvom jadrových reakcií a následne vytvárali nabité častice registrovateľné v aktívnom objeme detektora. Cieľom tohto experimentu bolo vyhodnotiť detekčnú účinnosť detektora pre rýchle neutróny a jeho schopnosť rozlišovať medzi neutrónmi indukovanými signálmi a inými typmi žiarenia.

Detekcia rýchlych neutrónov v detektore Timepix2 bola charakterizovaná podľa oblastí detektora s rôznymi konverznými vrstvami a bez konverznej vrstvy. Z meraní vyplývajú zistenia, že určité udalosti, najmä tie, ktoré produkujú ťažké nabité častice, sa vyskytli vo väčšej miere v oblastiach detektora s konverznými vrstvami, čo naznačuje, že reakcia detektora na neutróny bola sústredená v oblastiach, kde boli interakcie pravdepodobnejšie, a nie rovnomerne rozložené cez senzor.

Tieto neutróny vytvorili multipixelové stopy, známe ako klastre, s odlišným tvarom ako stopy iných druhov žiarenia, ktoré boli rovnomernejšie rozložené v celom detektore.

Vrstva konvertora v systémoch detekcie neutrónov, ako je napríklad ten, ktorý sa používa v detektore Timepix2, hrá rozhodujúcu úlohu pri zvyšovaní účinnosti detekcie a presnosti interakcií neutrónov. Jeho primárnou funkciou je zvýšiť pravdepodobnosť registrácie neutrónov detektorom, keďže neutróny sú neutrálne častice, ktoré priamo neionizujú materiál detektora. Vrstva konvertorov pokrývajúca spoločnú elektródu detektora Timepix2 bola použitá na premenu neutrónov na ionizujúce častice, ktoré môže detektor priamo registrovať. Vrstvy polyetylénu (PE) s hrúbkou 50 µm, 100 µm a 150 µm boli upevnené pomocou 10 µm hrubej polyamidovej fólie Kapton($C_{22}H_{10}N_2O_5$), na ktorej sa neutróny elasticky rozptyľujú na vodíku.(Obr.7, 8)[3]



Obr.7 - Diagram rozloženia konvertorov na senzore Kapton - 10 µm (1); PE - 50 µm (2), 100 µm (3) a 150 µm (4) [3]

Výber materiálov pre vrstvu konvertora sa líši v závislosti od typu detegovaných neutrónov. Pre tepelné neutróny sa používajú materiály ako lítium-6 (6Li) a bór-10 (10B), pretože pri absorpcii neutrónov vytvárajú nabité častice. Pre rýchle neutróny sa na ich registráciu používajú materiály bohaté na vodík, ako je polyetylén (PE), kde sa neutróny rozptyľujú na ľahkých jadrách, ktoré následne ionizujú atómy senzora.

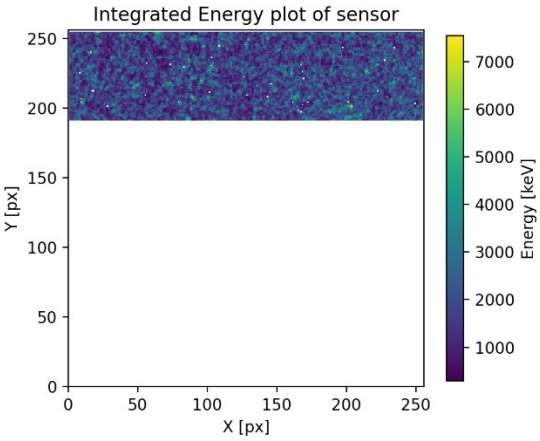
Pre experiment a následné spracovanie dát som použil špecifické filtre, vďaka ktorým sa nám podarilo selektívne odfiltrovať iba relevantné stopy žiarenia produkovaného interakciou neutrónov v detektore. Tieto filtre umožnili odstrániť nežiadúce signály, pozadie a šum, čo výrazne prispelo k presnejšej analýze získaných výsledkov.

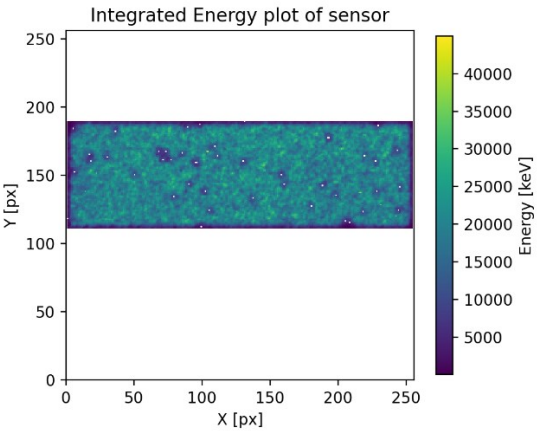
4 Výsledky

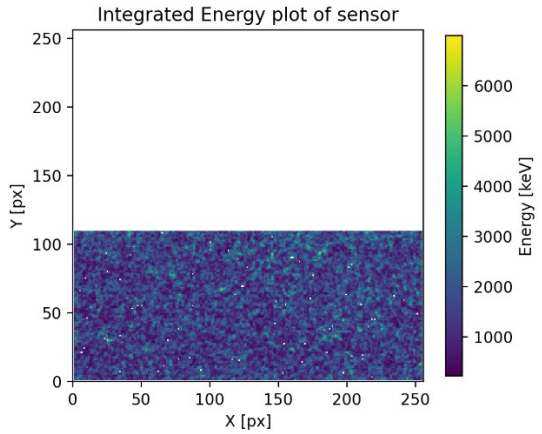
Experiment detekcie rýchlych neutrónov detektorom Timepix2 (Si) preukázal úspešnú detekciu a vizualizáciu radiačného poľa rýchlych neutrónov pri energiách 14 MeV. Pomocou detektora sme vizualizovali stopy jednotlivých registrovaných častíc ako klastre. Dôležitým krokom bola separácia neutrónmi indukovaných stôp z rôznych typov klastrov častíc na základe energie, veľkosti a tvaru, ktoré boli analyzované pomocou algoritmov na rozpoznávanie vzorov. Na základe týchto parametrov sme určili dve skupiny klastrov zodpovedajúcich interakciám neutrónov v materiáli detektora pre Timepix2.(Obr.9)

Tab.2: Parametre filtra stôp častíc

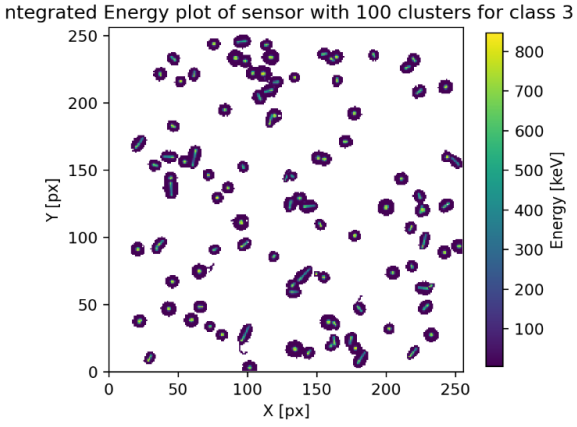
Energia [keV]	Veľkosť [px]	Linearita [0-1]	Oblosť [0-1]	Hrúbka [0-1]
50-20000	6-200	0-0.8	0.2-1.0	0-0.4

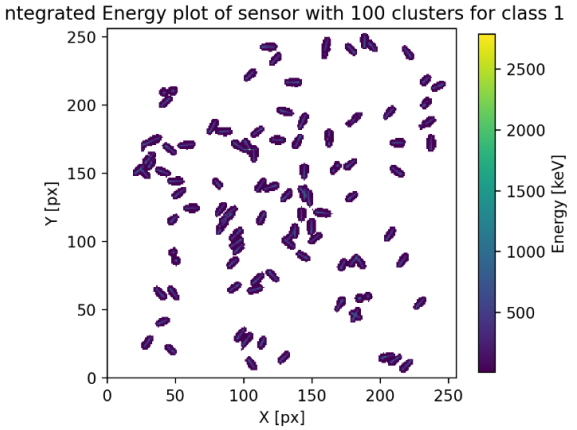






Obr.8 – Rozdelenie jednotlivých častí konvertora na Si detektore





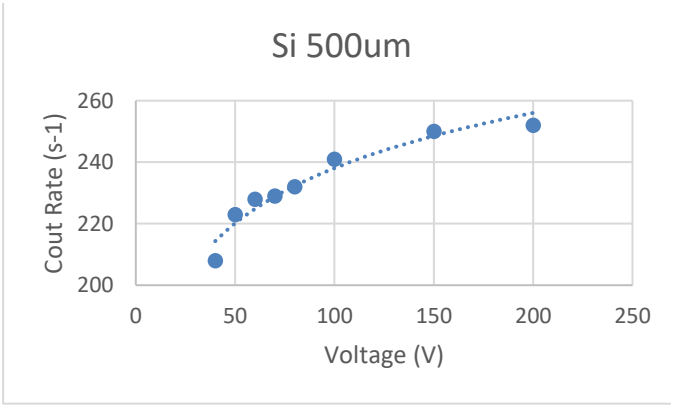
Obr. 9 Dva typy klastrov zodpovedajúcich interakciám neutrónov v materiáli detektora

Detekčná účinnosti jednotlivých vrstiev Timepix2 detektora pokrytého konverznými vrstvami sa menila podľa očakávania.(Obr.8) Použitá konvertorová vrstva zvýšila pravdepodobnosť detekcie rýchlych 14 MeV neutrónov o 31.8% v prípade PE vrstvy a o 17.1% v prípade Kapton fólie (Tab. 3). Výsledok experimentu teda potvrdzuje efektívnosť konverzných vrstiev pri zlepšovaní detekcie. Zároveň sa ukázalo, že hrúbka konvertorových vrstiev nemala výrazný vplyv na samotnú detekciu, čo naznačuje, že optimalizácia hrúbky konvertorov nie je kritickým faktorom pre tento typ merania.

Tab. 3: Porovnanie výsledkov dvoch experimentov

Energia neutrónov [MeV]	Pozícia (riadky) [px]	Početnosť [s^{-1}]	Počet udalostí na pixel [$s^{-1}px^{-1}$]	δ [%]
14.05	0-110	130.61	1.187	0
(CEA)	110-190	138.84	1.731	45.8
	190-255	89.57	1.378	16.1
14.70	0-110	108.05	0.982	0
(VdG)	110-190	90.23	1.127	14.8
	190-255	74.36	1.144	16.4

V rámci experimentu sme taktiež vyhodnotili vplyv napätia na množstvo detegovaných udalostí. Z napäťovej závislosti vidíme postupné zvyšovanie detekčnej účinnosti s rastúcim napätím (Obr. 10). So zvyšujúcim napätím počet udalostí postupne saturuje. S rastúcim záverným napätím sa prehlbuje šírka zakázaného pásma a teda sa rozširuje aktívna oblasť detektora. Napätie, pri ktorom je táto oblasť najväčšia volíme ako pracovné napätie.



Obr.10 – Vplyv napätia na detekciu udalostí

5 Záver:

V rámci experimentu sme vyhodnotili odozvu rýchlych neutrónov v Si detektore rodiny Timepix. Zdrojom detegovaných neutrónov bola D-T reakcia. Neutróny z tejto reakcie môžeme považovať za monoenergetické s energiou 14.7MeV a 14.05MeV. Vďaka výhodám hybridných pixelových detektorov Timepix sme dokázali zobraziť, analyzovať, odfiltrovať a potlačiť vplyv pozadia urýchľovača na meranie. Následne sme úspešne určili a zobrazili produkty interakcie neutrónov s detektorom. Medzi tieto produkty patrili častice alfa, nabité ióny alebo protóny. Odlíšenie detegovaného žiarenia bolo založené hlavne na morfológií a energií jeho stôp v detektore. Na detektore Timepix2 sme vyhodnocovali tri

oblasti plochy detektora, holú časť, časť s konvertorom z polyetylénu a časť s konvertorom typu Kapton. Kapton slúžil na fixovanie PE konverzných vrstiev, preto sme neočakávali významný vplyv na detekciu rýchlych neutrónov. Polyetylénová vrstva bola určená práve na zlepšenie detekcie rýchlych neutrónov. Výsledky sa zhodujú s našimi predpokladmi, v časti detektora bez konverznej vrstvy bolo zaznamenaných najmenej udalostí. Polyetylénová vrstva zvýšila počet udalostí oproti holej časti v priemere o 31.8%. Rovnako došlo k zlepšeniu detekcie aj v časti s vrstvou Kapton v ktorej sa v porovnaní s holou časťou zvýšil počet detegovaných udalostí o 17.1%. Pomocou detektora Timepix2 sme úspešne detegovali a zobrazili produkty interakcie rýchlych neutrónov. Tento experiment poukazuje na schopnosť detektorov kvantifikovať a zobrazovať radiačné polia ktorých súčasťou sú neutróny. Tento typ detektorov by mohol byť taktiež využitý pri zobrazovacích metódach pomocou neutrónov.

Literatúra

[1] GEIST, William H.; SANTI, Peter A.; SWINHOE, Martyn T. (ed.). *Nondestructive Assay of Nuclear Materials for Safeguards and Security*. Springer Nature, 2024.

[2] LIU, Linyue, et al. Radiation resistance of silicon carbide Schottky diode detectors in DT fusion neutron detection. *Scientific Reports*, 2017, 7.1: 13376.

[3] GRANJA, C., et al. Detection of fast neutrons with the pixel detector Timepix3. *Journal of Instrumentation*, 2023, 18.01: P01003.

[4] M. Lukas et al., *Data Processing Engine (DPE): data analysis tool for particle tracking and mixed radiation field characterization with pixel detectors Timepix*, 2024 *JINST* **19** C04026.

[5] HA, Jang Ho, et al. 4H-SiC PIN-type semiconductor detector for fast neutron detection. *Prog. Nucl. Sci. Technol*, 2011, 1: 237-239.

Nelineárna rezonancia magneticky budeného kyvadla

Roman Michelko[1], Peter Bokes[1]

[1]Ústav jadrového a fyzikálneho inžinierstva, Fakulta elektrotechniky a informatiky, Slovenská technická univerzita v Bratislave, Slovenská Republika

xmichelko@stuba.sk

Táto práca sa venuje dynamike fyzikálneho kyvadla s nelineárnym magnetickým budením. Takýto systém vykazuje pozoruhodné nelineárne javy, akými sú subharmonická rezonancia, samo-adaptívnosť a existencia diskrétnych amplitúd na ktorých sa kmitanie ustáli. Predstavená je jednoduchá experimentálna realizácia kyvadla umožňujúca pozorovanie spomínaných javov. Napokon je uvedený jednoduchý matematický model poskytujúci správne kvalitatívne predpovede pre pozorované javy.

1 Úvod

Bohatá dynamika nelineárnych procesov, nedosiahnuteľná vrámci lineárnych systémov a priblížení, sa prirodzene objavuje v mnohých odvetviach fyzikálneho bádania, od klasickej mechaniky, atmosferickej fyziky až po komplikované biofyzikálne procesy v živých organizmoch. Hoci už francúzsky matematik H. Poincaré na prelome 19. a 20. storočia identifikoval niektoré typické prejavy nelineárnych systémov (ako chaos a citlivosť na počiatočné podmienky), ich štúdium sa dostalo do popredia až s nástupom výkonnej výpočetnej techniky, o čom svedči aj dve Nobelove ceny udelené Nicolasovi Bloembergenovi v roku 1981 za prácu v oblasti nelineárnej optiky a Syukurovi Manabemu a Klausovi Hasselmannovi v roku 2021 za nelineárne modely Zemskej klímy.

Príkladom nelineárneho systému, na ktorom je možné demonštrovať nelineárne javy subharmonickej rezonancie, samo-adaptívnosti a existencie diskrétnych stabilných amplitúd kmitov, je fyzikálne kyvadlo budené nelineárnou magnetickou silou. Subharmonická rezonancia je jav, pri ktorom systém vykonáva stabilné kmity s frekvenciou blízkou vlastnej frekvencii ω_0, pri vyšších budiacich frekvecniách $\omega > \omega_0$. Samoadaptívnosť systému sa preukazuje nezávislosťou veľkosti stabilnej amplitúdy kmitov v širokom rozsahu intenzity budenia. Experimentálne štúdie poukázali na súvis medzi týmito javmi a symetriou budiacej sily [1]. Teoretickým aspektom nelineárnej dynamiky takéhoto systému sa venovala práca [2].

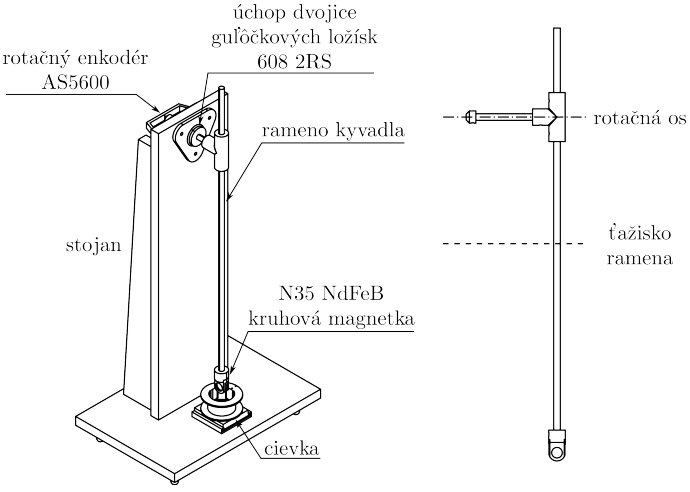
Prvá časť práce sa venuje experimentálnej realizácii magneticky budeného fyzikálneho kyvadla s diskusiou spomínaných pozorovaných nelineárnych javov. Zostrojené kyvadlo má potenciál slúžiť ako pomôcka pre demonštráciu bohatých prejavov nelineárnej dynamiky. Druhá časť práce je venovaná jednoduchému fyzikálnemu modelu kyvadla, ktorý dokáže poskytnúť správne kvalitatívne predpovede.

2 Experimentálna realizácia kyvadla

Za účelom experimentálneho pozorovania nelineárnej dynamiky magneticky budeného kyvadla, bola navrhnutá a realizovaná jednoduchá konštrukcia kyvadla pozostávajúca zo stojana, ramena kyvadla s malou magnetkou na jeho voľnom konci a cievky. Cievka je napájaná pomocou jednoduchého obvodu zosilnejúceho harmonický prúd z generátora signálov. Nasledovné časti práce sa venujú detailnejšiemu popisu jednotlivých častí kyvadla.

2.1 Stojan kyvadla

Vlastná konštrukcia kyvadla je zobrazená na izometrickom nákrese na obrázku 1. Stojan kyvadla je vytvorený z jedného kusa viacvrstvovej lepenej dosky. Podstava kyvadla má rozmery 200 × 295 mm. Doska kolmá na podstavu má výšku 500 mm a šírku 140 mm. Rameno kyvadla je na dosku uchopené pomocou dvojice jednoradových guľôčkových ložísk *608 2RS* s nízkym trením, umiestnených v plastových uchyteniach vytvorených pomocou 3D tlače. Rovina kmitov ramena sa nachádza vo vzdelanosti 55 až 60 mm od povrchu zvislej dosky. Celá konštrukcia kyvadla je postavená na štvorici nastaviteľných nožičiek umožňujúcich presné vodorovné umiestnenie kyvadla.



Obr. 1: Izometrický nákres magnetického kyvadla a bočný nákres ramena kyvadla

2.2 Rameno kyvadla

Konštrukcia ramena kyvadla, zobrazená na nákrese na obrázku 1, pozostáva z dvojice navzájom kolmých drevených osí v tvare valca s priemerom ⌀ 8 mm. Dlhšia os má dĺžku 500 mm, kratšia 120 mm. Osi sú spojené plastovou T-spojkou. Na spodnom konci dlhšej osi je v plastovom držiaku upevnená NdFeB kruhová magnetka *N35* orientovaná tak, aby jej rotačná os ležala v rovine kmitov. Na zadnom konci kratšej osi je v plastovom držiaku upevnená malá magnetka slúžiaca na meranie výchylky kyvadla magnetického rotačného senzora *AS5600*. Vďaka jednoduchej geometrii ramena kyvadla je možné na základe *Steinerovej vety* jednoducho odhadnúť moment zotrvačnosti ramena $I_{\mathrm{r},O}$ v mieste rotačnej osi O a vzdialenosť ťažiska od rotačnej osi d_{T}.

$$I_{\mathrm{r},O} = \sum_{i=1} I_i + m_i d_{O,i}^2, \quad d_{\mathrm{T}} = \frac{1}{M} \sum_{i=1} m_i d_{O,i}, \tag{1}$$

kde v sumách sčítavame cez všetky dielčie komponenty ramena, pričom M je celková hmotnosť kyvadla, $d_{O,i}$ označuje vzdialenosť ťažiska komponentu i od rotačnej osi O, I_i zodpovedá momentu zotrvačnosti i-tého komponentu v mieste jeho ťažiska a m_i zodpovedá jeho hmotnosti. Na základe týchto hodnôt je možné určiť vlastnú frekvenciu malých kmitov kyvadla ako:

$$\omega_0 = \sqrt{\frac{Mgd_{\mathrm{T}}}{I_{\mathrm{r},O}}}. \tag{2}$$

Takto určená hodnota vlastnej frekvencie sa nachádza v nasledovnej tabuľke.

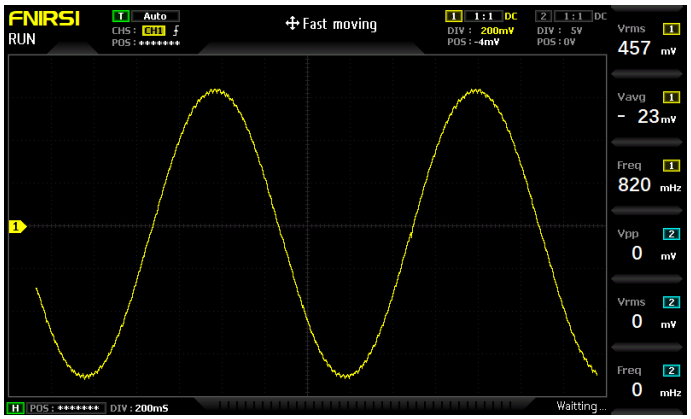
Tab. 1: Odhad vlastnej frekvencie fyzikálneho kyvadla

$I_{\mathrm{r},O}$ (kg.m^2)	d_{T} (mm)	T (s)	ω_0 (s^{-1})
$2{,}55 \times 10^{-3}$	170	1,22	5,17

2.3 Cievka a budiaci obvod

Ukazuje sa, že orientácia cievky voči magnetke zohráva kľúčovú úlohu v dynamike kyvadla [1] [2]. V prípade tejto práce bola zvolená orientácia cievky, pri ktorej je os cievky rovnobežná so zvislou osou. Cievka bola navinutá zo 100 závitov smaltovaného medeného drôtu s priemerom ⌀ 0,8 mm na plastovej špuľke. Vnútorný priemer cievky je ⌀ 37 mm a výška cievky je 22 mm. Pre návrh obvodu cievky bolo potrebné približne poznať jej odpor a indukčnosť. Odpor cievky určený meraním bol rovný približne $R_{\mathrm{cievka}} \approx 0{,}5\ \Omega$. Indukčnosť cievky $L_{\mathrm{cievka}} \approx 325\ \mu\mathrm{H}$ bola odhadnutá na základe empirického vzťahu uvedeného v článku [3].

Pre napájanie cievky nízkofrekvenčným harmonickým prúdom bol navrhnutý dvojstupňový zosilňovač triedy AB. Obvod bol navrhnutý tak, aby bol schopný dodávať cievke stály harmonický prúd s maximálnou efektívnou hodnotou 0,5 A, pri čo najmenšom možnom skreslení. Koncový stupeň obvodu pozostával z *push-pull* symetrického zapojenia páru výkonových bipolárnych tranzistorov NPN *BD139* a PNP *BD140*. Bázový obvod tranzistorov bol riadený pred-zosilňovacím stupňom, pozostávajúcim z duálneho *master-slave* zapojenia operačného zosilňovača *LM358*, s linearizačnou zápornou spätnou väzbou voči záťaži. Zosilňovacia časť obvodu bola napájaná napätím ±9 V z obvodu *virtuálnej zeme* pozostávajúceho z dvojice lineárnych regulátorov *LM317* a *LM337*. Pre napájanie obvodu jednosmerným napätím 18 V bol používaný laboratórny zdroj napätia, *WANPTEK NPS306W*. Ako zdroj harmonického signálu bol používný digitálny generátor signálov *FeelTech FY3200S*. Nameraný priebeh napätia meraného na 1 Ω rezistore zapojeného do série s cievkou, je zobrazený na nasledovnom obrázku.



Obr. 2: Priebeh napätia na predradenom 1 Ω rezistore, pripojenom na výstup budiaceho obvodu pri budiacej frekvencii $\omega = \omega_0$

2.4 Meranie výchylky kyvadla

Pre meranie výchylky amplitúdy sa využíva magnetický rotačný senzor *ams OSRAM AS5600*. Tento senzor integrovaný na malej doštičke je umiestnený voči malej magnetke s priemerom ⌀ 3,5 mm v plastom držiaku na konci rotačnej osi kyvadla. Senzor využíva Hallove senzory pre bezkontaktné meranie orientácie magnetky s 12-bitovým rozlíšením a maximálnou chybou v ideálnych podmienkach ±1° [4]. Senzor je pomocou *I^2C rozhrania* pripojený na vývojovú dosku *Arduino MICRO*, ktorá umožňuje okamžitú sériovú komunikáciu s počítačom pomocou *USB*.

Sériovú komunikáciu a spracovávanie údajov zo senzora zabezpečuje na strane vývojovej dosky kód `angle_mesurement.ino` v jazyku C. Tento kód po kalibrácii rovnovážnej polohy kyvadla, vykonáva v krátkych časových intervaloch $\Delta t = 10$ ms, merania výchylky kyvadla. Okamžitú nameranú výchylku θ v uhlovej miere spolu s časom, kedy bola táto hodnota meraná, v milisekundách následne odosiela do počítača.

Na strane počítača sú namerané hodnoty spracovávané Python kódom `pendulum_measurement.py`. Prijaté hodnoty sú v reálnom čase vynášané na graf okamžitej výchylky ramena kyvadla $\theta(t)$. Na základe dvojice po sebe idúcich meraní sa v každom časovom okamžiku numericky odhadne okamžitá rýchlosť ramena kyvadla ako:

$$\dot{\theta}(t) \approx \frac{\theta(t) - \theta(t - \Delta t)}{\Delta t}. \tag{3}$$

Hodnoty $\theta(t)$, $\dot{\theta}(t)$ a t sú napokon programom uložené do *CSV* dátového súboru.

2.5 Meranie malých kmitov

Pri malej počiatočnej výchylke $\theta_0 < 5°$ bude kyvadlo pri nulovom prúde cievkou vykonávať malé tlmené kmity. Jeho pohyb je opísaný pohybovou rovnicou v tvare:

$$\ddot{\theta} + 2\beta\dot{\theta} + \omega_0^2\theta = -\mu \operatorname{sign}\dot{\theta}, \tag{4}$$

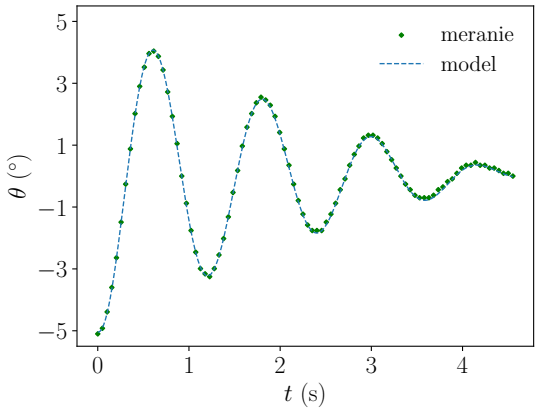
kde koeficient β súvisí s viskóznym trením a koeficient μ súvisí s valivým trením v ložiskách. V prípade pod-kritického tlmenia, kedy platí $\omega_0 > \beta > \mu > 0$, je možné približné riešenie rovnice vyjadriť ako [5]:

$$\theta(t) \approx \left(-\frac{2\mu}{\pi\beta\omega_0} + \left(\theta_0 + \frac{2\mu}{\pi\beta\omega_0}\right)e^{-\beta t}\right)\cos(\omega_0 t). \tag{5}$$

Analytický predpis približného riešenia bol použitý ako model pre určenie parametrov ω_0, β a μ na základe merania, metódou nelineárnych najmenších štvorcov využívajúcich *Levenberg–Marquardtov algoritmus*. Namerané body spolu s nájdenou optimálnou krivkou sa nachádzajú na grafe na obrázku 3. Nájdené optimálne parametre spolu s odhadmi neistôt určených z kovariančnej matice fitu sú uvedené v nasledovnej tabuľke.

Tab. 2: Parametre kyvadla určené z merania malých kmitov

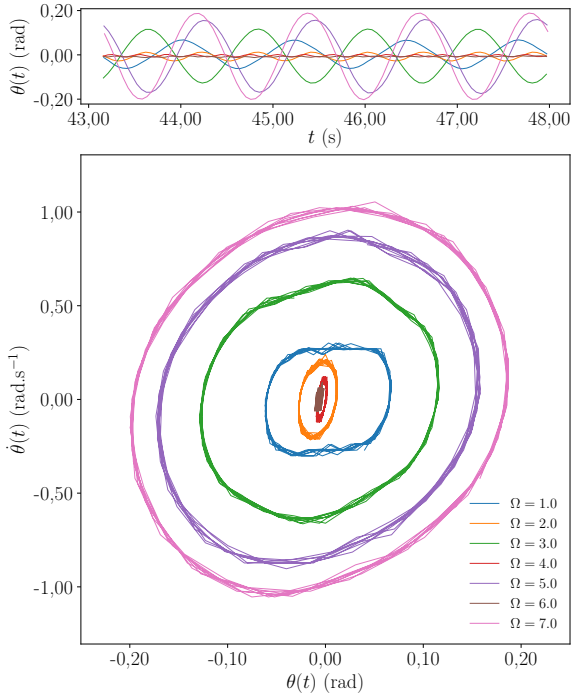
ω_0 (s^{-1})	σ_{ω_0} (s^{-1})	β (s^{-1})	σ_β (s^{-1})	μ (s^{-2})	σ_μ (s^{-2})
5,23	$3{,}48 \times 10^{-3}$	0,22	$8{,}68 \times 10^{-3}$	0,09	$2{,}82 \times 10^{-3}$



Obr. 3: Meranie malých kmitov kyvadla

2.6 Meranie kmitov s diskrétnymi stabilnými amplitúdami

Jedným z očakávaných prejavov nelineárnej dynamiky študovaného kyvadla je vyššie spomenutý jav subharmonickej rezonancie. Pri tomto jave sú pozorované stabilné kmity kyvadla pri frekvencii blízkej vlastnej frekvencii kyvadla ω_0 a budiacej frekvencii s frekvenciou $\omega = \Omega\omega_0$, kde pre pomer budiacej a vlastnej frekvencie platí $\Omega > 1$. Aby boli kmity kyvadla stabilné musia byť energetické straty kyvadla spôsobené trením



Obr. 4: Horný panel - Merané časové priebehy výchylky kyvadla v posledných 5 sekundách merania. Dolný panel - Merané fázové portréty kmitov kyvadla pri siedmich frekvenciách budenia

vykompenzované budiacou silou pôsobiacou na magnetku pri prelete malou oblasťou nad cievkou. Ako bude ukázané neskôr, budiaca sila $F(\theta)$ je v študovanom prípade párnou funkciou voči rovnovážnej polohe kyvadla $\theta = 0$. Ak je v tomto prípade magnetka pri prelete nad cievkou urýchľovaná v smere svojho pohybu, potom aby bola nad cievkou synchrónne urýchľovaná aj v opačnej pol-perióde svojho pohybu, musí byť pôsobiaca sila opačná. Preto sa fáza harmonického budiaceho prúdu v cievke musí posunúť o nepárny celočíselný násobok π. [2]. Stabilné kmity kyvadla potom môžu nastať pri frekvenčných pomeroch:

$$\Omega = 2n - 1, \quad \text{kde } n \in \mathbb{N}. \tag{6}$$

Pre ukážku nelineárnej rezonancie kyvadla bolo zaznamenaných sedem meraní s budiacimi frekvenciami $\omega = \Omega\omega_0$ v rozmedzí od 1 do 7. Merania kmitov sú zobrazené na spoločnom fázovom portréte na obrázku 4 v časoch po odznení prechodových javov. Namerané amplitúdy θ_{mer} a pomery meraných uhlových frekvencií ω_{mer} ustálených kmitov s hodnotou vlastnej frekvencie z tabuľky 2 sú pre všetky prípady uvedené v nasledovnej tabuľke.

Tab. 3: Namerané amplitúdy a uhlové frekvencie ustálených kmitov kyvadla

Ω	1	2	3	4	5	6	7
θ_{mer} (rad)	0,064	0,020	0,122	0,005	0,165	0,002	0,195
$\omega_{\text{mer}}/\omega_0$	0,98	2,01	1,01	4,01	0,99	6,01	1,01

Pre budiace frekvencie spĺňajúce rezonančnú podmienku (6) sa kmity ustália na diskrétnych hodnotách amplitúdy rastúcich s Ω s uhlovou frekvenciou blízkou vlastnej frekvencii ω_0. V prípade, ak rezonančná podmienka nie je splnená, kyvadlo vykonáva vynútené kmity s malými amplitúdami a uhlovou frekvenciou blízkou budiacej frekvencii.

3 Matematický model

Táto časť práce je venovaná opisu jednoduchého nelineárneho modelu kyvadla poskytujúceho kvalitatívne predpovede v súlade s meraniami.

3.1 Pohybová rovnica

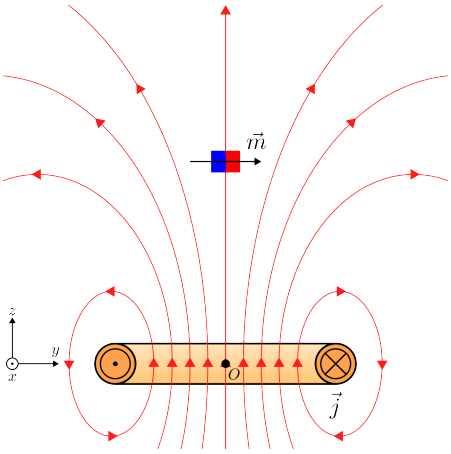
Úplnú pohybovú rovnicu magneticky budeného kyvadla je možné napísať v tvare:

$$\ddot{\theta} + 2\beta\dot{\theta} + \omega_0^2 \sin\theta + \mu \operatorname{sign}\dot{\theta} = \alpha(\theta)\cos(\omega t), \tag{7}$$

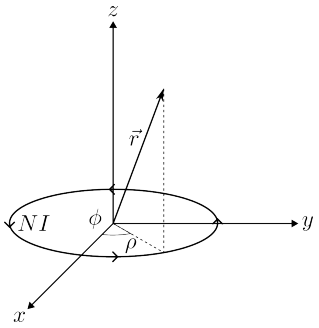
kde funkcia $\alpha(\theta)$ zodpovedá uhlovému zrýchleniu vyvolanému pôsobením nelineárnej budiacej sily na konci ramena kyvadla pri prelete magnetky nad cievkou na malom úseku v okolí rovnovážnej polohy.

3.2 Budiaca sila

V zjednodušenom prípade skutočnej situácie budenia kyvadla je možné pracovať s predstavou, kedy sa magnetka na konci ramena pohybuje v nehomogénnom magnetickom poli jedného závitu, ktorým preteká okamžitý prúd NI, kde N je počet závitov cievky, pričom magnetka je nahradená magnetickým dipólom $\vec{m}$ orientovaným pozdĺž osi y. Keďže vzdialenosť rotačnej osi a stredu magnetky ℓ je oveľa väčšia ako priemer cievky R a počiatok súradnicovej sústavy sa nachádza v strede cievky na jej osi, potom je možné uvážiť, že magnetka sa hýbe len pozdĺž osi y v nemenej zvislej vzdialenosti $z = d$, pričom platí $\theta \approx y/\ell$. Takáto situácia je naznačená na obrázku 5. Pre nájdenie sily pôsobiacej na magnetku je potrebné poznať zložku magnetického poľa v smere y v mieste $\vec{r} = (0, y, d)$. Pri



Obr. 5: Magnetický moment v nehomogénnom poli jedného závitu cievky



Obr. 6: Prúdový závit v cylindrických súradniciach

prechode do cylindrickej súradnicovej sústavy vďaka rotačnej symetrii voči osi z nezávisí pole cievky od azimutového uhla ϕ. Nenulovú zložku vektorového potenciálu sústavy $A_\phi(\rho, z)$ je možné vyjadriť v tvare [6]:

$$A_\phi(z,\rho) = C\frac{R}{k^2}\frac{(2-k^2)K(k) - 2E(k)}{\sqrt{(R+z)^2+\rho^2}}, \quad k^2 = \frac{4Rz}{(R+z)^2+\rho^2}, \tag{8}$$

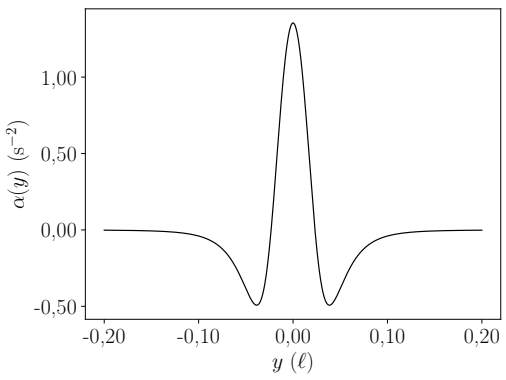
kde platí: $C = \mu_0 IN/\pi$, $\rho = \sqrt{x^2+y^2}$ a $K(k)$ a $E(k)$ sú úplné eliptické integrály prvého a druhého druhu [7]. Na základe vzťahu $\vec{B} = \nabla \times \vec{A}$ a transformácii medzi cylindrickou a karteziánskou súradnicovou sústavou je možné určiť zložku magnetického poľa v smere y ako:

$$B_y(\vec{r}) = -\frac{\partial A_\phi}{\partial z}\frac{y}{\rho} = \frac{C}{4\rho^2}\frac{yz}{\sqrt{(\rho+R)^2+z^2}}\left(\frac{2-k^2}{1-k^2}E(k) - 2K(k)\right). \tag{9}$$

Silu pôsobiacu na magnetku v smere osi y je potom možné určiť zo vzťahu $\vec{F} = \nabla(\vec{B}\cdot\vec{m})$. Pri priblížený malých uhlov $\theta \approx y/\ell$ je možné uvažovať, že sa orientácia magnetického momentu $\vec{m}$ pri pohybe nad cievkou nemení. Moment sily v osi rotácie vyvolaný pohybom magnetky v nehomogénnom poli cievky je potom možné odhadnúť ako $M_{x,O}(y) = I_{r,O}\alpha(y)$, odkiaľ pre funkciu $\alpha(y)$ platí:

$$\alpha(y) = \frac{F_y \ell}{I_{r,O}} = \frac{m_y \ell}{I_{r,O}}\frac{\partial B_y}{\partial y}. \tag{10}$$

Funkciu $\alpha(y)$, ktorej priebeh je zobrazený na grafe na obrázku 7, je možné vyčísľovať numericky.



Obr. 7: Vypočítaný priebeh budiacej funkcie $\alpha(y)$

3.3 Numerická simulácia

Pohybovú rovnicu (7) kyvadla je možné pri zadaní počiatočnej podmienky riešiť numericky. Pre numerickú simuláciu pohybu kyvadla pri siedmich budiacich frekvenciách totožných s meranými prípadmi bol vyvinutý Python kód `magnetic_pendulum.py`. Pre riešenie pohybovej rovnice bola využitá funkcia `solve_ivp()`, implementovaná v knižnici pre vedecké výpočty *SciPy 1.15.* [8]. Táto funkcia využíva pre integráciu rovníc pokročilú adaptívnu metódu *Runge-Kutta-Fehlberg* s globálnou chybou štvrtého rádu.

Parametre numerických simulácií, uvedené v následovnej tabuľke, boli na základe známych parametrov kyvadla zosúladené s meraniami kmitov v časti 2.6.

Tab. 4: Parametre numerickej simulácie

Kyvadlo		Cievka		Magnet	
ℓ (m)	0,4	N	100	m (A.m^2)	1,2
ω_0 (s^{-1})	5,23*	I_{rms} (A)	0,4	d (mm)	35
β (s^{-1})	0,22	R (mm)	20		
μ (s^{-2})	0,09	d (mm)	35		
I (kg.m^2)	2,55×10^{-3}				

$\theta(0)$ (rad)	$\dot{\theta}(0)$ (rad.s^{-1})	Max. čas (s)	Počet čas. krokov	Max. abs. chyba	Max. rel. chyba
$-\pi/12$	0	$40T_0$ †	1200	10^{-8}	10^{-6}

*Hodnota z tabuľky 2.
†Násobok periódy vlastných kmitov

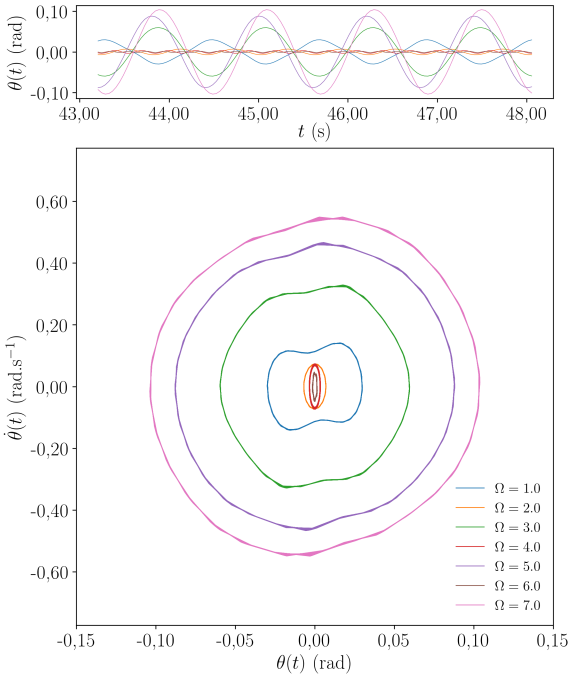
Simulácie pohybu kyvadla sú v časoch po odznení prechodných javov zobrazené, rovnako ako pri meraniach, na spoločnom fázovom portréte na obrázku 8. Amplitúdy stabilných kmitov θ_{sim} a pomery uhlových frekvencií ustálených kmitov voči prirodzenej frekvencii kyvadla $\omega_{\text{sim}}/\omega_0$ sú uvedené v nasledovnej tabuľke.

Tab. 5: Simulované amplitúdy a uhlové frekvencie ustálených kmitov kyvadla

Ω	1	2	3	4	5	6	7
θ_{sim} (rad)	0,030	0,007	0,060	0,003	0,088	0,001	0,104
$\omega_{\text{sim}}/\omega_0$	1,01	2,03	1,01	4,08	1,01	6,14	1,01

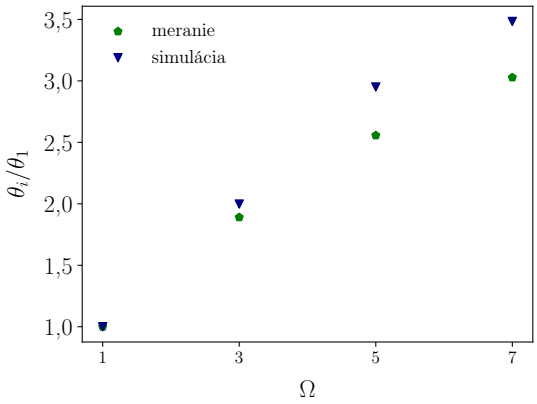
3.4 Porovnanie simulácie a merania

Kvalitatívnym porovnaním meraných a simulovaných fázových portrétov na obrázkoch 4 a 8, je možné skonštatovať, že fázové portréty majú v oboch prípadoch podobnú štruktúru. Takisto ako v meraných prípadoch je v prípade simulácií možné identifikovať očakávaný jav diskretizácie amplitúdy pri splnení podmienky subharmonickej rezonancie (6). Pri pomeroch frekvencií Ω nespĺňajúcich rezonančnú podmienku sú aj v prípade simulácii pozorované kmity s malými amplitúdami a frekvenciami blízkymi budiacim frekvenciám.



Obr. 8: Horný panel - Simulované časové priebehy výchylky kyvadla v posledných 5 sekundách simulácie. Dolný panel - Simulované fázové portréty kmitov kyvadla pri siedmich frekvenciách budenia

Pomer amplitúdy pri nepárnych celočíselných pomeroch Ω, voči amplitúde pri $\Omega = 1$ sú pre merania aj simulácie zobrazené na spoločnom grafe na obrázku 9.



Obr. 9: Amplitúdy meraných a simulovaných stabilných kmitov pri rezonančných budiacich frekvenciách $\omega = \Omega\omega_0$

4 Záver

Táto práca bola venovaná prezentácii zaujímavých fyzikálnych javov nelineárneho systému magneticky budeného kyvadla. V práci bol experimentálne a stimulačne demonštrovaný fenomén subharmonickej rezonancie, pri ktorej nastáva diskretizácia amplitúdy ustálených kmitov.

V rámci experimentálne časti práce, bola navrhnutá a zhotovená jednoduchá konštrukcia kyvadla umožňujúca bez-

odporové meranie výchylky kyvadla pomocou magnetického rotačného senzoru. Konštrukcia ramena kyvadla umožňovala odhad základných parametrov fyzikálneho kyvadla uvedených v tabuľke 1. Ostatné parametre kyvadla uvedené v tabuľke 2 boli určené na základe merania malých kmitov kyvadla. Vypočítaná hodnota vlastnej frekvencie malých kmitov kyvadla a hodnota určená z merania sa od seba odlišovali len o zhruba 1,1%.

Séria meraní pre sedem celočíselných pomerov budiacej a vlastnej frekvencie Ω bola zobrazená na spoločnom fázovom portréte na obrázku 4. Amplitúdy a frekvencie meraných stabilných kmitov kyvadla boli uvedené v tabuľke 3. Pri nepárnych pomeroch boli pozorované diskrétne amplitúdy kmitov s frekvenciami blízkymi vlastnej frekvencii. Pri párnych vykonával systém malé vynútené kmity s frekvenciou blízkou budiacej frekvencii.

Ďalšia časť práce bola venovaná jednoduchému matematickému modelu kyvadla. Na základe zjednodušení bolo možné nájsť tvar budiacej funkcie vystupujúcej v nelineárnej pohybovej rovnici kyvadla (7). Rovnica bola riešená numericky pri parametroch približne zodpovedajúcich skutočnému systému. Napriek značným zjednodušeniam poskytol model kvalitatívne správne výsledky.

Hlavným cieľom práce bola prezentácia zaujímavých vlastností experimentálne realizovateľného systému, ktorým je magnetické kyvadlo. Skonštruované magnetické kyvadlo má potenciál slúžiť ako účinný nástroj pre demonštráciu nelineárnej dynamiky v jednoduchých fyzikálnych systémoch. Za týmto účelom by bolo do budúcnosti možné navrhnúť výkonnejší budiaci obvod s autonómnym napájaním a generátorom signálov ovládaním užívateľom prostredníctvom PC, vďaka čomu by bolo možné pozorovať širšie spektrum nelineárnych javov, akými sú napríklad multistabilita alebo chaos. Ďalej by bolo možné vyvinúť užívateľsky prístupnú Python aplikáciu s grafickým rozhraním pre uľahčenie ovládania experimentu a zberu meraných dát.

Prezentovaný jednoduchý matematický model má slúžiť na vybudovanie predstavy o fyzikálnej podstate pozorovaných javov. V tomto ohľade je najdôležitejšie prepojenie medzi javom nelineárnej rezonancie a nelineárnym budením systému. Pokiaľ by sa ukázalo ako žiaduce spresniť model tak, aby bol v kvantitatívnej zhode s meraním, bolo by potrebné pracovať s realistickým model cievky a magnetky. Ďalším krokom v posunutí tohoto modelu by bolo uvažovanie rôznych orientácii cievky voči magnetke.



Obr. 10: Fotografia magnetického kyvadla

Poďakovanie

Táto práca bola podporená grantom KEGA 006STU-4/2025.

Literatúra

[1] Y. Luo, W. Fan, C. Feng, S. Wang, and Y. Wang. Subharmonic frequency response in a magnetic pendulum. *American Journal of Physics*, 88(2):115–120, Február 2020.

[2] A. I. Shumaev and Z. A. Maizelis. Distribution functions of argumental oscillations of the Duboshinskiy pendulum. *Journal of Applied Physics*, 121(15):154902, Apríl 2017.

[3] H. A. Wheeler. Simple inductance formulas for radio coils. *Proceedings of the Institute of Radio Engineers*, 16(10):1398–1400, Október 1928.

[4] ams AG. *AS5600 12-Bit Programmable Contactless Potentiometer – Datasheet*. Revízia 1.06, jún 2018. Dostupné na: `https://ams.com/as5600`

[5] M. V. Vitorino, A. Vieira, and M. S. Rodrigues. Effect of sliding friction in harmonic oscillators. *Scientific Reports*, 7:3726, Jún 2017.

[6] J. D. Jackson. *Classical Electrodynamics*. John Wiley & Sons, Inc., 1st edition, January 1962. ISBN: 978-0471431312, 641 s.

[7] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. National Bureau of Standards, Applied Mathematics Series 55, 1964.

[8] P. Virtanen et al. *scipy.integrate.solve_ivp — SciPy v1.10.1 Manual*. Online, dostupné na: `https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve_ivp.html` [cit. 25. 3. 2025].

Supravodivosť vs. prvák FEI STU & vice versa

Ján Virág

Ústav jadrového a fyzikálneho inžinierstva, FEI STU v Bratislave

xviragj@stuba.sk

Abstrakt – Tento príspevok sa zaoberá supravodivosťou ako javom, ktorý priťahuje pozornosť študentov všetkých stupňoch fyzikálnej erudície. Z pohľadu študenta prvého ročníka FEI STU sa venuje úplne základným poznatkom o supravodivosti a následnej aj podobne základným experimentom. Z tohto hľadiska sme sa pokúsili o prípravu štyroch vzoriek keramických vysokoteplotných supravodičov (HTS) zo starších práškových prekurzorov. Zmeraním ich $R(T)$ charakteristík sme overili, že tieto prekurzory nepodliehajú podstatnému starnutiu.

1 Úvod

Už od skorého veku ma fascinovali rôzne prírodné javy. Či už to bol odraz slnečných lúčov, vietor, teplo, alebo to, ako všetko padá k zemi, keď to pustím. No zo všetkého najviac ma očarila energia a jej rôzne formy. Pamätám si, ako som vždy pozoroval oblohu von z okna počas búrky, dúfajúc, že uvidím blesk. Blesky v sebe skrývajú obrovskú energiu, ktorá je schopná rozsvietiť celú oblohu. Vtedy som o nich takmer nič nevedel. Neskôr som zistil, že takejto energii sa hovorí elektrická energia. Okrem toho som sa vždy zaujímal aj o magnety a to, že ich fungovanie vie ovplyvniť elektrická energia, čo len viac zaujalo moju pozornosť.

Okrem prírody začali tiež viac a viac pútať moju pozornosť technológie. Zo začiatku šlo prevažne o autá, počítače a lietadlá. No vždy som sa zaujímal o jednu spoločnú vec, bez ktorej by ani jeden z týchto vynálezov nefungoval ktorou je práve energia, ktorá ich poháňa. Preto som sa rozhodol študovať na elektrotechnickej strednej škole, pretože som chcel vedieť viac. Na strednej škole som sa dozvedel o pojme elektromagnetizmus, tento pojem mi je síce stále celkom nejasný, no aj to málo, čo o tom viem, mi príde nesmierne zaujímavé a chcel by som sa v budúcnosti venovať aj tejto oblasti. Po strednej škole som nastúpil na FEI STU, odbor Jadrové a fyzikálne inžinierstvo, s vierou, že ak sa niekde môžem naučiť čo najviac o týchto oblastiach ktoré ma tak interesujú a ich využití v reálnom živote, tak je to práve tu. Netrvalo dlho a na prednáške z fyziky nám prednášajúci, aj keď len okrajovo, predstavil koncept supravodivosti. Nebolo toho povedané veľa, no fakt, že existuje niečo, čo v sebe skrýva všetky oblasti, ktoré ma najviac nadchýnajú, bolo príjemným zistením.

Supravodivosť má potenciál prenášať podobné množstvo energie ako blesk a technologický potenciál tohto javu je enormný. Začal som si o tejto oblasti hľadať viac informácií, najprv svojpomocne na internete, či už z rôznych článkov, alebo len náučných videí. No chcel som zistiť viac, v prvom rade som chcel vidieť tento jav na vlastné oči. Chcel som sa toho naučiť čo najviac. Keď som sa neskôr dozvedel, čo je to ŠVOČ, vedel som, že sa chcem zúčastniť. Vždy som chcel vedieť, aké to je podieľať sa na výskume. Na začiatku som váhal a nevedel som presne, o čom by moja práca mala byť. No potom mi bolo povedané, že náš prednášajúci je jedným z popredných odborníkov na supravodivosť. Nabral som odvahu, spýtal sa, či ma nepovedie na ŠVOČ, nevedel som síce ani navrhnúť, o čom by práca mala byť, no chcel som vedieť viac. Daroval mi svoju knihu o supravodivosti, ktorú sám napísal [1]. To, čo sa mi podarilo naučiť z tej knihy a z tejto celej práce, je uvedené nižšie v tomto dokumente. Aj keď som síce len prvák a neočakávam že táto práca bude prevratná. Rád by som sa toho čo najviac naučil a získal nové skúsenosti ktoré by som mohol v budúcnosti aplikovať pri nejakom už reálne dôležitom výskume.

2 Supravodivosť v skratke

Všetky tuhé (aj kvapalné) kovové látky vedú pomerne dobre elektrický prúd, pričom vykazujú istý elektrický odpor. Tento odpor je ovplyvňovaný mnohými faktormi a niektoré z nich si v tejto časti zhrnieme.

Jedným z podstatných faktorov ovplyvňujúcich odpor kovov je ich teplota. Ak teplota kovového vodiča stúpa, stúpa aj jeho odpor a naopak, ak teplota klesá, klesá aj odpor. Táto skutočnosť sa dá overiť aj bez laboratórnych podmienok. Stačí k tomu bežný ohmmeter, kovový vodič a nejaký spôsob na zmenu (zvyšovanie) teploty daného kovu. Z týchto zistení je možné prísť k záveru, že ak dostatočne ochladíme kov, odpor by teoreticky mohol poklesnúť až na nulovú hodnotu.

Závislosť odporu čistých kovov na teplote sa vyjadruje vzťahom

$$R=R_0\,(1+k\,(T-T_0)] \qquad (1)$$

kde R_0 je odpor pri teplote T_0 a k je teplotný súčiniteľ odporu.

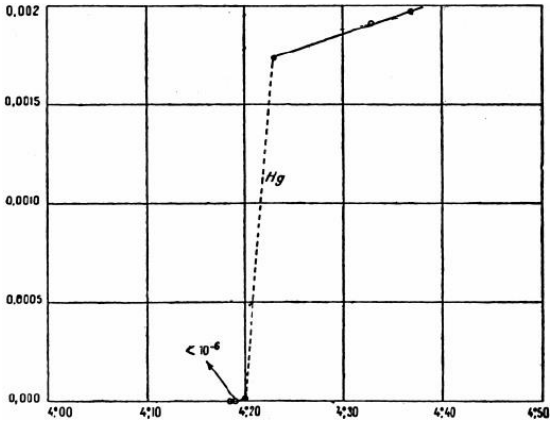
Okrem teploty môže ovplyvniť odpor vodičov ich čistota. Kryštalografická (počet porúch kryštalickej mriežky) aj

metalurgická (množstvo prímesí) čistota majú prakticky identický vplyv, keď odpor vzorky, pri zvolenej teplote, zvyšujú. Pri charakterizácii čistoty vzoriek sa používa tzv. RRR (Rezidual Resistance Ratio), čo je pomer odporu vzorky pri izbovej teplote a aproximovanej hodnote odporu pri termodynamickej nule. Čím väčší je RRR parameter, tým je vzorka čistejšia. Bežne dostupné medené drôty majú RRR na úrovni 100, zatiaľ čo pre najčistejšie monokryštalické vzorky presahuje hodnotu 10^5.

Supravodivosť je jav, ktorý s čistotou a teplotou materiálov samozrejme súvisí, avšak nie celkom podľa doteraz diskutovaných zjednodušených predpokladov. Ak postupne znižujeme teplotu kovových vzoriek (resp. vzoriek s kovových charakterom vodivosti) u niektorých z nich dôjde pri istej teplote (označovanej ako kritická teplota T_c) k skokovému poklesu odporu až na nulovú hodnotu. V tomto bode odpor kovov veľmi prudko poklesne k nule, a to aj keď len pri trocha vyššej teplote nadobúdal merateľné, často aj relatívne vysoké hodnoty. Tento jav označujeme ako supravodivosť a látky v tomto stave supravodiče. Avšak nie všetky kovy vykazujú podobný odporový skok. Je v skutočnosti pozoruhodné, že veľmi dobré vodiče ako meď, zlato a striebro sa nikdy nestávajú supravodivými, resp. ich kritická je nižšia ako experimentálne dostupné nízke teploty (T_c<10^{-4} K), pričom stále vykazujú teplotný priebeh odporu určený vzťahom (1). Naproti tomu, relatívne zlé vodiče ako olovo, cín, niób, etc. sa stávajú supravodičmi už pri teplotách niekoľkých K. Doteraz najvyššiu potvrdenú kritickú teplotu dosahujú tzv. keramické supravodiče a to na úrovni cca. 125 K.

2.1 Objav a výskyt supravodivosti

Z termodynamiky vieme, že najmenšia možná teplota akéhokoľvek fyzikálneho systému je absolútna nula, ktorej v Celsiovej tepelnej stupnici zodpovedá teplota -273,15 C°. Najúčinnejší spôsob na znižovanie teploty je mechanizmus odparovania kvapalín. Odparovanie kvapalín sa začalo používať ako hlavný spôsob chladenia vzoriek pri experimentoch už od prelomu 19. A 20.storočia. V roku 1898 James Dewar dokázal skvapalniť vodík. Jediný skvapalnený plyn, ktorý má nižšiu tepotu varu ako vodík, je tekuté hélium. Prvýkrát bolo hélium skvapalnené v roku 1908 na Univezite v Leydene. Dokázal to tím ktorý viedol holandský fyzik Heike Kamerlingh Onnes. A bol to práve Onnes ktorý ako prvý spozoroval jav supravodivosti pri chladení ortuti v kvapalnom héliu. .Stalo sa tak v roku 1911[3].



Obr.1 Historicky prvé pozorovanie supravodivosti.

3 Experimentálna časť

Hlavným cieľom experimentálnej časti tejto práce bolo oboznámiť sa s hlavnými črtami prípravy keramických vysokoteplotných supravodičov a overiť možné starnutie prekurzorov východzích práškových vzoriek.

3.1 Príprava vzoriek

Na obr. 1 sú znázornené nádoby obsahujúce štyri rôzne prekurzorové práškové formy zodpovedajúcich supravodičov. Vzorky označené ako „pre tenké vrstvy“ boli pripravené z tekutej formy rozpustených dusičnanov gadolínia, ytria a európia v 1990 roku [2]. Po odparení vody a vysušení, boli pri rýchlom výpale dusičnany pri teplote nad 750 °C rozložené a následne využité v minulosti pri príprave rôznych vzoriek typu Re $Ba_2Cu_3O_x$, kde Re je Y, Gd, Eu a x~6,5. Časť z týchto práškov bola dodnes uchovávaná pre účely podobné tejto práci. Tieto prášky sa vyznačujú malou granularitou, keď jednotlivé kryštality sú menšie ako 10 μm. Pre porovnanie sme pripravili aj vzorku $YBa_2Cu_3O_x$ z prášku zodpovedajúceho zloženia ale podstatne „mladšiu“ a pripravenú z oxidov obsiahnutých prvkov a podstatne väčšími (>25 μm) kryštalitmi.



Obr. 2. Nádoby s práškami jednotlivých typov supravodičov.

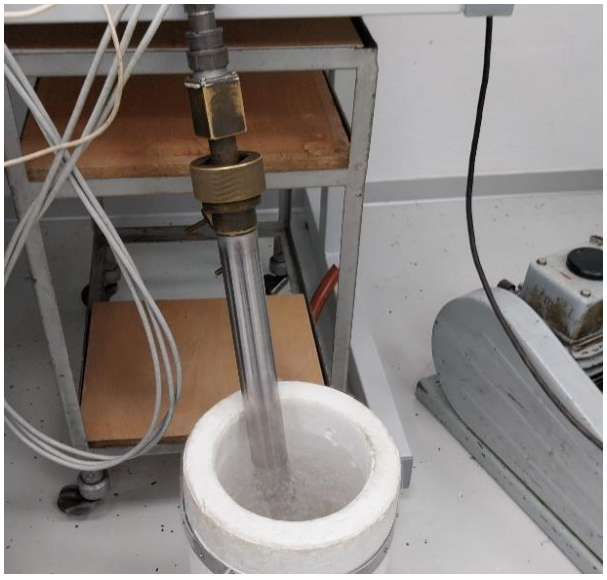
Prášky sme pri príprave vzoriek zmiešali s PVA pojidlom a denaturovaným alkoholom a následne sme ich ručne, metódou pripomínajúcou zjednodušenú formu sieťo-tlače, naniesli na MgO podložky. Takto pripravené vzorky boli následné spoločne umiestnené do muflovej pece a vypálené. Pre dosiahnutie lepšej väzby supravodiča a podložkou bol prvý výpal realizovaný relatívne rýchlym nábehom (250 °C/hod.) na teplotu 1050 °C. Po dosiahnutí tejto teploty boli vzorky v peci ochladené počas jednej hodiny (rýchlosť poklesu teploty 100 °C/hod.) na 950 °C pri ktorej výpal pokračoval 3 hodinovou výdržou. Pri tejto teplote sa typicky formuje tzv. 1:2:3 fáza, avšak samotné vzorky nie sú supravodivé, pretože pri tejto teplote je x~6 a vykazujú tetragonálnu kryštalickú štruktúru. Pre dosiahnutie potrebnej ortorombickej štruktúry s x>6,5, a s ňou spojenej vysoko-teplotnej supravodivosti, sme realizovali následný dlhodobý výpal pri teplotách medzi 550 až 400 °C s veľmi pomalým poklesom teploty (20 °C/hod), čím sme dosiahli ich dodatočnú oxidáciu vzdušným kyslíkom. Výsledné vzorky sú zobrazené na obr. 2.



Obr. 1. Vypálené vzorky vysokoteplotných supravodičov.

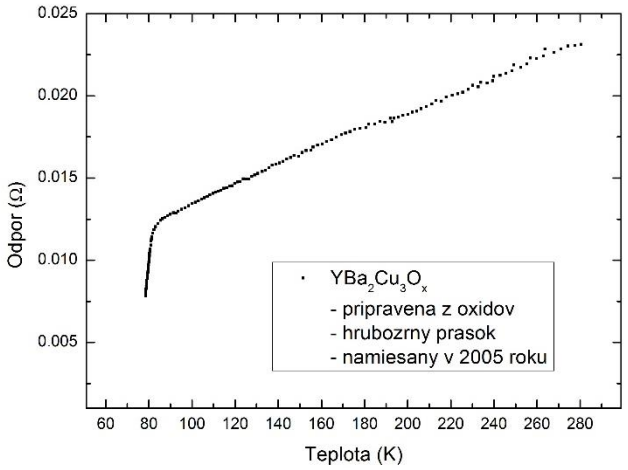
3.2 Meranie kritickej teploty vzoriek

Meranie kritickej teploty všetkých vzoriek sa uskutočnilo na pôde Elektrotechnického ústav SAV tzv. štvorbodovou kontaktnou metódou. Príprava na meranie bola nasledovná. Na každú vzorku boli pomocou pájky umiestnené 4 kontaktné body z india. Vzorka bola potom umiestnená a nalepená na podstavec ktorý bol pripojení na kovovú tyč, ktorou viedli kontaktne káble meracieho zariadenia. Celá tyč s odhalenou vzorkou bola vložená do kovovej rúry. Tento celok bol následne ponorený do nádoby s tekutým dusíkom (viď obr. 4). Rovnakým spôsobom boli merané aj ostatné vzorky. Metodologicky išlo o klasickú štvorbodovú metódu merania odporu vzoriek pri 10 mA konštantnom prúde. Uvádzaná teplota bola určená z odporu kalibrovanej polovodičovej diódy. Odčítanie spádu napätia na vzorkách (teda ich odporu) ako aj odpor tepelného senzora bolo zabezpečené počítačom riadeného systému.

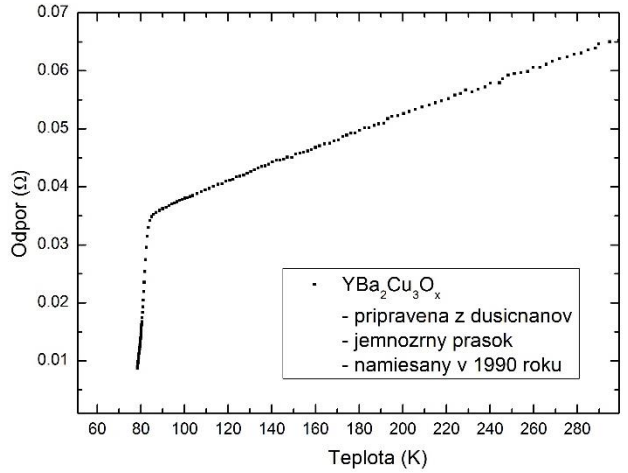


Obr.4 Meranie kritickej teploty vzoriek.

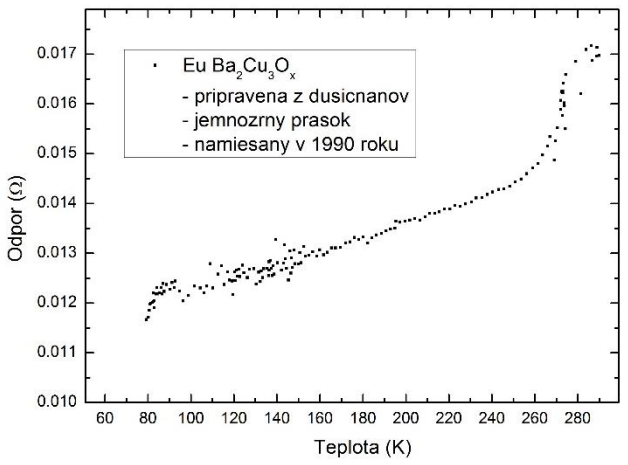
Meranie troch zo štyroch vzoriek prebehlo úspešne. Pri vzorke $EuBa_2Cu_3O_x$ sa počas merania vyskytla experimentálna chyba a vzorka sa odlepila od podstavca počas jej ponorenia do tekutého dusíka. Nasledujúce grafy zobrazujú jednotlivé výsledky týchto $R(T)$ meraní.



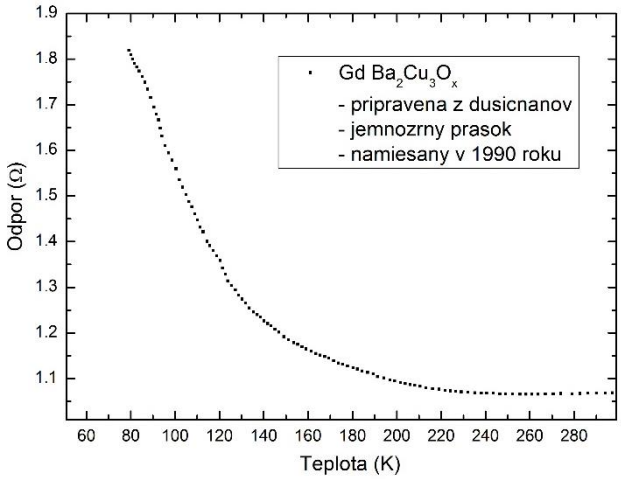
Obr. 5. $R(T)$ charakteristika vzorky $YBa_2Cu_3O_x$ pripravenej podľa legendy.



Obr. 6. $R(T)$ charakteristika vzorky $YBa_2Cu_3O_x$ pripravenej podľa legendy.



Obr. 7. $R(T)$ charakteristika vzorky $EuBa_2Cu_3O_x$ pripravenej podľa legendy.



Obr.8. R(T) charakteristika vzorky $GdBa_2Cu_3O_x$ pripravenej podľa legendy.

3.3 Pozorovanie diamagnetizmu supravodičov

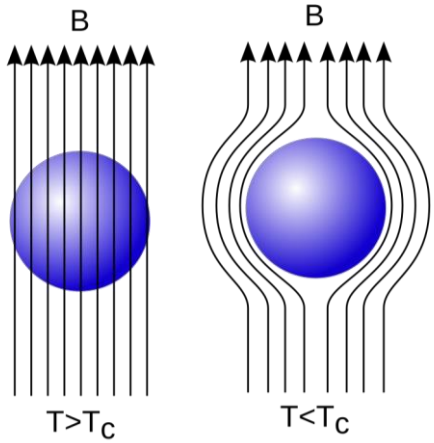
Aj keď sa supravodivosť najčastejšie spája s nulovým odporom pri vedení elektrického prúdu, v skutočnosti je to jav podstatne komplexnejší. Odhliadnuc od kvantovo-mechanických súvislosti supravodivosti, je nutné spomenúť aspoň ideálny diamagnetizmus supravodivých vzoriek, s ktorými je spojený tzv. Meissner-Ochsenfeldov jav, prvý krát pozorovaný a kvalitatívne vysvetlený v roku 1933. Jeho klasickým prejavom je „vznášanie“ sa magnetov nad supravodivými vzorkami, resp. opačne. My sme tento jav tiež kvalitatívne pozorovali, čo je znázornené na obr. 9.



Obr.9. Naše pozorovanie Meissner-Ochsenfeldovho javu.

Použili sme pritom väčšiu komerčnú tabletku keramického $YBa_2Cu_3O_x$, ktorú sme ochladili priamo v tekutom dusíku, nad ktorou sa vznáša silný samárium-kobaltový magnet.

Podstatou Meissner-Ochsenfeldovho javu je skutočnosť, že supravodiče bránia prieniku vonkajšieho magnetického poľa tak, že efektívne pôsobenie tzv. tieniacich supravodivých prúdov vytvorí „opačný magnet“. To sa prejaví odpudzovaním samotného magnetu, pričom supravodivý „antimagnet“ sa dynamicky prispôsobuje jeho polohe a orientácii. Výsledkom je nulové pole v supravodivej vzorke samotnej a zosilnené pole v jej okolí, viď obr. 10.



Obr.10. Magnetické pole v okolí normálnej vzorky (pri teplote vyššej ako kritická) a po jej ochladení a prechode do supravodivého stavu.

4 Diskusia

Ytriová vzorka pripravená z prekurzorov pochádzajúcich z roku 2005 pri chladení efektívne dosiahla kritickú teplotu a vstúpila do prechodu do supravodivého stavu. Avšak úplný stav supravodivosti (teda $R = 0$) nebol dosiahnutý. Tekutý dusík ma teplotu varu 77 K a aj pri tejto teplote bola hodnota odporu stále merateľná, ale pri zjavne pozorovanom prechode tento odpor prudko klesal. Môžeme preto oprávnene predpokladať, že pri len o niečo nižšej teplote by odpor klesol na nulu.

Druhá ytriová vzorka vykazovala „vzorný“ kovový charakter vodivosti pri teplotách nad kritickou teplotou, charakterizovaný lineárnym poklesom odporu. Po dosiahnutí kritickej teploty, ktorá bola prakticky rovnaká ako u predchádzajúcej vzorky (tesne pod 90 K), odpor rázne klesol na použitým experimentálnym zariadením nemerateľne malé hodnoty. Môžeme preto konštatovať, že vzorka je supravodivá.

Dá sa usúdiť že obe ytriové vzorky sú supravodivé, ale na hrane, keď pravdepodobne krátka doba oxidácie pri atmosférickom parciálnom tlaku kyslíka viedla k obsahu kyslíka (x) len tesne nad 6,5, pričom vzorka pripravená z dusičnanov je zjavne lepšia, keď aj pri podstatne menšom rozmere, má nižší odpor, pričom obe vzorky vykazujú kovový charakter vodivosti pri teplotách nad kritickou

Počas merania vzorky na báze európia došlo k experimentálnej chybe ktorá bola bližšie popísaná v časti 3.2. Z grafu vieme vydedukovať, že chyba nastala niekde v okolí teploty 150 K a ďalšie dáta sú preto nejednoznačne. Preto, aj keď sa po dosiahnutí nízkej teploty začali merané hodnoty odporu veľmi rýchlo blížiť k nule, nedá sa supravodivosť potvrdiť a pre túto vzorku je to len domnienka.

Gadolíniova vzorka mala na rozdiel od ostatných troch vzoriek úplne odlišný charakter. Elektrický odpor sa s meniacou teplotou najskôr držal na približne rovnakej hodnote. Pri nižších teplotách ale začal prudko stúpať. Gadoliniova vzorka je preto charakterom vodivosti úplne iná. Má zjavne polovodičový charakter teplotný priebeh odporu, teda zastúpenie kyslíka vo vzorke je výrazne nižšie $x < 6.5$.

Pokus o kvalitatívne pozorovanie Meissnerovho javu potvrdil existenciu tzv. ideálneho diamagnetizmu u supravodičov, keď použitý magnet sa nad supravodivou vzorkou vznášal až pokiaľ sa táto neohriala na teplotu vyššiu ako jej je kritická teplota.

5 Záver

Môžeme konštatovať, že obe ytriové vzorky mali takmer identický priebeh $R(T)$ charakteristík a že aj s odstupom značného času od prípravy ich prekurzorov, tieto zostávajú vhodné na prípravu ďalších vzoriek. Je tiež zrejmé, že vzorka (prekurzor) na báze gadolínia je podstatne citlivejšia na starnutie, alebo na teplotu (kvalitu) výpalu v potenciálne nie úplne homogénnom teplotnom poli použitej žíhacej pece. Prípadne aj kombinácia oboch faktorov mohla viesť k podstatne odlišnému charakteru vodivosti a absencie čo i len náznakov supravodivosti pri teplotách nad 77 K.

Experimentálnej chybe počas merania vzorky na báze európia sa síce dalo zabrániť, avšak aj takto je to dôležitá skúsenosť a ponaučenie, že pri nízkoteplotných experimentoch sa nehodno ponáhľať. Čo je však škoda, pretože inak namerané hodnoty vyzerajú sľubne avšak ich interpretácia je prakticky nemožná.

Kvalitatívne demonštračné pozorovanie diamagnetizmu supravodičov splnilo účel, keď zjavne indikovalo netriviálnu fyziku spojenú s javom supravodivosti.

Poďakovanie

Autor ďakuje Ing. Mariane Ušákovej, PhD. za podstatnú pomoc pri príprave vzoriek a Ing. Marcelovi Talackovi, PhD. pri meraní kritickej teploty vzoriek.

Literatúra

[1] VALKO P., Supravodivosť 2011, Kniha Zlín 2011
[2] VALAŠÍKOVÁ A. 1990, Príprava a sledovanie metalizovaných vysokoteplotných supravodičov, Diplomová práca EF SVŠT, 1990
[3] Kamerlingh-Onnes, H. (1911) Further Experiments with Liquid Helium. On the Change of Electric Resistance of Pure Metals at Very Low Temperatures, etc. IV. The Resistance of Pure Mercury at Helium Temperatures. Communications from the Physical Laboratory of the University of Leiden, 122, 13-15.

Optimalizácia procesu prípravy vrstiev pomocou rotačného nanášania

Mykhailo Olkhovyk[1], Patrik Novák[1]

[1]Ústav jadrového a fyzikálneho inžinierstva, FEI STU v Bratislave
[2] Názov inštitúcie/pracoviska autora

xolkhovyk@stuba.sk

Abstrakt –Táto práca sa zaoberá optimalizáciou procesu prípravy tenkých vrstiev pomocou rotačného nanášania (spin coating). Technika rotačného nanášania patrí medzi efektívne metódy tvorby homogénnych tenkých vrstiev v oblasti materiálového výskumu a technologických aplikácií. Hlavným cieľom práce je preskúmať vplyv procesných parametrov, najmä rýchlosti rotácie, viskozity roztoku a doby rotácie, na kvalitu a hrúbku vytvorených vrstiev. V rámci experimentu boli pripravené vrstvy z oxid grafénu (GO) deponovaných na kremíkových substrátov, pričom sa skúmal vplyv rôznych pomerov GO a destilovanej vody. Výsledky ukázali, že roztok s pomerom GO a destilovanej vody 1:7 začína „odstreďovať“ pri nižších otáčkach v porovnaní s čistým roztokom, čo naznačuje vplyv nízkej viskozity na priľnavosť vrstvy. Optimalizáciou týchto parametrov je možné dosiahnuť tvorbu homogénnych a kvalitných vrstiev vhodných pre rôzne aplikácie.

1 Úvod

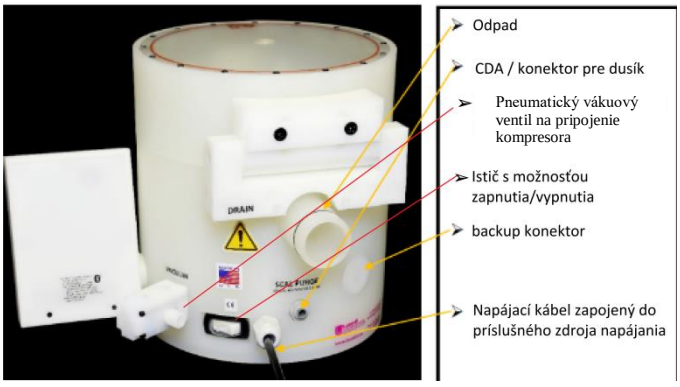
Technika rotačného nanášania, známa aj ako „spin coating“, patrí medzi najpoužívanejšie metódy prípravy tenkých vrstiev v oblasti materiálového výskumu a technologických aplikácií. Vyznačuje sa jednoduchosťou, rýchlosťou a vysokou reprodukovateľnosťou výsledkov, čo ju robí atraktívnou pre široké spektrum vedeckých a priemyselných aplikácií. Hlavným princípom techniky je nanesenie roztoku na povrch substrátu, ktorý sa následne roztočí vysokou rýchlosťou. Vďaka odstredivému zrýchleniu sa roztok rovnomerne rozloží po povrchu a vytvorí tenkú vrstvu [1].V literatúre sa často spomína využitie rotačného nanášania pri tvorbe polovodičových vrstiev, kde sa technika uplatňuje pri fotolitografii na nanášanie fotorezistov [2]. Ďalším významným použitím je nanášanie ochranných vrstiev na optické súčiastky, ako sú šošovky a optické filtre [3]. V oblasti biomedicíny sa táto technika používa na prípravu organických tenkých vrstiev a biofunkčných povlakov [4]. V súčasnosti sa výskum v oblasti rotačného nanášania sústreďuje najmä na optimalizáciu procesných parametrov s cieľom dosiahnuť požadovanú hrúbku vrstvy a jej homogénnosť. Tieto parametre zahŕňajú rýchlosť otáčania, zrýchlenie, viskozitu roztoku a jeho povrchové napätie [5].

Cieľom tejto študentskej práce je oboznámiť sa s technikou rotačného nanášania a vytvoriť ucelený manuál pre prípravu tenkých vrstiev touto metódou. Tento manuál bude slúžiť ako praktický návod pre študentov a výskumníkov pri príprave tenkých vrstiev, s dôrazom na optimalizáciu procesných parametrov a dosiahnutie vysokej kvality pripravených vrstiev.

2 Experimentálna časť

2.1 Opis aparatúry

Na prípravu vrstiev pomocou rotačného nanášania bol použitý Spin Coateru je *WS-650Hz-8NPPB*od spoločnosti Laurell, [6] ktorý pozostáva z rotujúcej základne, na ktorú sa umiestňuje substrát. Základňa je poháňaná motorom s nastaviteľnou rýchlosťou otáčania, pričom otáčky môžu dosahovať hodnoty od niekoľkých stoviek až po niekoľko tisíc otáčok za minútu. Zariadenie je vybavené dávkovacím systémom na presné nanášanie roztoku, ktorý sa aplikuje na stred substrátu pred začiatkom rotácie alebo počas programu. Dôležitým prvkom je riadiaca jednotka umožňujúca nastavenie rýchlosti, akcelerácie a času rotácie.



Obr. 2.1.1 Zariadenie Spin Coater

Nutnými podmienkami na spustenie akéhokoľvek programu Spin Coateru pre nanášanie vrstiev sú:

1. **Zatvorený horný kryt Spin Coateru**- kryt musí byť bezpečne uzavretý, aby bola zaistená ochrana používateľa a stabilita procesu.

2. **Pripojený zdroj dusíka alebo Clean Dry Air (CDA – čistý suchý vzduch) s tlakom vyšším ako 3,5 bar**- bez dostatočného tlaku vzduchu alebo dusíka nebude možné substrát bezpečne uchytiť na rotujúcu hlavu.
3. **Substrát správne umiestnený na hlavu Spin Coateru** - substrát musí byť umiestnený tak, aby zakrýval celú gumovú kružnicu na hlavičke zariadenia. V opačnom prípade nebude kompresor schopný pripevniť substrát, čo spôsobí, že program sa nespustí.
4. **Pripojený kompresor** - kompresor musí byť zapojený a funkčný, aby zabezpečil potrebné podtlakové uchytenie substrátu.

Program rotačného nanášania je možné nastavovať dvoma spôsobmi: **analógovo** alebo **digitálne**.

- **Analógové nastavenie** sa vykonáva priamo prostredníctvom ovládacieho rozhrania Spin Coateru (obr. 2.1.2). Toto rozhranie síce ponúka základné možnosti nastavenia, avšak je menej prehľadné a ťažkopádne. Napriek tomu je dôležité poznamenať, že aj keď je program plne nastavený digitálne pomocou počítača, niektoré základné prevádzkové podmienky je možné kontrolovať iba priamo cez analógové rozhranie.
- **Digitálne nastavenie** sa vykonáva pomocou počítača, kde je možné programovať jednotlivé kroky procesu. Po výbere konkrétneho programu a stlačení tlačidla **„EDIT“** sa zobrazí editor programu, v ktorom je možné definovať parametre jednotlivých krokov.



Obr. 2.1.2: Rozhranie Spin Coateru

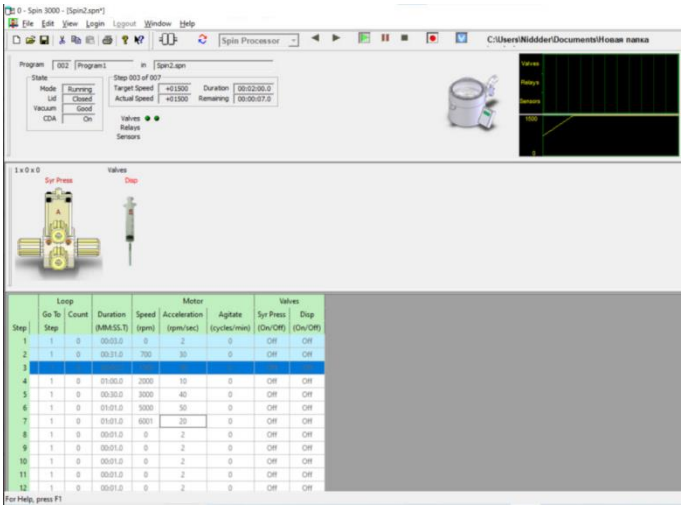
Medzi hlavné nastaviteľné parametre patria:

- **Rýchlosť otáčania (RPM – počet otáčok za minútu)** – určuje intenzitu odstredivej sily, ktorá ovplyvňuje hrúbku vrstvy.
- **Zrýchlenie** – určuje, ako rýchlo sa substrát roztočí na požadovanú rýchlosť.
- **Tlak v depozitáre (striekačke)** – reguluje množstvo nanášaného roztoku.
- **Stav ventilu depozitára** – určuje otvorenie alebo uzavretie dávkovacieho systému.

Správne nastavenie týchto parametrov je kľúčové pre dosiahnutie homogénnej a kvalitnej vrstvy. Kombinácia analógového a digitálneho riadenia umožňuje flexibilné nastavenie procesu s možnosťou presnej kontroly a úpravy počas nanášania.

Počas spúšťania programu rozhranie zobrazuje aktuálny stav každého parametra a zároveň indikuje, či sú splnené všetky nutné podmienky na spustenie programu. Ak sú všetky podmienky splnené, na obrazovke sa zobrazí hlásenie **„Ready“**. V prípade, že niektorá podmienka nie je splnená, rozhranie informuje používateľa o konkrétnom probléme. Napríklad, ak nie je pripojený dusík, systém zobrazí hlásenie **„CDA Required“**. Týmto spôsobom rozhranie zjednodušuje diagnostiku a urýchľuje prípravu zariadenia na nanášanie vrstvy.

Digitálne ovládanie zariadenia Spin Coater sa vykonáva prostredníctvom softvéru „**Spin 3000**“, ktoré sa pripojí k zariadeniu pomocou „bluetooth“ konektivity. Program umožňuje nastavovať rovnaké parametre ako ovládacie rozhranie priamo na zariadení, pričom poskytuje intuitívnejšie a prehľadnejšie prostredie.



Obr. 2.1.3 Program „Spin3000“

2.2 Princíp rotačného nanášania

Rotačné nanášanie je proces, pri ktorom sa roztok nanesie na povrch rotujúceho substrátu (obr. 2.2.1). Po nanesení kvapky na stred sa substrát roztočí vysokou rýchlosťou, čo spôsobí, že roztok sa rozleje v dôsledku odstredivých síl. Kľúčovým fyzikálnym princípom je pôsobenie odstredivého zrýchlenia, ktoré je dané vzťahom:

$$a_{\mathrm{d}} = \omega^2 r \qquad (1)$$

Kde je a_{d}odstredivé zrýchlenie,ωje uhlová rýchlosť rotácie (rad/s),$r$je vzdialenosť od stredu otáčania. Medzi najdôležitejšie faktory, ktoré ovplyvňujúce hrúbku vrstvy patria:

- **Rýchlosť rotácie:** Vyššia rýchlosť vedie k tenšej vrstve.
- **Doba rotácie:** Dlhšia doba umožňuje lepšie rozloženie roztoku.
- **Viskozita roztoku:** Vyššia viskozita spôsobuje hrubšie vrstvy.
- **Povrchové napätie:** Ovplyvňuje rozlievanie roztoku na substráte.
- **Odparovanie rozpúšťadla:** Počas rotácie sa rozpúšťadlo odparuje, čím sa mení koncentrácia roztoku.
- **Teplota prostredia:** Vyššia teplota urýchľuje odparovanie.
- **Vlhkosť vzduchu:** Ovplyvňuje rýchlosť odparovania.

Hrúbka vrstvy sa ustáli, keď sa odstredivé sily vyrovnajú s viskóznymi silami roztoku. Výsledná hrúbka závisí od rýchlosti otáčania podľa empirického vzťahu:

$$d \propto \frac{1}{\sqrt{\omega}} \qquad (2)$$

Zvýšenie rýchlosti rotácie vedie k tenšej vrstve, zatiaľ čo vyššia viskozita a povrchové napätie majú opačný efekt. Pri výrobe vrstiev je preto dôležité optimalizovať všetky tieto parametre.

Proces rotačného nanášania vrstiev možno rozdeliť do štyroch hlavných krokov: depozícia, roztočenie, odstránenie a vyparovanie. Každý krok je nevyhnutný pre správne vytvorenie homogénnej tenkej vrstvy.

1.**Depozícia**- V prvom kroku sa substrát umiestni na hlavičku Spin Coateru a roztok sa aplikuje blízko osi rotácie, teda do stredu substrátu (obr. 2.2.1).Existujú dva typy depozície v závislosti od metódy nanášania:

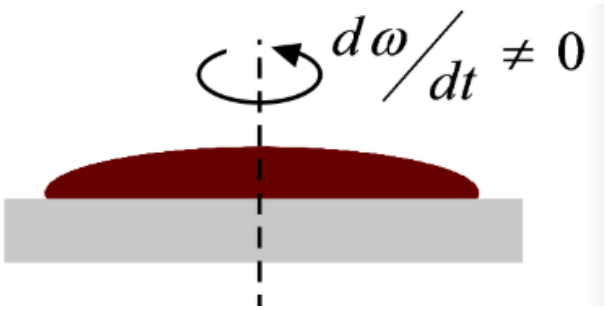
- **Statická depozícia** – Roztok sa aplikuje na substrát pred jeho roztočením.
- **Dynamická depozícia** – Roztok sa aplikuje na substrát počas jeho rotácie.



Obr. 2.2.1: Depozícia. [9]

Počas tohto kroku vplyvom odstredivých síl dochádza k disperzii roztoku smerom od stredu k okrajom, pričom sa hrúbka vrstvy znižuje na tzv. kritickú výšku.

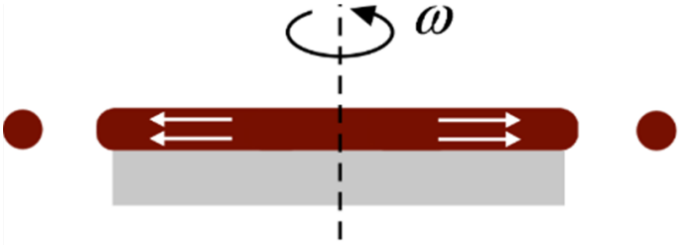
2. **Roztočenie** - Po aplikácii roztoku na substrát sa začína proces rotácie, pri ktorom sa substrát roztočí na požadovanú rýchlosť. Tento krok sa zvyčajne považuje za fázu intenzívneho odstredivého vyvrhnutia kvapaliny z povrchu substrátu. Odstredivá sila, ktorá je výsledkom zrýchlenia a uhlovej rýchlosti, spôsobuje odtekanie tekutiny smerom von z povrchu substrátu (obr. 2.2.2). Hrúbka vytvorenej vrstvy je významne ovplyvnená rýchlosťou rotácie a dobou trvania tohto kroku. Čím vyššia je rýchlosť rotácie, tým tenšia bude výsledná vrstva.



Obr. 2.2:2: Roztáčanie [9]

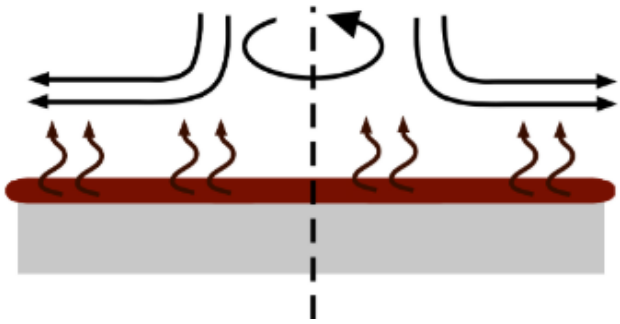
3. **Odstredenie** - Tento krok nastáva, keď sa substrát otáča konštantnou rýchlosťou. V tomto momente viskózne sily kvapaliny začínajú dominovať nad odstredivými silami, čo vedie k stabilnému riedeniu roztoku. Počas tejto fázy sa roztok homogénne rozloží po povrchu a prebytočná kvapalina sa odstraňuje na okraji substrátu (obr. 2.2.3).

Počas tejto fázy sa môže na okrajoch vytvárať kvapôčková štruktúra, ktorú je možné pozorovať pomocou mikroskopu. Tieto kvapôčky sú postupne odstredivo odstránené. Výsledná hrúbka vrstvy je závislá od povrchového napätia, viskozity a rýchlosti otáčania. Na okrajoch môže vzniknúť mierny rozdiel v hrúbke povlaku v dôsledku nerovnomerného odstránenia kvapôčok.



Obr. 2.2.3: Odstredenie [9]

4. **Vyparovanie** - Vyparovanie roztoku prebieha počas celého procesu vytvárania vrstvy, no v záverečnej fáze dominuje. V tejto fáze sa substrát otáča konštantnou rýchlosťou, pričom proces riedenia vrstvy prebieha najmä odparovaním rozpúšťadla (obr. 2.2.4).Odparovanie je podporované vysokou rotačnou rýchlosťou, čo zvyšuje efektívnosť sušenia. Hrúbka vrstvy môže byť ovplyvnená dĺžkou odparovania, pretože dlhší čas vedie k zvýšeniu koncentrácie roztoku a viskozity. Preto je dôležité správne nastaviť teplotné podmienky a čas rotácie tak, aby sa dosiahla požadovaná hrúbka filmu. Pri izbovej teplote sa roztok odparuje pomalšie, čo môže viesť k hrubšej vrstve. Na druhej strane, vyššia teplota urýchľuje proces odparovania, čím sa znižuje hrúbka výsledného filmu.



Obr.
2.2.4: Vyparovanie. [9]

2.3 Príprava vrstiev

Kvalita a homogenita tenkých vrstiev závisí nielen od procesu nanášania, ale aj od dôkladnej prípravy substrátu. V tomto experimente boli použité substráty z kremíka. Substráty boli najskôr čistené v destilovanej vode a následne sonifikované počas 15 minút v ultrazvukovom kúpeli s destilovanou vodou. Po tomto kroku boli opláchnuté etanolom a opäť sonifikované 15 minút v etanole, aby sa zabezpečilo dôkladné odstránenie organických nečistôt a prachu. Na zvýšenie priľnavosti oxid grafénu (GO) na povrch substrátu bol použitý (3-Aminopropyl)triethoxysilán (APTES), ktorý slúži ako adhézna vrstva. Čisté substráty v etanole boli priamo prenesené a ponorené na dobu 1 hodiny pri izbovej teplote do 1% roztoku APTES pripraveného v etanole, pričom počas celej doby boli chránené pred UV žiarením. Po skončení tohto kroku boli substráty dôkladne opláchnuté etanolom, aby sa odstránili zvyšky voľného APTES, a následne sušené prúdom dusíka (N_2) až do úplného vysušenia. Po vysušení boli substráty vystavené tepelnej úprave pri teplote 120 °C počas 1 hodiny. Tento krok je dôležitý pre stabilizáciu adhéznej vrstvy, čo zabezpečuje pevné ukotvenie oxid grafénu na povrchu substrátu. Po tepelnom spracovaní sme substráty nechali vychladnúť na izbovú teplotu, aby sa minimalizovalo pnutie v povlaku. Nakoniec, substráty boli 20 minút vystavené UV žiareniu pod UV-lampou [7,8].

Na očistené a pripravené substráty bol deponovaný 0,4% roztok oxid grafénu (GO) tak, aby pokrýval približne 2/3 plochy kremíkového substrátu. Následne boli tieto substráty roztočené pomocou jedného z troch zvolených programov(Tab. 1 – Tab. 3).

Krok	**Čas (M:SS)**	**Rychlosť (rpm)**	**Zrychlenie (rpm/sec)**
1.	2:00	500	30
2.	0:40	1000	50
3.	0:40	2000	50

Tab. 1: Program č. 1

Krok	**Čas (M:SS)**	**Rýchlosť (rpm)**	**Zrýchlenie (rpm/sec)**
1	2:00	500	30
2	0:40	1000	50
3	0:40	2000	50

Tab. 2: Program č. 2

Krok	**Čas (M:SS)**	**Rýchlosť (rpm)**	**Zrýchlenie (rpm/sec)**
1	2:00	500	30
2	0:40	1000	30
3	0:40	2000	50
4	1:20	4000	50

Tab. 3: Program č. 3

Vzorky substrátov s nanesenými GO vrstvami boli následne vysušené na horúcej platni pri teplote 100 °C počas 10 minút na vzduchu. Potom boli GO vzorky redukované vo výpare hydrazín monohydrátu (98, Sigma-Aldrich) pri teplote 80 °C počas 16 hodín. Na tento účel boli substráty umiestnené do stredu petriho misky a 1 ml hydrazínu bol rovnomerne rozmiestnený po okraji misky. Petriho misky boli následne hneď hermeticky uzavreté a položené na pec. Po 16 hodinách boli substráty ponechané na vzduchu z dôvodu odparenia prebytočných pár hydrazínu po dobu 1 hodiny a ochladené na izbovú teplotu. Potom substráty boli zahriate na teplotu 200 °C. Vďaka tomu procesu sme získali redukovaný oxid grafénu, kde klesol počet kyslíkatých skupín. Tieto vzorky sme následne analyzovali.

2.4 Analýza procesu nanášania

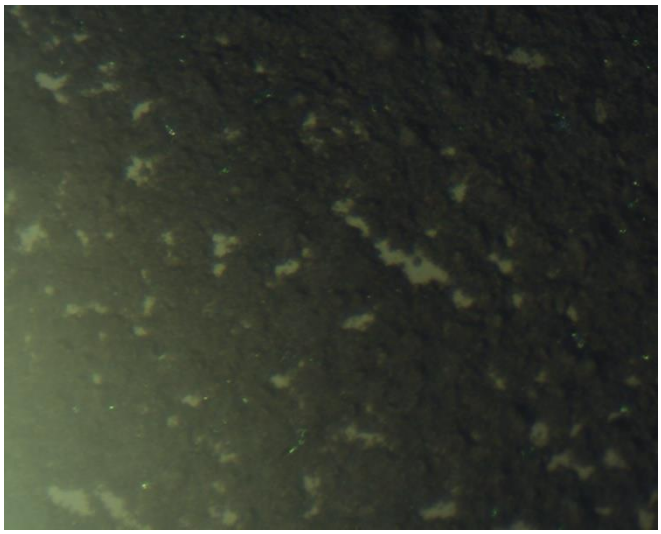
Jedným z kľúčových faktorov, ktorý ovplyvňuje proces rotačného nanášania, je aj viskozita roztoku. Viskozita určuje, akou rýchlosťou sa roztok rozlieva po povrchu substrátu a aký odpor kladie odstredivým silám počas rotácie. V praxi sa ukazuje, že roztoky s vyššou viskozitou vytvárajú hrubšie vrstvy, zatiaľ čo roztoky s nižšou viskozitou sa lepšie rozotierajú a vedú k tvorbe tenších vrstiev. Na základe experimentálnych pozorovaní bolo zistené, že roztok obsahujúci GO a destilovanú vodu v pomere 1:7 začína „odstreďovať“ a separovať sa od substrátu už pri nižších otáčkach v porovnaní s čistým roztokom. Tento jav je spôsobený nízkou viskozitou vodného roztoku v kombinácii s GO, čo znižuje priľnavosť vrstvy na substrát a spôsobuje predčasné odtrhnutie roztoku pri rotácii. Pre správnu kontrolu procesu je potrebné optimalizovať pomer medzi GO a destilovanou vodou tak, aby sa dosiahla požadovaná hrúbka vrstvy bez predčasného oddelenia. Skúmanie vplyvu rôznych pomerov GO a vody ukázalo, že zvýšenie viskozity (napr. pridaním väčšieho množstva GO) znižuje tendenciu roztoku „odstreďovať“, avšak zároveň môže viesť k hrubšej vrstve, čo je nežiadúce v prípade požiadavky na tenké vrstvy.

3 Diskusia a výsledky

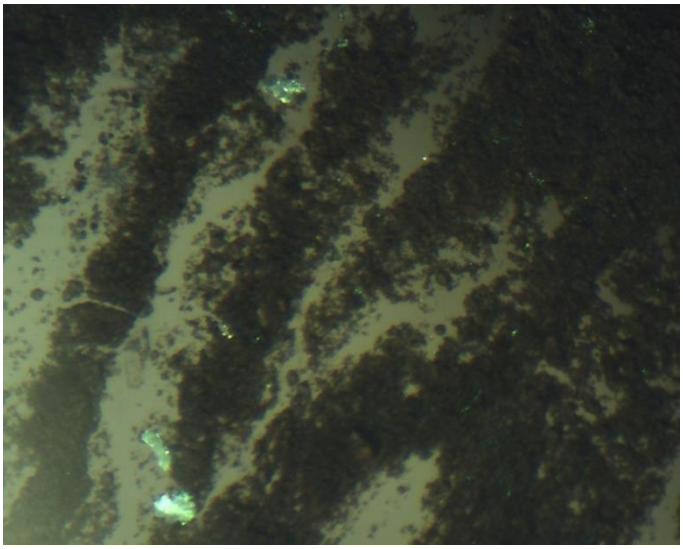
Povrchovú analýzu sme uskutočnili pomocou stereolupy OPTECH, pri zväčšení 7,5x a 34x. Získali sme detailné fotografie povrchov, ktoré sme pomocou softvéru „ImageJ“ analyzovali. Pri analýze povrchu vrstvy pripraveného programom č. 1 sme zistili, že rGO pokrýva 99.55% ± 0.05% celkovej plochy substrátu. Tento vysoký podiel pokrytia naznačuje, že pri nižších rýchlostiach rotácie dochádza k dostatočnému rozotretiu rGO po povrchu, čo vedie k vzniku súvislej vrstvy. Tento výsledok naznačuje, že aj napriek nižšej rotačnej rýchlosti je viskozita roztoku dostatočná na vytvorenie homogénnej vrstvy. Naopak, pri analýze vzorky pripravenej programom č. 3 sme zistili, že povrch vrstvy rGO na fotografii tvorí iba 80.92% ± 1.36% plochy, pričom zostávajúca plocha patrí kremíku. Tento výsledok naznačuje, že vyššia rýchlosť rotácie a zrýchlenie môžu viesť k príliš rýchlemu rozstreknutiu rGO, čo spôsobí diskontinuitu vrstvy. V tomto prípade zohráva úlohu nielen vyššia rotačná rýchlosť, ale aj samotná viskozita roztoku. Pri vyšších rýchlostiach sa roztok s nižšou viskozitou rozotiera nerovnomerne, čím vzniká diskontinuálna a nesúvislá vrstva oxid grafénu. Získané výsledky potvrdzujú význam optimalizácie procesných parametrov rotačného nanášania. Viskozita roztoku sa javí ako kľúčový faktor ovplyvňujúci priľnavosť a kvalitu výslednej vrstvy. Zvolený program významne ovplyvňuje výslednú kvalitu a pokrytie vrstvy, pričom rýchlosť a zrýchlenie otáčania sú tiež rozhodujúcimi faktormi pre zabezpečenie homogénnej vrstvy rGO.

Na potvrdenie kvality vrstiev sme vykonali merania na röntgenovom difraktometri D8 Advance od firmy Bruker s medenou lampou. Röntgenové žiarenie dopadalo na vzorku pri konštantnom uhle dopadu 0,3°. Výsledkom merania sú difrakčné záznamy, ktoré ukazujú závislosť intenzity od Braggovho uhla. Vzorky vykazovali amorfný charakter, čo naznačuje nepravidelné usporiadanie atómov v nanesenej vrstve. Zistili sme, že maximálna intenzita difrakčného maxima bola vyššia pre vzorky s väčším pokrytím rGO, čo môže súvisieť s hustotou a hrúbkou nanesenej vrstvy.

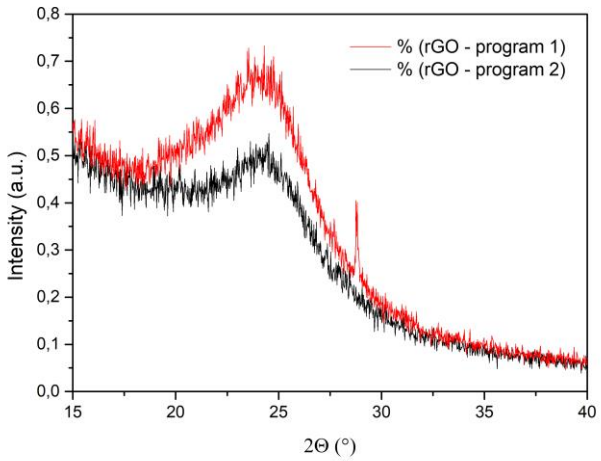
Pre budúce experimenty odporúčame vykonať ďalšie merania s rôznymi pomermi rGO a destilovanej vody, aby sa presnejšie určil optimálny rozsah procesných parametrov pre tvorbu kvalitných tenkých vrstiev.



Obr. 3.1: Vzorka rGO pripravená pomocou programu č.1 pri 34x zväčšení



Obr. 3.2:Vzorka rGO pripravená pomocou programu č.3 pri 34x zväčšení



Obr. 3.7: Rtg difrakčný záznam vzoriek rGO pripravených pomocou techniky rotačného nanášania

Poďakovanie

Práca bola vypracovaná s finančnou podporou projektu Kultúrnej a edukačnej agentúry Ministerstva školstva, výskumu, vývoja a mládeže Slovenskej republiky KEGA č. 006STU-4/2025.

Literatúra

[1] Kosar, A. (2020). SpinCoatingTechniquesforThin Film Preparation. Journal of MaterialsProcessingTechnology, 278, 116450.
[2] Smith, J., et al. (2019). PhotolithographicApplications of SpinCoating. SemiconductorScience and Technology, 34(8), 085002.
[3] Lee, C. & Kim, D. (2021). ProtectiveCoatingsforOpticalComponentsUsingSpinCoating. AppliedOptics, 60(4), 1258-1264.
[4] Nguyen, T. et al. (2023). BiomedicalCoatings by SpinCoatingTechnique. Journal of BiomedicalEngineering, 45(2), 190-197.
[5] Zhou, Y., et al. (2022). Optimization of SpinCoatingParametersforThin Film Homogeneity. ThinSolidFilms, 735, 138973.
[6] Laurell Technologies Corporation. (2024). User Manual: WS-650Hz-8NPPB SpinCoater.
[7] Zhao, S., Zhang, Y., Liu, X., &Wang, J. (2019). EnhancedPerformance of Graphene Oxide FilmsviaOptimizedSpinCoatingProcess. RSC Advances, 9(62), 36593-36602. DOI: 10.1039/C9RA06258K
[8] Chen, H., Li, J., Zhao, P., &Wang, Y. (2023). SpinCoating of ReducedGraphene Oxide forFlexible Electronics. IET Nanodielectrics, 6(1), 12-21. DOI: 10.1049/mna2.12070
[9] Basic Models of Spin Coating S. L. Hellstrom October 28, 2007 (Submitted as coursework for Physics 210, Stanford University, Autumn 2007) http://large.stanford.edu/courses/2007/ph210/hellstrom1
[10] Ciril Reiner-Rozman, et al. (2020). The top performer: Towards optimized parameters for reduced graphene oxide uniformity by spin coating. DOI: 10.1049/mna2.12070

Sekcia 6

Robotika a kybernetika

Riadenie mobilných robotov

Filip Kráľ[1], prof. Ing. František Duchoň, PhD.[1]

[1]Ústav robotiky a kybernetiky, FEI STU v Bratislave

xkralf@stuba.sk

Abstrakt – Tento príspevok sa zaoberá problematikou, a porovnaním SLAM systémov u mobilných robotov v systéme ROS a ich vzájomným porovnaním v simulácii. Pre porovnanie sme použili mobilného robota so simulovaným 2D lidarom a s jeho pomocou sme nazbierali dáta zo simulovaného prostredia domu. Na tomto datasete sme implementovali a analyzovali algoritmy Cartographer SLAM a SLAM Toolbox, pričom sme tiež porovnali synchrónne a asynchrónne nastavenie pre SLAM toolbox. Výsledné trajektórie sme porovnali pomocou python knižnice EVO.

1 Úvod

SLAM (Simultaneous Localization and Mapping) je výpočtový problém, pri ktorom sa vytvára alebo aktualizuje mapa neznámeho prostredia, pričom zároveň dochádza k lokalizácii mobilného robota v tomto prostredí.

Algoritmy SLAM vychádzajú zo známeho kinematického modelu mobilného robota, ktorý sa na začiatku nachádza v neznámej polohe a pohybuje sa v prostredí s prirodzenými alebo umelými orientačnými bodmi. V rámci SLAM sa poloha a dráha mobilného robota, ako aj značky v prostredí, vypočítavajú v reálnom čase.[1]

Cartographer sa skladá z dvoch podsystémov, lokálneho SLAMu (front-end) a globálneho SLAMu (back-end). Lokálny SLAM v reálnom čase vytvára submapy ktoré zarovnáva so skenom pomocou scan matchingu, ktorý využíva Ceres Scan Matching na nelineárnu optimalizáciu. Tento proces akumuluje chybu, ktorá je následne odstránena v globálnom SLAMe.

$$\arg\min_{\xi} \sum_{k=1}^{K} \left(1 - M_{\text{smooth}}(T_{\xi} h_k)\right)^2 \tag{1}$$

Globálny SLAM využíva Sparse Pose Adjustment (SPA) na optimalizáciu polohy všetkých submáp. Optimalizačný problém sa formuluje ako nelineárna metóda najmenších štvorcov.

$$\arg\min_{\Xi_m, \Xi_s} \frac{1}{2} \sum_{i,j} \rho\left(E^2(\xi_{m_i}, \xi_{s_j}; \Sigma_{ij}, \xi_{ij})\right) \tag{2}$$

Podrobnému opisu Cartographeru sa venovali Hess a kol. [2]

SLAM toolbox [3] poskytuje 3 základne funkcionality. Synchrónne mapovanie, ktoré ma využitie pokial je dôležitá kvalita mapy alebo offline spracovanie údajov, asynchrónne mapovanie, ktoré ma využitie v prípade že je dôležité mať kvalitnú lokalizáciu v reálnom čase a pure localization. Jednou z výhod je to, že SLAM toolbox umožňuje zlepšovať alebo rozširovať už existujúcu mapu, pretože dáta ukladá do pose-graph namiesto vytvárania submáp ako Cartographer.

2 Implementácia a porovnanie SLAM metód

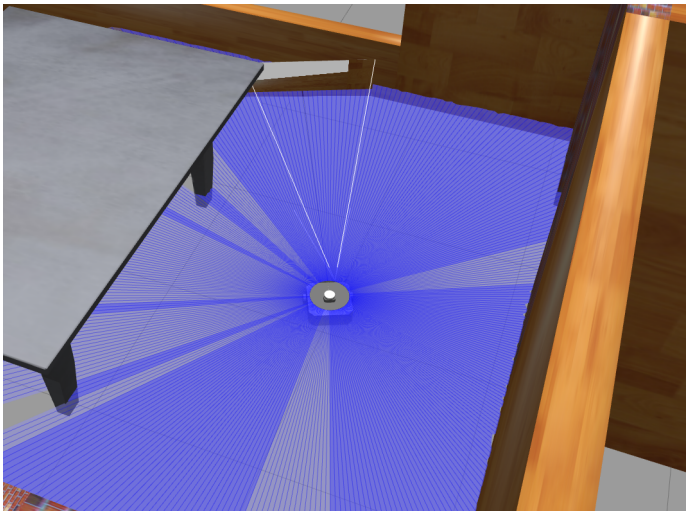
V tejto práci sme sa zaoberali implementáciou a porovnaním vybraných SLAM algoritmov. Cielom bolo analyzovať presnosť algoritmov pomocou absolútnej chyby trajektórie a relatívnej chyby polohy. Preto sme vytvorili experimentálny dataset v simulácii a následne aplikovali tri SLAM algoritmy: Cartographer SLAM a SLAM toolbox, pričom pri SLAM toolboxe sme odsimulovali asynchrónne a synchrónne spracovanie dát. Informácie ohladom PC zostavy sú v tabulke č. 1.

Tab. 1: Hardver špecifikácie PC zostavy

Komponenty	Model
CPU	AMD Ryzen 5 5600H
GPU	Nvidia Geforce RTX 3060 6 GB
RAM	32 GB
OS	Ubuntu 22.04 LTS
ROS2	Humble

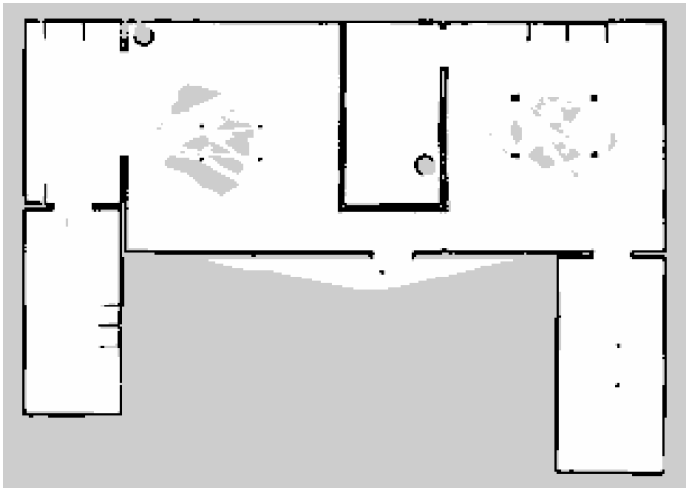
2.1 Príprava datasetu

Na vytvorenie datasetu sme použili balík Turtlebot3, konkrétne model WAFFLE, vybavený 2D lidarom LDS-01 ktorý umožňuje skenovať v rozsahu 360° do vzdialenosti 3,5 m.
Pre minimalizovanie chyby v porovnaní SLAM algoritmov, sme použili rovnaký dataset pre všetky 3 behy SLAMu. To sme dosiahli zaznamenaním dát zo simulácie do súboru ROSBAG. Simuláciu sme robili v simulátore Gazebo, kde sme skenovali prostredie turtlebot3_house.world, reprezentujúce dom so šiestimi izbami. Na ovládanie robota sme použili balík turtlebot3_teleop, ktorý nam umožnil ovládať mobilného robota pomocou klávesnice. Výstupné dáta sme následne vizualizovali v programe RViz.



Obr. 1: Vytváranie datasetu pomocou turtlebota

2.2 Implementácia Cartographer SLAM

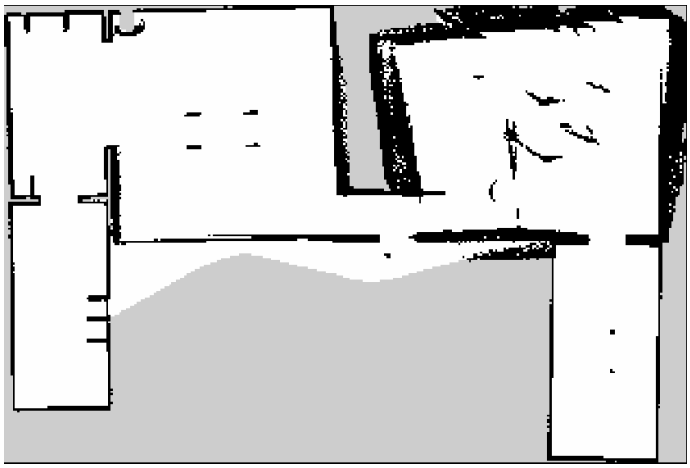


Obr. 2: Mapa získaná pomocou Cartographeru

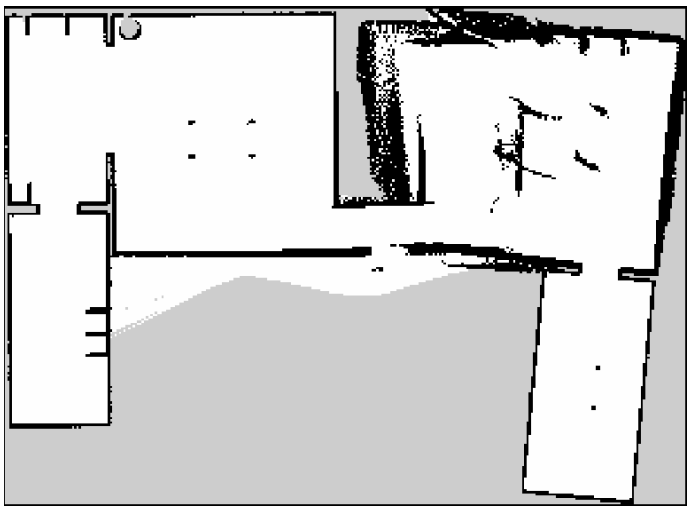
Algoritmus sme spustili zároveň s ROSBAG datasetom získaným zo sekcie 2.1, pripojili sme sa na topic /scan ktorý publikoval dáta sensor_msgs/msg/LaserScan a /odom ktorý publikoval dáta nav_msgs/msg/Odometry a publikovali sme topic /map. Výstupom bola occupancy grid mapa znázornená na Obr. 2.

2.3 Implementácia SLAM toolboxu

Pred samotným spustením SLAMu sme museli upraviť parametre transform_publish_period, minimum_travel_distance a stack_size_to_use v konfiguračnom súbore, aby sme docielili rovnakú frekvenciu publikovania dát ako pri Cartographeri. Potom sme spustili dataset získaný v sekcii 2.1 pomocou ROSBAG. Systém sa pripojil na topic /scan. Tento postup sme spravili najskôr pre asynchrónnu metódu a potom pre synchrónnu metódu. Vystupom bolo rovnako ako pri Cartographeri occupancy grid mapa publikovaná na /map topic, Obr. 3 a Obr. 4.



Obr. 3: Mapa získaná pomocou asynchrónneho SLAM toolboxu



Obr. 4: Mapa získaná pomocou synchrónneho SLAM toolboxu

2.4 Evaluácia dát

Keďže sme nemali k dispozícii skutočnú mapu (ground truth map), tak sme zvolili porovnanie pomocou trajektórii robota, kde sme pomocou python knižnice EVO porovnali skutočnú trajektóriu robota s odhadnutou trajektóriou robota, a vypočítali chybu strednej kvadratickej odchýlky (RMSE), strednú hodnotu (Median), priemernú hodnotu (Mean), štandartnú hodnotu (Std), minimum (Min) a maximum (Max). Pre evaluáciu vybraných metód SLAMu sme si zvolili metódy absolútnej chyby trajektórie (ATE) a relatívnej chyby polohy (RPE).

Absolútna chyba trajektórie sa vyhodnocuje ako porovnanie absolutnej vzdialenosti medzi skutočnou trajektóriou, a odhadnutou trajektóriou.

Relatívna chyba polohy meria lokálnu presnosť trajektórie za fixný časový interval. Tato chyba robota zodpovedá odchylke trajektórie mobilného robota.

Stredná kvadratická odchylka (RMSE) vyjadruje odchylku alebo chybu medzi skutočnými súradnicami trajektórie a odhadnutými súradnicami trajektórie:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2} \tag{3}$$

kde x_i su skutočné hodnoty, $\hat{x}_i$ sú odhadované hodnoty a n je počet hodnôt.

Podrobnejší opis týchto metód môžeme nájsť v prácach Kasára [4] a Filatova a kol. [5].

2.4.1 Porovnanie Cartographer SLAM

Pre porovnanie sme najskôr potrebovali získať skutočnú trajektóriu robota, tú sme získali pripojením sa na /odom topic ktorý publikuje nav_msgs/msg/Odometry ktorý sme získali pri vytváraní datasetu v sekcii 2.1. Tento topic publikuje polohu a orientáciu mobilného robota. Pred použitím dát sme konvertovali orientáciu z Eulerovej sústavy na kvaternióny, a následne sme dáta uložili do .tum súbora.

Následne počas bežania Cartographer SLAMu sme v konfigurácii aktivovali a pripojili sa na topic /tracked_pose, ktorý publikuje dáta geometry_msgs/msg/PoseStamped. Tieto dáta sme rovnakým spôsobom konvertovali a uložili do .tum súbora.

Výsledky porovnania je možné vidieť v Tab. 2 a na Obr. 5.

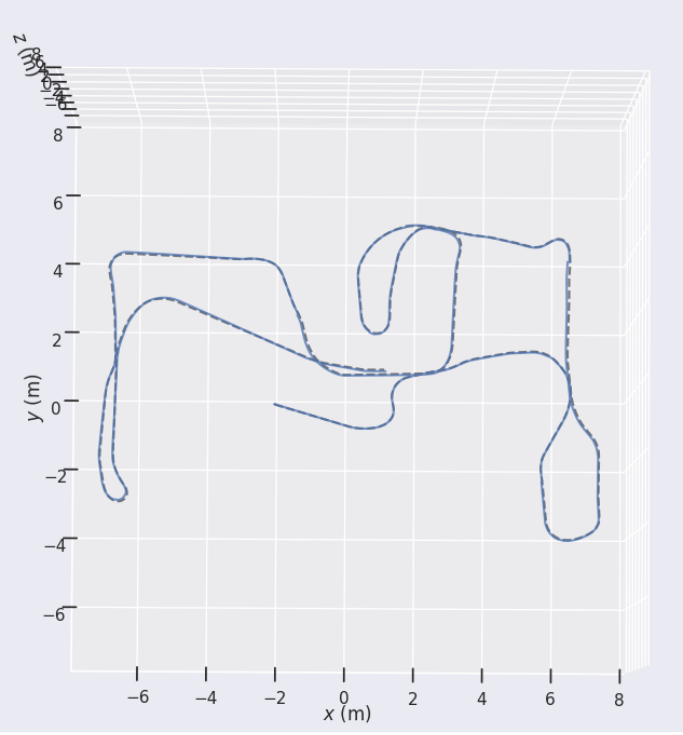
Tab. 2: tabulka ATE a RPE Cartographera a skutočnej trajektórie robota

Typ evaluácie	Absolútna Chyba Trajektorie	Relatívna Chyba Polohy	Jednotky
RMSE	0,128	0,002	m
Mean	0,124	0,001	m
Median	0,125	0,0004	m
Std.	0,034	0,002	m
Min	0,009	0,000	m
Max	0,194	0.063	m

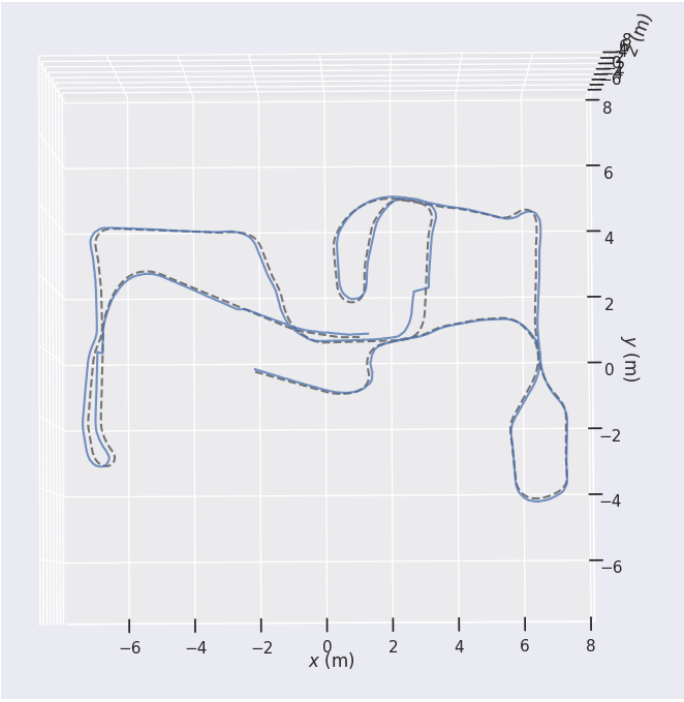
2.4.2 Porovnanie SLAM toolboxu

Podobne ako pri Cartographeri, aj tu sme sa počas behu SLAMu pripojili na topic /pose a dáta sme upravili a uložili do .tum súboru.

Výsledky porovnania pre asynchrónny mód sú v Tab. 3 a Obr. 6 a výsledky pre synchrónny mód sú v Tab. 4 a Obr. 7.



Obr. 5: porovnanie trajektórie Cartographera so skutočnou trajektóriou robota



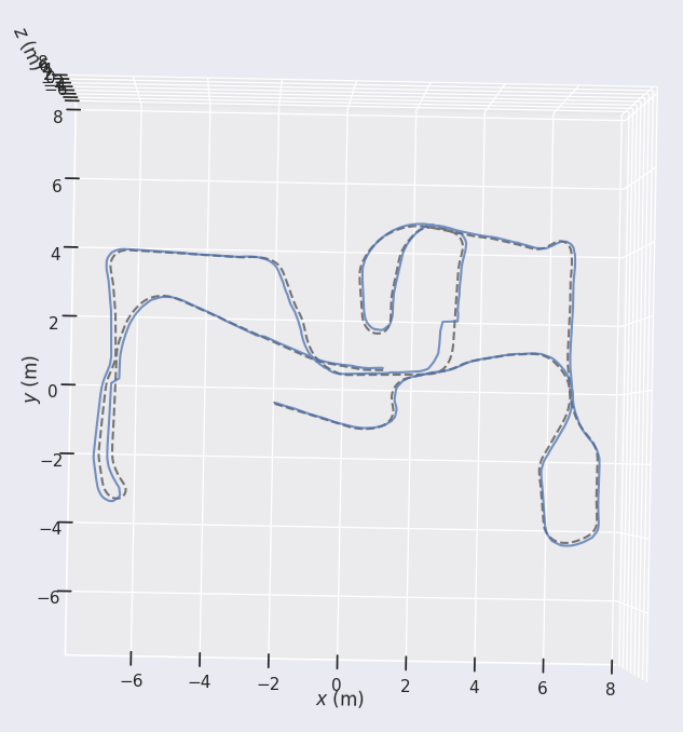
Obr. 6: porovnanie trajektórie asynchrónneho SLAM toolboxu so skutočnou trajektóriou robota

Tab. 3: tabulka ATE a RPE asynchrónneho SLAM toolboxu a skutočnej trajektórie robota

Typ evaluácie	Absolútna Chyba Trajektorie	Relatívna Chyba Polohy	Jednotky
RMSE	0,202	0,016	m
Mean	0,179	0,006	m
Median	0,192	0,003	m
Std.	0,092	0,015	m
Min	0,009	0,0002	m
Max	0,462	0,428	m

Tab. 4: tabulka ATE a RPE synchrónneho SLAM toolboxu a skutočnej trajektórie robota

Typ evaluácie	Absolútna Chyba Trajektorie	Relatívna Chyba Polohy	Jednotky
RMSE	0,135	0,021	m
Mean	0,124	0,007	m
Median	0,117	0,003	m
Std.	0,055	0,020	m
Min	0,020	0,001	m
Max	0,391	0.417	m



Obr. 7: porovnanie trajektórie synchrónneho SLAM toolboxu so skutočnou trajektóriou robota

3 Zhodnotenie

V tejto práci sme sa zamerali na porovnanie presnosti vybraných SLAM algoritmov. Cartographer SLAM a SLAM toolbox sme odtestovali v prostredí ROS2 pomocou simulovaného datasetu ktorý sme získali použitím turtlebot3 balíka. Cielom bolo analyzovanie rozdielov trajektórie medzi trajektóriou získanou pomocou odometrie (ground truth) a trajketóriami získanými zo SLAM algoritmov, pričom presnosť bola hodnotená pomocou metódy absolútnej chyby trajektórie (ATE) a relatívnej chyby polohy (RPE).

Výsledky nam ukázali, že Cartographer SLAM dosiahol najnižšiu absolútnu aj relatívnu chybu trajektórie,
a vyprodukoval aj najlepšiu occupancy grid mapu. SLAM toolbox mal problémy s uzatváraním slučky, čo spôsobilo výraznú odchylku - posun celej miestnosti a drift v trajektorii mobilného robota, ktorý je možné vidieť na Obr. 6 a Obr. 7. Z porovnania asynchrónnej a synchrónnej metódy SLAM toolboxu sme zistili, že synchrónna metóda je presnejšia, avšak stále zaostáva za Cartographerom.

Tieto výsledky naznačujú, že Cartographer je najpresnejšou volbou spomedzi testovaných SLAM algoritmov, avšak Cartographer už nieje aktívne udržiavaný, čo može v buducnosti spôsobiť problémi s kompatibilitou s novšími ROS2 verziami.

Literatúra

[1] DUCHOŇ, František; TÖLGYESSY, Michal; DEKAN, Martin; RODINA, Jozef; CHOVANEC, Ľuboš et al. Riadenie mobilných robotov. Bratislava: FELIA, 2023. ISBN 978-80-89824-14-4.

[2] W. Hess, D. Kohler, H. Rapp and D. Andor, "Real-time loop closure in 2D LIDAR SLAM,"2016 IEEE International Conference on Robotics and Automation (ICRA), Stockholm, Sweden, 2016, pp. 1271-1278, doi: 10.1109/ICRA.2016.7487258

[3] Macenski et al., (2021). SLAM Toolbox: SLAM for the dynamic world. Journal of Open Source Software, 6(61), 2783, https://doi.org/10.21105/joss.02783

[4] Amey Kasar, *Benchmarking and Comparing Popular Visual SLAM Algorithms*, 2018, arXiv preprint `https://arxiv.org/abs/1811.09895`.

[5] Anton Filatov, Artyom Filatov, Kirill Krinkin, Baian Chen, Diana Molodan, *2D SLAM quality evaluation methods*, in Proceedings of the *FRUCT 2017*, pages 120–126, November 2017, DOI: `https://doi.org/10.23919/FRUCT.2017.8250173`.

Študentská vedecká a odborná činnosť
Fakultná prehliadka prác, 8. apríl 2025
Fakulta elektrotechniky a informatiky STU v Bratislave Sekcia 6. Robotika a kybernetika

Generation of AI images in ophthalmology

Dávid Szépvölgyi, Jarmila Pavlovičová, Michal Kováč

Institute of Robotics and Cybernetics, FEI STU v Bratislave
Faculty of Electrical Engineering and Information Technology of STU in Bratislava

xszepvolgyi@stuba.sk

Abstract

Generative models, particularly diffusion models, have recently emerged as state-of-the-art techniques, making them a promising choice for generating medical images. This article focuses on the generation of retina (fundus) scans that can be used in the detection of diabetes. Such an approach effectively addresses the challenges of collecting real-world medical data by offering a scalable solution for synthetic dataset creation.
In this work, we train a basic diffusion model and explore the impact of modifying its architecture and key hyperparameters. Our results demonstrate how these changes influence the model's performance. The findings confirm that diffusion models hold significant potential for medical image synthesis and diagnostic support.

1 Introduction

In the current era, there is an increasing demand for new and diverse data. Despite ongoing technological advancements, fresh datasets remain crucial for the continued development of artificial intelligence (AI) across various sectors. Modern AI systems, particularly large models, rely heavily on comprehensive, high-quality datasets to learn patterns and make accurate predictions. The broader and more representative these datasets are, the better the resulting model's performance becomes.

However, in the medical field, the acquisition of diverse patient data poses a unique set of challenges. Rare diseases are especially difficult to study because there are far fewer patients to draw samples from in the first place, making it hard to gather sufficient data for meaningful AI research. Strict privacy regulations, such as the General Data Protection Regulation (GDPR).

A promising way to solve these challenges involves generating artificial images that closely resemble real patient data. By creating synthetic medical images through advanced AI techniques, researchers can potentially sidestep some of the limitations posed by data scarcity or privacy regulations. Synthetic datasets can be shared more freely for training large AI models, facilitating further progress in medical research while still protecting patient confidentiality. This approach not only opens new possibilities for studying rare diseases but also helps accelerate the overall development of AI-driven solutions in healthcare.

High-quality and diverse medical datasets are essential for developing robust deep learning models in healthcare. However, obtaining sufficient patient data, especially for rare diseases, poses significant challenges as mentioned above. To address these limitations, this article explores the potential of synthetic data generation, focusing on *fundus images* used for diagnosing diabetes. By creating realistic synthetic fundus images, researchers can enrich existing datasets, bolster model training, and mitigate concerns tied to sensitive patient information.

This work centers on the use of diffusion models, specifically Denoising Diffusion Probabilistic Models (DDPM), as a state-of-the-art method in image synthesis [2]. Diffusion models have recently gained prominence due to their flexibility and high-fidelity outputs. That is why diffusion models are particularly well-suited for medical image synthesis.

Our primary **goals** are:

- **Generation of Fundus Images:** Train models with various configurations to identify optimal settings and assess the impact of different model components or sectors.
- **Evaluation of Generated Images:** Use the Fréchet Inception Distance (FID) score as a continuous benchmark throughout training. Upon completion, perform a visual assessment to evaluate the realism of the generated images and determine whether they achieve a trustworthy and usable quality level.

We will discuss the impact of different model architectures and explain the use of various loss functions. Improvements and problems observed throughout the process will be documented and contextualized based on the experiments conducted during training. Finally, our conclusions will be presented in light of these comparative findings.

2 Generative AI

In recent years, GANs (Generative Adversarial Networks) [1] and their variations have been considered state-of-the-art. However, diffusion models have now reached state-of-the-art benchmarks [2] and present a significant challenge to other methods. Their greatest advantage in image generation is the ability to produce highly detailed images with real-life accuracy, effectively combining the strengths of both VAE generators and GANs.

The main disadvantage of diffusion models is their relatively slow image generation process compared to competitors. However, the multi-step denoising process, as mentioned earlier, enables the creation of exceptionally detailed images.

The basic principle of diffusion models is to learn the data distribution, which can later be recreated using probability and structure infusion during the generation process [3].

In this work, the focus will be on Denoising Diffusion Probabilistic Models (DDPM).

3 Diffusion models

Diffusion models are derived from the diffusion process in physics, which, in simple terms, describes the movement of particles based on concentration gradients [3]. In these models, we reconstruct a structure by removing noise and gradually restoring the image. However, due to their probabilistic nature, the process results in variations of the learned structure [9].

3.1 Training Process of Diffusion Models

The training consists of two processes:

The forward process is where the data is prepared; no network is involved during this step. First, we take images from a batch and select a time step, which determines how much noise we add to the original image. To embed noise into clean images, we require a noise scheduler [4].

A noise scheduler is an algorithm that determines the noise value and its distribution depending on the time step. The noise addition at each step depends on the previous time step, as shown in the equation below:

$$q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}\left(\mathbf{x}_t;\ \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}\right) \tag{1}$$

Where q is our forward process, $\mathbf{x}_t$ is the output of the forward process at step t (naturally $\mathbf{x}_{t-1}$ is an input at step t). $\mathcal{N}$ is a *normal distribution* [4]

Fortunately, the noise scheduler handles this step-by-step process for us. By following the corresponding equation, we can account for the noise added at each time step without having to iterate through the entire sequence.

$$q(\mathbf{x}_t|\mathbf{x}_0) = \sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon} = \mathcal{N}\left(\mathbf{x}_t;\ \sqrt{\bar{\alpha}_t}\mathbf{x}_0, (1-\bar{\alpha}_t)\mathbf{I}\right) \tag{2}$$

Calculation of noise at any arbitrary step t ($\bar{\alpha}_t$ is known because β_t is known). Now we can precisely track what the added noise looks like at each timestamp without iterating from the first one.

After these steps and the addition of noise to the images, it is time for the denoising step, where a convolutional network becomes essential.

An important point to consider is that during training, no images are generated. Instead, the algorithm uses the forward process to create noised images. These images are then fed into the network, which predicts the noise present in each sample from the batch. The difference between the actual noise and the predicted noise is calculated and passed to the loss function. After processing the entire batch, a gradient step is performed [4].

$$\nabla_\theta \left\| \epsilon - \epsilon_\theta(\sqrt{\overline{\alpha}_t}\, x_0 + \sqrt{1-\overline{\alpha}_t}\, \epsilon,\ t) \right\|^2 \tag{3}$$

The training process is repeated until stated by number of epochs or by other condition.

The simplified training process on the algorithm 1.

Algorithm 1 Training

1: **repeat**
2: $x_0 \sim q(x_0)$
3: $t \sim \text{Uniform}(\{1, \ldots, T\})$
4: $\epsilon \sim \mathcal{N}(0, I)$
5: **Take gradient descent step on**

$$\nabla_\theta \left\| \epsilon - \epsilon_\theta\Big(\sqrt{\overline{\alpha}_t}\, x_0 + \sqrt{1-\overline{\alpha}_t}\, \epsilon,\ t \Big) \right\|^2$$

6: **until** *converged*

3.2 Image generation process of diffusion models

The generation process differs from training in several key aspects, as no images are generated during training. During generation, the input image passes through the network, where the noise present in the image is predicted and subtracted. However, this does not immediately produce the final image; instead, the image undergoes multiple iterations through the network, corresponding to the defined number of time steps.

At each iteration, for time steps greater than zero, the network predicts the noise in the current input image, removes it, and then adds a smaller amount of noise. This updated image becomes the input for the next timestep, which is one step lower than the previous one. This iterative refinement continues until the final timestep is reached, gradually reconstructing a clean and structured image.

A simplified visualization of the generation process is shown here:

3.3 Structure of diffusion models

In this method, a convolutional network is used. Specifically, for diffusion models, we use a model structure called a U-Net. This structure processes the image using convolutions to extract features. It consists of three main parts:

Algorithm 2 Sampling

$\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$
for $t = T, \ldots, 1$ **do**
  $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = 0$
  $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}} \epsilon_\theta(\mathbf{x}_t, t) \right) + \sigma_t \mathbf{z}$
end for
return $\mathbf{x}_0$

Downsampling block: This part uses convolutions to extract features at each level of the architecture. These extracted features are later used in the upsampling block.
Middle block (also called the bottleneck): This block is responsible for extracting features at the lowest resolution level, which is particularly useful for capturing global similarities.
Upsampling block: This block reconstructs the image, using the previously extracted features to guide the reconstruction. The final image is obtained after this block.
ResNet blocks: Serve as the primary building components of each layer, as they contain the necessary convolutional layers. Additionally, multiple ResNet blocks can be stacked in series to increase the network's capacity and depth [10].
Skip connections: Allow the features extracted during downsampling to be used directly in the upsampling process, improving detail preservation.

Additionally, the network uses an embedder, which keeps track of the current denoising timestep and helps the model condition its predictions accordingly.

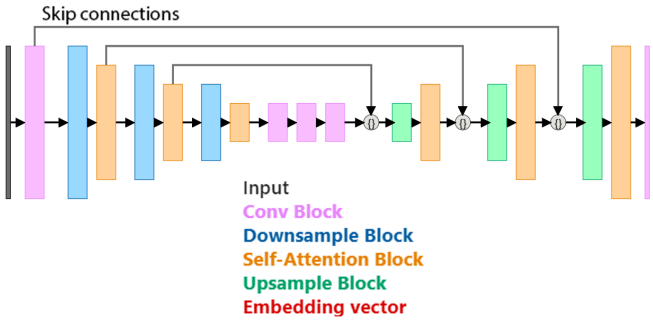


Figure 1: U-net Basic structure [5].

4 Image quality benchmarking

For the generated images, we can perform an analysis of their quality and resemblance to reality. There are many techniques that can be used, varying based on the specific perspective of image adequacy. In this work, we will use the Fréchet Inception Distance (FID).

4.1 Fréchet Inception Distance

Fréchet Inception Distance (FID) provides a principled way to measure the similarity between distributions of real and generated images. By working in a high-level feature space and comparing Gaussian approximations of these distributions, FID offers a reliable and interpretable metric that correlates well with human judgment and the level of image perturbation [6]. Its wide adoption in the research community underscores its importance as a benchmark for evaluating generative models.We call the Fréchet distance $d(\cdot, \cdot)$ between the Gaussian with mean $(\mathbf{m}, \mathbf{C})$ obtained from $p(\cdot)$ and the Gaussian with mean $(\mathbf{m}_w, \mathbf{C}_w)$ obtained from $p_w(\cdot)$ the "Fréchet Inception Distance" (FID), which is given by:

$$d^2\left((\mathbf{m}, \mathbf{C}), (\mathbf{m}_w, \mathbf{C}_w)\right) = \|\mathbf{m} - \mathbf{m}_w\|_2^2 + \mathrm{Tr}\left(\mathbf{C} + \mathbf{C}_w - 2\left(\mathbf{C}\mathbf{C}_w\right)^{1/2}\right). \quad (4)$$

4.2 Evaluation by observer

We will also evaluate the images based on personal perception of the details that should exhibit specific characteristics, locations, and shapes. These details often serve as key indicators of the model's ability to preserve the necessary image structures. Additionally, they provide confirmation that the model's architecture is sufficiently complex to accurately generate and maintain these features.

5 Metodology

- We begin with a simple baseline architecture to observe the effect of different training datasets on the quality of generated images.
- We systematically modify the network structure to evaluate how architectural changes influence model performance.
- We experiment with various loss functions to analyze their impact on the training process and the realism of generated outputs.

To meet these requirements, we must define the model's properties and address any issues that arise by adjusting them based on the outcomes.

6 Experiments on real-life datasets

We conducted various experiments, which will be described below. The differences among them lay in structure, hyper parameters, and datasets.

Correlations between experiments were analyzed, and if any issues arose, they were identified and examined. Potential causes of these problems were then targeted to explore new findings, either confirming or disproving assumptions about the areas responsible for the issues.

6.1 Signs that we focus on

The images we aim to generate are images of the fundus eye containing numerous detailed features. The main areas of focus

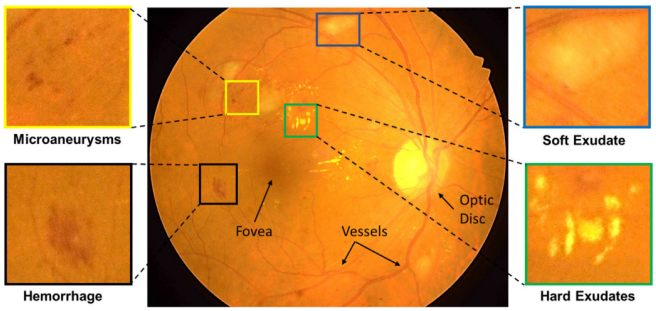


Figure 2: Classifications of fundus image [11]

include the veins, the optic disk and its structure, the fovea, and, importantly, symptoms such as exudates, hemorrhages, and microaneurysms. These features are illustrated in Figure 2.

For the veins, we want them to appear naturally connected, rather than randomly starting or disappearing at arbitrary points, which would be biologically impossible. Additionally, the veins should converge at the optic disk, a crucial feature we want to observe in the generated images.

Regarding symptoms, we are focusing on unconditioned generation, meaning they may or may not appear in the images.

6.2 Hyper parameters

Table 1: Hyperparameter configuration

Hyperparameter	Value
Batch size	20
Number of epochs	1 000
Learning rate	1e−4
Diffusion time steps	1 000
Beta start / end	0.0001 / 0.02
Dropout scale	0.00
Down / Up block count	3
Mid block count	2
Resnet block per layer	1
Group norm size	2
Image size	256
Loss function	MSE

6.3 Used datasets

Two datasets with high-resolution images were used during training. The first one, the DDR dataset, contains a large number of samples (6,260). However, there are noticeable differences between the images, such as varying markings on the eyes or differences in cropping at the top and bottom of the eye scans. These inconsistencies are caused by the images being captured with different machines.

The second dataset, IDRID, has a noticeably smaller number of samples (376). However, the images are more uniform, with consistent markings and cropping patterns.

6.4 Experiment with datasets

The choice of dataset resulted in the most significant observable differences. As expected, the larger but less uniform DDR dataset achieved better results in generating detailed sections. Both models had exactly 2,587,491 parameters. This relatively small number was chosen due to training time constraints. Taining on the larger dataset took around 17 minutes per epoch for this configuration and approximately 50 minutes for a larger network structure with around 16,266,941 parameters.

On the other hand, the model trained on the IDRID dataset learned the overall shape of the structures very well but failed to capture finer details, resulting in improper vein generation. This training took only around 2 minutes per epoch.

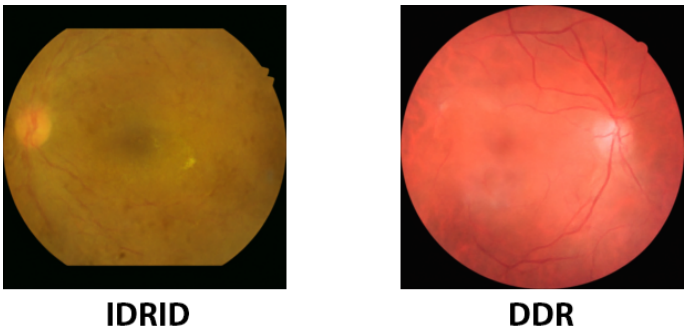


Figure 3: Comparition of DDR and IDRID dataset with 2 mil parameters

6.5 Stucture of the U-net

As one of the first experiments, we tested different model structures that varied in the number of levels and the number of layers per level.

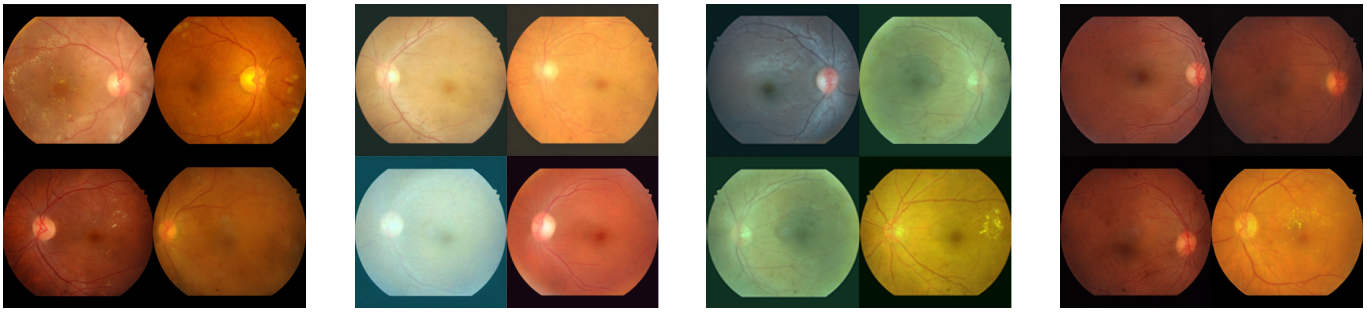
The first structure we used had four levels, with two ResNet blocks per level. However, this structure lacked attention layers, and we suspected this contributed to the generated images being tinted with a color hue. This problem persisted throughout all training epochs.

Table 2: 1. Structure

Hyperparameter	Value
Down / Up block count	5 : 0x attention layers
Resnet block per layer	2

The next structure introduced attention layers in an attempt to preserve global features and eliminate the hue tint issue. Unfortunately, the problem did not disappear, and this structure did not result in significant improvements. Both the initial and this modified structure had around 50 million trainable parameters, which, based on previous training experience, should have been sufficient.

Finally, we tested a third structure with 57 million trainable parameters, consisting of five downsampling and five upsampling layers, with attention layers incorporated after every ResNet block in the second and fourth layers. The mid-layer



(a) Train dataset (b) First structure (c) Second sstructure (d) Final structure

Figure 4: Training on different structuresn: 4a shows the training dataset, 4b shows generated images using second architecture, 4c shows generated images using first architecture, 4d shows generated images using final architecture

Table 3: 2. Structure

Hyperparameter	Value
Down / Up block count	5 : 1x attention layers
Resnet block per layer	2

included two ResNet blocks with an attention layer between them. This structure proved to be sufficient for the network to learn properly. Additionally, we introduced dropout with a value of 0.08 to help prevent overfitting, which was likely contributing to the persistent hue tint problem.

Table 4: 3. Structure

Hyperparameter	Value
Down / Up block count	5 : 2x attention layers
Resnet block per layer	2
Dropout scale	0.08

6.6 Loss functions

For loss functions we chose to compare these loss functions:

6.6.1 MSE loss

Also known as L2 loss function is widely used in AI training, particularly for regression tasks. It is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{5}$$

where $\hat{y}_i$ represents the actual values, y_i the predicted values, and n the number of data points. MSE is beneficial because it penalizes large errors more heavily due to the squared term, ensuring the model prioritizes reducing significant deviations. Smooth and differentiable nature makes it suitable for gradient-based optimization, leading to stable convergence. However, MSE is sensitive to outliers. [7].

6.6.2 L1 loss

Also known as Mean absolute error is summation of absolute difference between actual and predicted values.

$$L1 = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \tag{6}$$

where $\hat{y}_i$ represents the actual values, y_i the predicted values, and n the number of data points. It have advantages in convergence when it actually reaches zero and convergence to it bit faster. Also unlike MSE outliers do not disproportionately affect the loss [7].

6.6.3 Hubert loss

It is combination of above mentioned loss functions that combines their advantages. It uses a coefficient that determines influence of them, which introduces new hyperparameter that needs to be adjusted.

$$L_i = \begin{cases} \frac{1}{2}z^2, & \text{for } |z^{(i)}| \le \delta \\ \delta|z| - \frac{1}{2}\delta^2, & \text{for } |z^{(i)}| > \delta \end{cases} \tag{7}$$

Where δ is the coefficient to determine when the quadratic component of the Huber loss penalizes outliers and is applied to errors equal to or smaller than δ, ensuring a smooth and differentiable loss function in this range. On the other hand, the linear component prevents the model from being over-penalized for large errors by applying a less aggressive penalty to errors greater than δ, thereby balancing robustness and sensitivity in training [8].

6.6.4 Change of loss function

Regarding the different loss functions tested, not much difference was observed. All results were very similar, with no significant improvements or declines in image quality. This can be illustrated by comparing the FID curves, which remained nearly identical across the different loss functions.

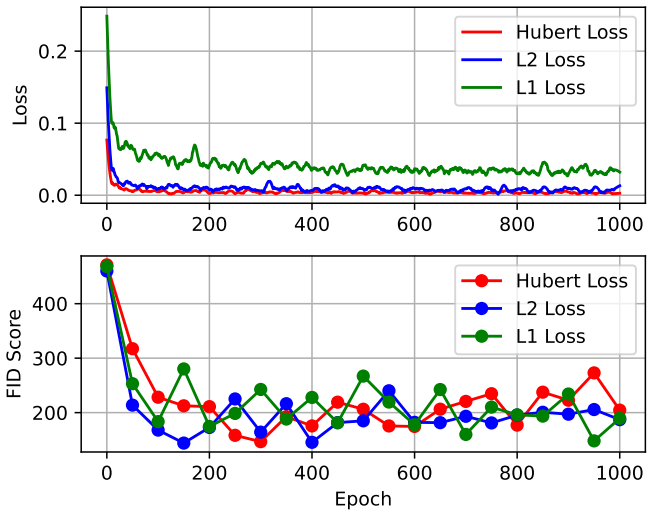


Figure 5: Loss affection

7 Conclusion

The experiments demonstrated how changes in hyperparameters and the structure of the DDPM model influence its performance. The final results, after applying all observations and adjustments, led to the successful generation of realistic images. This was achieved despite relatively high FID scores, highlighting the limitations of relying solely on such metrics for diffusion models. Evaluating diffusion models based on quantitative scores is challenging due to the probabilistic nature of the generation process and the variability within datasets.

The final outcomes were primarily evaluated through visual inspection, showing promising results and indicating potential for practical application in the medical field. The generated images displayed realistic anatomical features, making this approach a valuable tool for synthetic dataset creation in areas such as diabetic retinopathy research.

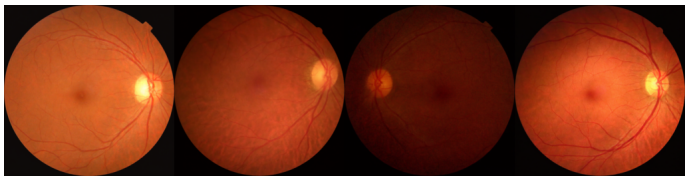


Figure 6: Generated image samples in 800. epoch

Thanks

I would like to sincerely thank Michal Kovač for being an excellent supervisor and for always being there to help whenever I got stuck in my research. I am also very grateful to Jarmila Pavlovičová for her kind assistance with all the official matters and paperwork. Finally, I would like to express my thanks to Danish professor Dan Witzner Hansen for his valuable guidance and help with the theory of generating images and for answering all my questions around this topic.

References

[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio Generative Adversarial Networks. *arXiv preprint arXiv:1406.2661*, 2014.

[2] Ryan Po, Wang Yifan, Vladislav Golyanik, Kfir Aberman, Jonathan T. Barron, Amit H. Bermano, Eric Ryan Chan, Tali Dekel, Aleksander Holynski, Angjoo Kanazawa, C. Karen Liu, Lingjie Liu, Ben Mildenhall, Matthias Nießner, Björn Ommer, Christian Theobalt, Peter Wonka and Gordon Wetzstein State of the Art on Diffusion Models for Visual Computing. *arXiv preprint arXiv:2310.07204*, 2023.

[3] Jascha Sohl-Dickstein and Eric A. Weiss and Niru Maheswaranathan and Surya Ganguli. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. *arXiv preprint arXiv:1503.03585*, 2015.

[4] Jonathan Ho and Ajay Jain and Pieter Abbeel Denoising Diffusion Probabilistic Models. *arXiv eprint arXiv:2006.11239*, 2020.

[5] Erdem, Baris Step-by-step Visual Introduction to Diffusion Models. *Medium, Accessed: 2025-03-27*, 2023.

[6] Yuli Wu and Fucheng Liu and Rüveyda Yilmaz and Henning Konermann and Peter Walter and Johannes Stegmaier A Pragmatic Note on Evaluating Generative Models with Fréchet Inception Distance for Retinal Image Synthesis. *arXiv preprint arXiv:2502.17160*, 2025.

[7] Juan Terven, Diana M. Cordova-Esparza, Alfonso Ramirez-Pedraza, Edgar A. Chavez-Urbiola and Julio A. Romero-Gonzalez Loss Functions and Metrics in Deep Learning. *arXiv preprint arXiv:2307.02694*, 2024.

[8] K. Gokcesu and H. Gokcesu. Generalized Huber loss for robust learning and its efficient minimization for robust statistics. *arXiv preprint arXiv:2108.12627*, 2021.

[9] Zhenning Shi, Haoshuai Zheng, Chen Xu, Changsheng Dong, Bin Pan, Xie xueshuo, Along He, Tao Li and Huazhu Fu Resfusion: Denoising Diffusion Probabilistic Models for Image Restoration Based on Prior Residual Noise. *https://openreview.net/forum?id=JrIPBXWiS8*, 2024.

[10] Kaiming He and Xiangyu Zhang and Shaoqing Ren and Jian Sun Deep Residual Learning for Image Recognition *https://arxiv.org/abs/1512.033858*, 2015.

[11] Prasanna Porwal, Samiksha Pachade, Manesh Kokare, Girish Deshmukh, Jaemin Son, etc. IDRiD: Diabetic Retinopathy – Segmentation and Grading Challenge *Medical Image Analysis*, 59:101561, 2020.

Simulation Analysis of a Sensor Utilizing Phototransistors for Measuring the Gravitational Constant in a Microsatellite

Viktor Hrúz[1], Ľuboš Chovanec[2]

[1]Institute of Science and Research, FEI STU in Bratislava
[2]Institude of robotics and Cybernetics

xhruzv@stuba.sk

Abstract – Gravitational constant is one of the fundamental physical constants that are crucial for the quantitative description of gravitational interaction. It was first introduced by Isaac Newton as part of his law of universal gravitation, which defines the gravitational attraction between bodies. This constant not only enables the description of gravitational forces but also allows for the calculation of the masses of celestial bodies and objects on Earth. Despite its significance, accurately measuring it is extremely challenging due to its high sensitivity to experimental conditions. This work focuses on the design and simulation of a sensor intended for measuring the gravitational constant. The experimental concept is based on analyzing the gravitational interaction between two spherical bodies placed onboard a microsatellite, with the goal of achieving higher measurement accuracy compared to ground-based experiments.

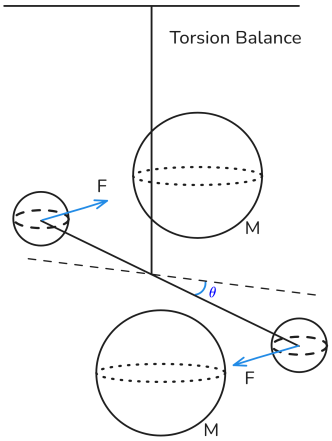


Fig. 1: Torsion balance experiment

1 Measurement of the Gravitational Constant

In 1687, Newton, in his work Philosophiae Naturalis Principia Mathematica, formulated a law stating that every particle attracts every other particle in the universe with a force that is proportional to the product of their masses and inversely proportional to the square of the distance between them.

$$F = G\frac{m_1 m_2}{r^2} \tag{1}$$

Where F represents the gravitational force between two bodies with masses m_1 and m_2 at a distance r, and G is the gravitational constant.

To experimentally determine the value of G, it is essential to precisely know the masses of the interacting objects and measure the gravitational interaction between them. In laboratory experiments, where the masses of objects are known with high accuracy, the gravitational interaction is extremely weak, making its detection highly challenging. On the other hand, for large astronomical bodies such as planets, gravitational effects are observable, but accurately determining their masses presents a significant challenge.

Modern methods for measuring the gravitational constant are based on the principle of Henry Cavendish's experiment, which enabled the quantification of the gravitational force between two laboratory objects [1]. His experiment utilized a torsion balance, consisting of a thin torsion wire suspending two smaller spheres of known mass. The torsion wire allowed for highly sensitive detection of rotational motion caused by the gravitational influence of larger spheres present in the experiment.

Based on Cavendish's results, the value of the gravitational constant was determined to be $6.754 \times 10^{-11} m^3 kg^{-1} s^{-2}$ [2]. According to the latest CODATA 2022 recommendations, its value is set at $6.67430 \times 10^{-11} m^3 kg^{-1} s^{-2}$ [3], representing a difference of approximately 1.2%. Despite more than 200 years of experimental research, the accuracy of measuring the gravitational constant has not improved as significantly as that of other fundamental constants.

1.1 Sources of Disturbance

Most modern experiments use a torsion balance or a torsion pendulum, monitoring the frequency or amplitude of their motion. One of the main sources of experimental errors is interference caused by air movement, which is why modern experiments are often conducted in a vacuum. However, torsion balances and similar devices are extremely sensitive to vibrations, which can be induced by vacuum pumps or external mechanical influences, including the movement of people near

the experiment.

In addition to mechanical disturbances, another issue is electromagnetic radiation, which can affect sensitive detection systems. To minimize these effects, thorough shielding of the apparatus against electromagnetic interference and isolation from vibration sources are essential.[4]

Another challenge in determining the value of G is the need for precise measurement of all experimental parameters, such as the masses of test bodies and gravitational sources, the period of the mechanical oscillator, and the distance between objects. Any uncertainty in these parameters leads to the accumulation of systematic and random errors, causing significant variability in the measured values of G. This sensitivity to experimental conditions is one of the reasons why the accuracy of measuring the gravitational constant remains relatively low even after more than two centuries of research.[4]

1.2 Measurement of the Gravitational Constant in space

Measuring the gravitational constant in space would eliminate many sources of experimental disturbance present in Earth-based measurements. One possible solution is the use of a satellite with a drag-free control system.

A drag-free satellite operates on the principle of isolating an inner mass, known as the proof mass, from external forces such as aerodynamic drag or solar radiation pressure. The proof mass is allowed to float freely inside the satellite, with its relative position to the outer shell continuously measured. The outer shell is equipped with an active compensation system that adjusts its trajectory to closely follow the motion of the proof mass, ensuring no contact occurs[5]. In an ideal case, the inner reference mass is influenced solely by gravitational forces, which enables precise gravitational experiments.

In the experiment presented in [4] and [6], [7] the authors propose the use of the drag-free satellite principle to analyze the gravitational interaction between two test spheres in a CubeSat-type satellite environment. Precise tracking of the position of these test spheres is essential and must be based on an optoelectronic principle to minimize disturbance to the measured objects.

The thesis [6] proposed the use of TOF sensors. This approach requires a complex calibration process, including determining the sensor's orientation relative to the measured objects. After successful calibration of the sensors in the sphere's coordinate system, the position of the objects was determined using trilateration. This method achieved an accuracy of less than 1 mm when tracking the trajectory of the objects.

In this experiment, three TOF sensors were used to measure the sphere's position. The disadvantage of this approach is the need for precise calibration, which requires an accurate determination of the sphere's initial position. According to the authors, the sensor measurements exhibited significant noise, and the moving average method was not sufficiently effective in eliminating unwanted influences.

Thesis [7] primarily focused on the design of a three-axis magnetic levitation system utilizing six electromagnets. In the

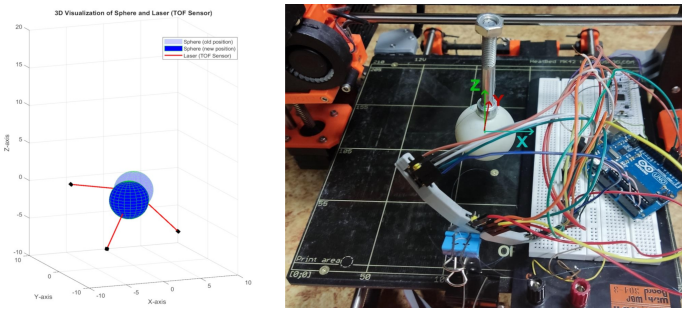


Fig. 2: Use of TOF Sensors for Sphere Position Measurement [6]

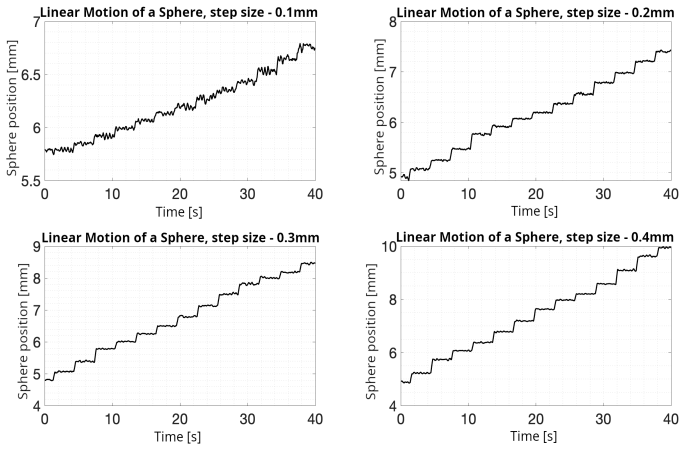


Fig. 3: Resolution of the Phototransistor Sensor [7]

experimental part, a stand with an electromagnet was designed to maintain a metal sphere in a suspended equilibrium state. Additionally, a method for measuring the sphere's position using phototransistors was developed, achieving a resolution of 0.2 mm. Experiments tested the movement of the sphere along one and two axes. During these measurements, the sensor exhibited a maximum deviation of 0.4 mm with an average deviation of 0.17 mm.

In this work, we plan to extend the concept of a position sensor based on phototransistors through simulation. The simulation model will allow us to better understand the sensor's functionality and optimize its parameters to make it suitable for use in a CubeSat environment.

2 Phototransistor-Based Sensor

2.1 Position Sensing

The method for detecting an object's position is based on creating a shadow on the phototransistors. When an object is placed between an LED and a phototransistor, the intensity of light reaching the phototransistor decreases, causing a voltage drop in the output circuit. This voltage drop can be interpreted as the presence of the object in front of a specific phototransistor.

Since the only available information from the phototransistors is the presence or absence of the object, methods such

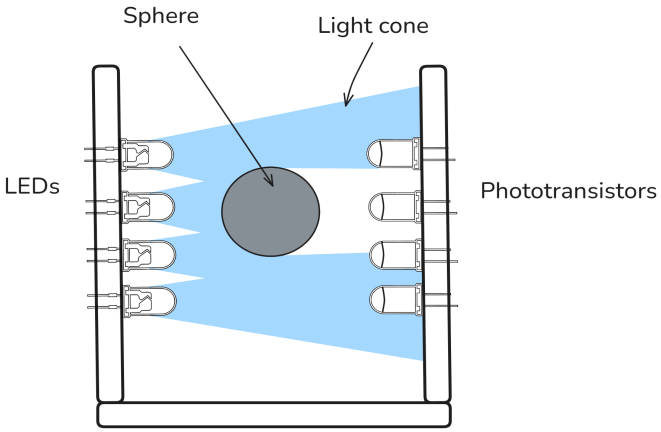


Fig. 4: Sphere Position Sensing

as trilateration, used in thesis [6], are not applicable, as the exact distance of the object from individual phototransistors is unknown. However, we can estimate the position using the known locations of the sensing components.

During sensor calibration, the operating range of each phototransistor can be determined and then converted into shading percentage, where 0% represents full illumination and 100% represents complete occlusion, meaning the object blocks all light reaching the phototransistor. During measurement, the detected voltage values can be transformed into this range and used as a weighting factor in position calculation. The shading percentage thus determines the weight of each phototransistor's contribution to the object's position estimation. The final position is then calculated as the weighted average of the intensities detected by the individual phototransistors.

In thesis [7], a similar approach was used, where phototransistors were arranged in a row with variable spacing. For our purposes, we propose extending this method to a two-dimensional $n \times n$ array of components. Each component is assigned coordinates within the matrix, allowing for the localization of the object's center along two axes.

2.2 Definition of Simulation Parameters

A CubeSat defines a single unit as $1U$, corresponding to a cube with a volume of $10cm^3$. [8]. For effective experimental testing of gravitational interactions, the test spheres must have sufficient mass and be made of a non-ferromagnetic metal or alloy. In our simulation, we will consider test spheres with a diameter of $5cm$.With a density of approximately $20g/cm^3$, corresponding to materials such as tungsten or similar metals, the total mass (along with additional equipment such as batteries, communication devices, and other components) would exceed the theoretical mass limit of a $2U$ CubeSat [8]. However, for the purposes of our simulation, we will disregard this limitation and allow for adjustable sphere dimensions. To simplify the model, we will consider only a single $1U$ section of the CubeSat with one sphere placed inside.

As mentioned in the previous chapter, we will use sensor arrays to detect the position of the object. We consider two

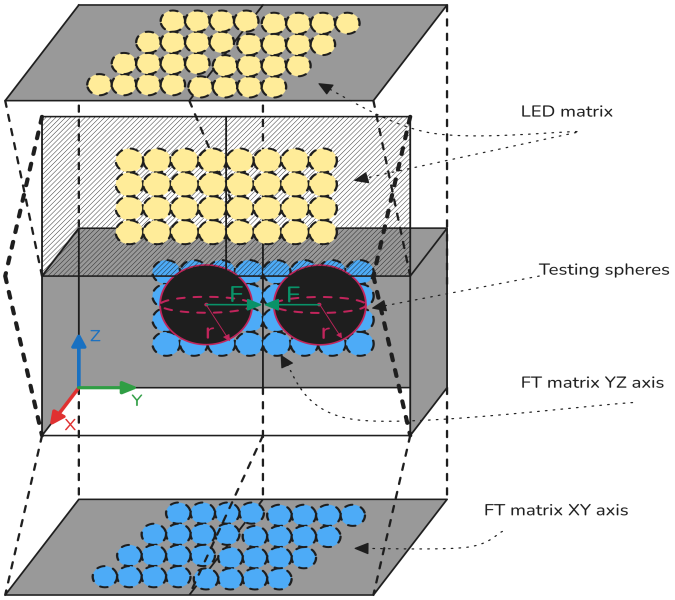


Fig. 5: Tracking the position of a spheres in 3 axes

configurations: two sets of arrays placed opposite each other on the cube's walls, with one set used to detect the object's position in the XY plane and the other in the YZ plane. We also introduce parameters such as sensor spacing, which determines the distance between individual sensors, and matrix start offset, which defines the initial shift of the array. The goal is to precisely specify the area where the sensors will be placed, particularly by defining empty spaces that allow for the free placement of components. This structure is designed with future optimization algorithms in mind, enabling the refinement of sensor placement and other experimental parameters.

Next, we define the parameters viewing angle and reception angle. The viewing angle represents the angle at which the light cone reaches 50% of its maximum intensity. This parameter is often referred to as the half-power angle or half-intensity angle. Its value depends on the shape of the lens and the structural design of the light source. Similarly, for the phototransistor, we define the reception angle, which determines the range of angles around the component's normal within which the sensor can effectively capture incoming light. The final variable parameters will be the radius of the test sphere and the position of its center.

2.3 Visibility Calculation

After defining the parameters, the matrix dimensions and the number of empty positions within the matrix are calculated. Once the occupied positions are selected, each component is assigned its exact spatial coordinates within the cube's coordinate system.

Next, it is necessary to determine the visibility of each phototransistor, meaning identifying which LED sources it can respond to. This step is crucial since the opposing matrices may not be fully populated.

Given the known emission angle of the LED, θ_{LED}, we

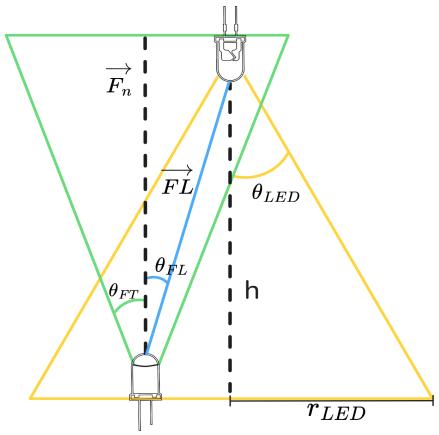


Fig. 6: Determining LED Visibility

can calculate the radius of the light cone on the opposite wall, r_{LED}.This radius defines the illuminated area in which a phototransistor must be located to detect the emitted light.

In addition to identifying the illuminated area, it is also necessary to verify whether the light reaches the phototransistor at an angle it can detect. To achieve this, we define the vector between the center of the phototransistor and the LED, $\overrightarrow{FL}$, as well as the normal vector of the phototransistor, $\overrightarrow{F_n}$. Using the dot product of these vectors, we compute the incidence angle of the light, θ_{FL}.

$$\theta_{FL} = \cos^{-1}\left(\frac{\overrightarrow{FL} \cdot \overrightarrow{F_n}}{|\overrightarrow{FL}| \cdot |\overrightarrow{F_n}|}\right) \tag{2}$$

The computed angle is compared with the phototransistor's reception angle, θ_{FT}, to ensure it falls within the sensor's field of view for successful detection.

Among all visible LEDs, we select the one that produces the smallest angle of incidence on the phototransistor, as this LED provides the highest light intensity. The chosen LED is then used to compute the shading percentage of the phototransistor, determining how much of the incoming light is blocked by the object.

2.4 Calculation of Shading Percentage

To accurately compute the shading percentage of a phototransistor, we must select an appropriate modeling approach. We identified two possible methods:

- Empirical Modeling Based on Input-Output Data
- Geometric Modeling of Phototransistor Shading

If we assume that the phototransistor operates in its active region, the voltage between the collector and emitter V_{CE} is proportional to the base current, which depends on the intensity of the incident light. To verify this relationship, we can design a test circuit where the LED current is controlled by a PWM signal from a microcontroller, gradually decreasing the light intensity from full brightness to the lowest possible level, thus covering the entire operating range of the phototransistor..

The collected data allows us to construct an input-output characteristic by measuring the phototransistor's output voltage as a function of the current flowing through the LED. However, this alone is insufficient for precise phototransistor modeling, as it only describes the relationship between the V_{CE} voltage and the LED current. In our system, we require the input variable to be light intensity. To determine this dependency, we would need to measure illumination intensity, which would require an additional sensor.

Although this approach is feasible, it comes with several limitations. The model would capture only the characteristics of a specific phototransistor, making it unreliable when using different components. Additionally, if individual phototransistors in the matrix exhibit varying maximum illumination levels, a separate model would have to be created for each one, significantly increasing system complexity.

We have therefore decided to model shading geometrically, directly determining the extent to which each phototransistor is obscured rather than physically modeling the component itself. In a real setup, this approach is equivalent to bypassing calibration, as we can directly read the shading percentage without the need for additional calibration, with values ranging from 0% to 100%.

As an example, consider the situation illustrated in the figure 7. The phototransistor is positioned on the bottom wall of the cube, while its corresponding LED is located on the top wall.

The first step is to determine the two-dimensional plane in which shading will be computed - in this case, the XY plane. The height for this calculation corresponds to the center of the sphere, where its diameter is largest. At this height level, we compute the radius of the LED's light cone and the phototransistor's reception cone using trigonometric formulas. This method simplifies the shading calculation by reducing a three-dimensional problem to a two-dimensional one. The final step involves solving the intersection of three circles.

There are many ways in which three circles can overlap. According to [9], there are 13 different scenarios, excluding

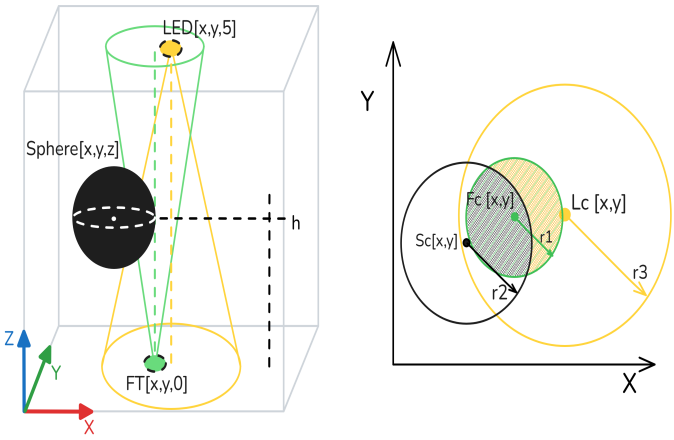


Fig. 7: Determining Shading percentage Based on Cone Overlap

cases where the circles are tangent. For an accurate evaluation of all possible cases (where both the radius and position of each circle vary) it is necessary to design an algorithm capable of precisely resolving these scenarios.

To determine the shading percentage, we first calculate the intersection area between the phototransistor's circle and the LED's circle, representing the fully illuminated region. Next, we need to compute the overlap of this intersection with the sphere's circle, which corresponds to the shaded area.

If one of the circles is completely inside the other, we can simplify the problem to the intersection of two circles. As an example, we use figure 7. If the circle of the phototransistor is entirely inside the circle of the LED diode, the entire area of the phototransistor's circle represents the illuminated area. In this case, if the circle of the sphere intersects with the circle of the phototransistor, it is sufficient to calculate the intersection of these two circles. The equation for calculating the area of intersection of two circles can be expressed as:

$$A = r_1^2 \cos^{-1}\left(\frac{d^2 + r_1^2 - r_2^2}{2dr_1}\right) + r_2^2 \cos^{-1}\left(\frac{d^2 + r_2^2 - r_1^2}{2dr_2}\right) - \frac{1}{2}\sqrt{(-d + r_1 + r_2)(d + r_1 - r_2)(d - r_1 + r_2)(d + r_1 + r_2)} \tag{3}$$

where d is the distance between the centers of the circles F_C and S_C. Similarly, we proceed in the opposite cases, for example, when the LED diode circle is completely inside the phototransistor circle.

This approach, however, does not cover all scenarios, as we still need to calculate the intersection of three circles when neither of the circles is completely contained within the other. While searching for a solution to this problem, we came across an article that describes how to solve this problem. [10]. The author of the article presents two approximation methods:

- Monte Carlo method
- Approximation using a quadtree

Both methods work on the principle of bounding the area around the common points of intersection and then filling the space with points or squares according to the given resolution, thus accumulating the calculated intersection area. Although these methods provide a good foundation, due to the computational complexity of the simulation, they prove to be inefficient.

The author also presents a third approach, called the Exact solution [10]. This method utilizes the fact that from the common points of intersection, we can form a polygon. From each segment of this polygon, we then determine the arcs that form the remainder of the intersection.

Since each segment can lead to two possible arcs for each circle, it is important to always analyze them in the same direction, thus reducing the number of possibilities to three. Finally, we calculate the arcs for all three options, determine the center of the arc, and include the one with the smallest height for each segment in the final area. The area of the arc can be calculated by finding the area of the circular segment and subtracting the area of the triangle, where the base is the chord of the arc.

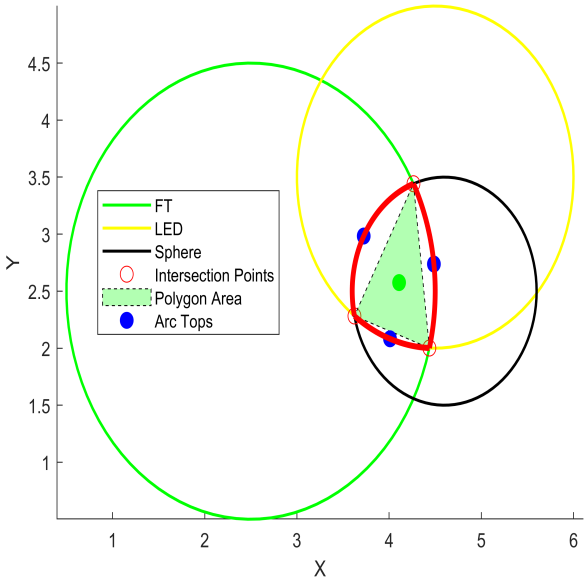


Fig. 8: Area of the intersection of three circles

2.5 Position Calculation

After determining the shading percentage for each phototransistor, we can proceed with calculating the object's position. As mentioned in section 2.1, the final position is determined as the weighted average of the positions of individual phototransistors, where the weights correspond to the degree of occlusion. The equation for calculating the object's position along the x-axis is:

$$X_{Obj} = \frac{\sum_i x_i . S_i}{\sum_i S_i} \tag{4}$$

where x_i is the coordinate of a given phototransistor, and S_i represents its percentage of occlusion.

To visualize the simulation, a companion program was created that displays the setup, the object's position, and visualizes the shading of individual matrices using a heatmap (fig. 9). The dimensions used were half the size of a 1U CubeSat, $5 \times 5 \times 5cm$. The radius of the test sphere was set to $1.25cm$, and the sensor spacing along with the initial offset was set to $1cm$.

2.6 Application of Simulation in Evolutionary Algorithms

The simulation provides a tool for analyzing the behavior of the sensor with a test sphere in various positions and simulated trajectories. Additionally, its structure allows for integration into evolutionary algorithms, enabling the optimization of the setup and testing of different component parameters. As an example, we present the application of the simulation in a genetic algorithm.

One of the key factors for optimization is the number of components. Testing has shown that increasing the density of sensor arrays can improve accuracy to below $0.05mm$. However, too many components can make the circuit more complex, which is not always beneficial. Therefore, finding the optimal balance

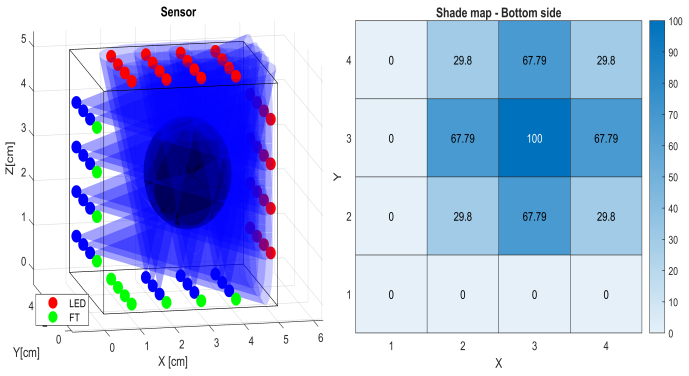


Fig. 9: Simulation Visualization

between the placement and number of components is well-suited for a genetic algorithm.

Since we define possible component placements rather than their exact configuration, we can create a dense matrix with a sensor spacing of $2,5mm$. For a cube with dimensions $5 \times 5 \times 5cm$ this results in a 19×19 matrix, providing 1444 available positions for the entire sensor. These positions form the chromosome of an individual in the genetic algorithm's population.

Another key step is defining the evaluation function (fitness), which determines the quality of each solution. Typically, a single optimal solution is sought, but due to the nature of this problem, we have chosen a Pareto-optimal approach, where we search for so-called non-dominated solutions. Each individual is evaluated based on two criteria:

$$Fit_1 = \sum_{i=1}^{n} \left((x_{obj,i} - x_{est,i})^2 + (y_{obj,i} - y_{est,i})^2 + (z_{obj,i} - z_{est,i})^2 \right) \tag{5}$$

$$Fit_2 = \alpha \cdot \text{Unused}_{\text{LED}} + \beta \cdot \text{Unused}_{\text{FT}} + \text{Num}_{\text{LED}} + \text{Num}_{\text{FT}}$$

The first fitness function evaluates the accuracy of position estimation, while the second focuses on the efficient use of components, such as minimizing the total number of elements and the weighted penalty for unused ones. A solution is considered better than another if it is at least as good in all criteria and better in at least one. For each individual, we calculate how many others are better than it, and this ranking score is then used in the evolutionary process of the genetic algorithm. This approach allows us to select a solution that best balances our priorities - whether we aim for higher accuracy or a simpler sensor design.

3 Conclusion

In this work, we focused on the design and simulation of a phototransistor-based position sensor. The simulation was developed with the primary goal of analyzing the sensor's theoretical behavior under a wide range of conditions, including variations in light intensity, distance, and the angle of incidence. However, while the simulation results provide a solid foundation, to fully validate the accuracy and practical feasibility of the proposed concept, it will be necessary to build a functional prototype. This prototype will allow us to compare its performance with the simulation data and evaluate the consistency between theory and real-world behavior.

The simulation framework was also designed to support optimization using evolutionary algorithms, aiming to improve both the measurement accuracy and the efficient use of components. These algorithms enable the fine-tuning of various parameters, such as sensor alignment, component selection, and circuit configuration. This provides a systematic approach to testing different configurations and identifying the most suitable design.

The next crucial step in this research is to construct the sensor prototype and subject it to a series of experimental tests. These tests will serve to confirm whether the proposed concept can be practically implemented in a microsatellite environment.

References

1. LERNER, K. L. The Cavendish Experiment and the Quest to Determine a Gravitational Constant. *DRAFT COPY subsequently published in Science and Its Times: Understanding the Social Significance of Scientific Discovery. Thomson Gale*. 2001.
2. POYNTING, J. *The Mean Density of the Earth: An Essay to which the Adams Prize was Adjudged in 1893 in the University of Cambridge*. C. Griffin, limited, 1894.
3. TIESINGA, E.; MOHR, P.; D.B., N.; B.N., T. *CODATA Recommended Values of the Fundamental Physical Constants: 2022*. 2024. Available from DOI: 10.48550/arXiv.2409.03787.
4. BOTTIN, M.; GOUBEAUX, A.; VALKO, P.; HUDEC, R. Determination of Newton's gravitational constant(GN) via space based experiment (cubesat). 2021.
5. LANGE, B. The drag-free satellite. *AIAA journal*. 1964, **2**(9), 1590–1606.
6. KIS, D. *Design of a satellite module for gravitational constant measurement*. 2024. Masters Thesis. Bratislava: URK FEI.
7. PARILLA, J. *Návrh mechaniky pre meranie gravitačnej konštanty*. 2023. Masters Thesis. Bratislava: URK FEI.
8. JOHNSTONE, A. *CubeSat Design Specification* $1U-12U$, *Revision 14*. The CubeSat Program, Cal Poly SLO, 2020.
9. AMBRSOFT. *Intersection of Two Circles - Trigonometry Calculator*. 2025. Available also from: https://www.ambrbit.com/TrigoCalc/Circles3/Intersection.htm. Accessed: 2025-03-20.
10. FREDERICKSON, B. *Calculating the Intersection of 3 or More Circles*. 2023. Available also from: https://www.benfrederickson.com/calculating-the-intersection-of-3-or-more-circles/. Accessed: 2025-03-20.

Polohovanie dotykovej ultrazvukovej sondy priemyselným kolaboratívnym robotom

Juraj Kováč[1], Ľuboš Chovanec[2]

[1]Ústav vedy a výskumu, FEI STU v Bratislave
[2]Ústav robotiky a kybernetiky (FEI)

xkovacj1@stuba.sk

Abstrakt – V tejto práci sme sa, v spolupráci s firmou VÚEZ, zaoberali polohovaním citlivého nástroja, pomocou kolaboratívneho robota. Objavili sme však, že doterajšie komerčné riešenia v oblasti silovo-momentového riadenia sú pre naše účely nedostatočné. Preto sme sa rozhodli v tejto práci navrhnúť a otestovať naše alternatívne riešenie. Experimentálne sme porovnali naše riešenie s doterajším riešením od firmy Fanuc a ukázali sme, že naše riešenie má oproti nemu mnohé výhody.

1 Úvod

V tejto práci sa, v spolupráci s firmou VÚEZ, budeme zaoberať polohovaním citlivého nástroja, ako je v našom prípade dotyková ultrazvuková sonda, pomocou kolaboratívneho robota. Cieľom je zabezpečiť, aby robot preskúmal žiadanú plochu, sériou bezpečných dotykov nástrojom. To prináša pomerne veľké nároky na polohovanie, pretože musíme zabezpečiť dostatočný kontakt nástroja s povrchom, no zároveň robot nesmie pôsobiť príliš veľkou silou, ktorá by znížila životnosť nástroja, alebo ho priamo poškodila. Objavili sme však, že doterajšie komerčné riešenia v oblasti silovo-momentového riadenia sú pre naše účely nedostatočné a nespoľahlivé. Často prekračujú žiadanú silu aj napriek tomu, že vyžadujú pomerne dlhý čas na vykonanie dotyku. Preto sme sa rozhodli v tejto práci navrhnúť a otestovať naše alternatívne riešenie, o ktorom si myslíme, že prináša mnohé výhody oproti doterajšiemu riešeniu.

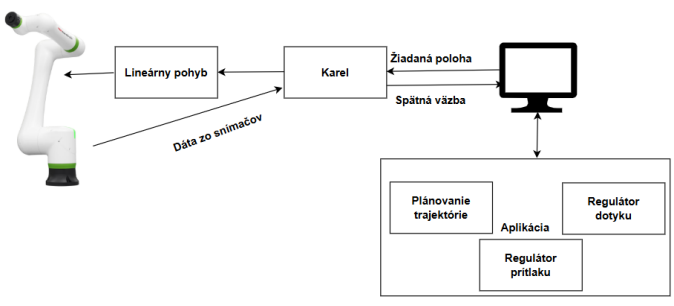
1.1 Požiadavky na polohovanie

Je potrebné, aby sme zabezpečili bezpečné a kvalitné polohovanie nástroja. Nástroj musí byť pri kontakte s plochou, ktorej sa dotýka, orientovaný kolmo na danú plochu. Taktiež je potrebné zabezpečiť dostatočný prítlak nástroja na danú plochu, aby sme zabezpečili kvalitný kontakt a teda prenos ultrazvukových vĺn. Na druhú stranu musíme zabezpečiť minimálne preregulovanie tohto prítlaku, pretože priveľký prítlak by mohol znížiť životnosť, alebo poškodiť samotnú ultrazvukovú sondu. Preto musíme zabezpečiť kvalitné a spoľahlivé polohovanie a reguláciu prítlaku. Ďalším cieľom je minimalizovanie času potrebného na vykonanie dotyku. Toto je dôležité pre použitia, kde je potrebné vykonať veľké množstvo dotykov, ako napríklad dotyková ultrazvuková diagnostika.
V tejto práci sa budeme konkrétne venovať vykonávaniu dotykov na rovine, no naše riešenie by sa dalo taktiež upraviť aj na iné situácie. Bolo by však potrebné samozrejme upraviť generátor trajektórie a odhad vzdialenosti nástroja od povrchu.

2 Doterajšie riešenie

Testovali sme doterajšie komerčné riešenie v podobe Fanuc Force controll funkcií, konkrétne funkcie force push. Táto funkcia pohybuje nástrojom v zadanom smere, konštantnou rýchlosťou, pokiaľ nezdeteguje prítlak žiadanou silou. Následne sa pokúsi udržať tento prítlak na nastavený čas. Táto funkcia na prvý pohľad spĺňa naše požiadavky, no pri testovaní sme objavili mnohé nedostatky. Hlavným z nich, je výrazné prekračovanie žiadaných síl prítlaku, najmä pri nižších žiadaných silách. Tento problém pravdepodobne vychádza so samotného princípu funkcie, pričom funkcia pohybuje nástrojom konštantnou rýchlosťou, pokiaľ nezdeteguje dostatočný prítlak. Robot však nedokáže zastaviť okamžite a teda žiadnu silu presiahne. Výška preregulovania sily je úmerná rýchlosti pohybu. Keďže je táto rýchlosť konštantná, sme nútený spraviť pomerne nepriaznivý kompromis medzi rýchlosťou a teda časom potrebným na dotyk a výškou preregulovania. Táto funkcia je taktiež pomerne uzavretá a nemáme nad ňou teda vonkajšiu kontrolu, čo je z vývojového hľadiska obmedzujúce.

3 Naše navrhnuté riešenie



Obr. 1: Jednoduchý diagram nášho riešenia

Naše navrhnuté riešenie spočíva z dvoch hlavných častí. Prvá časť je KAREL program, ktorý beží na samotnom počítači robota. Druhá časť je Python polohovacia aplikácia, bežiaca na externom počítači. Ovládacie počítače robotov majú často pomerne obmedzený výpočtový výkon a takto dokážeme presunúť väčšinu výpočtov na externý počítač. Tieto 2 programy sú navzájom prepojené pomocou TCP/IP spojenia. Karel program slúži na sprostredkovávanie žiadaných polôh z počítačovej aplikácie do robota a sprostredkovávanie dát zo senzorov robota do počítača. Po nadviazaní spojenia si KAREL program od aplikácie vyžiada polohu, ktorú má robot nástrojom dosiahnuť, pričom pošle svoju aktuálnu polohu a sily pôsobiace na nástroj. Python aplikácia následne vypočíta novú žiadanú polohu a pošle ju naspäť do KAREL programu. Ten potom vykoná pohyb na túto polohu. Po dokončení tohto pohybu sa celý proces opakuje, až do prejdenia celej trajektórie. Toto riešenie prináša niekoľko výhod oproti komerčnému riešeniu. Presúva časť kontroly nad robotom z výpočtovo slabšieho a uzavretého Fanuc počítača na výkonnejší externý počítač, nad ktorým máme väčšiu kontrolu, čo je užitočné pri vývoji nových robotických aplikácií. Zároveň nám umožňuje dynamicky meniť krok, o ktorý sa robot pohne. Robot sa tak môže hýbať rýchlejšie ak je to bezpečné a pomalšie ak je to potrebné. Bohužiaľ však nemáme schopnosť priamo regulovať túto rýchlosť, iba žiadanú polohu robota, pretože nám to uzavretá architektúra robota v tejto forme nedovolí. Musíme sa teda spoliehať na pohybové príkazy už implementované v robote. Aj napriek tomu veríme, že naše riešenie prináša mnoho zlepšení oproti doterajším riešeniam.

4 Python polohovacia aplikácia

Naša navrhnutá aplikácia sa skladá z troch hlavných častí, generátor trajektórie, regulátor dotyku a regulátor prítlaku.

4.1 Generátor trajektórie

Účelom generátora trajektórie je vytvoriť vhodnú trajektóriu, teda súbor polôh, na ktorých má robot vykonať dotyk nástrojom. (Pre účely tejto práce uvažujeme nad trajektóriou na rovine.) Tieto dotykové body by mali pokryť celú zadanú oblasť. Prehľadávaná oblasť je vopred zadaná troma bodmi. Robot následne prejde na tieto body a začne nástrojom posúvať vpred konštantným (dostatočne malým) krokom v Z-ovej osi nástroja, pokiaľ nezaznamená dotyk. Po detegovaní dotyku je zaznamenaná poloho dotyku a robot tento proces zopakuje na ďalšom bode. Po zaznamenaní pozície dotyku s povrchom na všetkých troch bodoch, sú tieto pozície predané do generátoru trajektórie.
Využijeme fakt, že 3 nekolineárne body v priestore definujú rovinu. Generátor trajektórie z týchto bodov p1,p2 a p3 najprv vytvorí 2 vektory.

$$\overrightarrow{v_1} = p_1 - p_2 \quad \overrightarrow{v_2} = p_1 - p_3 \tag{1}$$

Následne vypočítame spoločnú normálu vektorov pomocou vektorového súčinu

$$\overrightarrow{n} = (n_x, n_y, n_z) = \overrightarrow{v_1} \times \overrightarrow{v_2} \tag{2}$$

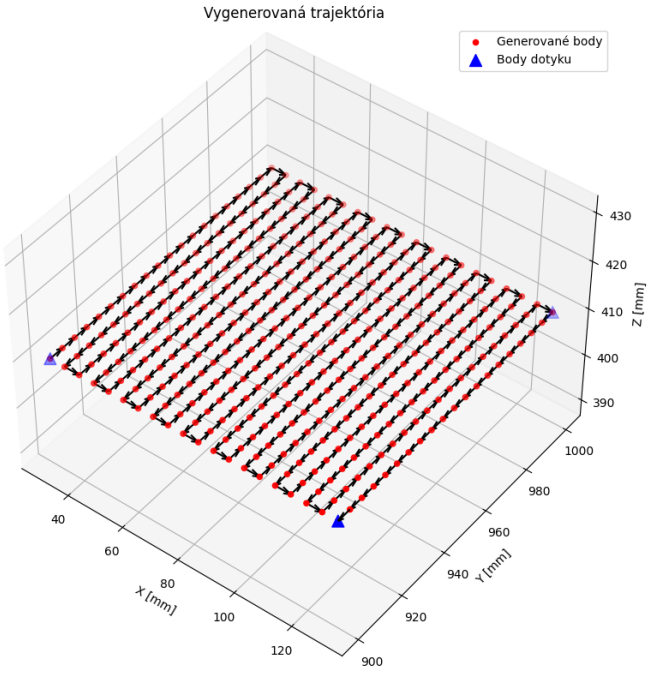
Potom pomocou normálového vektora a jedného z bodov ležiaceho na našej rovine, vypočítame rovnicu roviny,

$$Ax + By + Cz = D \tag{3}$$

pričom A,B,C sú prvky normálového vektoru a D vypočítame skalárnym súčinom normálového vektora a jedného z bodov, ležiaceho na rovine.

$$n_x = A \quad n_y = B \quad n_z = C \quad \overrightarrow{n} \cdot P_1 = D \tag{4}$$

Na základe tejto rovnice dokážeme určiť, predpokladanú výšku dotyku (Z súradnicu), pre akúkoľvek X,Y polohu. Generátor trajektórie následne vygeneruje sériu bodov na našej rovine, ktoré pokryjú celú plochu medzi našimi zadanými bodmi, s nastavenou maximálnou vzdialenosťou medzi vygenerovanými bodmi. Príklad vygenerovanej trajektórie môžeme vidieť na obrázku 2.
Keďže jednou z požiadaviek je, aby bol nástroj pri dotyku



Obr. 2: Príklad vygenerovanej trajektórie

kolmý na povrch, pre body trajektórie vypočítame taktiež aj také natočenie nástroja, aby bol zarovnaný s normálovým vektorom našej roviny. Vďaka tomu bude nástroj kolmý na povrch aj v prípadoch, kde je povrch naklonený. Natočenie robota je reprezentované vo forme w,p,r, teda eulerovych uhlov reprezentujúcich natočenie okolo osí x,y,z svetovej súradnicovej sústavy robota.

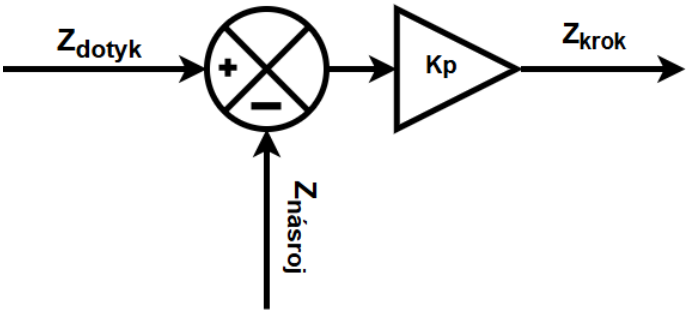
$$w = arctan2(n_y, n_z) \quad p = arctan2(-n_x, \sqrt{n_y^2 + n_z^2}) \tag{5}$$

Pričom r zatiaľ zostáva nezmenené.
Nakoniec je vykonaná translácia v osi Z súradnicovej sústavy nástroja o -10mm, aby boli body o 1cm posunuté od povrchu a zamedzili sme tak nežiadúcej kolízii s povrchom pri presune nástroja medzi bodmi trajektórie.

4.2 Regulátor dotyku

Po tom, čo sa nástroj dostane na jeden z dotykových bodov trajektórie, aplikácia aktivuje náš regulátor dotyku. Princíp tohto regulátoru je, že generuje pozície, ktoré sú posielané do robota a ten sa na tieto pozície pokúša dostať. Generované pozície sa postupne približujú k povrchu, ktorého sa pokúšame dotknúť, pričom dokážeme dynamicky meniť vzdialenosť, o ktorú sa žiadaná pozícia priblíži k povrchu. Vďaka tomu dokážeme nástrojom pohybovať rýchlejšie ak sme ďalej od povrchu a pomalšie ak sme blízko, čím minimalizujeme čas potrebný na dotyk aj prekročeniu žiadanej sily prítlaku. (Ako už bolo spomenuté regulácia rýchlosti je len nepriama pomocou zmeny kroku o ktorý sa nástroj posúva.)

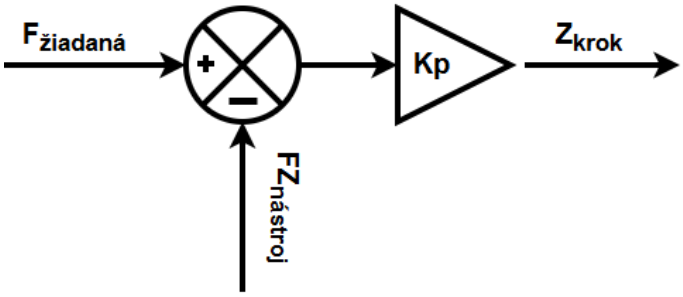
4.2.1 Štruktúra regulátora



Obr. 3: Jednoduchý diagram regulátora dotyku

Na tento účel sme zvolili proporcionálny regulátor, pre jeho jednoduchosť. Keďže samotný robot zabezpečuje polohovanie nástroja na žiadanú pozíciu, nemusíme uvažovať o externých poruchách, alebo „odpore“, ktorý by spôsoboval priveľkú trvalú regulačnú odchýlku a teda sme usúdili, že proporcionálna zložka je dostačujúca. Zároveň, ak by sme v tomto prípade použili Integračnú zložku, je veľmi pravdepodobné, že by spôsobila preregulovanie, pretože by sa akčný zásah z nej zvyšoval pri tom, ako by sa nástroj približoval k povrchu. Pre naše účely potrebujeme, aby sa akčný zásah zmenšoval spolu s spriblížením nástroja k povrchu. Derivačná zložka taktiež pre tento účel nemá veľký význam. Akčným zásahom nášho regulátora je vzdialenosť (krok) o ktorú sa priblíži žiadaná poloha k povrchu, ktorého sa pokúšame dotknúť. Regulačná odchýlka je vzdialenosť nástroja od tohto povrchu. Túto vzdialenosť v realite samozrejme presne nepoznáme a teda ju musíme odhadnúť. Na odhad používame rovnicu roviny, získanú z 3 kalibračných dotykov na vygenerovanie trajektórie. Pre každý bod v trajektórií je vypočítaná jeho výška. Táto výška nie je vždy presná, keďže v realite nie je žiadny povrch perfektná rovina a taktiež je závislá na presnosti nameranej polohy dotyku, ale pre jednoduchšie prípady je takýto odhad postačujúci. Pre zložitejšie prípady by bolo potrebné použiť komplxnejšiu aproximáciu očakávaných polôh dotyku. Akčný zásah je nesymetricky ohraničený saturáciou z oboch strán. Spodná saturácia je kladná, keďže ide o vzdialenosť, o ktorú sa približujeme k povrchu a ak by bola záporná, začali by sme sa vzďaľovať. Táto saturácia slúži na to, aby sa nástroj ďalej približoval v prípade, keď sme dosiahli odhadovanú výšku povrchu, ale stále nenastal dotyk. V tomto prípade sa bude nástroj približovať, minimálnou rýchlosťou, pokiaľ nenastane dotyk. Samozrejme je vhodné, po prejdení určitej vzdialenosti bez dotyku, tento pohyb z bezpečnostných dôvodov zastaviť. Horná saturácia nie je príliš dôležitá. Slúži len na zamedzenie príliš veľkého akčného zásahu a pre nižšie hodnoty zosilnenia regulátora (Kp <1), nemusí byť horná saturácia potrebná. V prípadoch kde je zosilnenie Kp>=1 by bez saturácie nastal problém, kde by sa robot nástrojom pokúsili dotknúť jediným pohybom, alebo by krok presiahol vzdialenosť nástroja od povrchu a teda by sme stratili kontrolu. Zosilnenia nad 1 pre naše účely ale aj tak nemajú príliš veľký význam.

4.3 Regulátor prítlaku



Obr. 4: Jednoduchý diagram regulátora prítlaku

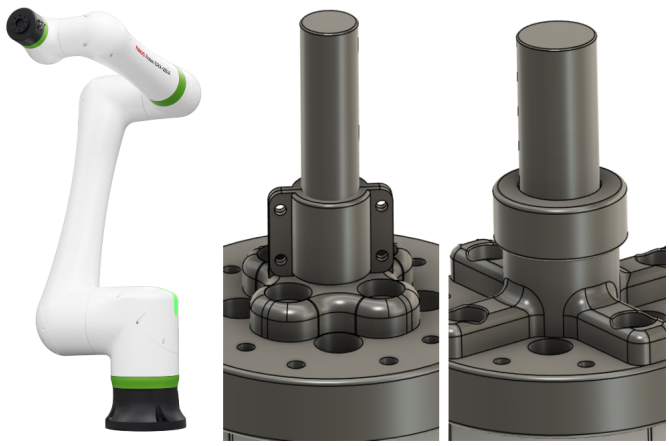
Po zaznamenaní dotyku sa aplikácia prepne z regulátoru dotyku na regulátor prítlaku. Tento regulátor slúži na udržanie nastaveného prítlaku, na nastavený čas. Taktiež, v prípadoch, kde regulátor dotyku prereguluje prítlak, regulátor prítlaku ho doreguluje naspäť na žiadanú hodnotu. Keďže pre naše účely potrebujeme len krátke dotyky, tento regulátor nie je tak dôležitý, ako regulátor dotyku. Na tento účel sme si taktiež zvolili proporcionálny regulátor. Malá integračná zložka obmedzená saturáciou by taktiež mohla pomôcť urýchliť reguláciu, avšak za ceny zvýšenia zložitosti. Derivačná zložka by pre účely regulácie prítlaku mohla byť prospešná, na zamedzenie preregulovania prítlaku, no taktiež prináša so sebou ďalšiu komplexnosť a P regulátor je pre naše účely postačujúci. Akčným zásahom je vzdialenosť o ktorú sa nástroj posunie vo svojej Z-ovej osi, rovnako ako pri regulátore dotyku. V tomto prípade môže byť táto vzdialenosť aj záporná a teda regulátor dokáže pohybovať nástrojom v oboch smeroch. Regulačná odchýlka je rozdiel žiadanej sily a nameranej sily pôsobiacej v Z-ovej osi nástroja.

5 Experimentálne overovanie riešení

Experimentálne sme overili kvalitu polohovania komerčného riešenia Fanuc force push a našej navrhnutej polohovacej aplikácie.

5.1 Použitý hardware

Experimenty sme vykonali na kolaboratívnom robote Fanuc crx-25ia, ktorý nám vypožičala firma VÚEZ. Tento robot je vybavený momentovými snímačmi v každom z jeho šiestich kĺbov. Robot si z týchto snímačov dokáže prepočítať sily a momenty pôsobiace na nastroj. Kvalitu polohovania sme overovali na dvoch 3d vytlačených testovacích nástrojoch. Prvý nástoj bol valec pevne uchytený na robot. Druhý nástroj sa skladal z piestu na pružine, voľne pohybujúceho sa v Z-ovej osi nástroja, teda osi v ktorej bude nastávať dotyk. Vďaka tomu bol tento nástroj pružný v osi dotyku. Cieľom je otestovať vplyv pružnosti nástroja na kvalitu polohovania.



Obr. 5: Fanuc crx-25ia, tvrdý nástroj a pružný nástroj

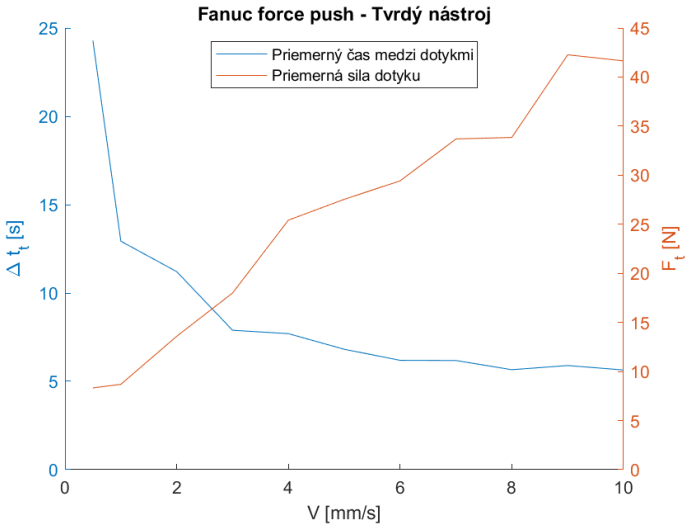
5.2 Experimenty

Vykonali sme 4 sety experimentov, v ktorých sme prešli trajektóriu zobrazenú na obrázku 2. Táto trajektória pokrýva oblasť 100 mm x 100 mm na pevnom stole, s 20mm vzdialenosťou medzi vygenerovanými bodmi dotyku, čo nám dáva 25 dotykov v každom experimente. Experimenty sme vykonali pre obe navrhnuté riešenia s pevným aj pružným nástrojom a pre rôzne nastavenia rýchlosti približovania pre Fanuc force push a pre rôzne zosilnenia nášho regulátora dotyku. Pri všetkých experimentoch bola nastavená žiadaná sila 5N a čas prítlaku 1s. Pre Fanuc force push sme testovali rýchlosti 0,5mm/s, 1mm/s až po 10mm/s s krokom 1mm/s. Pre našu polohovaciu aplikáciu sme testovali hodnoty zosilnenia regulátora dotyku od 0,1 po 1 s krokom 0,1. Zosilnenie regulátora prítlaku sme nastavili na $K_p = 5 * 10^{-5}$ pre pevný nástroj a $K_p = 5 * 10^{-2}$ pre pružný nástroj. Vyhodnocovali sme 2 hlavné metriky, priemerná perióda dotykov (reprezentujúca čas potrebný na vykonanie dotyku) a priemerná sila dotyku.

5.3 Experimenty s pevným nástrojom

Ako prvé sme vykonali experimenty s pevným nástrojom. Sila prítlaku pri dotyku s pevným nástrojom veľmi rýchlo stúpa, preto je regulovanie sily dotyku pre takýto nástroj pomerne náročné.

5.3.1 *Fanuc force push*

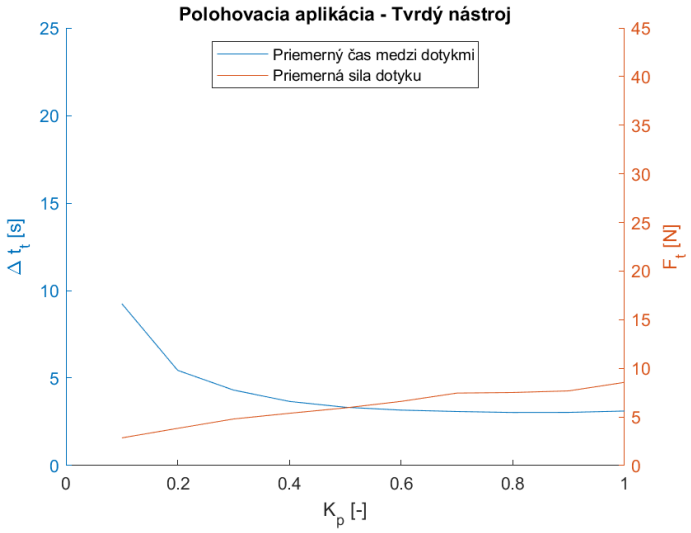


Obr. 6: Výsledky experimentov s pevným nástrojom s použitím funkcie Fanuc force push

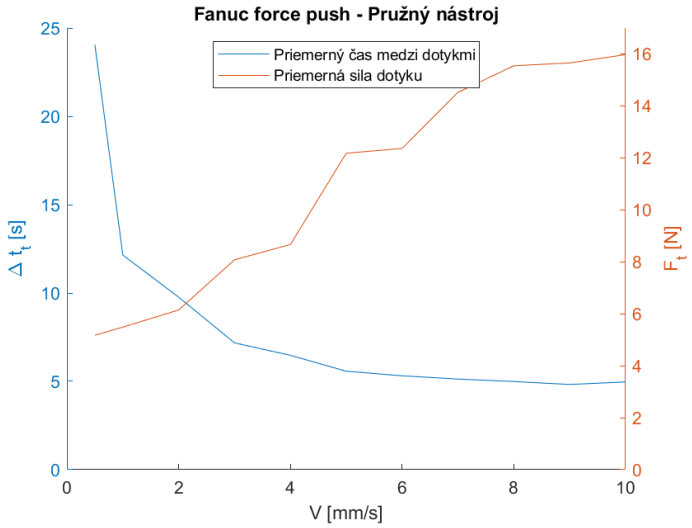
Na obrázku 6 môžeme vidieť závislosť priemernej periódy dotykov a priemernej sily dotykov na rýchlosti pohybu. So zvyšujúcou sa rýchlosťou sa samozrejme zmenšuje priemerná perióda dotykov, no zároveň výrazne stúpa sila dotykov. Tu môžeme pozorovať hlavný problém riešenia od Fanuc. Na to aby bola sila prítlaku v okolí žiadanej hodnoty, je potrebné použiť veľmi malé rýchlosti. To však významne predlžuje čas potrebný na vykonanie dotyku. To je problematické, keďže na dostatočné otestovanie objektu ultrazvukovou diagnostikou, môžu byť potrebné stovky dotykov, pričom sa toto spomalenie výrazne prejaví. Preto nás toto riešenie núti urobiť kompromis medzi rýchlosťou dotyku a preregulovaním sily, kde buď musíme nástrojom pohybovať príliš malou rýchlosťou, alebo tolerovať výrazné prekročenie žiadanej sily prítlaku.

5.3.2 *Polohovacia aplikácia*

V tomto experimente sme pozorovali, že so zvyšujúcim sa zosilnením regulátora dotyku taktiež klesá priemerná perióda dotykov a stúpa priemerná sila dotykov. Tento vplyv je však výrazne menší v porovnaní s riešením od Fanuc. Vďaka dynamickému meneniu vzdialenosti, o ktorú sa nástroj približuje k povrchu a teda nepriamo aj dynamickému meneniu rýchlosti, dokážeme znížiť periódu dotykov, bez toho, aby sme výrazne zvýšili preregulovanie prítlaku.



Obr. 7: Výsledky experimentov s pevným nástrojom s použitím našej polohovacej aplikácie



Obr. 8: Výsledky experimentov s pružným nástrojom s použitím funkcie Fanuc force push

5.3.3 Vyhodnotenie výsledkov

Pri experimentoch s pevným nástrojom sme pozorovali výrazné výhody nášho riešenia oproti riešenia od Fanuc. Pri použití našej polohovacej aplikácie sme pozorovali niekoľkonásobne menšie preregulovanie sily prítlaku a zároveň konzistentne výrazne menšiu priemernú periódu dotykov. Môžeme teda vidieť, že naša aplikácia sa dokáže vysporiadať výrazne lepšie s kontrolovanými dotykmi pevným nástrojom.

5.4 Experimenty s pružným nástrojom

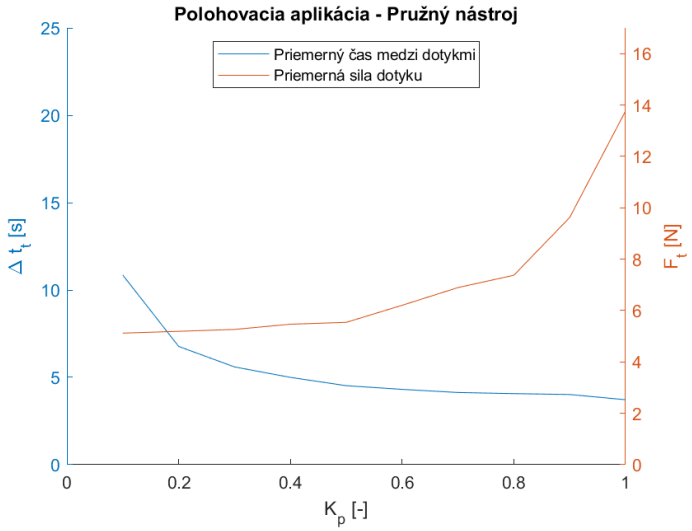
Vykonali sme rovnaké experimenty pružným nástrojom v osi dotyku. Našou hypotézou bolo, že pružnosť nástroja čiastočne zakryje nedostatky polohovania a zmierni náraz a teda zmenší požiadavky na polohovanie.

5.4.1 Fanuc force push

Na obrázku 8 môžeme hneď vidieť, že pružnosť nástroja výrazne zlepšila kvalitu polohovania z pohľadu preregulovania sily prítlaku. Priemerná perióda dotykov, zostala však prakticky rovnaká, keďže boli použité rovnaké rýchlosti. Preregulovanie stále pomerne lineárne, aj keď pomalšie ako pre pevný nástroj, stúpa so zvyšovaním rýchlosti. Zároveň tieto výsledky stále neprekonali našu polohovaciu aplikáciu s pevným nástrojom.

5.4.2 Polohovacia aplikácia

Experimentom sme pozorovali zlepšenie výsledkov oproti Fanuc force push. Prekvapivo sme však taktiež pozorovali zhoršenie oproti experimentu s pevným nástrojom. Toto neočakávané zhoršenie môže byť spôsobené rôznymi faktormi. Hlavným z nich je pravdepodobne nekonzistentný, alebo nepresný odhad vzdialenosti nástroja od povrchu. Pružnosť nástroja zmenšuje presnosť odhadu tejto vzdialenosti tým, že zmenší presnosť kalibračných meraní polohy dotykov a teda zhoršuje



Obr. 9: Výsledky experimentov s pružným nástrojom s použitím našej polohovacej aplikácie

kvalitu regulácie. Ďalšia potenciálna príčina je nekonzistentná pružnosť testovacieho nástroja, čo môže taktiež zhoršovať odhad vzdialenosti od povrchu. Veríme, že tento problém sa dá zmierniť lepším dizajnom, zmenšením tolerancií nástroja a použitím kvalitnejšej pružiny.

5.4.3 Vyhodnotenie výsledkov

Aj napriek problémom, mala naša polohovacia aplikácia konzistentne lepšie výsledky pri polohovaní pružného nástroja, oproti doterajšieho riešenia od Fanuc. Rovnako ako pri pevnom nástroji, mala naša aplikácia konzistentne menšiu priemernú periódu medzi dotykmi. Sila prítlaku bola taktiež konzistentne menšia pri našom riešení, no rozdiel medzi riešeniami nebol tak výrazný ako pri pevnom nástroji. To potvrdzuje našu hypotézu, pričom pružnosť nástroja čiastočne zakryla nedostatky riešenia

od Fanuc. Zároveň to ale odkrýva nedostatky nášho riešenia. Veríme však, že sa tieto nedostatky dajú odstrániť a tým ešte výrazne zlepšiť vlastnosti našej polohovacej aplikácie.

6 Výhody a nedostatky nášho riešenia

Na výsledkoch z experimentov sme pozorovali, že naše riešenie je konzistentne lepšie z pohľadu minimalizácie preregulovanie sily prítlaku a času potrebného na dotyk. Vďaka dynamickému regulovaniu vzdialenosti o ktorú sa približujeme k povrchu dokáže naša aplikácia pohybovať nástrojom rýchlejšie, keď je nástroj ďalej od povrchu a pomalšie keď sa k nemu priblíži, čím skrátime čas potrebný na dotyk a minimalizujeme prekročenie žiadanej sily pri dotyku. Preto veríme, že naše riešenie je vhodnejšie na použitia, kde je potrebné vykonávať dotyky s krehkým nástrojom, hlavne ak by tento nástoj nebol odpružený v osi dotyku. Na druhú stranu má naše riešenie aj niekoľko nedostatkov. Hlavným z nich je nutnosť poznať, respektíve čo najpresnejšie odhadnúť vzdialenosť nástroja od povrchu. Toto môže byť v niektorých prípadoch pomerne náročné a nepresný odhad môže mať výrazný vplyv na kvalitu polohovania. Tento problém má viacero možných riešení. Jedným z nich je zlepšenie kvality odhadu zväčšením počtu kalibračných dotykov. Taktiež môžu byť implementované korekcie v reálnom čase, pričom by si aplikácia neustále upresňovala odhady vzdialeností na základe predošlých dotykov. Ďalším riešením môže byť použitie prídavných senzorov, ako napríklad laserový diaľkomer, ktorým by sme mohli presne namerať vzdialenosť od povrchu a nemuseli sa spoliehať na odhad. Iným nedostatkom nášho riešenia je potreba použitia externého počítača, čo spôsobuje oneskorenie regulácie. Na druhú stranu vďaka tomu máme väčšiu kontrolu nad našim robotickým systémom. Taktiež nemáme priamy prístup k rýchlostnému ovládaniu robota a preto ho musíme ovládať iba polohovo, čo zhoršuje kvalitu regulácie. Aj napriek týmto limitáciám a nedostatkom veríme, že naša polohovacia aplikácia prináša zlepšenie oproti doterajšiemu komerčnému riešeniu.

7 Záver

V spolupráci s firmou VÚEZ sme navrhli riešenie polohovania citlivého nástoja, ako je napríklad ultrazvuková sonda, pre vykonávanie dotykov. Následne sme naše riešenie porovnali s doterajším komerčným riešením v podobe funkcie Fanuc force push. Porovnanie sme vykonali na pevnom nástroji a na nástroji, ktorý bol pružný v osi dotyku. Experimentálne sme ukázali, že naše riešenie je lepšie z pohľadu minimalizovania prekročenia žiadanej sily prítlaku a z pohľadu minimalizovania času potrebného na dotyk. Experimentami sme taktiež odhalili niektoré nedostatky našej polohovacej aplikácie a navrhli sme možné vylepšenia, po ktorých by sa mohli ešte výrazne zlepšiť jej výsledky.

Poďakovanie

Rád by som sa týmto poďakoval firme VÚEZ za spoluprácu a dodanie potrebného hardwaru, vedúcemu práce Ing. Ľubošovi Chovancovi, PhD. za celkovú podporu a taktiež Ing. Marekovi Trebuľovi za technickú podporu a pomoc pri práci v jazyku KAREL.

Literatúra

[1] Weisstein, Eric W. Plane. From MathWorld–A Wolfram Web Resource. URL: https://mathworld.wolfram.com/Plane.html.

[2] D. Eberly. Euler angle formulas. Geometric Tools, LLC, Technical Report, 1–18, September 2008.

[3] Gregory G Slabaugh. Computing Euler angles from a rotation matrix. URL: https://eecs.qmul.ac.uk/ gslabaugh/publications/euler.pdf,1999.

Variable Structure Control for a Motion System: Synthesis and Performance Assessment

Radovan Jakubčík[1], Marián Tárník[1]

[1]Ústav robotiky a kybernetiky, FEI STU v Bratislave

xjakubcikr@stuba.sk

Abstract – This paper explores the synthesis and performance assessment of Variable Structure Control (VSC) applied to a second-order motion system without inherent astatism and breaking potential. The control strategy is formulated using symmetric control law. The proposed approach is evaluated in MATLAB/Simulink environment and validated through experiments on a real-world system.

1 Introduction

The synthesis of the VSC for a motion system depends on the inherent astatic nature of the *controlled system (CS)*, or plant, to remove the permanent control error in a finite time. *Sliding Mode Control (SMC)* will be used as the *control law (CL)* for the CS. The measured plant is a *nonlinear-time-variant (NTV)* system; thus, we are required to identify a linearized model before attempting CL synthesis.

The proposed CL will attempt to solve the nonlinear and time-variant nature of the plant. By forcing specific dynamics on the plant. To implement the SMC, we will need to identify the plant as a second or higher order system; in the case of this article, a *second order (2nd order)* system. The loss of higher order dynamics will be compensated for by the CL. For the missing inherent astatic nature of the plant, a modified CL will be proposed. A separate CL will be implemented to account for this missing characteristic; in order to zero out the permanent control deviation.

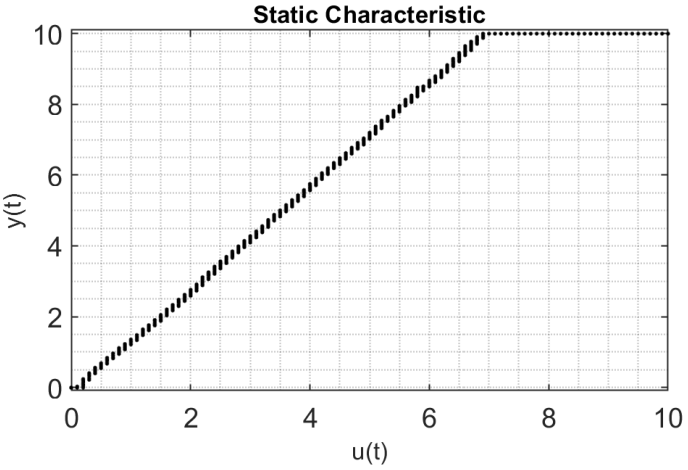
Pulsing signal and a double monotonically increasing step signal will be used for performance assessment. This will simulate the change of the dynamics, and an external signal simulates the time-variance. The change in dynamics of the plant will be simulated by the change in the operating point (OP) in time. This will render the identified model obsolete because of the nonlinear nature of the system.

2 Identification

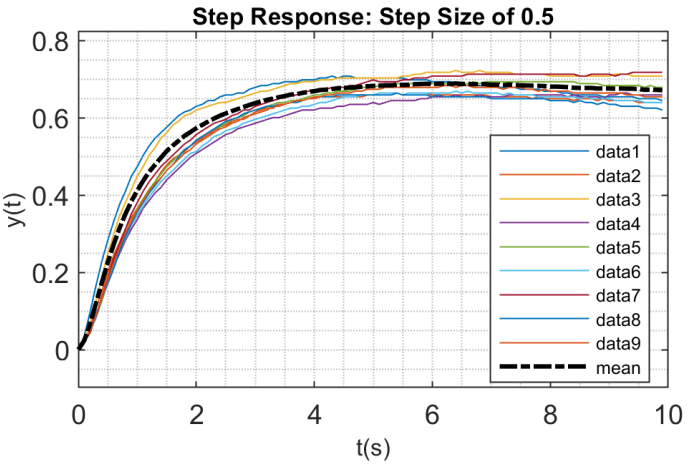
The identification process approximates a 2nd order system without astatism with varying time constants in the *monic transfer function form (transfer function)*, or in the Frobenius form as can be seen in (1a), and (1b).[1]

$$G(s) = \frac{b_0}{s^2 + a_1 s + a_0} \tag{1a}$$

$$\dot{\mathbf{x}}(t) = \begin{bmatrix} 0 & 1 \\ -a_0 & -a_1 \end{bmatrix} \mathbf{x}(t) + \begin{bmatrix} 0 \\ b_0 \end{bmatrix} u(t) \tag{1b}$$



(a) Static Characteristic of the experimental motion system, measured at 0.1 step increments of the system input signal.



(b) Step response of the experimental motion system.

Figure 1: Measured system characteristics

Approximated *linear-time-invariant (LTI)* models were identified using measurements in various OPs and further used

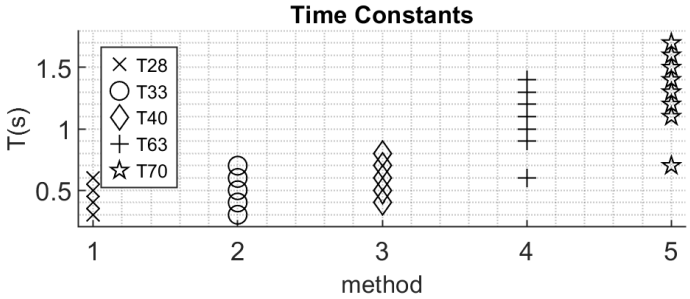


Figure 2: T interval of the identified LTI models using different identification methods.

to set the *time constant interval (T interval)*. The identified T intervals are necessary in the synthesis process of the SMC for its robustness. The T interval can be seen in fig. 2. Various identification methods have been applied to the measured data in order to choose the most suitable model.[2, 3]

This process requires measuring the static characteristic which can be seen in fig. 1a of the plant to determine the nonlinearities and linear sections. To identify an LTI model, only the linear sections are to be used to estimate multiple models at different OPs. Otherwise, the OP linear neighborhood shrinks based on the selected OP, at which the model was linearized. The step response of this particular system can be seen on fig. 1b for multiple different OPs.

3 Methods

SMC method heavily depends on being able to slow down the system using "braking"; direct reference to negative control output. In the case of this motion system, deceleration magnitude is immutable. Therefore, the SMC uses the inherent fall time characteristic (step response in decreasing direction) of the motion system in order to cause deceleration. The step response of this motion system differs depending on the direction of the step. For a better description of the system's dynamics, a need to implement an identified model selection method depending on the current situation arises.

3.1 SMC

SMC structure shown in the fig. 3 which uses control error defined in (2). This *control algorithm (CA)* uses actual error to "generate" actual control output to control the plant. The problem with this method is the measurement of the actual output. It contains disturbances that are unaccounted for in our identified model. This results in scaled noise and incorrect derivative of measured data. Furthermore, leaving a negative impact on the control output of the control system. Possibly leading to unexpected behavior of the controlled system.

$$e(t) = w(t) - y(t) \tag{2}$$

Where $e(t)$ is the control error, $w(t)$ is the setpoint (reference signal), and $y(t)$ is the measured output of the system.

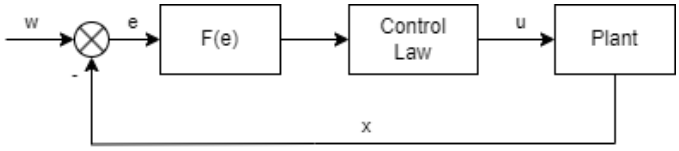


Figure 3: Diagram visualizing the SMC in block scheme.

The goal of this control method is to cancel the control error in a way that can be predicted in advance. The benefit of using the SMC is such that the control process is known at all possible initial conditions of the system. The main idea of this control method is for the system to slide along a sliding curve to a steady state. In order to reach the sliding surface, the system's trajectory has to be modified by the actions of the control law in order for the system to reach and start sliding along the sliding curve. Thus, effectively following the prescribed dynamics for the control process. This control method is forcing the plant to trace a certain trajectory to a steady state of the system. To introduce the *state space (SS)* structure of the SMC we can define it as follows (3), when taking into consideration the structure of our modeled system.

$$\dot{e}(t) = \dot{y} \tag{3a}$$
$$\ddot{e}(t) = -\left(a_1\dot{e}(t) + a_0 e(t) + b_0 u(t)\right) \tag{3b}$$

Where we can define the parameters $a_1, a_0,$ and b_0 in the following form, considering the model having a unit gain.

$$a_1 = \frac{T_1 + T_2}{T_1 T_2} \tag{4a}$$
$$a_0 = \frac{1}{T_1 T_2} \tag{4b}$$
$$b_0 = \frac{K}{T_1 T_2} \tag{4c}$$

As can be seen in (4) the T interval values are still present in the SS representation. This is from where we will derive the slope of the sliding curve using the identified T interval and the control law (10) proposed for this SMC.

3.1.1 Switching Function

Before defining the CL, first we define the *switching function (SF)* which will force the system to always approach the sliding curve. The implemented SF is the following relation.[4]

$$F(e(t)) = \dot{e}(t) + \alpha e(t) \tag{5}$$

If we let the (5) equal 0, we get the following equation.

$$\dot{e}(t) + \alpha e(t) = 0 \Rightarrow \dot{e}(t) = -\alpha e(t) \implies \dot{e}(t)e(t) < 0 \tag{6a}$$
$$\frac{\dot{e}}{e} = -\alpha \implies \frac{d\dot{e}}{de} = -\alpha \tag{6b}$$

The implication in (6a) defines the inverse relation between the change in control error and actual control error. When

the actual control error is greater than zero, thus implying an overshoot, the needed change in control error will be negative, thus forcing the error to zero by decreasing the control error and vice-versa. Furthermore, the equation (6b) gives us the necessary relation for the derivation of the optimal α parameter, thus the slope of the sliding surface. From this relation we can further derive; using a crude assumption, the functional relation of the isocline of the system. This provides all the points at which the trajectory of the system has the slope equal to α, thus it can smoothly transition from its trajectory to the sliding curve, without overshooting the sliding curve, or overshooting the setpoint and possibly causing an unwanted accident.[5]

If we set (6b) equal to (3) we get the functional relation between the slope of the sliding curve and the controlled system as follows.

$$\alpha = \frac{a_1\dot{e}(t) + a_0 e(t) + b_0 u(t)}{\dot{e}(t)} \tag{7}$$

To further progress with solving the (7) equation, we would assume the worst-case scenario. This would mean the assumption of the longest T interval, largest control output, change in control error, and actual control error. Which can be solved, but for our motion system without the possibility of controlled deceleration we will use the slowest SMC. Which implies the use of such α value such that the sliding curve is always reached. This means the use of the following (8) functional relationship for determining the value of the sliding curve slope.

$$\alpha = \frac{1}{T_{max}} \tag{8}$$

3.1.2 Control Law

To define the control law, we have to know the limitations of the plant. For example, in the case of our motion system, the control input is a PWM signal. This allows the use of relay control law, but this would lead to an artifact known as chattering in the near neighborhood of the sliding curve while the system is tracking its trajectory along the sliding curve. Such a control law has the following form.

$$u(t) = M\text{sign}(F(e)) \tag{9}$$

This control law is similar to a Bang-Bang control method. This type of control process is not favored when controlling a motion system, because of the rapid switching frequency. This brings us to the topic of shaping the control output using a different nonlinear control technique. We can change the relay characteristic for a different function. We propose a control law with an equivalent control output magnitude to the relay controller defined by the equation (9), but with a smooth continuous transition between control outputs. Therefore, an attempt to remove the Bang-Bang nature of the SMC with a simple relay control law, the proposed control law can be seen in equation (10).

$$u(t) = G\left(M\tanh\frac{F(e)}{\phi} + \gamma u_{-1}(t)\right) = G(\Psi) \tag{10a}$$

$$G(\Psi) = M\begin{cases}\text{sign}(\Psi) & |\Psi| > M\\ \Psi & |e| > \varphi\\ \beta\Psi & |e| \le \varphi\end{cases} \tag{10b}$$

The proposed control law uses direct shaping of the switching function using the hyperbolic tangent with a gain to reduce the slope of the hyperbolic tangent function at its origin. It uses the previous control output to help with the continuity in the control output. The last layer of the control law is constructed using a dynamic saturation with a boundary interval around the sliding curve. Upon reaching the boundary, the control output is scaled down to reduce the chattering in the vicinity of the sliding curve.

3.2 Observed SMC

Observer SMC refers to an indirect SMC method of controlling the plant, meaning that the real-world system is being observed using an observer. The output from the observer; let us call it observed output, is used as control error and further down the control circuit the control output is computed using the observed output. The identified models from sec. 2 can be further modified and used as reference models for controlling the real-world plant. Implementing this control method would require the use of either lookup tables, gain scheduling, fuzzy logic, neural network, or a different selection method of the proper reference model depending on the OP. In which case, the reference model would need to have faster dynamics in comparison with the real-world system.[6]

This method helps reduce the noise featured in the data. It helps with the use of the derived output signal in the control law, with less noise to be scaled, thus helping in the reduction of the computed control output. This approach comes at the cost of setting up an observer. In order to implement this one-step prediction process, the model has to accurately approximate the system. Otherwise, the observed output is different from the actual output, thus the whole control process is flawed and cannot be used without any further modifications. For example, an online identification, or an adaption algorithm can be implemented to help the observer better predict the systems behavior. Otherwise, the observer becomes obsolete as it is observing a reality which does not exist.

To implement an observer into the scheme mentioned in fig. 3, we introduce a new block observer, with the inputs for control output and actual output, or other necessary signals for the observer of choice. We will provide a simple guide for implementing an observer and a measurement of the control process when using an observer.

In this case, the controlled system changes from the real-world system to the observer, through which we indirectly control the real-world system. The SS model for a classical state observer is in the following form (11)

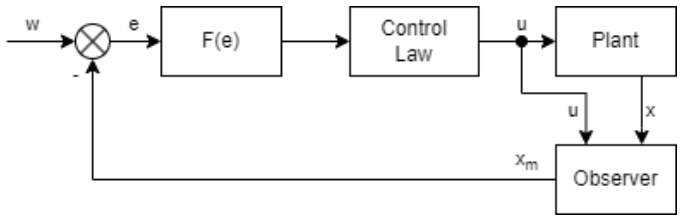


Figure 4: Diagram visualizing the SMC with an observer in block scheme.

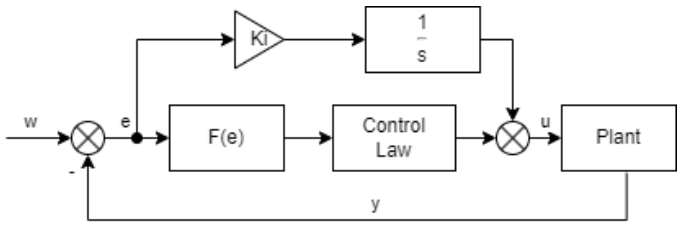


Figure 5: Diagram visualizing the SMC with the added integrator in block scheme.

$$\dot{x}_m = \mathbf{A}_0\mathbf{x}_m + \mathbf{b}_0 u + \mathbf{k}e_y \tag{11a}$$
$$y_m = \mathbf{c}^T\mathbf{x}_m \tag{11b}$$
$$e_y = y - y_m \tag{11c}$$

Where e_y is the difference between the actual output and the observer output, it can be defined using the equation (11c). The parameters $\mathbf{A}_0$, $\mathbf{b}_0$ are the parameters of the identified model, when expressed in an SS model. The unknown $\mathbf{k}$ vector is to be picked; for example, using the *pole placement method*, such that the final observer model is faster than the system being controlled. The following characteristic polynomial (12) needs to be solved to satisfy the condition.[7]

$$s\mathbf{I}_n - \left(\mathbf{A}_0 - \mathbf{k}\mathbf{c}^T\right) = 0 \tag{12}$$

4 Implementation

The implementation of the previously mentioned control structure is straightforward (see fig. 3), thus it will not be derived. The only part of the controller that has yet to be revealed is the separate integrator control law. This simple control law is an active part of the control method can be seen on fig. 5. It is used to remove the permanent control deviation, whereas the SMC is a supporting controller when the reference signal changes in a step-like manner. We implemented a simple integrator with a gain where the input is the control error scaled by the integrator gain. The output is added together with the output of the SMC method, computing the control output. The computed output is then saturated to fit the needs of the system.

The parameters of the control law are as follows in the equations (5) and (10). In the first place $\alpha = 2.5165$, which at the same time is the integrator gain in fig. 5 presented as the gain *Ki*. Moving on to the parameters: $M = 10$, $\gamma = -0.2$, $\varphi = 0.01$, $\beta = 0.5$, and $\phi = 1$.

4.0.1 Observer

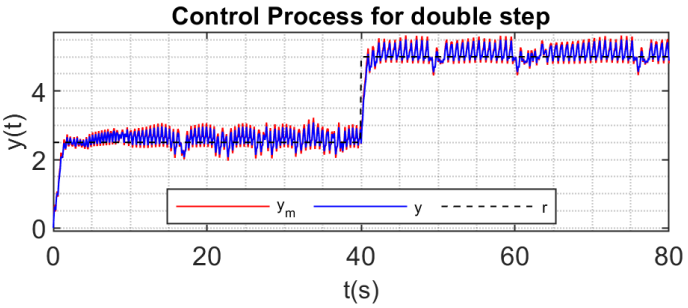
The implemented observer is in the SS model, with the following parameters:

$$\mathbf{A}_0 = \begin{bmatrix} 0 & 1 \\ -12.4183 & -12.2581 \end{bmatrix};\ \mathbf{b}_0 = \begin{bmatrix} 0 \\ 16.9356 \end{bmatrix};\ \mathbf{k} = \begin{bmatrix} 1.5804 \\ 0.8171 \end{bmatrix}.$$

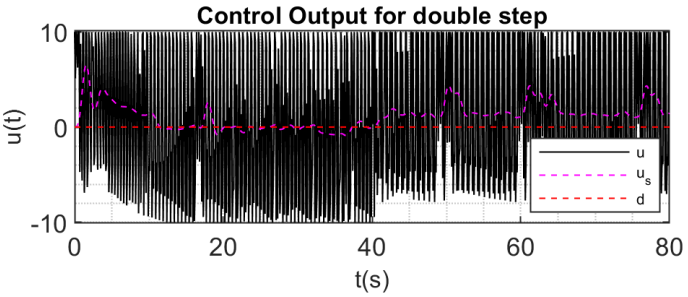
The $\mathbf{c}^\mathbf{T}$ vector remains in the standard form of $\begin{bmatrix} 1 & 0 \end{bmatrix}$.

5 Verification

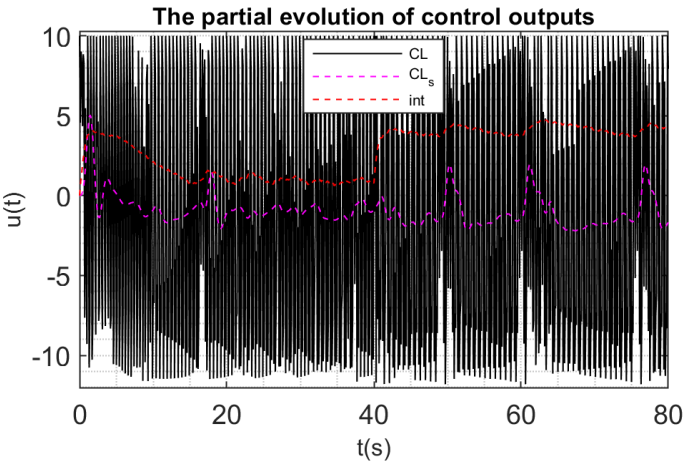
This section will mainly contain the numerical simulations and the measured experiments with *sampling period* (T_s) of 100ms.



(a) Actual output with the observed output for varying external signal.

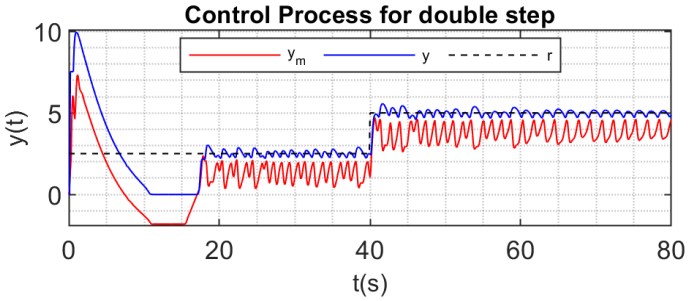


(b) Computed control input of the system for varying external signal.

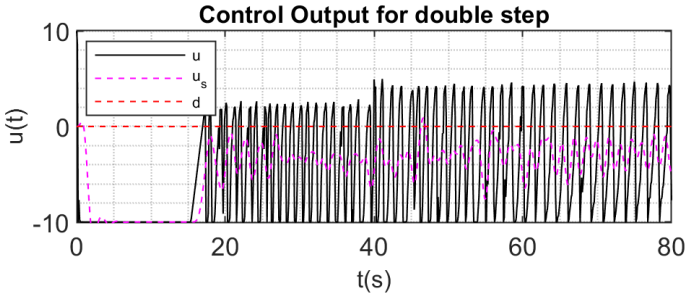


(c) The partial contributions of the control law elements.

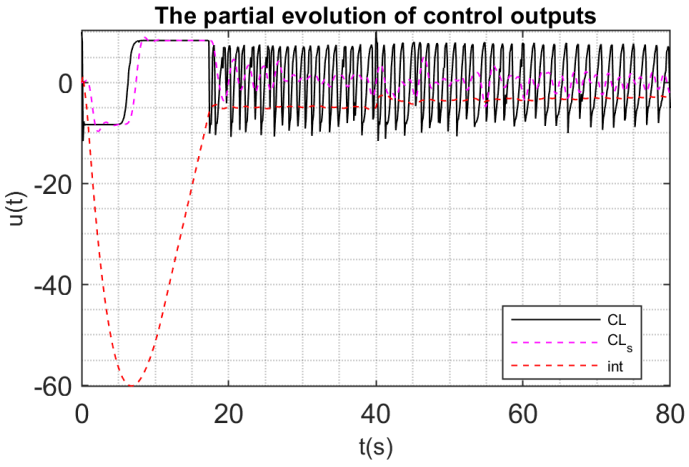
Figure 6: Simulated models actual output and computed control input with fastest dynamics.



(a) Actual output with the observed output.



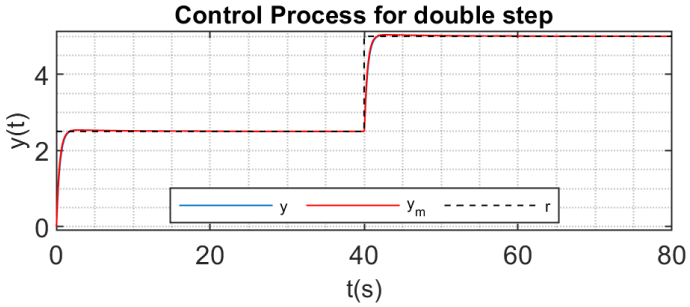
(b) Computed control input of the system.



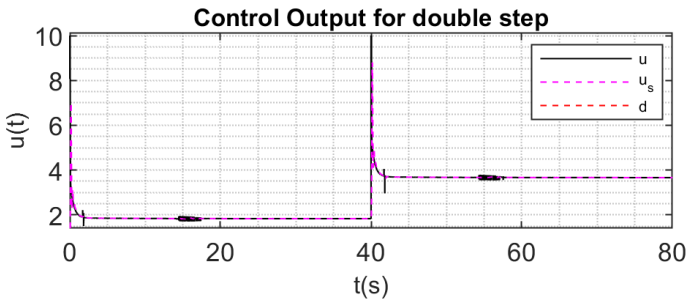
(c) The partial contributions of the control law elements.

Figure 7: Measured systems actual output and computed control input with quickest dynamics.

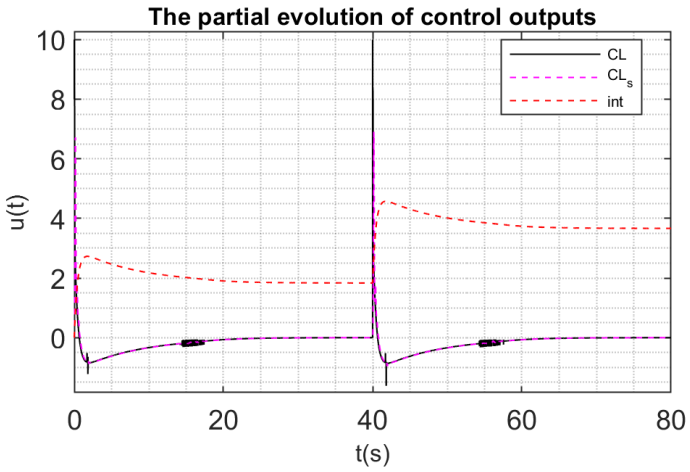
The fig. 8 uses the identified model of the system to simulate the control process at T_s set to 10ms to show the effect of T_s on the control process. As we can see, the control output tends to be similar in the case of the fig. 7 and fig. 6. The control process shown on fig. 8a shows the ideal control process for the proposed control structure. From the data, we can deduce that the real-world plant is slower than the modeled system. This is caused by the identification process and the measuring process of the step response, which has been conducted using the step size of 0.5, while in the case of the experiment, we increased the step size by a factor of 5. We introduced a Butterworth filter to better visualize similarities in the computed control outputs, denoted by u_s. In the figures 7c, 6c, and 8c we can see the roles of each controller. The integral controller is used to get to the OP, while the SMC handles quick transfers in the reference signal. Finally, the ideal control process for the proposed SMC is on fig. 8d.



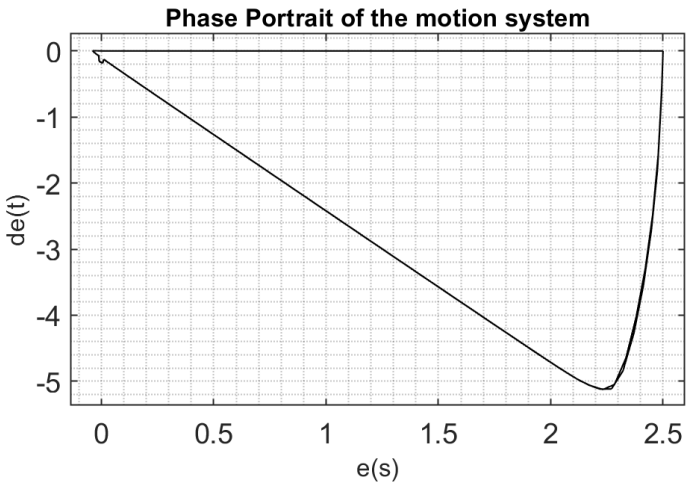
(a) Actual output with the observed output.



(b) Computed control input of the system.

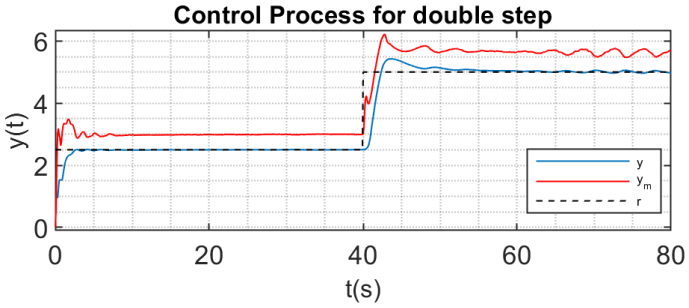


(c) The partial contributions of the control law elements.

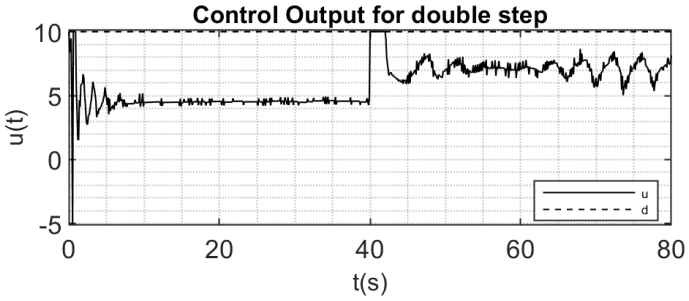


(d) Phase portrait for the identified model, showing the chosen sliding curve.

Figure 8: Simulated systems actual output and computed control input with quickest dynamics with T_s of 10ms.

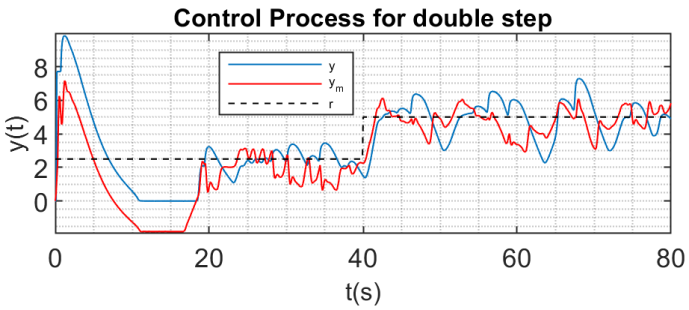


(a) Actual output with the observed output for external signal set to 10.

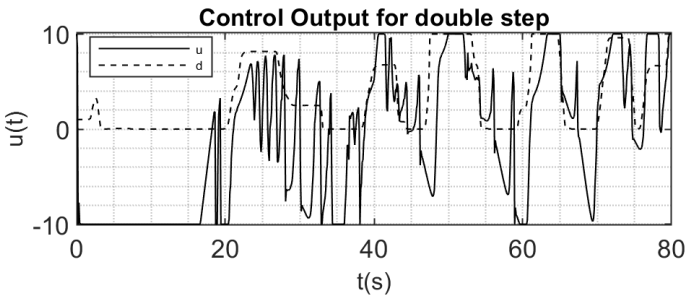


(b) Computed control input of the system for external signal set to 10.

Figure 9: Measured systems actual output and computed control input with the slowest dynamics.



(a) Actual output with the observed output for varying external signal.



(b) Computed control input of the system for varying external signal.

Figure 10: Measured systems actual output and computed control input with varying dynamics.

After modifying the system dynamics to be slower, measurements are presented in fig. 9 and fig. 10. These results can be compared to those in fig. 8, showing improved agreement. All the provided graphs include the signal y_m, which represents the observer's output. However, due to the continuously varying system dynamics influenced by external disturbances, the observer output deviates from the real-world plant behavior. In contrast, numerical simulations show that the observer model closely predicts the system's future state with high accuracy. In fig. 10, a specialized control process is observed, where the system dynamics are modified randomly. Initially, the system deviates from the reference signal because the integrated control action is smaller than the control output's saturation limit. As a result, the control process temporarily halts, allowing the integral component to gradually reach the OP. This waiting phase is illustrated in fig. 7c, which displays the individual elements of the complete controller. The observed overshoot is primarily due to the system's rapid dynamics and its inability to decelerate effectively. Once the integral controller reaches the OP, the system stabilizes, successfully handling random disturbances while tracking the reference signal and approaching steady-state operation.

6 Conclusion

In this article, we have successfully implemented the SMC method with the proposed control law and further modified it to compensate for the system's lack of inherent astatism. Numerical simulations and real-world experiments have demonstrated the feasibility of this approach. Through our analysis, we identified key challenges affecting the SMC method, including its failure to reach a steady state when tracking the reference signal, the inability to brake effectively, and a sampling period that is slower than the system's fastest dynamics. Additionally, we highlighted the limitations of a simple state observer and proposed a solution to address them. The performance of the proposed control structure was evaluated through both numerical simulations and experimental validation, showing a strong correlation between simulation results and real-world behavior, even without complete knowledge of the system's dynamics.

References

[1] Shlomo Engelberg. *Mathematical Introduction To Control Theory, A*. Vol. 8. World Scientific, 2024. ISBN: 9781800615540.

[2] Jer-Nan Juang. *Applied system identification*. Prentice-Hall, Inc., 1994.

[3] Peter Hudzovič. *Identifikácia a modelovanie*. Slovenská vysoká škola technická v Bratislave, 1986.

[4] Ján Kardoš. "Pohybové systémy s premenlivou štruktúrou". In: *Automatizace* 52.12 (2009), pp. 688–692.

[5] JP Barbot, M Djemai, and T Boukhobza. "Sliding mode observers". In: *Sliding mode control in engineering* 11 (2002), p. 33.

[6] Yuri Shtessel et al. *Sliding mode control and observation*. Vol. 10. Springer, 2014.

[7] Abdelkrim Benchaib et al. "Real-time sliding-mode observer and control of an induction motor". In: *IEEE Transactions on Industrial Electronics* 46.1 (1999), pp. 128–138.

Model Reference Adaptive Control: Design and Implementation for a Laboratory Dynamic System

Adrián Jakubčík[1], Ing. Marián Tárník, PhD.[1]

[1]Institute of Robotics and Cybernetics, FEI STU, Bratislava

xjakubcik@stuba.sk

Abstract – This paper explores the challenges associated with designing and implementing advanced control methods for dynamic systems with time-varying dynamics. Specifically, it focuses on the development and implementation of a Model Reference Adaptive Control (MRAC) system. The study aims to highlight the advantages of adaptive control in scenarios where traditional control method's performance may become insufficient, demonstrating its necessity in specific applications.

1 Introduction

Designing and implementing a control method for a dynamic system presents significant challenges, particularly when the system is time-varying—that is, its dynamics change over time.

In such cases, the performance of conventional control methods, such as a *PID controller*, may be insufficient due to their inherent time-invariance. To address this limitation, advanced control methods become necessary. In our work, we employ adaptive control, specifically *model reference adaptive control (MRAC)*, to better accommodate the system's varying dynamics.

Alternatively, robust control methods are often used to design time-invariant controllers that perform satisfactorily under multiple operating conditions. Additionally, gain scheduling—sometimes considered a form of adaptive control—can be implemented to achieve desired control objectives by adjusting controller parameters in response to changes in the operating environment.

2 System Identification

Before designing the control method, system identification was necessary to approximate the system's dynamics. Given that the system is initially unknown, a black-box approach was chosen to estimate its behavior, more about system identification can be found in [1].

To model the identified system, *linear time-invariant (LTI)* transfer functions were used to simulate the dynamics at a selected *operating point*.

As shown in Fig. 1, the system exhibits significant noise and variability. Repeated step-response measurements (10 trials from the same steady state) yielded different outputs, indicating

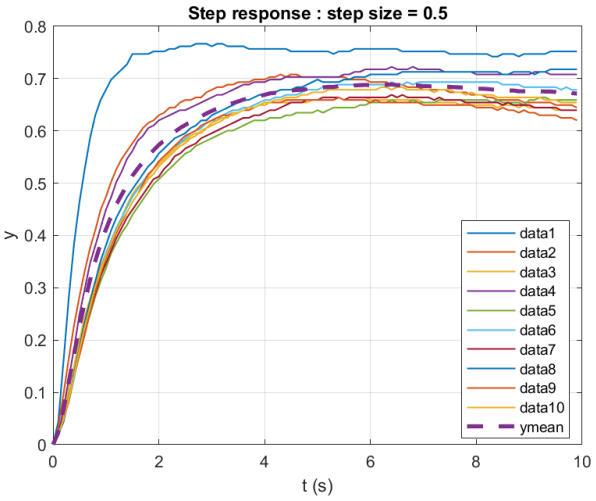


Figure 1: Measured data step-responses of the system

a high degree of measurement dispersion. Aggregating the data produced a data which where used for identification and model comparison.

2.1 First-order LTI model

As shown in Fig. 2, the first-order LTI transfer function (1b), obtained from system identification using aggregated data, provides an approximate representation of the system with a relatively small *sum of squares error (SSE)*.

$$\dot{y}(t) = -a_0 y(t) + b_0 u(t) \tag{1a}$$

$$G(s) = \frac{b_0}{s + a_0} \tag{1b}$$

2.2 Second-order LTI model

The results in Fig. 2 suggest that the first-order model does not fully capture the system dynamics. A second-order LTI transfer function (2b), derived from the same aggregated data, provides a more accurate representation of the system, as indicated by its

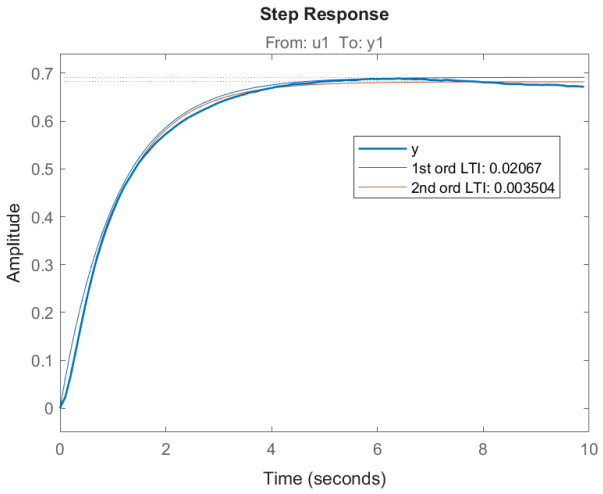


Figure 2: Measured step-response data and simulated models' step response

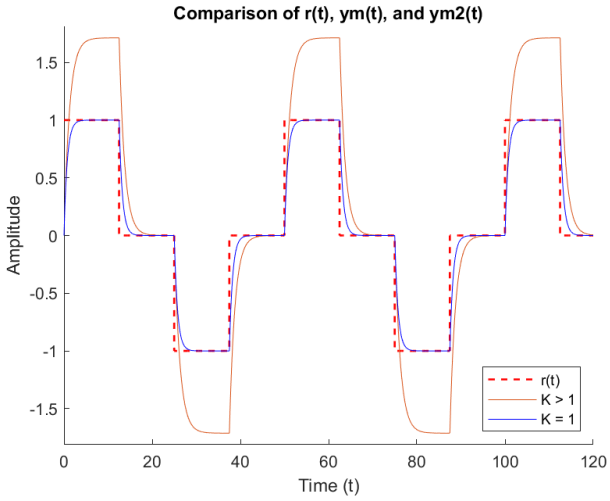


Figure 3: *r(t)* signal amplification in the reference model for different values of *K*.

lower *SSE*. w

$$\ddot{y}(t) = -a_1\dot{y}(t) - a_0 y(t) + b_0 u(t) \tag{2a}$$

$$G(s) = \frac{b_0}{s^2 + a_1 s + a_0} \tag{2b}$$

3 Control Algorithm Design

Designing *Model Reference Adaptive Control* (MRAC) requires appropriately choosing a *reference model* that defines the desired system dynamics for the algorithm to track and adapt to, see [2].

Beyond selecting a suitable reference model, a *control law* must be designed to generate the *controller output*. The reference model and control law must satisfy the *matching condition* to ensure proper adaptation and stability.

Lastly, an appropriate *adaptation law* is required to dynamically adjust the parameter vector Θ during runtime. Under ideal conditions and with reasonable adaptation law, the Θ parameters in the control law should converge to the optimal Θ^* parameters given by the *matching condition*, ensuring that the system output closely follows the reference model.

3.1 Reference Model

When selecting an appropriate *reference model* that defines the desired dynamics for the system. It is critical that the static gain is unity (i.e., *K=1*), so that the reference signal *r(t)* is not inadvertently amplified. If *K* were different from 1, the steady state value of the model output would be amplified, as illustrated in Fig. 3 and shown in (3c).

$$\frac{y_m(s)}{r(s)} = \frac{b_m}{s + a_m} \tag{3a}$$

$$y_m(\infty) = \frac{b_m}{a_m} r(\infty) \tag{3b}$$

$$K = \frac{b_m}{a_m} \tag{3c}$$

Apart from setting *K=1*, the *matching condition* must be met to ensure that the reference model is compatible with the control law. The following reference models for the identified systems ensure both of these conditions are satisfied.

3.1.1 First-order Reference Model

For the first-order system, the reference model is defined as:

$$\frac{y_m(s)}{r(s)} = \frac{b_m}{s + a_m}, \quad \text{where } b_m = a_m \tag{4}$$

3.1.2 Second-order Reference Model

For the second-order system, the reference model is given by:

$$\frac{y_m(s)}{r(s)} = \frac{b_{0m}}{s^2 + a_{1m} s + a_{0m}}, \quad \text{where } b_{0m} = a_{0m} \tag{5}$$

3.2 Control Law

The design of the control law in MRAC is constrained by the set of measurable signals. After substituting the control law, the *matching condition* must be satisfied, as defined in [3], [4], and [5]. This condition ensures the existence of an ideal parameter vector Θ^* such that the tracking error, defined as

$$e(t) = y(t) - y_m(t) \tag{6}$$

converges to zero as

$$\lim_{t\to\infty} e(t) = 0 \tag{7}$$

Achieving this convergence ensures that the output of the controlled system, $y(t)$, asymptotically replicates the output of the reference model, $y_m(t)$, thus yielding the desired closed-loop behavior.

Assuming that the matching condition is satisfied, the general form of the control law [5][4] is given by:

$$u(t) = \Theta^T(t)\omega(t), \tag{8}$$

where the regressor vector $\omega(t)$ contains the measurable signals (e.g., $y(t)$, its derivatives, $r(t)$, and or their combinations).

3.2.1 First-order Control Law

For a first-order system, the regressor vector can be chosen as:

$$\omega(t) = \begin{bmatrix} y(t) \\ r(t) \end{bmatrix} \tag{9}$$

leading to the control law:

$$u(t) = \Theta_1 y(t) + \Theta_2 r(t) \tag{10}$$

In this formulation, the ideal parameter vector $\Theta^* = \begin{bmatrix} \Theta_1^* & \Theta_2^* \end{bmatrix}^T$ is determined such that the *closed-loop* behavior meets the reference model dynamics :

$$y(s)(s + a_0) = b_0(\Theta_1 y(s) + \Theta_2 r(s)) \tag{11}$$

Then closed-loop (CL) transfer function (TF) is given as

$$\frac{y(s)}{r(s)} = \frac{b_0\Theta_2}{s + (a_0 - b_0\Theta_1)} \tag{12}$$

Hence yielding the ideal parameter vector Θ^*

$$\Theta_1^* = \frac{a_0 - a_m}{b_0} \tag{13a}$$

$$\Theta_2^* = \frac{b_m}{b_0} \tag{13b}$$

3.2.2 Second-order Control Law

For a second-order system, additional dynamic information is often necessary. One common choice is to include the derivative of the output in the regressor vector:

$$\omega(t) = \begin{bmatrix} y(t) \\ \dot{y}(t) \\ r(t) \end{bmatrix} \tag{14}$$

Thus, the control law becomes:

$$u(t) = \Theta_1 y(t) + \Theta_2 \dot{y}(t) + \Theta_3 r(t). \tag{15}$$

Here, the ideal parameter vector $\Theta^* = \begin{bmatrix} \Theta_1^* & \Theta_2^* & \Theta_3^* \end{bmatrix}^T$ is chosen to satisfy the matching condition:

$$y(s)(s^2 + (a_1 - b_0\Theta_2)s + a_0 - b_0\Theta_1) = b_0\Theta_3 r(s) \tag{16}$$

Yielding the closed-loop transfer function as

$$\frac{y(s)}{r(s)} = \frac{b_0\Theta_3}{s^2 + (a_1 - b_0\Theta_2)s + (a_0 - b_0\Theta_1)} \tag{17}$$

Giving the *matching condition* as

$$\Theta_1^* = \frac{a_0 - a_{0m}}{b_0} \tag{18a}$$

$$\Theta_2^* = \frac{a_1 - a_{1m}}{b_0} \tag{18b}$$

$$\Theta_3^* = \frac{b_{0m}}{b_0} \tag{18c}$$

3.3 Adaptation Law

To adapt the parameter vector Θ, we employ the MIT rule as explained here [5]. The cost function to be minimized is defined as

$$J(\Theta) = \frac{1}{2}e^2 \tag{19}$$

To minimize $J(\Theta)$ over time, we adjust Θ in the direction of the negative gradient of the cost function, see [4]. The *MIT* rule is then formulated as:

$$\frac{d\Theta}{dt} = -\alpha\frac{\partial J}{\partial \Theta} = -\alpha e\frac{\partial e}{\partial \Theta} \tag{20}$$

where $\alpha > 0$ is the learning rate. Since $e(t) = y(t) - y_m(t)$ and the reference model output $y_m(t)$ is independent of Θ, we have:

$$\frac{d\Theta}{dt} = -\alpha e\frac{\partial y}{\partial \Theta} \tag{21}$$

This expression shows that the rate of change of the parameter vector Θ is proportional to both the error $e(t)$ and the sensitivity of the system output y with respect to Θ.

3.3.1 First-order Adaptation Law

Solving the sensitivity function for the first-order model yields:

$$\frac{\partial y}{\partial \Theta_1} = \frac{b_0}{(s + a_0 - b_0\Theta_1)}\, y \tag{22a}$$

$$\frac{\partial y}{\partial \Theta_2} = \frac{b_0}{(s + a_0 - b_0\Theta_1)}\, r \tag{22b}$$

Hence, the adaptation laws for the first-order parameters become:

$$\dot{\Theta_1} = -\alpha e\,\frac{\text{sign}(b_0)}{(s + a_m)}\, y \tag{23a}$$

$$\dot{\Theta_2} = -\alpha e\,\frac{\text{sign}(b_0)}{(s + a_m)}\, r \tag{23b}$$

$$\begin{bmatrix} \dot{\Theta_1} \\ \dot{\Theta_2} \end{bmatrix} = -\alpha e \begin{bmatrix} \frac{\text{sign}(b_0)}{(s+a_m)}\, y \\ \frac{\text{sign}(b_0)}{(s+a_m)}\, r \end{bmatrix} \tag{23c}$$

Here, the term $\frac{\text{sign}(b_0)}{(s+a_m)}$ serves as an effective sensitivity filter. Importantly, this formulation only requires knowledge of the $sign(b_0)$; the exact value of b_0 is not needed.

3.3.2 Second-order Adaptation Law

For a second-order system, the dynamics are more complex, and the sensitivity functions must capture the effects of both the

output and its derivative. Hence, the following adaptations laws:

$$\dot{\Theta}_1 = -\alpha_1 e \frac{\text{sign}(b_0)}{(s^2 + a_{1m}s + a_{0m})} y \tag{24a}$$

$$\dot{\Theta}_2 = -\alpha_2 e \frac{\text{sign}(b_0)s}{(s^2 + a_{1m}s + a_{0m})} y \tag{24b}$$

$$\dot{\Theta}_3 = -\alpha_3 e \frac{\text{sign}(b_0)}{(s^2 + a_{1m}s + a_{0m})} r \tag{24c}$$

In this case, the adaptation law includes a term with s (the Laplace variable) in the sensitivity of Θ_2 to account for the dynamic effect of the derivative of y present in the control law.

4 Implementation

The implementation of the MRAC controller in Simulink for the second-order system consists of several components. The Simulink schema is depicted in Fig. 4. More about simulink can be found in [6].

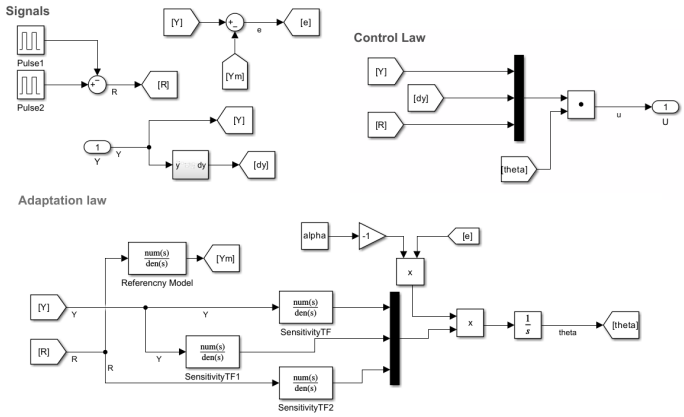


Figure 4: Simulink implementation of the MRAC controller for the second-order system

Choosing the appropriate learning rate coefficient vector α is crucial for ensuring system stability. A small learning rate may lead to slow convergence, whereas an excessively large learning rate can cause instability—especially in the presence of noise, where the derivative in the control law may amplify disturbances.

To determine the optimal α, we formulate the following minimization problem, if the cost function $J(\alpha)$ is defined as

$$J(\alpha) = \int_0^{t_f} e^2(t)\, dt \tag{25}$$

Then minimization is given as

$$\min J(\alpha) = \min \int_0^{t_f} e^2(t)\, dt = \min \int_0^{t_f} (y(t) - y_m(t))^2\, dt \tag{26}$$

where the objective is to minimize the sum of squared errors $e^2(t)$.

Initializing Θ with the final value from the previous simulation ensures that the system starts from a known state, which can be beneficial in certain scenarios.

The reference model chosen for the simulation is a second-order system with the transfer function:

$$\frac{y_m(s)}{r(s)} = \frac{\omega^2}{s^2 + 2\zeta\omega s + \omega^2}, \quad \text{where } \zeta = 1, \omega = 10 \tag{27}$$

and the learning rate coefficients are set as:

$$\alpha = \begin{bmatrix} 210 & 12 & 300 \end{bmatrix} \tag{28}$$

5 Simulation

5.1 Second-order System

The simulation of the second-order system with the MRAC controller produced the results shown in Fig. 5.

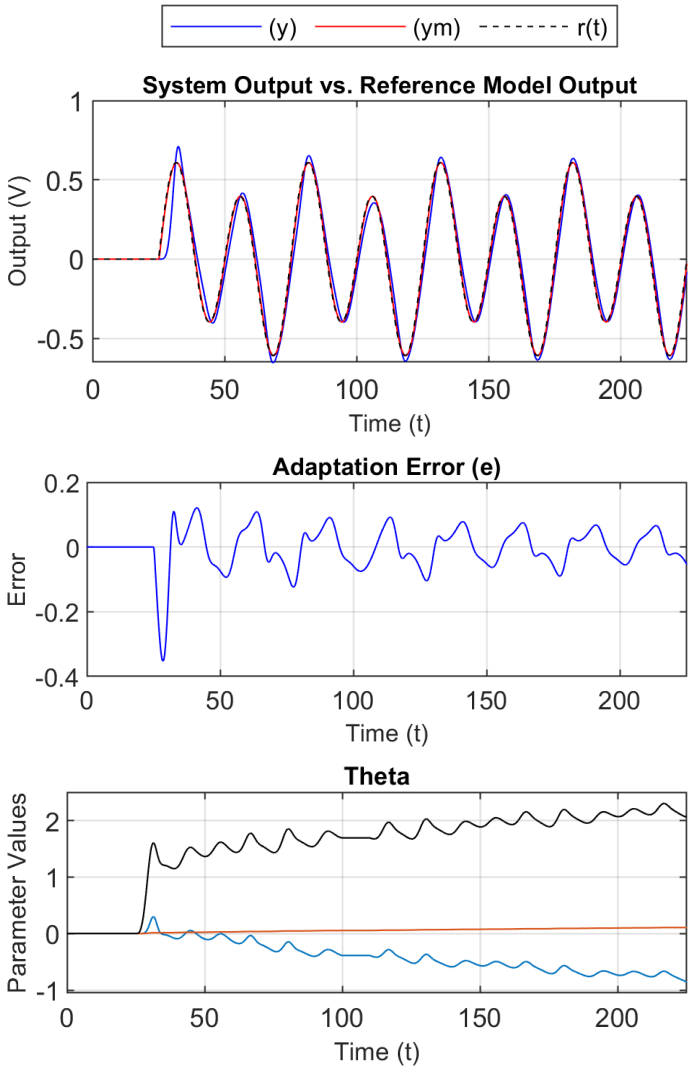


Figure 5: Simulation results for the second-order system with the MRAC controller

As shown in Fig. 5, the system output successfully adapts from an initial condition of $\Theta = 0$, leading to the convergence of the adaptation error $e(t)$ to zero, as indicated in (7).

In this case, the adaptation process takes longer to converge compared to the first-order system. This slower convergence is attributed to the higher complexity of the dynamics in the second-order control system and the presence of the derivative of y in the control law, which introduces additional dynamic effects that must be accommodated in the adaptation law.

6 Verification

6.1 On Laboratory Dynamic System

The laboratory dynamic system is a single-input single-output (SISO) setup that features an electronic circuit with operational amplifiers engineered to exhibit a specific dynamic behavior. Essentially, this circuit acts as an adjustable electronic filter whose dynamics and gain can be finely tuned using a potentiometer. A small DC motor, equipped with a tachogenerator mounted on its shaft, is connected to the circuit. In this configuration, the input signal is first filtered by the circuit, and the resulting voltage sets the supply voltage for the DC motor. The motor's rotational speed is then measured by the tachogenerator, which produces an output voltage proportional to that speed.

All voltage signals involved—the filter input, the motor supply, and the tachogenerator output—range from 0 to 10 volts. Additionally, the operating condition of the system is manually adjusted via the potentiometer, whose position is represented by a voltage signal varying between 0 V and 10 V.

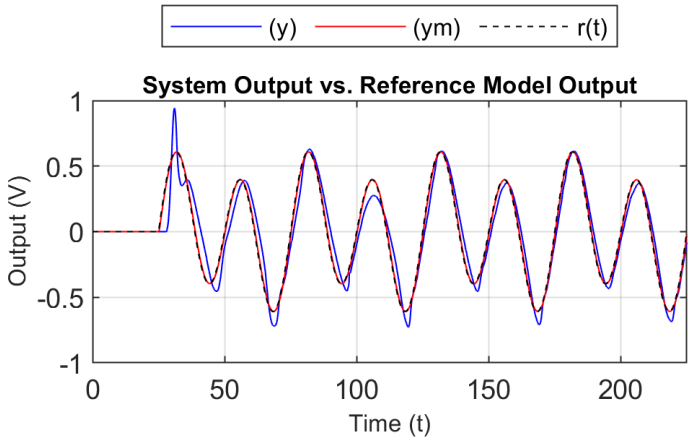
Figures Fig. 6a and Fig. 6b demonstrate that the MRAC controller tracks the reference model output, with the adaptation error $e(t)$ gradually converging to zero. The evolution of the parameter vector Θ over time is illustrated in Fig. 6c.

To validate the MRAC approach in a real-world setting, a laboratory dynamic system was used to generate the measured results presented in Fig. 6.

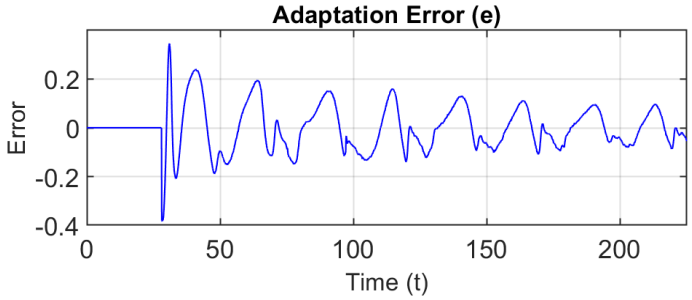
To further assess the controller's adaptability, it was tested under different operating conditions. After an initial evaluation in one regime, the system's dynamics were deliberately altered, requiring the controller to adapt in real-time. The results of this test are shown in Fig. 7.

A noticeable shift in the system's dynamics occurs at approximately the 100th second, as depicted in Fig. 7. In response, the MRAC controller adjusts its parameter vector Θ, with the rate of adaptation depending on the chosen learning rate coefficients. This adaptation, illustrated in Fig. 7, demonstrates the controller's ability to handle time-varying conditions effectively.

(a) System output and reference model output



(b) Adaptation Error (e)



(c) Adaptation of the parameter vector Θ

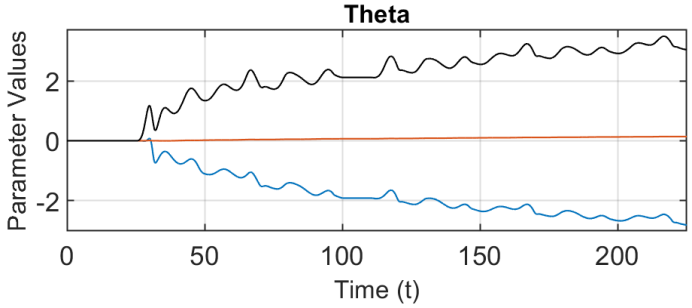


Figure 6: Measured response of the laboratory dynamic system with MRAC control

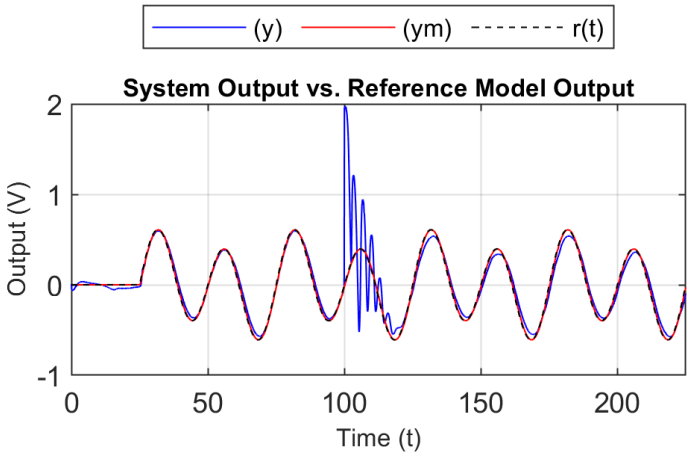
Overall, the results confirm that the MRAC controller successfully adapts to changes in system dynamics, showcasing its robustness even in real-world scenarios.

7 Conclusion

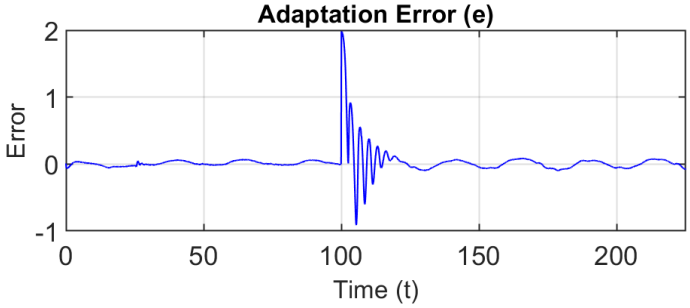
In conclusion, we have identified a dynamic system and successfully designed and implemented a Model Reference Adaptive Control (MRAC) controller for a laboratory dynamic system, that is able to adapt to and handle time-varying dynamics.

In future work, we plan to explore the implementation of different types of MRAC for better performance. Additionally, we aim to investigate the use of PID in symbiose with MRAC such that the controller and system can adapt to time-varying dynamics.

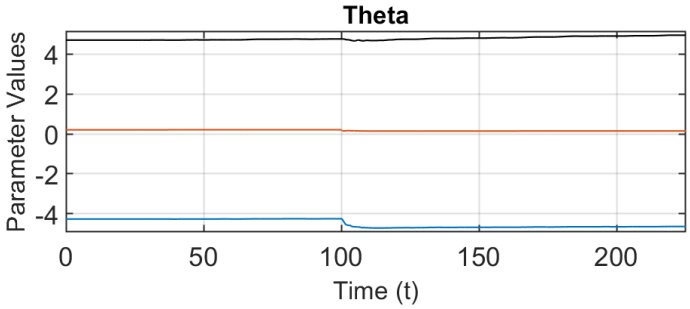
(a) System output and reference model output



(b) Adaptation Error (e)



(c) Adaptation of the parameter vector Θ



(d) Potentiometer output

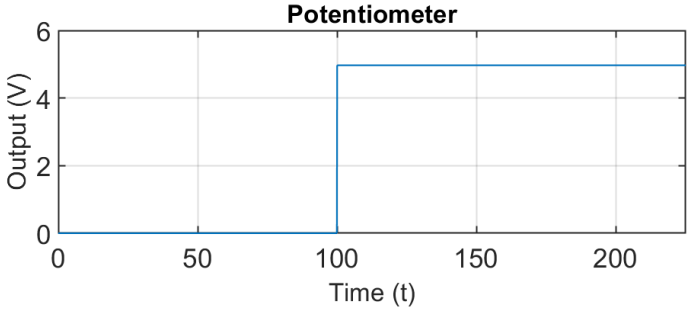


Figure 7: Measured response of the laboratory dynamic system with MRAC control under time-varying conditions

Lastly, we will explore and consider the implementation of such control method in an industrial setting, using industrial hardware and software such as *PLC* due to its *low latency* and high reliability.

References

[1] Peter Hudzovič. *Identifikácia a modelovanie*. Slovenská vysoká škola technická v Bratislave, 1986.

[2] Shlomo Engelberg. *Mathematical Introduction To Control Theory, A*. Vol. 8. World Scientific, 2024. ISBN: 9781800615540.

[3] P.A. Ioannou and J. Sun. *Robust Adaptive Control*. Dover Books on Electrical Engineering Series. Dover Publications, Incorporated, 2012. ISBN: 9780486498171.

[4] Ján Murgaš and Ivan Hejda. *Adaptívne riadenie technologických procesov*. Slovenská technická univerzita v Bratislave, 1993. ISBN: 80-227-0529-2.

[5] Marián Tárník. *Adaptívne riadenie MRAC gradientný*. Online course material. [Online; accessed 2025-03-12]. 2025. URL: `https://github.com/PracovnyBod/ADRIA/blob/master/doc/AR04_txt_MRAC_gradientny/AR04_txt_MRAC_gradientny.pdf`.

[6] MathWorks. *MATLAB Documentation*. [Online; accessed 2023-10-27]. 2023. URL: `https://www.mathworks.com/help/matlab/`.

Comparative Analysis of Neural Network Architectures for Sentiment Classification in Slovak Textual Reviews

Matej Medveczky, Jarmila Pavlovičová

Institute of Robotics and Cybernetics, FEI STU in Bratislava

xmedveczky@stuba.sk

Abstract – Customer sentiment evaluation helps companies better understand their needs and improve products or services based on real feedback. Successful automatic sentiment analysis using machine learning methods enables rapid problem identification and increased customer satisfaction. This paper deals with comparing different neural network architectures for sentiment analysis of Slovak text reviews. We examine the performance of Multilayer Perceptron (MLP), Convolutional Neural Networks (CNN), and Recurrent Neural Networks (LSTM) in sentiment classification. Experiments are conducted on a dataset of Slovak reviews using language models and various machine learning techniques. The results highlight differences in the performance of individual architectures and provide a comparison of the effectiveness of different natural language processing approaches.

1 Introduction

Sentiment analysis represents an important area of Natural Language Processing (NLP), which focuses on the automatic identification and categorization of emotional coloration of text. In the context of the growing importance of digital communication, the ability of machine sentiment recognition is becoming increasingly significant for various applications, from customer feedback analysis to social media monitoring. [1]

The Slovak language presents a specific challenge in terms of natural language processing. It is characterized by a rich morphological structure, complex grammar, and a relatively limited amount of available language resources compared to other world languages. These characteristics place specific demands on the design and implementation of machine learning models focused on sentiment analysis.

The main goal of this paper is to implement various neural architectures for sentiment classification and compare the performance of the implemented neural architectures on a dataset of Slovak text reviews.

2 Methodology

The following subchapters will gradually discuss the individual methodological steps, starting with dataset preparation and augmentation, through embedding generation, to the implementation of neural networks themselves and their evaluation.

2.1 Dataset Preparation and Augmentation

Our dataset consisted of reviews from the Slovak environment, originally containing 10,200 records. This dataset was created by reducing the original dataset of 178,000 customer reviews from a certain Slovak online store. We filtered out duplicates and reviews shorter than 50 characters. The original customer five-star ratings were transformed into a three-level sentiment scale, where negative sentiment represents ratings below 3 stars, neutral sentiment corresponds to a 3-star rating, and positive sentiment represents ratings above 3 stars.

Recognizing the insufficient scope of the already filtered dataset, we implemented the following text augmentation procedures.

The first approach utilized multilingual translation of reviews through Google Translator, where we set the translation process from the original Slovak through English to German and back to Slovak. The aim of this process was to preserve semantic meaning and generate text variations. The assumption was, that since both these languages belong to the same language family, but their word order is significantly different, the translator would be able to change the sentence while maintaining its original sentiment and meaning.

The second augmentation approach consisted of masked language augmentation using the XLM-RoBERTa [2] model for masked language modelling. This technique allowed random selective replacement of words in the original reviews and generation of new text variants through prediction of masked sentence parts. This approach replaced original words with their synonyms. [3]

Thanks to these two approaches, we arrived at a balanced dataset of 50,001 reviews with 16,667 reviews from each sentiment class.

This dataset was divided for training purposes into training, testing, and validation data in an 80:15:5 ratio.

2.2 Embedding Generation

We chose the SBERT model to convert textual data into a multidimensional vector space, specifically the *stsb-xlm-r-multilingual* [4] version. The SBERT (Sentence-BERT) approach fundamentally differs from standard BERT embeddings by optimizing sentence-level semantic representations. While traditional BERT models generate context-dependent word embeddings, SBERT focuses on creating meaningful sentence-level embeddings that capture nuanced semantic relationships. This is particularly crucial for

the Slovak language, which features complex morphological structures and rich grammatical variations. Comparative studies have shown that this approach significantly outperforms traditional embedding techniques, especially for morphologically rich languages like Slovak. The ability to capture semantic relationships beyond simple word-level translations makes it particularly suited for sentiment analysis tasks that require understanding subtle emotional coloration. [5]. The input to the encoder is a sentence, and its output is a 768-dimensional vector field, which will serve as input for neural networks

2.3 Neural Network Architectures

Within the research, we implemented five different neural architectures, each with its own specific characteristics and data processing approaches.

A methodological problem in comparing the performance of different neural network architectures was setting equivalent parameters. Manually determining optimal parameters for each architecture would have been extremely challenging and could have led to result bias and non-objective comparison. For this reason, we took an innovative approach -- using the large language model (LLM) Claude 3.5 [6] for impartial and systematic parameter setting. We presented the research problem to the model and asked it to determine optimal and consistent parameters for individual neural network architectures. Next, we described the selected NN, we used in you comparison

2.3.1 Multilayer Perceptron (MLP)

The basic MLP architecture consists of multiple fully connected layers that transform input embedding vectors through non-linear activation functions. Our model uses a three-layer structure with ReLU activation function and dropout layers to prevent overfitting. The first layer contains 128 neurons, the second 128 neurons, and the third 64 neurons, with the final layer of three neurons implementing softmax activation for three-class sentiment classification.

2.3.2 Deeper Multilayer Perceptron

The extended MLP version adds additional layers and mechanisms to increase model complexity [7]. In addition to increasing the number of neurons (256 → 128 → 64 → 32), we implemented batch normalization, which stabilizes the training process and accelerates convergence. Dropout layers are maintained for model regulation and prevention of overfitting.

2.3.3 Convolutional Neural Network (CNN)

The Convolutional Neural Network applies 1D convolutional layers to extract local features from the embedding space. The model contains two convolutional layers (128 and 64 filters) with a max-pooling layer to reduce dimensionality. A global max-pooling layer aggregates the most important features, which are subsequently processed by a fully connected layer and output softmax layer

2.3.4 Long Short-Term Memory - LSTM

The Long Short-Term Memory (LSTM) network is designed to capture sequential dependencies in data. Our model uses one LSTM layer with 128 neurons that transforms input embedding vectors. A dropout layer prevents overfitting, and a dense layer with ReLU activation serves for final classification. LSTM is particularly suitable for sentiment analysis as it can capture contextual relationships between words.

2.3.5 Multi-layer LSTM

The extended LSTM version adds a second recurrent layer for deeper feature extraction [8]. The model contains two LSTM layers (128 and 64 neurons) with dropout layers between them. An added dense layer with ReLU activation allows the model to capture more complex patterns in the data. This architecture is designed to capture finer nuances in text.

2.4 Training Process

The neural network models were implemented in Python using the TensorFlow/Keras library. Python was employed for data pre-processing and evaluation, while TensorFlow/Keras provided the necessary framework for building and training the neural architectures. This approach facilitated systematic experimentation and the integration of multiple network configurations.

All models were trained with identical optimization parameter settings. We used the Adam optimizer with a learning rate of 0.0001 and sparse categorical cross entropy loss function, which is suitable for multi-level classification tasks. The maximum number of epochs was set to 300, and we implemented an early stopping mechanism with a patience of 5 epochs to prevent overfitting. (Fig. 2.)

During training, we monitored accuracy and loss for the training and validation sets. The early stopping mechanism allowed terminating training the moment when validation accuracy stopped showing improvement, thereby preventing model overfitting.

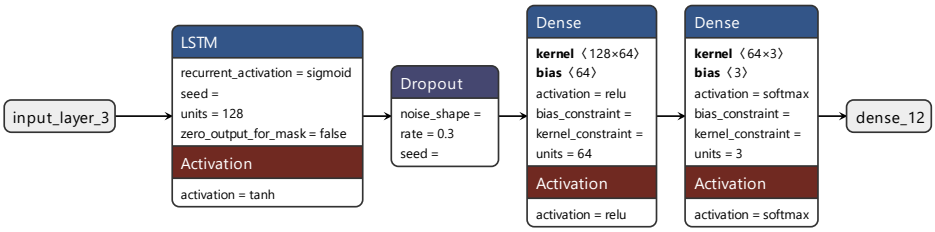


Fig. 1. Architecture of the LSTM network

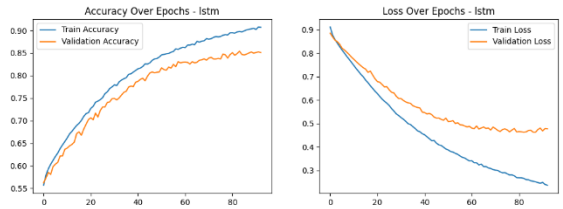


Fig. 2. Accuracy and loss during the training of the LSTM network

After training completion, we evaluated each model on the test data set using a confusion matrix, accuracy, recall, and F1 score for each sentiment class (negative, neutral, positive). Results were visualized through accuracy and loss graphs during the training process and confusion matrices for individual models.

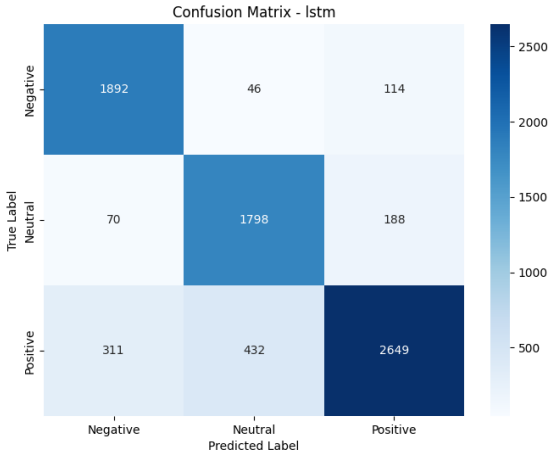


Fig. 3. Confusion matrix of the LSTM network

2.5 Model Evaluation

To determine the quality of individual models, we used indicators such as accuracy, recall, F1 score, and loss. These metrics provide a comprehensive view of the model's ability to correctly classify samples and identify patterns in the data. We used sparse categorical cross entropy as the loss function, which measures the difference between predicted probabilities and actual classes. The following equations and definitions more detailed explain individual metrics.

Accuracy is defined as the proportion of correctly classified samples to the total number of samples:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP are true positives, TN true negatives, FP false positives, and FN false negatives.

Recall expresses the proportion of correctly identified positives from the total number of actual positives:

$$recall = \frac{TP}{TP + FN}$$

F1 score is the harmonic mean of precision and recall, which provides a balanced evaluation of the model, especially with imbalanced class distribution:

$$F1 = \frac{2 * precision * recall}{precision + recall}$$

where **precision** is defined as:

$$precision = \frac{TP}{TP + FN}$$

Loss in sparse categorical cross entropy measures the difference between predicted probabilities and actual classes. For N samples and C classes, it is defined as:

$$loss = -\frac{1}{N}\sum_{i=1}^{N} \log p_{i,y_i}$$

where p_{i,y_i} is the predicted probability of the correct class y_i for sample i

Using these indicators, we can determine the quality of individual models.

Tab. 1. Results comparison

Model	Accuracy	F1 - Score
MLP	0,7867	0,7867
Deeper MLP	0,8091	0,8085
CNN	0,5553	0,5034
LSTM	0,8452	0,8447
Multi-layer LSTM	0,8060	0,8057

2.5.1 Architecture Comparison:

Results analysis revealed significant differences in the performance of individual neural architectures. The LSTM network proved to be the most successful, achieving the highest accuracy of 0.8452 and F1 score of 0.8447, confirming its ability to effectively capture contextual relationships in language data. On the other hand, the CNN model significantly lagged with an accuracy of only 0.5553, which is likely related to the specifics of the Slovak language and limited ability to capture text sequential dependencies.

An interesting finding was that the deeper MLP model with an added layer managed to improve accuracy from 0.7867 to 0.8091 compared to the basic architecture. However, this trend is not universal - in the case of LSTM networks, we observed the opposite, where the simpler architecture achieved better results than the multi-layer version. These findings point to the complexity of designing neural networks and the importance of testing of various architectures.

3 Conclusion

Our research into sentiment analysis for Slovak language text reviews has produced significant insights into the performance of different neural network architectures. Among the analysed networks, recurrent neural networks, specifically the LSTM model, showed as the most promising approach for sentiment classification in Slovak sentences.

The LSTM network's performance, achieving an accuracy of 0.8452 and an F1 score of 0.8447, was particularly noteworthy. This result is not just a numerical achievement but represents a meaningful step forward in natural language processing for the Slovak language, as this is one of the highest scores achieved. The model's ability to capture contextual relationships and nuanced linguistic patterns suggests promising directions for future research in sentiment analysis.

Some unexpected findings emerged during our research. The deeper MLP model showed improved accuracy compared to its simpler counterpart, while the multi-layer LSTM performed slightly worse than the single-layer version. These results remind us that neural network design is not a straightforward, predictable process. Each architectural choice requires empirical validation, and intuitive assumptions about model complexity do not always hold true.

While our results are promising, we acknowledge the limitations of our study. The research was conducted on our specific dataset of customer reviews, and the findings may not entirely generalize to all types of Slovak language text. Future research could expand the dataset to include more diverse text sources, develop more sophisticated augmentation technique and fine-tune the existing neural network architectures.

References

[1] A. C. S. R. Mayur Wankhade, „A survey on sentiment analysis methods, applications, and challenges,“ *Artificial Intelligence Review*, zv. 55, p. 5731–5780, 2022.

[2] „Hugging Face,“ [Online]. Available: https://huggingface.co/docs/transformers/model_doc/xlm-roberta. [Cit. 03 2025].

[3] K. Z. Jason Wei, „EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks,“ 2019.

[4] “Hugging Face,” 28 03 2025. [Online]. Available: https://huggingface.co/sentence-transformers/stsb-xlm-r-multilingual.

[5] I. G. Nils Reimers, „Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks,“ rev. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2019.

[6] „Claude LLM,“ Anthropic, 03 2025. [Online]. Available: https://claude.ai/new.

[7] „Turing,“ [Online]. Available: https://www.turing.com/kb/explanation-of-deep-neural-network-multilayer-perceptron-deep-q-network. [Cit. 03 2025].

[8] Y. H. E. A. W. S. Afan Galih Salman, „Single Layer & Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting,“ rev. *3rd International Conference on Computer Science and Computational Intelligence 2018*, 2018.

Riadenie mobilného robota

Yehor Kiriienko

[1] Ústav robotiky a kybernetiky, FEI STU v Bratislave
xkiriienkoy@stuba.sk

Abstrakt – Táto práca sa zaoberá návrhom a implementáciou autonómneho mobilného robota, ktorý využíva senzory typu LIDAR a kameru na detekciu okolia a vyhýbanie sa prekážkam v reálnom čase. Robot je schopný vytvárať mapu prostredia, lokalizovať sa v nej a efektívne plánovať svoju trajektóriu od východiskovej polohy až k cieľu.

1 Riadenie mobilného robota

Riadenie mobilného robota je kľúčovou oblasťou robotiky, ktorá zahŕňa techniky pre efektívnu a bezpečnú navigáciu v rôznych prostrediach. Lokalizačné riadenie umožňuje robotu presne určiť svoju polohu a orientovať sa v priestore. V práci sa zameriavame na vývoj a testovanie algoritmov pre lokálnu navigáciu, ktoré umožnia robotom vykonávať úlohy ako navigácia v skladoch alebo doručovanie tovaru. Na simuláciu bol použitý ROS (Robot Operating System), ktorý poskytuje knižnice a nástroje na simuláciu robotických systémov. Pre vizualizáciu sa využíval simulátor Gazebo, ktorý umožňuje realistickú simuláciu v 3D prostredí. Simulácia senzorov bola realizovaná pomocou modelu LIDAR, ktorý meria vzdialenosť od prekážok v okolí robota.

Tab. 1. Porovnanie simulacnych prostredi

Kritérium	ROS	Webots	V-REP/ CoppeliaSim
Integračná podpora	Veľmi silná podpora pre rôzne robotické systémy	Stredne silná podpora, zamerané na edukáciu	Vysoká podpora pre rôzne roboty a senzory
Realistickosť simulácie	Vysoká – veľmi realistické simulácie s Gazebo	Stredná – dobré, ale menej realistické ako Gazebo	Vysoká – realistické simulácie robotov
Podpora algoritmov a nástrojov	Vysoká – podpora pre-SLAM, DWA, A* a mnoho ďalších	Obmedzená podpora pre pokročilé algoritmy	Podpora pre pokročilé algoritmy a robotické systémy

2 Mobilné roboty a ich aplikácie v navigácii

Mobilné roboty sa dnes široko využívajú v rôznych oblastiach, ako je logistika, zdravotná starostlivosť, priemysel a doručovanie. Ich hlavnou výhodou je schopnosť vykonávať autonómnu navigáciu v reálnych prostrediach. V tomto projekte sa zameriavame na robota vybaveného senzormi ako LIDAR, ktorý vykonáva autonómne pohyby v prostredí a vyhýba sa prekážkam.

2.1 Používané technológie a algoritmy

Kľúčovou súčasťou robotov je navigácia, ktorá využíva algoritmy ako SLAM (Simultánna lokalizácia a mapovanie) a AMCL (Adaptívna Monte Carlo lokalizácia) na vytváranie máp prostredia a určovanie polohy robota. Na efektívne pohyby robota sa používajú algoritmy ako Dynamic Window Approach (DWA), A-star (A*), D-star (D*), ktoré umožňujú navigáciu v reálnom čase na rôznych mapách, ako sú statické (napríklad pre sklady, kde je možné vytvoriť statickú mapu) alebo dynamické (kde nie je možné vytvoriť mapu a robot musí samostatne a v reálnom čase vytvárať mapu a pohybovať sa v nej).

V rámci tohto projektu som implementoval vlastný algoritmus lokálnej navigácie na základe algoritmu TangentBug, ktorý umožňuje robotu pohybovať sa k cieľu, vyhýbať sa prekážkam a prispôsobovať trajektóriu podľa aktuálnych podmienok.

2.2 Simulácia

Pre overenie svojich algoritmov som využil ROS (Robot Operating System), pretože je to silný nástroj, ktorý poskytuje rozsiahlu podporu pre simuláciu rôznych robotických systémov. ROS umožňuje efektívne testovanie algoritmov pre navigáciu, riadenie, ako aj prácu so senzormi a akčnými členmi. Na simuláciu robota a testovanie týchto algoritmov som využil Gazebo, ktoré je plne integrované s ROS a umožňuje realistickú simuláciu robotov v 3D prostredí.

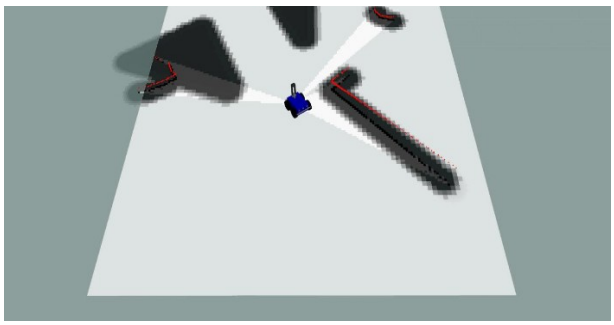


Obr. 1. Model robota v Gazebo

Pre základy svojho robota som využil model robota MRVK (Mobilný robot vývojový kit). Tento robot je vybavený štyrmi kolesami a LIDAR-om. Senzor poskytuje robotovi dôležité informácie o jeho okolí, ktoré sú nevyhnutné pre vykonávanie úloh, ako je vyhýbanie sa prekážkam a navigácia v neznámom prostredí. LIDAR sa používa na získavanie dát o priestore, čo umožňuje robotovi orientovať sa v prostredí a efektívne sa pohybovať.

2.3 Model

Pre model robota bol implementovaný model skutočného LIDAR-u **Hokuyo URG-04LX-UG01**. Tento laserový diaľkomer používa polovodičovú laserovú diódu s infračerveným žiarením o vlnovej dĺžke **λ = 785 nm**. Meranie vzdialenosti sa vykonáva pomocou **fázového posunu**.



Obr. 2. Vizualizácia dát z LIDAR senzora

Rozsah detekcie: **D** ∈ ⟨**0,06; 4**⟩ **m**
Rozsah detekcie v 2D priestore: **240°**, krok **0,36°**

Tento výpočet D využíva rýchlosť svetla c a dobu, ktorú lúč potrebuje na prekonanie danej vzdialenosti. Preto je potrebné vydeliť hodnotu doby letu t hodnotou 2, pretože lúč prechádza vzdialenosť D dvakrát: raz pri ceste k objektu a druhýkrát pri návrate.

$$D = \frac{c \cdot t}{2}$$

$$t = \frac{\Delta \varphi}{\omega}$$

Kde t je doba letu lúča a c je rýchlosť svetla, 1/2 kvôli prejdeniu vzdialenosti D lúčom 2-krát. Dobu letu lúča je možné vypočítať za pomoci fázového posunu Δφ.

Výpočet doby letu lúča pomocou fázového posunu Δφ:

$$\omega = 2\pi f$$

$$f = \frac{c}{\lambda}$$

Kde ω je uhlová frekvencia, f je frekvencia a λ je vlnová dĺžka vysielaného lúča.

Vzdialenosť od prekážky je potom možné vypočítať v závislosti od fázového posunu:

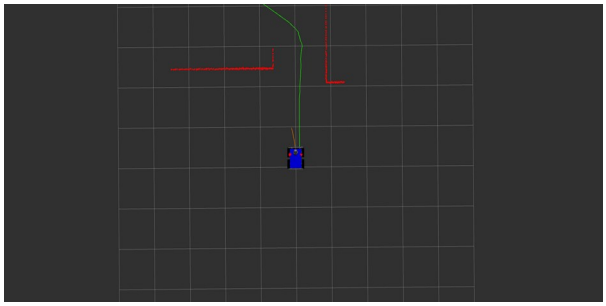
$$D = \frac{\lambda}{4\pi} \Delta \varphi$$

2.4 Výsledky lokálnej navigácie

Implementovaný systém lokálnej navigácie umožňuje mobilnému robotu pohybovať sa smerom k zvolenej cieľovej pozícii, ktorá je určená priamo v prostredí RViz. Navigačný algoritmus je založený na princípe algoritmu TangentBug, ktorý je známy svojou schopnosťou efektívne kombinovať priamy pohyb k cieľu s vyhýbaním sa prekážkam.

Vďaka využitiu senzora typu LIDAR je robot schopný v reálnom čase analyzovať svoje okolie a identifikovať potenciálne prekážky v dráhe. Na základe získaných údajov dokáže dynamicky upraviť svoju trajektóriu, čím efektívne obchádza prekážky.

Testovanie prebiehalo v rôznych typoch simulovaných prostredí – od jednoduchých statických máp až po dynamické a nepravidelné rozmiestnenie prekážok. V každom z týchto scenárov robot úspešne dosiahol stanovený cieľ, pričom jeho trajektória bola plynulá a adaptívna vzhľadom na aktuálne podmienky v okolí. Celý proces prebiehal v reálnom čase a bez výskytu kolízií, čo potvrdzuje robustnosť a praktickú použiteľnosť navrhnutého riešenia.



Obr. 3. Pohyb robota k cieľu

3 Záver

V tejto práci sme sa zamerali na vývoj algoritmov pre lokálnu navigáciu mobilného robota, ktoré umožňujú efektívnu a bezpečnú navigáciu v rôznych prostrediach, ako sú sklady a doručovanie tovaru. Použili sme ROS a Gazebo na testovanie a simuláciu týchto algoritmov, čo nám umožnilo otestovať navigáciu v realistických podmienkach bez potreby fyzického robota.

Výsledky ukázali, že naše algoritmy sú efektívne pre navigáciu a vyhýbanie sa prekážkam v dynamických prostrediach. Tento vývoj prispieva k zlepšeniu autonómnych systémov a ich aplikácií v priemysle a každodennom živote. V budúcnosti by bolo vhodné tieto algoritmy vylepšiť a otestovať v reálnych podmienkach.

Literatúra

[1] [s. n.]. Hokuyo URG-04LX-UG01 Scanning Laser Range Finder. 2009. Specification. [online] [cit. 2021-05-22].
[2] [s. n.]. Gazebo. 2014. [online]
[3] [cit. 2021-05-22]. [4] [s. n.]. Gazebo. 2014. URDF in Gazebo. [online] [cit. 2021-05-22].
[4] [s.n.]. ROS Wiki. 2014. ROS/Wiki/Patterns/Communication. [online]cit. 2021-05-22].
[5] [s. n.]. ROS Wiki. 2020. move_base. [online] [cit. 2021-05-22].
[6] [s. n.]. ROS Wiki. 2019. base_local_planner. [online][cit. 2021-05-22].

Sekcia 7

Informačné a komunikačné technológie

Data science application in ICT

Dmitrii Savin,

[1]Institute of Multimedia Information and Communication Technologies, FEI STU in Bratislava

xsavin@stuba.sk

Abstract – This paper demonstrates the vital role of Data Science in ICT by solving complex problems through advanced analytics. Using the "5G Resource Allocation Dataset" [1] we explore resource distribution in 5G networks. The study applies data cleaning, feature engineering, and machine learning to optimize allocation. The practical section includes Python implementations with pandas and the RandomForest [3] algorithm to calculate resource shares, sort application types, and predict allocations based on network parameters.

1 Introduction

Data Science has emerged as a pivotal field within Information and Communication Technologies (ICT), empowering professionals to extract insights and make data-driven decisions. With the increasing complexity of digital infrastructures and the explosive growth of data, Data Science offers robust methods to analyze, model, and optimize operations across diverse sectors.

The advent of 5G networks brings revolutionary improvements in connectivity, speed, and latency. However, dynamically allocating network resources in such environments remains a complex challenge.

1.1 Data Science Methodology

At the heart of this work is a Data Science workflow that includes key steps: Data Collection and Cleaning, which involves handling raw datasets and ensuring that missing or inconsistent data are managed effectively; Feature Engineering, which focuses on extracting and transforming relevant features from the dataset such as signal strength, latency, and bandwidth requirements; and Statistical Analysis and Modeling, which utilizes statistical methods and machine learning algorithms (like RandomForest) to build predictive models.

1.2 Present usage of 5G networks

The analysis of the dataset reveals current trends in mobile network usage. The diversity of application types reflects the wide range of tasks that 5G networks must support.

- **Video Communication and Multimedia:** Applications for video calls and streaming video require high bandwidth, minimal latency, and a reliable connection due to the real-time transmission of large volumes of data.
- **IoT Devices:** Applications related to the Internet of Things exhibit stable but less bandwidth-demanding resource allocation. Although these devices transmit relatively small amounts of data, they require continuous connectivity and energy efficiency.
- **Online Gaming and Social Networks:** These application categories have unique requirements — from low latency to high-speed data transmission — which necessitate dynamic allocation of network resources.

1.3 Rapid Development of 5G Networks

Since early trials in the 2010s and the establishment of standards by organizations like 3GPP, 5G quickly evolved into a global network by 2019, driving innovations with its unprecedented speeds and low latency that now power smart cities, IoT, and autonomous systems.

2 Theoretical part and it's components

In the theoretical part of this study, fundamental concepts of dynamic resource allocation in ICT were employed, along with statistical analysis and machine learning methods. Core theoretical materials include a dataset with different key parameters such as signal strength, latency, and bandwidth requirements.

These concepts guided the development of our analytical approach and informed the design of the Python code, which implements functions for calculating resource allocation percentages, sorting by application type, and training a RandomForest model for predictive analytics.

2.1 Software used

The study primarily relies on Python as the main software environment. We used the **Pandas** [2] library that simplifies data manipulation and analysis. It introduces two primary data structures: the DataFrame, which organizes data in a tabular format with labeled axes, and the Series, a one-dimensional array with an associated index. These structures allow for intuitive data handling, enabling users to filter, merge, reshape, and aggregate data efficiently. Pandas also offers extensive support for reading from and writing to various file formats such as CSV, Excel, and SQL databases. Its seamless integration with libraries like NumPy and scikit-learn makes it an essential tool

for building data-driven applications and performing exploratory data analysis.

RandomForest, a popular ensemble method available in scikit-learn, builds multiple decision trees during training and aggregates their predictions. For regression tasks, it computes the average of all tree outputs, while for classification tasks, it uses majority voting. This ensemble approach helps mitigate overfitting and improves prediction accuracy. In our work, we leverage the **RandomForestRegressor** to predict network resource allocation based on features like signal strength, latency, and bandwidth requirements, demonstrating the practical application of machine learning in optimizing ICT systems.

3 Experimental approach and results

3.1 Key Analytical Functions

Below is a list of essential functions used in the analysis, along with a brief description of their roles:

```python
def train_random_forest_model(df, target_col='Resource Allocation', feature_cols=None, test_size=0.2, random_state=42):
    """
    Trains a RandomForestRegressor and prints the R  score.
    Returns the trained model and test data.
    """
    if feature_cols is None:
        feature_cols = df.select_dtypes(include=['float64', 'int64']).columns.drop(target_col).tolist()
    data = df.dropna(subset=[target_col] + feature_cols)
    X, y = data[feature_cols], data[target_col]
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=test_size, random_state=random_state)
    model = RandomForestRegressor(random_state=random_state)
    model.fit(X_train, y_train)
    score = r2_score(y_test, model.predict(X_test))
    print(f"R  score: {score:.2f}")
    return model, (X_test, y_test), score
```

Listing 1: Random Forest Model Training

3.2 Results Visualization and Analysis

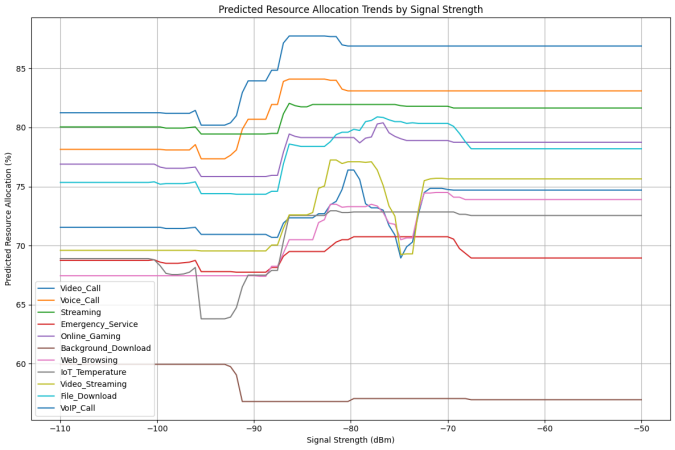


Figure 1: Predicted resource allocation trends by signal strength.

As signal strength improves (from about –110 to –70 dBm), most applications receive more resources, with the largest increases for critical and interactive services (Video Call, Emergency Service, Online Gaming). Lower-priority apps (IoT, File Download) rise more gradually.

The stepped changes indicate threshold splits from the RandomForest model. Overall, the model effectively reallocates resources, prioritizing high-demand applications as signal quality improves.

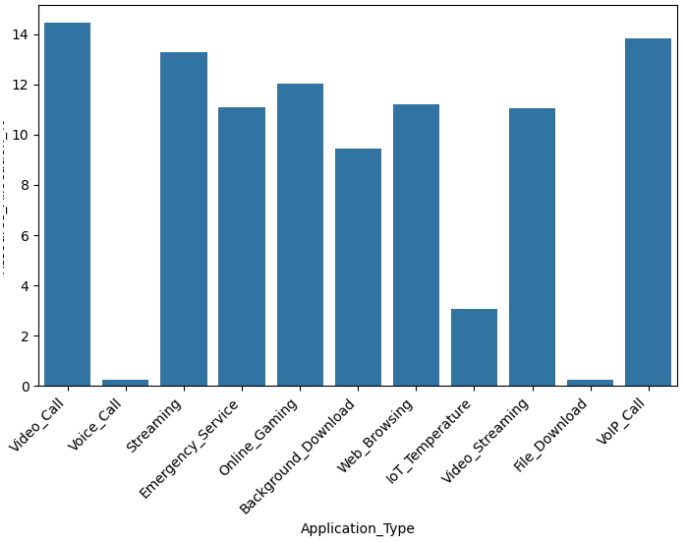


Figure 2: Resource allocation distribution by application type.

This bar chart shows how resources are distributed across different application types. **VoIP Call**, **Video Call**, receive the highest allocations, suggesting a priority on real-time and critical services. Meanwhile, lower demands are allocated to IoT, file downloads, and certain streaming tasks, indicating that less urgent applications get fewer resources in this model.

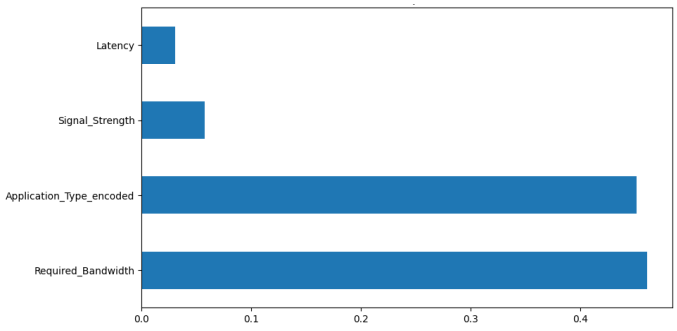


Figure 3: Feature Importance

This bar chart highlights the importance of each feature in predicting resource allocation. The model relies most heavily on **Required Bandwidth** and **Application Type**, indicating that the nature of the application and its bandwidth needs drive the majority of allocation decisions. Signal Strength has a moderate impact, while Latency plays a comparatively smaller role in this particular model's predictions.

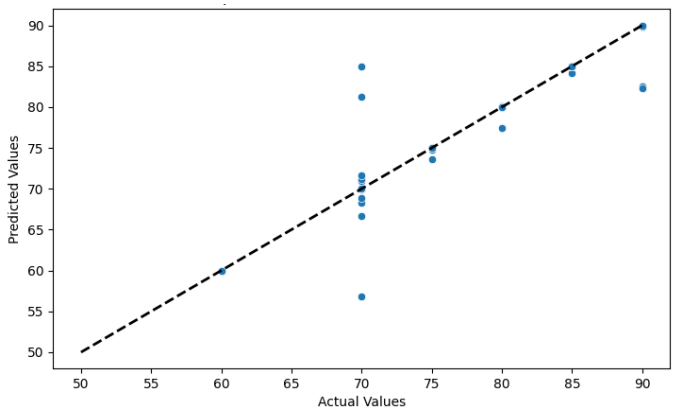


Figure 4: Comparison of Predictions and Actual Values

This scatter plot compares predicted resource allocation to actual values, with the dashed line representing perfect predictions. Most points cluster near the line, indicating the model generally performs well. However, a few outliers deviate noticeably, suggesting some underestimation or overestimation in those cases. Overall, the distribution indicates a solid predictive capability, though there is room for improvement in handling outliers.

4 Conclusion

This study demonstrates that advanced Data Science techniques are essential for optimizing resource allocation in 5G networks — a key aspect of ICT. Using the "5G Resource Allocation Dataset" and Python functions to compute allocation percentages, sort by application type, and train a RandomForestRegressor, our results show that the model effectively prioritizes high-demand, real-time services as signal quality improves.

Overall, our integrated approach of data cleaning, feature engineering, and machine learning provides a scalable framework for dynamic network management, with future work aimed at refining model accuracy and addressing outliers.

Acknowledgments

I thank Mykola Melnyk, a third-year ICT student at the Faculty of Electrical Engineering and Information Technology, for his valuable help with the project's practical part and his insightful advice.

References

[1] SOBHY, O. (2023). *5G Resource Allocation Dataset: Optimizing Band*. Kaggle.

Available at: `kaggle.com/datasets/omarsobhy14/5g-quality-of-service`.

[2] KAGGLE. (n.d.). *Learn Python, Data Viz, Pandas & More | Tutorials | Kaggle*.

Available at: `kaggle.com/learn`.

[3] *RandomForestRegressor — scikit-learn 1.6.1 documentation*.

Available at: `scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html`.

Creation of interactable environment for Virtual Reality using Unreal Engine 5

Yelyzaveta Marchenko

Ústav multimediálnych informačných a komunikačných technológií

xmarchenkoy@stuba.sk

This project is dedicated to the creation of an interactive environment for virtual reality (VR) using advanced technologies of the Unreal Engine 5. The main goal is to create an interactive and realistic virtual space that allows the user to interact with objects and the environment in real time. The project includes the stages of design, modeling, texturing, programming interactive elements and optimization for VR platforms. Implementation of mechanical interaction, including manipulation of objects, and adaptive animations. The project can be used for educational, simulation purposes, demonstrating the potential of modern tools for developing VR spaces. This project can be used to demonstrate specific objects in a virtual environment, competing with objects in reality.

1 Modern development of VR and UE 5

The current capabilities of Virtual Reality and Unreal Engine 5 are impressive. These technologies are developing very quickly and gaining popularity. They have many applications, which I will now talk about.

1.1 VR applications in the present

The Virtual reality is an illusion of reality created using computer systems that provide visual, auditory and other sensations. Virtual reality glasses and helmets consist of two small screens, each located in front of the eye. The screens show images of the same object, slightly shifted relative to each other, creating the illusion of three-dimensional space. The set also includes glasses motion controllers and sensors that track the user's movements and gestures in real time.

The main areas of use of VR in the modern world:

Tab. 1. The main areas of use of VR.

Entertainment and gaming	Creating a unique experience of immersion, wide opportunities for interaction, interactive pastime.
Education	New interactive opportunities in learning, an environment for complex experiments and experiments in medicine and physics, modeling dangerous situations and finding effective solutions to overcome them.
Medicine	The training environment for surgery, new approaches to treating psychological trauma, etc.
Military sphere	Practicing realistic combat scenarios.
Design	Creation of the most realistic projects in the field of building design, interiors, clothing, etc.
Aviation	Training of pilots of passenger and transport aircraft, emergency simulators.

1.2 Fast progress of UE 5 for development

Unreal Engine 5 is a game engine developed and supported by Epic Games. The first game on this engine was the first-person shooter Unreal, released in 1998. Although the engine was originally intended for the development of first-person shooters, its subsequent versions have been successfully used in games of various genres.

Written in C++, the engine allows you to create games for most operating systems and platforms: Microsoft Windows, Linux, Mac OS, Xbox consoles and others.

All elements of the game engine are presented in the form of objects with a set of characteristics and a class that defines the available characteristics. In turn, any class is a “child” class of object.



(a)



(b)

Fig. 1. Virtual reality glasses (a).
Fig. 2. Logo Unreal Engine 5 (b).

2 The theoretical part and its components

Having familiarized ourselves with the main technologies used in the project, we pay attention to the title “Creation of interactable environment for Virtual Reality using Unreal Engine 5”, namely “Creation of interactive environment”. It is this part that will be created virtually and tested for performance in reality.

2.1 The range of the study

The main goal of the study was to create an object in Unreal Engine 5 with which a person could interact in real time using VR glasses. As an example, I took a car that was created in the UE 5 program, then edited and improved. After completing the program, I moved on to the practical part of using the glasses. Putting them on and seeing the car in front of me, I could interact with it using the controllers, which I was holding in my hands.

2.2 The theoretical part of the work VR and UE 5

The theoretical part of the project needed to study the basis for working with the UE 5 and VR glasses. Official websites and additional sources helped me with this.

To create this project, I had to work on: Engine Terminology, Tools and Editors, Editor Interface, Working with Content, and knowing the Mathematical Foundations. These components have a main influence on the model created. Since the car model was downloaded from an online resource, I mostly worked with tools and editors that changed the look of the car and its textures.

For the work were used VR glasses Meta Quest 3. Glasses are based on displays, sensors, and motion tracking systems. Display: VR glasses use two displays located directly in front of the eyes. These displays create a stereoscopic effect, allowing you to perceive the virtual world in 3D. Sensors and motion tracking: To track head movements and changes in perspective in a virtual environment, glasses are equipped with different sensors such as accelerometers, gyroscopes, and magnetometers. It also uses outside sensors and cameras to better track body movements.

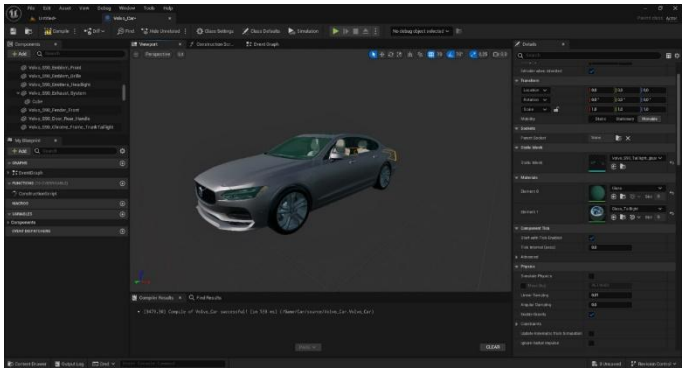


Fig. 3. Development of the model in the UE 5 engine.

2.3 The principle of VR glasses functioning

VR glasses work on the principle of creating a stereoscopic image and tracking the position of the user's head to allow immersion in virtual reality. Most of the principles have already been written in the previous paragraph of the project, but here is in-depth information about VR glasses.

Basic principles of operation:

Tab. 2. The list of basic principles

Displays and lenses	Lenses change focus to create the effect of viewing the real world.
Stereoscopic image	A little changed image (parallax) is presented to each eye, simulating depth as in human vision. The scene is visualized from two perspectives.
Head tracking	Built-in gyroscopes, accelerometers, and magnetometers track head movements and tilts. More expensive models use front-facing cameras to accurately identify the position in space.
Hand and controller tracking	Cameras or sensors track hand movements. Controllers provide tactile feedback through vibration.
3D sound	Speakers or headphones create a 3D sound effect by changing the direction of the sound as your head moves.

Also, the image reproduction and field of view are important factors. A high refresh rate (90 Hz and above) is important to avoid lag and glitches. The angular resolution is measured in pixels per degree (ppd) and is a main factor in the clarity and realism of a virtual world. From a helmet design perspective, the angular resolution is determined by two factors - the resolution of the display and the FOV (Field Of View). For optical systems, the pupils positions are fixed and the size of the display is the defining limiter of the field of view for a given lens. In these cases, the FOV is described horizontally, vertically and diagonally.

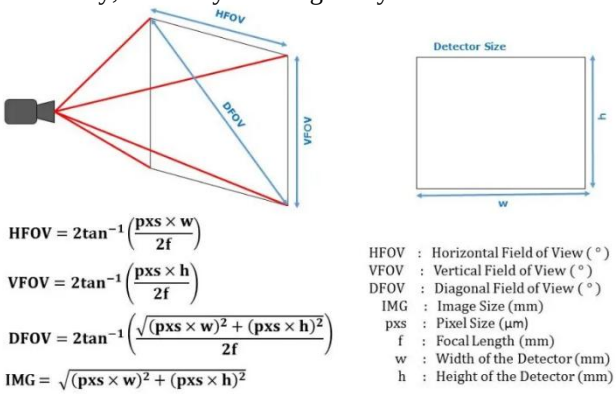


Fig. 4. Formulas for measuring the field of view.

3 The practical part and its description

The practical part of the project began with a search for a special car model on the Internet. The model had to be like a construction set, consisting of many elements and layers for better work with it. This is necessary so that it can be edited and animated. Volvo C90 was chosen for the work. This model was created in UE 5 and consisted of many parts, layers, details, and textures.After it was downloaded to the computer, it had to be assembled like a construction set. Some parts needed to be edited and changed, processing and changing textures took a lot of time, but the job was done.

After the car was finished, this part was saved and transferred to the virtual room so that it could be viewed in VR glasses in the future.

The next stage was fascinating and showed many benefits and pluses of my project and proved the importance of this research in the present.

I made all the necessary manipulations to connect the VR glasses of the Meta Quest 3 model and set them up, then connected the file with the finished car. I moved on to the main part - a virtual inspection of the car. Putting on the glasses and taking the controllers in my hands, I could see a basic room with the simplest textures and the main car.

Its look was accurately reproduced from the theoretical part. All the details of the body and interior, appearance and color with textures, exactly matched the specified parameters in UE 5. With the glasses on, you could see the car in real size (as in life). Walking around, you could fully see the size and examine the details. Looking inside the car, you could see all the buttons and levers. Virtual reality glasses fully conveyed all the textures, which made it possible to better study the object in front of you.



Fig. 5. The appearance of the car in the Virtual reality.

3.1 Present application of UE 5 and VR

People already combine UE 5 and VR glasses and use them together. Most often, these are games. Playing games and quests, online streams, playing online with friends, etc. But this is only the entertainment part of their possibilities.

VR glasses are often used to simulate combat operations for military training. In fact, this is very progressive and gives young soldiers important training for combat, but it is still analogous to a regular shooter game.

Young pilots can also take similar training courses. It simulates flight and unpredictable situations that pilots have to deal with. But this application is similar to completing a task in a game, too.

In today's realities, VR are in great demand on the battlefield of modern wars. VR glasses are now helping the military to hit their targets. Modern drones with explosives are equipped with mini-cameras connected to the goggles. This way, by controlling the drone with joysticks, hitting the target increases by 80%.

But the main question is how can we combine UE 5 and VR to make it useful in the future?

3.2 Future opportunities

The possibilities of using such a pair as UE and VR are very wide. But based on the indescribable present, I will emphasize two ways.

The first way is very useful for people. When choosing a new car in a car dealership, we always want to see as many options as possible. But this is not always possible, because there cannot be so many car models and so many different configurations in the showroom. In this case, people don't have full access to the overview of possible offers. And this is where Unreal Engine 5 and Virtual reality can help. When creating any car, it can be digitized and transferred to the UE for online work to create an identical copy. The doors, hood, and trunk should open, all buttons and levers should be in place, the engine should start with a sound, and other details should be correct.

And when a person wants to choose a new car, he or she will be able to make a full inspection even virtually. This will make the choice much easier. But the main condition will be an identical copy of this car, which will be created in the UE 5 engine.

The second application is in the military sector. The production of military equipment is increasing every year. Looking at the present military situation in the world and the Russian-Ukrainian war, every country wants to have its own armament. That's why everyone creates their own military equipment. But in some cases, it is not always practical. A good example is the situation with the training of the Ukrainian military in wartime. In order to teach soldiers how to operate new equipment from other countries, they were sent to those countries. Traveling and the training took too much time. In other cases, they promised to send tanks and other equipment for familiarization and training. But the transfer process could take more than three months. This time is extremely important in a time of war, but it was lost due to waiting.

That is why such an option as familiarization and training through the VR is extremely necessary. Before allowing the military to drive real military machines, they must first take virtual training. This will make it much easier in the future. The soldiers will already know the correct position of all the buttons and levers, the view from the inside and outside. Also, returning to the first point about cars, it is impossible to get familiar with all tank models. But if there is an identical copy of them created by a special remote control and connected to the glasses, it will be possible and will speed up the process of adaptation in the future.

4 Conclusions

Although Unreal Engine 5 and Virtual Reality are already popular and used in many areas as supporting tools. They still have many applications that we have not reached and are not using yet. So in this project, I proposed a few of them. My conclusions and opinions are unique, and I was mainly motivated by modern developments and inventions. I hope that these suggestions for use will become real and will help the world in specific industries.

Acknowledgments

I would like to thank Mykola Melnyk, a third-year student of the Information and Communication Technologies program at the Faculty of Electrotechnics and Informatics, for his help with the practical part of the project and useful advices. I also thank the student parliament of FEI STU for providing VR glasses.

References

[1] https://www.vrwiki.cs.
[2] https://uk.wikipedia.org
[3] https://vr-okuliary-virtualnoi-realnosti
[4] https://dev.epicgames.com/documentation
[5] https://wiki/Unreal_Engine_5
[6] https://dev.epicgames.com/documentation

Simplistic simulation model for impairments in DNA strings in selected system

Pavlo Melnyk

Ústav multimediálnych informačných a komunikačných technológií, FEI STU v Bratislave

xmelnykp@stuba.sk

Abstract – In this paper, we describe the creation of a simulation model for the selected system. Our model was designed based on principles of mathematical probability and statistics. To verify its implementation, we conducted testing to assess its accuracy. The assessment was done by calculating a 95 % Confidence Interval (CI), Mean Absolute Percentage Error (MAPE) and accuracy per impairment, which was later used to determine overall precision.

1 The rationale of the simulation model

Simulation models are key components in many research projects as they can provide valuable insights into the problematic issues scientists face daily. The most significant advantage of modelling is that it requires no specific hardware except a computational server or desktop computer. While simple models can be hardcoded, most advanced models rely on probabilities and statistics to address uncertainty, variability, and real-world approximations.

In our work, we aimed to recreate the selected system by applying concepts from mathematical probabilities and statistics. Specifically, we will compare the results we received from a series of conducted tests with those from the selected system. Since some parts of our selected system lack detailed data, we decided to simplify the model by using probabilities represented by error rates of respective impairments.

1.1 The description of the selected system

The chosen system is from research published in Nature: Communications, where researchers experimented with DNA assembly methods to enhance nanopore sequencing's real-time reading capacities [1]. Those DNA sequences were assembled using two core methods widely used in bioinformatics. The first one was a Consecutive Gibson Assembly, which helped to make nanopore sequencing throughput higher, but assembling more than 6 DNA fragments was problematic. The second one was an Overlap-Extension PCR (OE-PCR) which proved helpful as it not only created no side products but also required less sample preparation compared to Gibson assembly. After DNA fragments are assembled, the decoding process can start. The decoding algorithm was based on a consensus algorithm with respective adjustments that were needed for it to work correctly. After decoding, researchers found out that nanopore sequencing struggles with homopolymers. For our research, we are interested only in simulating errors during the decoding process. This means that assembly methods and everything before decoding don't have significant relevance to our research, and yet it is crucial to know all the details about the system.

2 Our model design

As described in the section dedicated to our selected system, we dealt with the biological process of DNA sequences being decoded back into information. During that process, impairments occur randomly in different spots of the sequence. This information provided us with valuable data, which we later used to define the type of model for our research. For systems with high inherent randomness, such as weather modelling, biological processes, etc., it is common to utilise stochastic models. The randomness of impairments in single DNA sequences was simulated using random sampling. Random sampling is a feature of the Monte Carlo simulation model.

Therefore, our simulation model (later model) combines characteristics from stochastic and Monte Carlo simulation models, ensuring a more precise system recreation. We defined the error rate of respective impairment as values of the probability vector. To properly simulate substitution impairment, we created a matrix where we stored probabilities of each nucleotide being replaced by another.

2.1 Our experiment design

Our experiment consisted of three stages. The modelling stage generated data that was later used during our testing. The last stage was the calculation of the precision of our model using statistical methods. To simplify the model's usage and save some time, we implemented an algorithm that copies the original file with DNA sequences and saves the model's result in a separate file.

This straightforward approach allowed us to simplify the testing stage, where we compared two files. During that, we logged every comparison step for future analysis and displayed results. Later, those results were saved to the spreadsheet to organise the data correctly. After we had done two runs for two files, we calculated our statistical data. We were particularly interested in the same file used in our selected system to receive the most precise result.

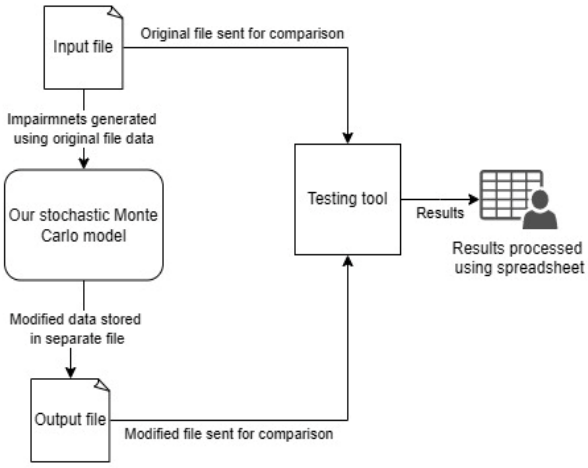


Fig. 1 A diagram that showcases the experiment design

2.2 Software we used

Our model was programmed utilising MATLAB programming language and its IDE (Integrated Development Environment). The language itself is a high-level one, meaning it has a lot of pre-programmed functions that are helpful because they streamline the process of developing the model. The biggest advantage of MATLAB is the vast number of plugins known as toolboxes. These toolboxes possess a set of unique functions that are related to a specific topic. However, according to our observations, MATLAB utilizes many system memory resources. This was overcome by modifying the program code to run efficiently and use fewer system memory resources. Occasionally, we used Python as a programming language and its respective IDE. It was helpful during data preparation.

Tab. 1 Technical data about the software used

Name of programming language or toolbox	IDE (Integrated Development Environment)	The version of the programming language or toolbox
MATLAB	Matlab R2024b	R2024b
Bioinformatics toolbox	Matlab R2024b	24.2
Python	JetBrains PyCharm Professional edition 2024.2.3	3.9

After testing the model, the received data is written into a spreadsheet. We decided to do all later-stage calculations using MATLAB, which has all the necessary functionalities to work with spreadsheets.

3 Data preparation for the experiment

Before implementing our model, we had to prepare data such as key variables and files containing DNA sequences. Those variables were crucial for successfully implementing program code that generates impairments. The variables contain error rates of respective impairment and substitution error rates of one nucleotide by another stored in a matrix. Such design ensured precise impairment generation and did not require complicated implementation. The files we have taken for our experiment were images encoded in DNA sequences. We have taken those files from the official repository of the system that served as a basis for our model [2]. There were four images converted to the files containing DNA sequences. Despite that, we decided to focus only on two files because both utilise different assembly methods, and we have the original results for one presented in the research paper. Those results were used as a crosscheck reference for later comparison.

Both files were text document file types. Originally, files contained a single DNA sequence on each line, which was structured to start at the beginning with notation 5' and end with 3'. However, those notations were irrelevant to us and may have caused unexpected behaviour in our model. We decided to remove them by writing a simple Python program that automatically deletes them on each line. After that, files were prepared to be used as input for our model.

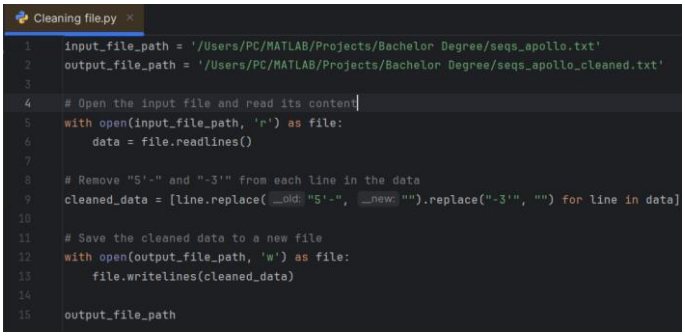


Fig 2. The example of program code to clear all notations using Python in its respective IDE.

3.1 Implementation of our model

The implementation of our model functionality was divided into three stages. In the first stage, we implemented file processing and initialised variables for our error rates of respective impairments. To ensure that the program code functioned effectively, we opened the original file (one that contains already cleared DNA sequences) in the read mode and the mutated file (one that will comprise a copy of the original file with generated impairments) in the write mode. This method reduced the resources used for reading and writing operations. As the data preparation paragraph mentioned, our error rates were converted from percentage to probabilistic form. After that, the probabilistic error rates were stored inside of variables. The unique approach was used due to the specific nature of substitution simulation. We solved this by implementing a matrix that contained error rates for all possible substitution combinations. This approach made using our data easier during the later stages of implementation.

During the second stage, we implemented an algorithm for simulating impairments based on our previously initialised variables. The algorithm consisted of two cycles. The first cycle reads the file line by line and then converts it into the

character array. A precaution was taken to avoid reading empty files by applying error handling based on conditional checks. The next step was to use our character array containing a single line (DNA sequence) in the second cycle. All impairment simulation was done in the second cycle by determining whether the impairment was an insertion, deletion, substitution, or the nucleotide (our character in an array) left unchanged.

The process of figuring out the type of impairment is based on random sampling and probability distribution. If our random number matched the defined limits of uniformly distributed range for impairments, we would generate the corresponding one. The limits revolved around impairment type, where they were determined either by a single probability of corresponding impairment or their sum. In the case of impairment called deletion, its range was determined through Eq.1 and by using only its respective error rate (in the form of probability).

$$r < [0, p_{del}] \tag{1}$$

The limits for insertion and substitution were defined by the sum of previous probability (or probabilities) and the error rate of our current impairment. For instance, the limits for insertion and substitution were determined through Eq. 2 and Eq .3

$$r < [p_{del}, p_{del} + p_{ins}] \tag{2}$$

The r character represents the value from random sampling. When that value falls within one of the defined limits, the simulation of the corresponding impairment occurs.

$$r < [p_{del} + p_{ins}, p_{del} + p_{ins} + p_{subs}] \tag{3}$$

For the third stage, we implemented the writing of the DNA sequence line with simulated impairments into the file for future analysis. The process continued until the last line of the original file was simulated. After that, we closed our files, and we could now compare those files in our testing tool.

3.2 Implementation of our testing tool

The result we received from our model was in the form of a text document file containing lines of mutated DNA sequences. To test whether our model functions properly, the creation of a testing tool was necessary. The testing tool performed basic data analysis to determine impairment types and calculated their error rates. One of the biggest challenges had been the difference in length between the original sequence and the mutated one. The length difference could have caused improper comparison between sequences, leading to false identification of mutations. To address this issue, we referred to a method called sequence alignment. Sequence alignment is a method widely used in bioinformatics as it allows scientists to compare and analyse DNA, RNA sequences, or proteins based on regions of similarity. Therefore, there exist many alignment methods, the most popular of which are Smith-Waterman alignment and Needleman-Wunsch alignment. Local alignment is represented by Smith-Waterman alignment, while Needleman-Wunsch alignment represents global. For example, the functionality of global alignment can be seen in Fig. 3.

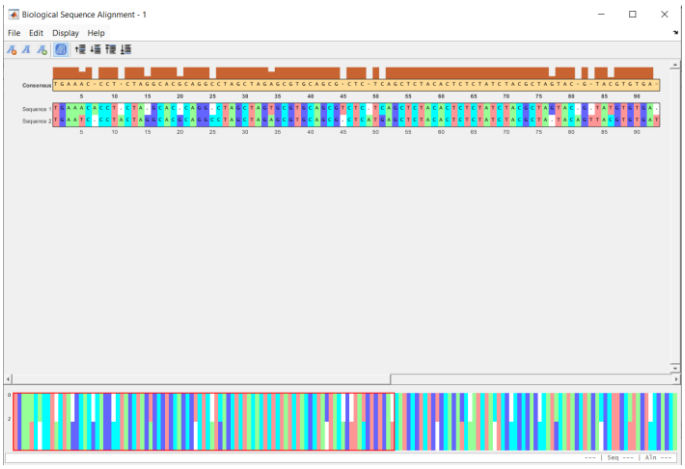


Fig 3. The graphic example of globally aligned sequences in MATLAB using Bioinformatics toolbox.

Writing complex data analysis algorithms is a complicated task, so to do it without mistakes is impossible for beginners like me. As a solution, logging the analysis results in separate files can provide valuable insights into where the problem originates. All those precautions were taken while writing program code for our data analysis tool, leading to faster work time and better precision during the process.

The implementation started with both files being opened in read mode to optimize the resources used. After that, all necessary variables and arrays were initialized. When all fundamental steps were finished, we declared a cycle as the basis for our parsing method. This method is the same as in our model, comparing files line-by-line for better precision and resource management. The part responsible for identifying mutations is designed to compare two sequences based on matching characters (nucleotide bases). Based on algorithm logic, insertion occurs when there is a '-' character (later gaps) in the original sequence compared to the mutated one. This character represents the missing or inserted nucleotide base in the original or mutated sequence. We implemented methods to avoid double-counting neighbouring insertions and grouping methods to count consecutive insertions as a single one.

Deletion is identified when gaps are placed at the character's position in the original sequence. When potential insertion or deletion is found, the counter for the corresponding error is updated, and the result is logged.

Identification of substitution required a unique approach, because of the shifted characters inside of the sequence. Due to characters shifting one position to the right or left compared to the original sequence, it potentially causes falsely identifying

all nucleotide bases as substitutions. The decision was made to use an algorithm that checks whether neighbouring nucleotide bases match the one in the original sequence. If that step succeeds, we will check for gaps near our character to eliminate the possibility of falsely identifying insertion or deletion as a substitution. The script won't confirm substitution when gaps are nearby.

At the end of the script, we saved our log data in the text document files and visualised the comparison data we received.

3.2.1 Verifying testing tool functionality

Before using our tool to compare the output file from the model with the input file, we decided to use smaller files to test whether our testing tool functions correctly. The files contained two lines of DNA sequences that were manually created to include a small number of mutations picked by hand, and the amount is logged to crosscheck results later. We put there a total of nine insertions, seven deletions, and ten substitutions.

Results were almost identical, with a minor exception of substitutions that were two mutations short. Our mutation counts were displayed only during the testing and refinement process because the exact number of mutations would have been insignificant. Logging our data provided necessary feedback because it gave insight into where the problem in the script originated. All mutations had predetermined positions, which ensured faster debugging of the code.

3.3 Results and their Statistical Validation

When we were sure that our data analysis tool functioned within the desired accuracy range, we started to test our sequences. The sequences we used were sourced from the research article that served as the foundation for our model.

The first one was an Apollo image converted to the DNA sequence, and the second one was the book 365 Dishes, which was converted too. The Apollo sequence has a size of 15 Mb, which was ten times larger than the size documented in the research article. It was the same situation for our second sequence. The amount of payload decoded in our model can't be calculated due to the simplistic nature of the simulation, which allowed us to neglect it. The number of lines of the original sequence represents the payload amount. After conducting experiments with both sequences, we received the following results for the first run.

The Apollo file had an overall error rate of 6.2425%, where the insertion rate was 2.0221%, the deletion rate was 2.5296%, and the substitution rate was 1.6908%. The 365 Dishes book sequence had an overall rate of 6.2383%, where the insertion rate was 2.0023%, the deletion rate was 2.5468%, and the substitution rate was 1.6836%. The results for other runs can be seen in Tab. 2

Tab. 2 Results from the experiment

	Overall rate (%)	Insertion rate (%)	Deletion rate (%)	Substitution rate (%)
Selected System	6.87	2.16	2.95	1.75
Run 1 Apollo	6.2425	2.0221	2.5296	1.6908
Run 1 365 Dishes	6.2383	2.0023	2.5468	1.6836
Run 2 Apollo	6.2404	2.0201	2.5293	1.6905
Run 2 365 Dishes	6.2661	2.0513	2.5168	1.6980

After receiving the results, we observed slight deviations compared to the values from the selected system. Overall, we see a tendency for the error rate to be below the systems' error rate, where all stay within range. We didn't notice significant bias between the impairments. That suggests our data analysis tool can produce relevant results, yet the deviation from the values of the selected system can be observed. During the debugging process, we were actively checking our logs for necessary insight into how to solve our problem. The algorithm could be refined further to achieve better results.

3.3.1 Statistical Validation of Accuracy

To estimate the accuracy of the model and our testing tool, we calculated accuracy per impairment, Mean Absolute Percentage Error (later MAPE), 95% Confidence interval (CI), and overall accuracy. We determined accuracy per impairment through Eq. 4.

$$a = 1 - \frac{|k - n|}{k} \tag{4}$$

In this formula character a represents accuracy, while k is the error rate of the selected system and n is our simulated value. The calculation of MAPE was required to get a percentage deviation. We calculated it using Eq.5.

$$MAPE = \frac{100}{N} \sum_{i=1}^{N} \left| \frac{k_i - n_i}{k_i} \right| \tag{5}$$

As the last step of determining accuracy, we calculated a 95% confidence interval. It was done according to the Eq.6. Where $\overline{x}$ is mean accuracy across simulations, s is the standard deviation of accuracy, n is the number of simulation runs, and Z is 1.96 (for a 95% confidence level).

$$CI = \overline{x} \pm \mathrm{Z}\frac{s}{\sqrt{n}} \quad (6)$$

To save time and obtain more accurate results, all calculations were done in MATLAB. The results are shown in Fig. 4. Based on previously defined statistical methods, we can observe individual simulation accuracy for each impairment and overall accuracy.

```
=== Accuracy Per Impairment ===
Insertion Accuracy: 93.63%
MAPE: 6.37%
95% Confidence Interval: [92.70%, 94.57%]
Deletion Accuracy: 85.51%
MAPE: 14.49%
95% Confidence Interval: [85.09%, 85.93%]
Substitution Accuracy: 96.43%
MAPE: 3.57%
95% Confidence Interval: [96.15%, 96.71%]
Overall Accuracy: 90.87%
MAPE: 9.13%
95% Confidence Interval: [90.69%, 91.05%]
================================
Overall Simulation Accuracy: 91.61%
```

Fig.4 The screenshot from MATLAB showcases the accuracy calculation results for each impairment and the systems' overall accuracy.

4 Conclusion

We verified that the model, based on principles of mathematical probabilities and statistics, could be used to simulate the selected system. Although the model accuracy was 91.61 %, it is worth mentioning that the precision of our data analysis algorithm could have influenced the results. On the other hand, our model would be helpful for educational purposes, requiring knowledge of statistics, simplistic data analysis, and data interpretation.

Acknowledgements

We also recognize our supervisor, Prof. Ing. Peter Farkaš, DrSc, who supports us by providing all the necessary resources and valuable advice. Funded by the EU NextGenerationEU through the Recovery and Resilience Plan for Slovakia under the project No. 09I05-03-V02-00051.

References

[1] https://www.nature.com/articles/s41467-019-10978-4.
[2] https://github.com/uwmisl/data-ncomms19- nanopore

Translation codes construction supported by computer

Mykola Melnyk

[1] Institute of Multimedia Information and Communication Technologies, FEI STU in Bratislava

xmelnyk@stuba.sk

Abstract – This paper presents the construction and analysis of translation codes using computational methods. We explore both mathematical and algorithmic approaches, including matrix-based encoding and the Freiman-Wyner technique, implemented primarily in Python. Performance comparisons across programming languages highlight trade-offs in speed and memory efficiency. While matrix methods offer theoretical rigor, they are computationally intensive. Experimental improvements to prefix/suffix generation show modest speed gains but remain impractical for real-time use. The results provide a foundation for efficient translation code design in digital communication systems.

1 The importance of the right programming language

Programming language is essential for modern information technology scientific field. It is understandably widely used in different fields as one of the major problem-solving tool. There are an enormous amount of programming languages to choose from, but it certainly depends on task at hand. Some languages are more suitable for certain problems, than others. Therefore, the choice of programming language sometimes decides the flow which algorithms are built, but there are factors that help to make decision easier, such as memory requirements, data amount, time complexity and etc.

In our work we aimed to achieve greater computational speeds for translation codes, while optimizing their usage on machines with any amount of computational power. Specifically, to construct our codes using from classic based method, Freiman-Wyner[1]. Since the amount of data we were going to compute, was enormous and there was needed versatility, while still achieving decent computational speeds, the Python was chosen as main language for this task.

1.1 Basics of translation codes construction

Translation codes, itself is group of EC-RLL codes that consist of different encoding and decoding information methods. TC introduce redundancy, in order to translate any digital sequence at the input of TC encoder to such output sequence, which fulfills constraints deduced from practical requirements. To construct TC codes a variety of methods can be used, therefore each of the methods have its own requirements.

2 Fundamentals of Computerized Construction

When constructing TC codes, there are two methods of computerized construction. In the most common method, mathematical, based on usage of integrated python libraries such as NumPy and SumPy, and calculating matrices while also their roots. Less known, but practically more crude is method based on string concatenation, though the method takes its roots from data science it is still applicable for codeword construction. Both approaches illustrates Fig.1.

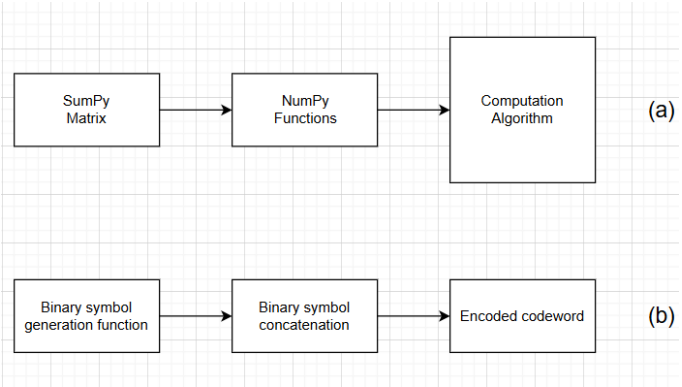


Fig. 1.Computation method: mathematical (a) and crude (b).

Both methods are technically feasible, while possessing certain pros and cons. Mathematical method require mathematical understanding of algorithms and also library usage, though it significantly improves implementation. The same is true for concatenation method, while majority of scientific data is processed with help of sorting algorithms. Therefore this algorithms can be repurposed for codeword generation. For the matrix construction method there are two existence requirements, that should be met. Such as, the sum of row should never be more or equal to 2^n, where **n** is equal to the length of encoded codeword. Code speed **R** should not be greater than Channel capacity **C.** Where **C** is equal to Eq.1 and λ_{max} is sum of roots in our matrix.

$$C = log_2(\lambda_{\max}) \quad (1)$$

The Freiman-Wyner method requires the construction of prefix and suffix, which are constructed by two different means for even or odd scenarios as seen in Eq.2

$$\frac{n}{2} \quad (2)$$

and Eq.3 for odd.

$$\frac{n \pm 1}{2} \quad (3)$$

2.1 Equipment used

Our system uses high-end Personal computer (details see Tab. 1.). Also, JetBrains PyCharm was used as IDE of choice. The Operation System was Windows 11 Pro stable build.

Tab. 1. Technical specifications of the PC

Processor	Intel Core i7-13700k
RAM	Kingston fury ddr5 4800MT/s
Graphics card	Nvidia RTX4080 16GB
Cooling System	NZXT Kraken 280
Storage	Samsung 980 pro 2tb
Operating system	Windows 11 Pro
Motherboard	Gigabyte z790 Aourus
IDE of choice:	
PyCharm	2021.2.3
CLion	2022..3.2

The IDEs were required by applying for student license[2] on JetBrains website.

3 Experimental approach and results

For the purpose of our experiment, we constructed several program iterations of different methods. Therefore, as intended we used Python to recreate matrix method with help of libraries and Freiman-Wyner technique that was realized and tested through concatenation method.

3.1 Freiman-Wyner technique

In this part we tried implementing Freiman-Wyner technique. Therefore, we tried implementing different solutions for computerized construction. The problem at hand we had was that technique itself was rather visual than mathematical. Therefore, the test program was written to test the number of existing and popular programming languages. Program itself is prototype implementation of technique with certain limitations. This script is non-adaptive, meaning the number of prefixes and suffixes cannot be changed during runtime. For the program to function correctly, prefixes and suffixes must be precomputed, and the maximum allowable number of repeated symbols must be defined in advance. The program generates codewords by appending suffixes to a given prefix until two criteria are met: a) The binary symbols in the code do not exceed the specified maximum length constraint. b) The length of the generated codeword matches the desired binary length.

This approach produces more codewords than the original technique, allowing a selection process to identify specific codewords for encoding. Codewords are chosen randomly from the set of valid results.

Then we started our testing phase with implementing technique into C++ programming language. The results are shown on Fig.2.

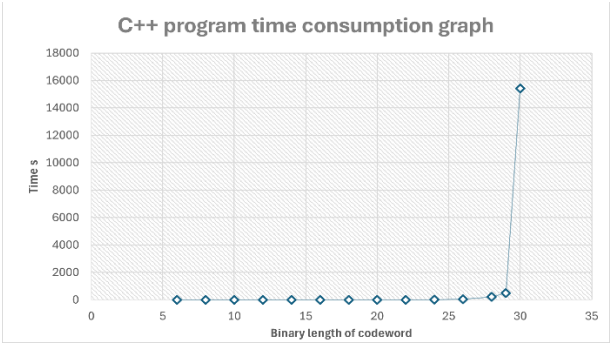


Fig. 2. C++ complete time consumption graph for binary length of 30.

As graph shows us, computation speed significantly decreases at 30 binary digits. Mostly it corresponds to large amounts of operations required and memory limitations.

Although Java's implementation is generally more straightforward due to its user-friendly syntax, the language's inherent memory limitations became apparent during the testing process Fig.3.

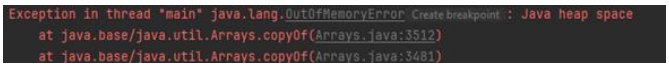


Fig. 3. Java memory issue.

The performance analysis, as illustrated in the graph seen in Fig.4., shows that the Python script was most effective among all tested implementations. Specifically, for encoding codewords of 30 binary bits, the script took approximately 2973.1284 seconds. Which is faster than C++ implementation for the same binary length.

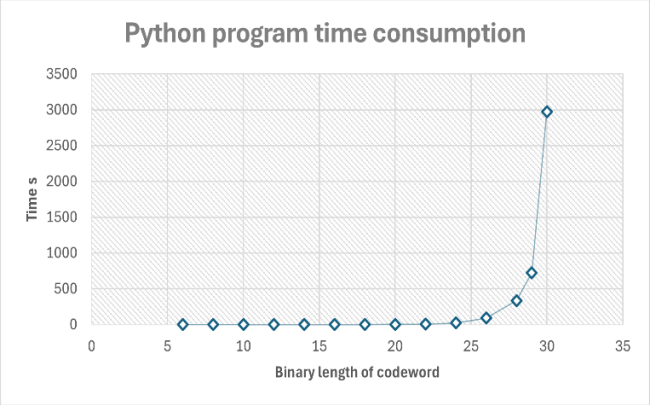


Fig.3. Speed comparison for Python script.

After the initial testing was conducted, we got to the point where this version of Freiman-Wyner technique was not seen effective enough. The peak speed results of each implemented method can be seen in the table below.

Tab. 2. Peak speed comparison table for Implemented solutions.

Programming language	Binary length(n)	Top speed
C++	29	495.48s
Python	30	2973s
Java	20	Not available

Therefore, some experimental solutions were created, which is described in next part.

3.2 Experimental approach for Freiman-Wyner

An experimental approach was developed for the Freiman-Wyner encoding technique, utilizing an alternative formula for suffix and prefix generation. Two experiments were conducted, one with division by 4 and the other with division by 8.

The first experiment, which applied division by 4, demonstrated improved computational speeds compared to the classical method that employs a prefix/suffix generation function. Specifically, the computational speed increased by 8.71% relative to the classical approach. However, despite this improvement, the method remains impractical for real-time implementation due to computational constraints (see Fig.4.).

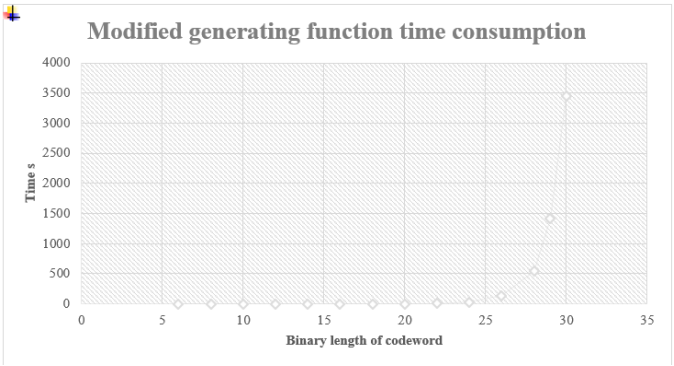


Fig.4. Prefix generation function time consumption graph division by 4.

In contrast, the second experiment, which involved division by 8, resulted in slower computation and a memory overload when generating codewords exceeding 28 binary symbols. The computational speed decreased by 11.65% compared to the first experiment and by 10.96% relative to the classical method (see Fig.5.).

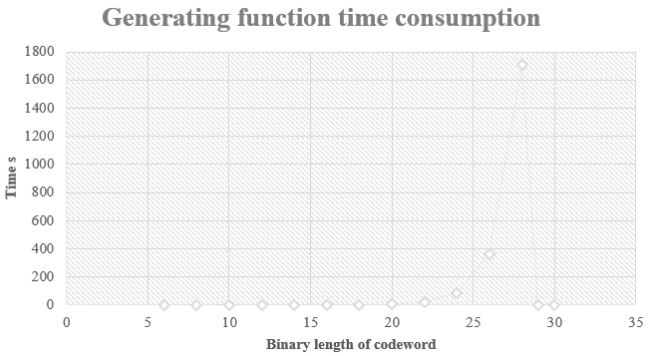


Fig.5. Prefix generation function time consumption graph division by 8.

These results indicate that increasing the division factor beyond 4 is inefficient and impractical for implementation.

3.3 Classic translation code construction (matrix method)

The matrix method for translation code construction was used as a theoretical baseline. It defines allowed symbol transitions through a state matrix and estimates channel capacity via eigenvalue analysis. While mathematically rigorous, this approach proved less practical in real-world scenarios.

Even with a modest number of states, computing higher matrix powers and solving symbolic characteristic polynomials became computationally expensive. Complex roots introduced further numerical challenges, requiring high-precision arithmetic for stable results.

Although the method accurately reflects theoretical limits, it scales poorly and is difficult to implement efficiently. As a result, it is unsuitable for real-time encoding or applications that demand speed and low resource usage. However, it remains useful as a reference point for evaluating more practical coding techniques.

4 Conclusions

We verified that it is indeed possible to create and adapt translation code construction algorithms for computer-based implementation. With access to high-performance computing systems, it becomes feasible to construct effective translation codes using matrix-based methods and symbolic tools.. The proposed method shows strong potential for educational use, prototyping, and future experiments. It also opens up possibilities for further research and advancements in the field

of telecommunications, particularly in constrained coding and capacity analysis.

Acknowledgments

We recognize also our supervisor prof. Ing. Peter Farkaš, DrSc. for support with providing all the necessary resources and valuable advice.

References

[1] Transcontrol codes Peter Farkaš
[2] https://www.jetbrains.com/community/education/#students Educational license JetBrains

Analýza nákladov ochrany prevádzky v pasívnych optických sieťach WDM-PON

Bc. Filip Fuňák[1]

[1] Ústav multimediálnych informačných a komunikačných technológií, FEI v Bratislave
xfunak@stuba.sk

Abstrakt – Siete WDM-PON patria medzi pokročilé pasívne optické siete bez aktívnych prvkov, využívajúce široké spektrum vlnových dĺžok na zvýšenie spoľahlivosti, škálovateľnosti a kapacity. Tým zabezpečujú vysokú kvalitu, rýchlosť a dostupnosť služieb pre koncových používateľov.
Z hľadiska topológie sa delia na kruhové a bod-mnohobodové siete. Ochrana v bod-mnohobodových sieťach zahŕňa typy B, C a duálne B, zatiaľ čo kruhové topológie zahŕňajú ochranu prístupových a metropolitno-prístupových sietí.
Dôležitým aspektom návrhu sietí WDM-PON sú investičné (*CAPEX*) a prevádzkové (*OPEX*) náklady, ktoré zohrávajú kľúčovú úlohu pri optimalizácii siete. Ich efektívne riadenie umožňuje dosiahnuť ekonomicky udržateľnú a výkonnú infraštruktúru. V práci sme sa zamerali na vytvorenie modelu na výpočet prevádzkových a operačných nákladov a následnú analýzu.

1 Úvod

Siete PON (Passive Optical Network) predstavujú kľúčovú technológiu v telekomunikáciách, využívajúcu topológiu P2MP (Point-to-Multipoint) v sieti ODN (Optical Distribution Network) bez aktívnych prvkov. Táto koncepcia umožňuje efektívne a ekonomicky výhodné pre riešenie FTTH (Fiber to the Home) [1].

Hlavnou nevýhodou sietí PON je útlm signálu v dôsledku absencie regenerácie, čo obmedzuje dosah. Alternatívou sú siete AON (Active Optical Network), ktoré eliminujú tento problém za cenu vyššej zložitosti a nákladov. Hybridné riešenia, ako siete LR-PON (Long Reach PON), kombinujú výhody oboch systémov [2,3].

Moderné optické prístupové siete GPON (Gigabit PON) a EPON (Ethernet PON), vrátane ich rozšírených verzií, sa čoraz viac implementujú najmä v mestských aglomeráciách [1]. Sieť WDM-PON (Wavelenght Division Multiplex) zahŕňa CO (Central Office), cyklické prvky AWG, vlákna FF(Feeder Fiber) a DF(Distribution Fiber) a jednotky ONU(Optical Network Unit). Kanály sú vlnovodĺžkovo oddelené s rozstupom 100 GHz (0,8 nm)

Kľúčovými výhodami WDM-PON sú bit-rate transparentnosť, vysoká bezpečnosť, jednoduchá správa a škálovateľnosť. Multiplexovanie viacerých optických kanálov do jedného vlákna umožňuje vytvorenie vysokokapacitných sietí, pričom kombináciou vlnového a časového delenia vznikajú hybridné WDM-TDMA PON [4].

Technológie WDM-PON sa delia podľa šírky kanálových rozostupov[1].

Technológia CWDM (Coarse Wavelength Division Multiplexing) je štandardizovaná podľa ITU-T G.694.2 a využíva široké kanály s rozostupom 20 nm v rozsahu 1271–1611 nm. Definovaných je 18 kanálov, pričom hlavnými výhodami sú jednoduchá implementácia a nižšie náklady. Technológia sa uplatňuje v metropolitných sieťach MAN (Metropolitan Area Network) s dosahom desiatok kilometrov [5].

Technológia DWDM (Dense Wavelength Division Multiplexing), štandardizovaná ITU-T G.694.1, pracuje v pásmach C, L a čiastočne S, v rozsahu 1490–1620 nm. Kanálové rozostupy sú výrazne menšie (0,1–0,8 nm), čím umožňuje vyššiu kapacitu prenosu. Frekvencie sú definované vzťahmi závislými od rozostupov 12,5–100 GHz. Združovanie DWDM podporuje aj flexibilné kanály, ktoré je možné kombinovať so štandardnými, avšak bez vzájomného rušenia [6].

2 Pasívne optické siete WDM-PON

2.1 Architektúry sietí WDM-PON

Siete WDM-PON poskytujú škálovateľnosť podporou viacerých vlnových dĺžok v jednom optickom vlákne, čím zabezpečujú transparentnosť v dátovej priepustnosti a výkonovom rozpočte bez pasívnych optických strát. Pre flexibilitu šírky pásma a počtu používateľov musia byť optické komponenty ekonomicky efektívne pre rôzne typy architektúr [7].

2.1.1 Architektúra P2MP siete WDM-PON

V tejto architektúre je každému komponentu ONU priradený samostatný vlnový kanál v oboch smeroch prenosu. Vzniká tak bod-bod spojenie medzi OLT (Optical Line Terminal) a ONU, pričom každá jednotka ONU môže pracovať na dátovej rýchlosti určenej jej vlnovým kanálom. Kanály môžu mať rôzne prenosové rýchlosti, čo umožňuje variabilitu služieb [7].

2.1.2 Kruhová (RING) architektúra prístupovej WDM-PON siete

Jednotky ONU sú prepojené v kruhovej topológii s obojsmerným pripojením na terminál OLT. Táto architektúra umožňuje re-moduláciu optických signálov pre efektívnejší prenos a podporuje flexibilnú alokáciu vlnových dĺžok, čím sa zvyšuje kapacita a škálovateľnosť siete [8,9].

2.1.3 Kruhová (RING) architektúra metropolitno-prístupovej WDM-PON siete

Nazývaná aj LR-PON, využíva kruh WDM s pasívnymi prvkami AWG a následne sa prepája na prístupovú vetvu TDM (Time Division Multiplex). Použitím zosilňovačov EDFA (Erbium-Doped Fiber Amplifier) sa môžu využívať L- a C-pásma. Modularita tejto architektúry umožňuje pripojenie veľkého počtu používateľov pri zachovaní prijateľných útlmových hodnôt. Optické výkonové parametre sa nastavujú podľa dĺžky vlákien a výkonových charakteristík laserov [10]

2.2 Architektúry ochrany prevádzky v sieťach WDM-PON

Siete WDM-PON zaznamenávajú rast napriek vysokým kapitálovým nákladom, pričom je nevyhnutné zabezpečiť dostatočnú sieťovú odolnosť. Budúce architektúry závisia od spôsobu nasadenia – zelené polia (greenfield) ako evolúcia z TDM-PON alebo hnedé polia (brownfield) ako rozšírenie HPON sietí. Pri postupnom prechode zo sietí TDM-PON na hybridné pasívne optické siete zohrávajú čoraz dôležitejšiu úlohu aj ochranné schémy prevádzky ako vidno v Tab .1.

2.3 Architektúry ochrany prevádzky P2MP WDM-PON

Ochranná schéma P2MP v sieti WDM-PON sa realizuje duplikáciou sieťových komponentov (OLT, RN, ONU) a optických vlákien podľa štandardu ITU-T pre siete TDM-PON [11]. Cieľom je zabezpečiť vysokú dostupnosť širokopásmovej siete a minimalizovať výpadky pri poruchách. Na zníženie počtu optických vlákien je potrebná agregácia medzi OLT a koncovými bodmi [12].

Tab. 1- Súhrn uvažovaných architektúr ochrán prevádzky v sieťach WDM-PON

P2MP	RING
Nechránená sieť P2MP	Nechránená prístupová sieť
Ochrana typu B	Chránená prístupová sieť
Duálna ochrana typu B	Nechránená metropolitno -prístupová sieť
Ochrana typu C	Chránená metropolitno- prístupová sieť

2.3.1 Ochrana typu B

Chráni primárne vlákno FF, ktorého zlyhanie ovplyvní veľký počet užívateľov. Je cenovo výhodná pre operátorov so stredným až nízkym zaťažením siete [13].

2.3.2 Dual-parented ochrana typu B

Duplikuje terminálu OLT a vlákno FF. Keďže sieť WDM-PON prepája viac účastníkov na jedno OLT ako TDM-PON, výpadok jednotky OLT alebo vlákna FF má väčší vplyv na spoľahlivosť siete [13].

2.3.3 Ochrana typu C

Duplikuje všetky komponenty okrem terminálu OLT, čím zabezpečuje vysokú dostupnosť. Vhodná pre vidiecke oblasti, ale vyžaduje vysoké kapitálové investície [13].

2.4 Architektúry ochrany prevádzky RING WDM-PON

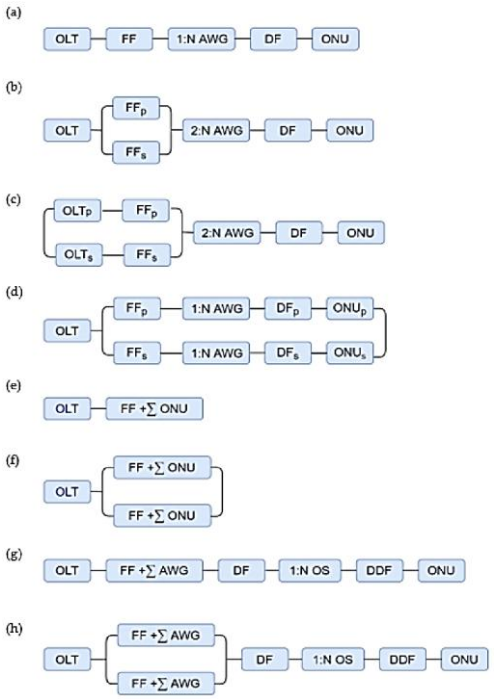
Architektúra RING ponúka efektívnu ochranu s nižšími nákladmi. Umožňuje redundanciu signálu, čím zvyšuje spoľahlivosť pripojenia [14,9]. Možné scenáre:

2.4.1 Chránená prístupová sieť

Zabezpečuje ochranu proti poruche vlákna a potlačenie šumu Rayleighovho rozptylu. Podporuje škálovateľnosť siete a vyššiu kapacitu kanála pre jednotku ONU. Pri poruche sa komunikácia obnovuje optickými prepínačmi [9]

2.4.2 Chránená metropolitno-prístupová sieť

Využíva duplikované vlákna FF a synchronizované optické prepínače. Umožňuje prenos až 1 Tbit/s, podporuje veľké geografické oblasti a vysoký počet užívateľov. Vysoká dostupnosť je zabezpečená pri nižších nákladoch oproti iným schémam [15].



Obr. 1 – Diagramy RBD (Reliability Block Diagram) (a)–(d) predstavujú scenáre ochrany prevádzky siete P2MP WDM. a) nechránená sieť P2MP, b) Ochrana typu B c) Duálna ochrana typu B d)Ochrana typu C Diagramy RBD (e)–(h) špecifikujú kruhové WDM-PON architektúry ochrany prevádzky. e) nechránená prístupová sieť f) chránená prístupová sieť g) nechránená metropolitno-prístupová sieť h) chránená metropolitno-prístupová sieť [16].

3 Analýza nákladov

Nasadenie a prevádzka sietí PON závisia od zložitosti ich architektúry. Ako vyplýva z diagramov RBD, niektoré ochranné schémy využívajú redundantné prvky, ako terminál OLT a jednotku ONU, čím sa zabraňuje potrebe ochrany **1:1** alebo **1+1**. Hoci tieto techniky zabezpečujú stabilnú obnovu prevádzky medzi terminálom OLT a jednotkou ONU, výrazne zvyšujú počiatočné investície do siete.

Duplikácia terminálu OLT predstavuje výrazne vyššie náklady oproti ochrane **1:1** alebo **1+1**, čo vyžaduje kompromis medzi počiatočnými nákladmi a dobou obnovy pri zlyhaní. Minimalizácia zložitosti systému je preto kľúčová, pre neohrozenie detekcie a obnovu siete.

Analýza zahŕňa údaje o nákladoch jednotlivých architektúr ochrany prevádzky, ktoré predstavujú významný faktor pri hodnotení hospodárnosti.Náklady v sieťach WDM-PON možno rozdeliť na *kapitálové náklady (CAPEX) a operačné náklady (OPEX)* [14].

3.1 Kapitálové náklady *Capex*

Náklady CAPEX predstavujú investície do nasadenia siete PON v prístupovej doméne. Zahŕňajú nákup sieťových zariadení (OLT, ONU, AWG, OS), náklady na prenosové médium (DF, FF) a práce spojené s inštaláciou optickej infraštruktúry, vrátane výkopových prác a zavádzania optických vlákien [14,17].

Celkové kapitálové náklady možno vyjadriť ako vidno v rovnici (1):

(1)

$$Capex = C_Z + C_M + C_P$$

Môžeme ich teda vyjadriť všeobecnou rovnicou, kde *Cz* sú náklady na zariadenia C_M náklady na optické médium a C_P náklady na práce spojené s realizáciou optickej siete ako výkopové práce, nafúkavanie optických vlákien a podobne

Rovnicu však môžeme vyjadriť aj pomocou jednotlivých diagramov RBD podľa [16] aj z Obr. 1:

(2)

$$Capex = C_{OLT} + C_{FF} + C_{AWG} + C_{DF} + C_{ONU} + C_P$$

Pri výpočte musíme uvažovať aj počet účastníkov, nakoľko to má vplyv na výsledný počet zariadení a počet vlákien DF pre každého účastníka v sieti [17].

3.1.1 Výpočet kapitálových nákladov pre nechránenú sieť

(3)

$$Capex_{P2MP\_N} = C_{OLT} + C_{FF} + C_{AWG} + C_{DFMP} + C_{ONU} + C_P$$

3.1.2 Výpočet kapitálových nákladov pre architektúry ochrany typu B

(4)

$$Capex_{P2MP\_B} = C_{OLT} + 2 * C_{FF} + C_{AWG} + C_{DFMP} + C_{ONU} + C_P$$

3.1.3 Výpočet kapitálových nákladov pre architektúry duálnej ochrany typu B

(5)

$$Capex_{P2MP\_DB} = 2 * C_{OLT} + 2 * C_{FF} + C_{AWG} + C_{DFMP} + C_{ONU} + C_P$$

3.1.4 Výpočet kapitálových nákladov pre architektúry ochrany typu C

(6)

$$Capex_{P2MP\_C} = C_{OLT} + 2 * C_{FF} + 2 * C_{AWG} + 2 * C_{DFMP} + 2 * C_{ONU} + C_P$$

3.1.5 Výpočet kapitálových nákladov pre architektúry kruhovej prístupovej nechránenej siete

(7)

$$Capex_{K\_PNECHRAN\_} = C_{OLT} + C_{FF} + \sum_{i}^{N} C_{ONU}$$

3.1.6 Výpočet kapitálových nákladov pre architektúry kruhovej prístupovej chránenej siete

(8)

$$Capex_{K\_PCHRAN} = C_{OLT} + 2 * C_{FF} + \sum_{i}^{N} C_{ONU}$$

3.1.7 Výpočet kapitálových nákladov pre architektúry kruhovej metropolitno-prístupovej nechránenej siete

(9)

$$Capex_{K\_MPNECHRAN} = C_{OLT} + C_{FF} + C_{OS} + C_{DDF} + C_{ONU} + C_P + \sum_{i}^{N} C_{ONU}$$

3.1.8 Výpočet kapitálových nákladov pre architektúry kruhovej metropolitno-prístupovej chránenej siete

(11)

$$Capex_{K\_MPCHRAN} = C_{OLT} + 2 * C_{FF} + C_{OS} + C_{DDF} + C_{ONU} + C_P + \sum_{i}^{N} C_{ONU}$$

3.2 Operačné náklady *Opex*

Náklady OPEX predstavujú neinvestičné výdavky na prevádzku siete od jej nasadenia až po výmenu za novú technológiou [14]. Zahŕňajú:

1. Opravy a výmeny zariadení, vrátane opravy optických vlákien a práce technikov [14].
2. Pokuty za výpadky služby, definované v zmluve o poskytnutí služieb SLA (Service Level Agreement) medzi operátormi a zákazníkmi [7].
3. Spotrebu elektrickej energie aktívnych prvkov siete (OLT, ONU), vypočítanú ako súčin jednotkovej ceny elektriny a celkovej spotreby siete počas jej životnosti [14,7]

(12)

$$Opex = (O_O + O_P + O_S) * počet\ rokov$$

Operačné náklady v sieťach WDM-PON teda môžeme vypočítať podľa (12), kde *Oo* sú náklady na opravy, O_P sú náklady na sankcie za nefunkčnosť služieb a O_S sú náklady vynaložené na spotrebu elektrickej energie. Všetky spomenuté náklady sú za časové obdobie jedného roka a preto ich následne treba prenásobiť počtom rokov prevádzky siete.[14,7].

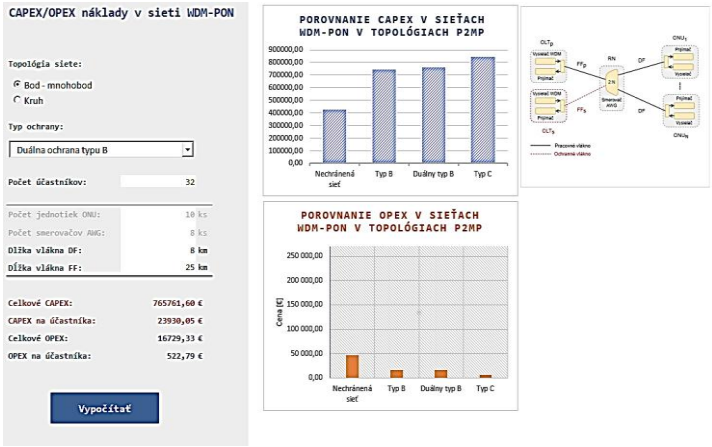
Pri výpočte operačných nákladov na opravu O_o sme vychádzali z doby dostupnosti siete podľa [16]. a následne sme vypočítali dobu nedostupnosti siete a z nej náklady na opravu s použitím špecifických parametrov použitých v Tab.2 v publikácii [16].

3.3 CAPEX/OPEX výpočtový model

Na výpočet Capex a Opex sme použili výpočtový model Obr.2 , ktorý pozostáva z modulu na výpočet týchto nákladov. Model má používateľské rozhranie pozostávajúce z 3 častí v pravej si používateľ vie vybrať z rôznych topológií P2MP a kruh v sieťach WDM-PON. Používateľ si vie pomocou tohto rozhrania navoliť o aký druh topológie a typ ochrany má záujem. Následne mu program po zadaní voliteľných parametrov a stlačení tlačidla „VYPOČÍTAŤ“ vypočíta operačné a kapitálové náklady a pre lepšiu predstavu aj náklady na jedného účastníka. V strednej časti obrazovky je porovnanie CAPEX/OPEX nákladov pri zadaných parametrov pre všetky typu ochrany v danej topológií v grafickej podobe. Tretia časť programu zobrazuje schematickú reprezentáciu vybratej topológie. Program ďalej pracuje s údajmi v Tab.2 a Tab. 3 a hodnotami v [16]. Program je teda ľahko modifikovateľný aj pre iné ceny vstupných parametrov ako nami definované.

Tab. 2- Ceny prác, komponentov a zariadení pre siete WDM-PON pre výpočet CAPEX [18]

	položka	porty	cena (€)	merná jednotka
C_{OLT}	OLT základňa		792,00	ks
	OLT prepínač vl.		509,00	ks
	OLT MODUL 10G		829,00	ks
	OLT modul 40/100G		2001,00	ks/port
	OLT prijímač		319,00	ks
	RACK komplet		800	ks
C_{FF}	FF		1400,00	km
C_{DF}	DF		300	km
C_{AWG}	AWG	4	262,80	ks
		8	657,60	ks
		16	802,60	ks
		40	2037,60	ks
		64	2700,00	ks
		96	12471,40	ks
C_{ONU}	ONU		47,00	ks
C_{OS}	OS		2059,00	ks
C_P	výkopové práce		10000,00	km
	chránička		400,00	km
	nafúkanie káblu		1000,00	km



Obr. 2 - Úvodná obrazovka výpočtového modelu CAPEX/OPEX sietí WDM-PON

3.4 Vyhodnotenie Capex a Opex v architektúrach ochrany WDM-PON

V simulácií vyhodnocujeme kapitálové a operačné náklady pre siete WDM-PON v závislosti od zvolených typov ochrany prevádzky r v topológiách bod-mnohobod a kruh. Analyzujeme siete bez ochrany aj s ochrannými mechanizmami. Zvolili sme parametre pre 32 účastníkov s dĺžkami optických vlákien: pri topológii bod-mnohobod predpokladáme vlákno FF s dĺžkou 25 km a vlákno DF 8 km, zatiaľ čo v kruhovej topológii má vlákno FF dĺžku 10 km a vlákno DF 3 km.

Ako možno vidieť v Tab. 4 a následne aj na Obr. 3 zavedením ochrany do siete sa zvýšia výrazným spôsobom kapitálové náklady na implementovanie siete ale naopak výrazným spôsobom sa zvýši dostupnosť a bezporuchovosť siete, čo má za následok výrazné zníženie operatívnych nákladov pre jednotlivé architektúry ochrany v sieti

Tab.3- Ceny prác, komponentov a zariadení pre siete WDM-PON pre výpočet OPEX [18,17,14]

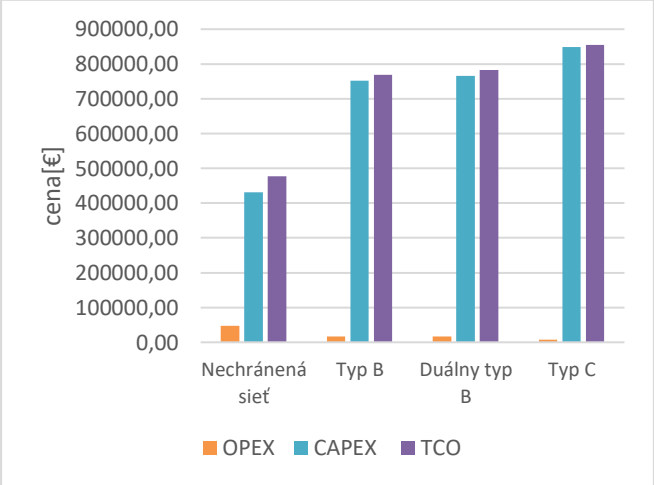
	položka		hodnota	merná jednotka
O_S	Spotreba	OLT	0,235	kWh
		ONU	0,01	kWh
	Cena elektriny		0,142	kWh
O_O	Cena opráv		1000	€/hod
O_P	Sankcie	p2mp	-	€/hod
		kruh	100	€/hod
	Počet rokov		10	rok

Tab. 4-Náklady na CAPEX/OPEX v architektúrach ochrany prevádzky WDM-PON

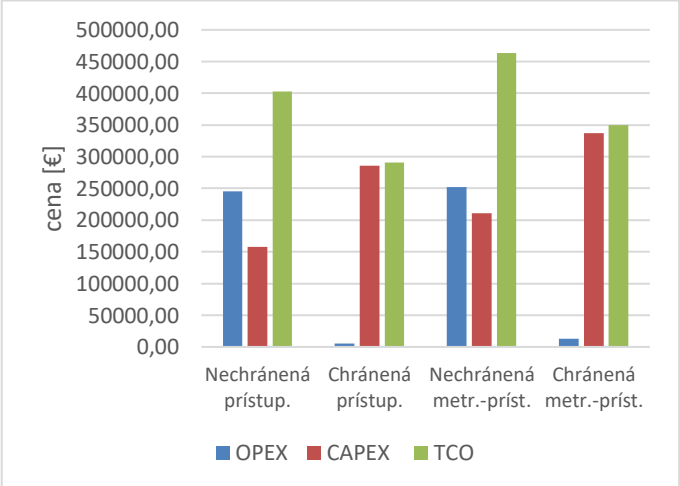
P2MP	**OPEX [€]**	**CAPEX [€]**	**TCO [€]**
Nechránená	46 723,13	430 942,60	477 665,73
Typ B	16 774,18	751 451,60	768 225,78
Duálny typ B	16 729,33	765 252,60	781 981,93
Typ C	6 948,61	848 593,20	855 541,81
Kruhová	**OPEX[€]**	**CAPEX[€]**	**TCO[€]**
Nechrán.prístupová	245 344,96	157 350,00	402 694,96
Chrán. prístupová	5 736,95	285 350,00	291 086,95
Nechrán metr.-prist	252 322,54	210 809,80	463 132,34
Chrán metr.-prist	12 714,52	336 750,80	349 465,32

Pri sieťach s topológiou kruh sa z Tab. 4 a Obr. 4 dá odčítať, že kapitálové náklady so zavedením ochrany v metropolitno-prístupovej ale aj prístupovej narastajú avšak operatívne náklady výrazne klesajú pri oboch chránených topológiách. Naopak pri nechránených topológiách sú síce počiatočné náklady nižšie, ale výrazne vyššie následné operatívne náklady. V porovnaní celkových nákladov teda chránené architektúry metropolitno-prístupovej a prístupovej siete sú nižšie ako celkové vynaložené náklady v sieťach bez nasadenej architektúry ochrany prevádzky

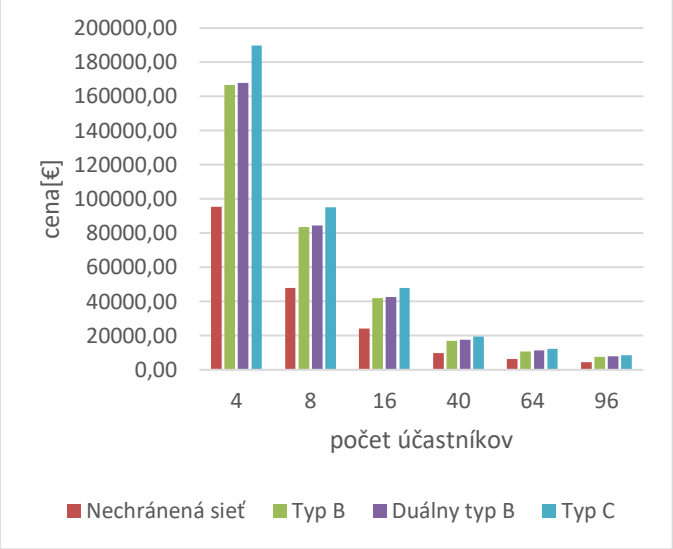
Pri výpočte kapitálových nákladov v topológií P2MP možno ako vidno na Obr. 5, že s rastúcim počtom účastníkov sa znižujú aj kapitálové náklady pri všetkých typoch architektúr ochrany prevádzky rozpočítané na jedného účastníka, z tohto dôvodu musí pri plánovaní siete brať do úvahy zhotoviteľ aj predpokladaný počet budúcich účastníkov v sieti.



Obr. 3 - Porovnanie hodnôt Capex/Opex pre jednotlivé ochrany prevádzky v sieťach P2MP WDM-PON



Obr. 4- Porovnanie hodnôt Capex/Opex pre jednotlivé ochrany prevádzky v sieťach typu KRUH WDM-PON



Obr. 5- Kapitálové výdavky na jedného účastníka v architektúrach P2MP pre jednotlivé ochrany prevádzky

4 Záver

V tejto práci sme sa zaoberali analýzou architektúr ochrany prevádzky v sieťach WDM-PON, pričom sme sa zamerali na hodnotenie kapitálových a operačných nákladov v závislosti od zvolenej topológie a typu ochrany prevádzky. Výsledky simulácií potvrdili, že implementácia ochranných mechanizmov síce zvyšuje počiatočné investičné náklady, avšak významne prispieva k zvýšeniu spoľahlivosti siete a zníženiu prevádzkových výdavkov v dlhodobom horizonte.

Porovnanie bod-mnohobodových a kruhových topológií ukázalo, že kruhové metropolitno-prístupové siete poskytujú lepší pomer medzi nákladovou efektívnosťou a dostupnosťou služieb. Ochranné architektúry v bod-mnohobodových sieťach, najmä ochrana typu C, ponúkajú vyššiu redundanciu, avšak za cenu zvýšených kapitálových nákladov. Naopak, nechránené siete vykazujú nižšie investičné náklady, avšak vyššie operačné náklady v dôsledku zvýšenej pravdepodobnosti porúch a dlhšieho času obnovy prevádzky.

Získané výsledky poukazujú na dôležitosť dôkladnej analýzy a optimalizácie ochranných mechanizmov v závislosti od špecifických požiadaviek siete a počtu koncových účastníkov. Pri návrhu a nasadzovaní sietí WDM-PON je potrebné brať do úvahy nielen technické parametre a výkonové charakteristiky, ale aj ekonomické faktory ovplyvňujúce celkové náklady na výstavbu a prevádzku siete. Práca zároveň poskytuje základ pre ďalší výskum v oblasti ekonomicky efektívneho návrhu optických sietí a možností optimalizácie ochranných mechanizmov v kontexte budúcich telekomunikačných požiadaviek.

Literatúra

[1] FILKA, M. Optické sítě. Elektr. skripta. VUT FEKT, Brno 2012.
[2] LAFATA, P.-VODRÁŽKA, J. Optické přístupové sítě a přípojky FTTx. Praha: Česká technika - nakladatelství ČVUT, 2014. 294 s. ISBN 978-80-01-05463-5.
[3] BANERJEE, A.-PARK, Y.- CLARKE, F.- SONG, H.- YANG, S.- KRAMER, G.- KIM, K.- MUKHERIEE, B. Wavelength-division-multiplexed passive optical network technologies for broadband access: A review. J. Opt. Netw. 2005, 4, 737–758.
[4] EL-HALABI, Eng.Mahmoud. (2015). Wavelength Division Multiplexing Passive Optical Network (WDM-PON) technologies for future access networks
[5] ITU-T SERIES G.694.2, Spectral grids for WDM applications: CWDM wavelength grid, 12/2003
[6] ITU-T Series G.694.1, Spectral grids for WDM applications: DWDM frequency grid, 10/2020
[7] CHEN, J.- WOSINSKA, L.- MAS M.- CARMEN-JAEGER, M. (2010). Cost vs. Reliability Performance Study of Fiber Access Network Architectures. Communications Magazine, IEEE. 48. 56 - 65. 10.1109/MCOM.2010.5402664.
[8] RAMANTAS, K.-VLACHOS,K.- ELLINAS, G.; HADJIANTONIS, A. Efficient resource management via dynamic bandwidth sharing in aWDM-PON ring-based architecture. In Proceedings of the 14th International Conference on Transparent Optical Networks, Coventry, UK, 2–5 July 2012; pp. 1–4.
[9] YEH, C.- CHOW, C.- HUANG, S.- SUNG, J.- LIU, Y.- PAN, C. Ring-based WDM access network providing both Rayleigh backscattering noise mitigation and fiber-fault protection. J. Lightwave Technol. 2012, 30, 3211–3218.
[10] RÓKA, R. Hybrid PON Networks—Features, Architectures and Configuration; LAP Lambert Academic Publishing: Saarbrücken, Germany, 2015; ISBN 978-3-659-43686-4.
[11] ITU-T. Series G—Supplement 51: Passive Optical Network Protection Considerations; International Telecommunication Union: Geneva, Switzerland, 2016.
[12] LARSEN, C.P.- GAVLER, A.- WANG, K. Comparison of active and passive optical access networks. In Proceedings of the 9th Conference on Telecommunication, Media and Internet, Ghent, Belgium, 7–9 June 2010; pp. 1–5.
[13] MACUCHA, C.M.-WOSINSKA, L.; Chen, J. Assessment methodology of protection schemes for next generation optical access networks. Opt. Fiber Technol. 2015, 26, 82–93.
[14] IMTIAZ, W.- IQBAL, J.- QAMAR, A.- ALI, H.- IDRUS. Impact of Fiber Duplication on Protection Architectures Feasibility for Passive Optical Networks, Sevia,.2017
[15] ESMAIL, M.- FATHALLAH, H. Optical coding for next-generation survivable long-reach passive optical networks. J. Opt. Commun. Netw. 2012, 4, 1062–1074.
[16] RÓKA, R. Performance Analysis of Wavelength Division Multiplexing-Based Passive Optical Network Protection Schemes by Means of the Network Availability Evaluator 2022
[17] DAVIM, J.P.-PINTO,A.P. "CAPEX Model for PON Technology," 2010 2nd International Conference on Evolving Internet, Valencia, Spain, 2010, pp. 192-196, doi: 10.1109/INTERNET.2010.41.
[18] FS EUROPE [online], © 2009-2023 Neufahrn, Germany [cit. 2023-1-10]. Dostupné na: < https://www.fs.com/de-en >

Optimizing signal transmission through cables using machine learning

Valerii Cherniatiev

Institute of Multimedia Information and Communication Technologies, FEI STU in Bratislava

Institute of Computer Science and Mathematics – Applied Informatics, FEI STU in Bratislava

xcherniatiev@stuba.sk

Abstract – This paper explores signal restoration using convolutional neural networks (CNNs) in wireless communication models. It examines the Shannon-Weaver model with noise, signal attenuation, multipath propagation, and wireless fading. A CNN is trained in MATLAB to reconstruct distorted signals, achieving high accuracy. The study highlights CNNs' efficiency in mitigating signal degradation, enhancing communication reliability.

1 Basics of Communication models

1.1 The Shannon-Weaver model with noise.

The Shannon-Weaver model with noise is a foundational framework in communication theory that describes how information is transmitted through a channel while being subjected to external interference, referred to as "noise." Noise introduces random disturbances that degrade the quality of the transmitted signal, reducing the signal-to-noise ratio (SNR).

The SNR quantifies the relative strength of the signal compared to the noise and is critical in determining the reliability of communication. According to Shannon's channel capacity theorem, the maximum data rate *C* (in bits per second) that can be transmitted over a noisy channel is given by:

$$C = B \cdot \log_2(1 + SNR) \quad (1)$$

where B is the bandwidth of the channel (in Hz), and SNR is the signal-to-noise ratio.

This equation establishes the theoretical upper limit on the rate of error-free data transmission for a given channel. As noise increases, the SNR decreases, thereby reducing the channel capacity. Conversely, increasing the bandwidth or improving the SNR enhances the maximum achievable data rate.

This model underscores the importance of designing systems that mitigate noise effects, such as error-correcting codes and signal amplification techniques, to approach the theoretical channel capacity. The Shannon-Weaver model remains pivotal in modern communication systems, including wireless networks and satellite communications.

1.2 Signal Attenuation

Signal attenuation is a critical phenomenon in communication systems, describing the exponential reduction in signal power as it propagates through a medium over a distance. This degradation is primarily caused by absorption, scattering, and other losses inherent to the transmission medium. The received power P_r is related to the transmitted power *Pt* by the equation:

$$P_r = P_t \cdot e^{-\alpha d} \quad (2)$$

where α is the attenuation coefficient (measured in dB per unit distance), and d is the propagation distance.

The attenuation coefficient depends on the properties of the medium, such as its conductivity and frequency-dependent behavior. As d increases, P_r decreases exponentially, highlighting the importance of amplification or repeaters in long-distance communication systems. Signal attenuation imposes fundamental limits on the range and quality of communication links, necessitating careful system design to compensate for power loss

1.3 Multipath Propagation

Multipath propagation arises when a transmitted signal reaches the receiver via multiple paths due to reflections, diffractions, or scattering caused by obstacles in the environment.

These multiple components interfere at the receiver, leading to constructive or destructive interference, commonly referred to as fading. The total amplitude of the received signal is given by:

$$A_{total} = \sum A_i \cdot e^{j\phi_i} \quad (3)$$

where A_i and ϕ_i represent the amplitude and phase of each path component, respectively.

Constructive interference occurs when the phases align, enhancing the signal strength, while destructive interference results in signal cancellation, reducing the overall amplitude. Multipath effects are particularly pronounced in wireless environments such as urban areas or indoor settings, where signals encounter numerous reflective surfaces.

1.4 Wireless Fading Channel Model

The Wireless Fading Channel Model characterizes a communication channel where the transmitted signal undergoes various impairments, including noise, attenuation, and multipath propagation.

These effects cause fluctuations in the received signal's amplitude, phase, and power, collectively referred to as fading. The combined effect of all distortions can be mathematically expressed as:

$$r(t) = \sum_{i=1}^{N} \alpha_i \cdot s(t - \tau_i) \cdot e^{j\phi_i} + n(t) \quad (4)$$

where α_i, τ_i, and ϕ_i represent the attenuation, delay, and phase shift of the *i*-th path, respectively; *s*(*t*) is the transmitted signal; and *n*(*t*) denotes additive noise.

Fading channels are categorized into two types: large-scale fading (e.g., path loss and shadowing) and small-scale fading (e.g., rapid fluctuations due to multipath). Understanding and modeling these effects are crucial for designing robust wireless systems capable of maintaining reliable communication under varying conditions.

2 Convolutional Neural Networks (CNNs)

Convolutional Neural Networks (CNNs) are specialized deep learning models primarily used for image recognition and classification. CNNs operate by applying convolutional filters to input data, extracting spatial hierarchies of features through multiple layers. The core operation is the convolution, defined as:

$$(S(i,j)) = (I \cdot K)(i,j = m \sum n \sum I(i+m, j+n) K(m,n) \quad (5)$$

where *I* is the input, *K* is the kernel (filter), and *S* is the resulting feature map.

CNNs consist of convolutional layers, pooling layers, and fully connected layers. Convolutional layers detect local features using learnable kernels. Pooling layers (e.g., max pooling) down sample feature maps to reduce dimensionality, preserving important features while reducing computational cost. Fully connected layers then integrate extracted features for classification.

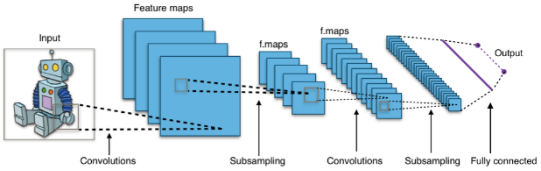


Image 1. CNN Scheme.

Key advantages of CNNs include parameter sharing (same kernel across the input) and spatial invariance (robustness to translations). Activation functions like ReLU ($f(x)=\max(0,x)$) introduce non-linearity, enabling complex pattern recognition.

CNNs minimize loss (e.g., cross-entropy) via backpropagation, updating weights using gradient descent:

$$r(t) = w := w - \eta \frac{\partial L}{\partial w} \quad (6)$$

where *w* is the weight, η is the learning rate, and *L* is the loss function.

3 Algorithm for optimization

The provided algorithm demonstrates a neural network-based approach to classify points in a 2D plane using a simple feedforward architecture. The core concept involves generating synthetic data, training a neural network on this data, and evaluating its performance on unseen test cases.

First, random points are generated within a defined range, and their labels (0 or 1) are determined based on their position relative to the line (y = x). This creates a binary classification problem where the model learns to distinguish between two classes. The neural network architecture consists of an input layer, a hidden layer with ReLU activation, and an output layer with sigmoid activation, suitable for binary classification tasks.

During training, the Adam optimizer minimizes the loss function, enabling the network to learn the decision boundary. The trained model is then tested on predefined points to predict their class labels. The key idea is to evaluate the network's ability to generalize and accurately classify new data points.

This implementation showcases the fundamental process of supervised learning: preparing labeled data, designing a neural network, training it, and assessing its

performance. Such an approach can be extended to more complex datasets and tasks, making it a foundational example of machine learning applications.

3 Practical realization in MATLAB

Firsly, we will simulate the signal for each of our connectivity models.

Then MATLAB code aims to restore distorted signals using a convolutional neural network (CNN). The process begins by iterating through a collection of distorted signals. For each signal, the original and distorted versions are normalized to ensure consistent scaling, which is critical for effective training. Normalization is achieved by dividing the signal values by their maximum absolute value. Once normalized, the distorted signal serves as input data, while the normalized original signal acts as the target output for supervised learning.

A 1D CNN is constructed using MATLAB's Deep Learning Toolbox. The network architecture includes a “sequenceInputLayer” for handling time-series data, a “convolution1dLayer” to extract features via convolutional filters, a “reluLayer” for introducing non-linearity, a “fullyConnectedLayer” to map features to outputs, and a “regressionLayer” for predicting continuous values. Training options are configured using the “trainingOptions” function, specifying the Adam optimizer, a maximum of 200 epochs, a mini-batch size of 32, and an initial learning rate of 0.001. Training progress is visualized using plots.

The “trainNetwork” function trains the CNN, after which the “predict” function reconstructs the signal from the distorted input. The reconstructed signal is denormalized to its original scale. The difference between the original and restored signals is computed, and the Mean Squared Error (MSE) quantifies restoration accuracy. This workflow leverages MATLAB’s Deep Learning Toolbox for neural network creation and training, alongside basic signal processing tools for normalization and error calculation. After finished learning, process we got next results:

- For Shannon-Weaver model with noise:

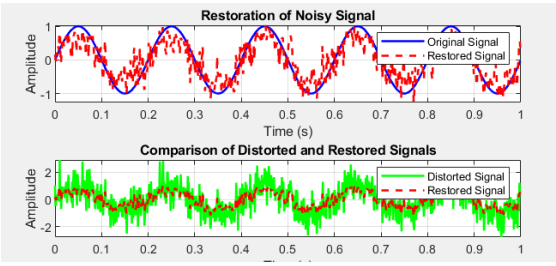


Image 2. Results for Shannon-Weaver model with noise

- For Signal Attenuation:

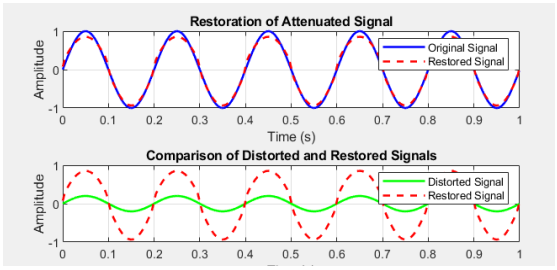


Image 3. Results for Signal Attenuation

- Multipath Propagation:

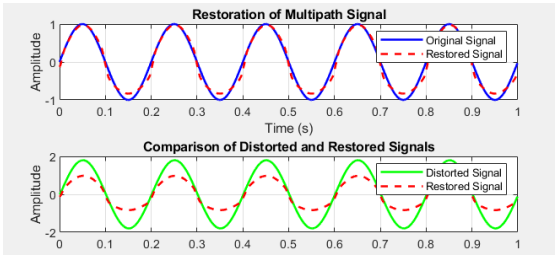


Image 4. Results for Multipath Propagation

- Wireless Fading Channel Model:

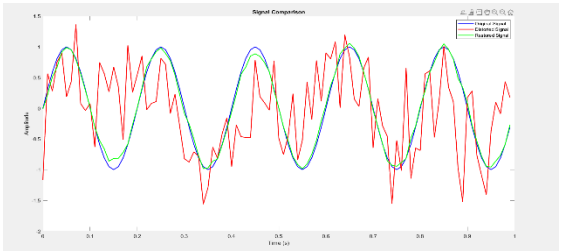


Image 5. Results for Wireless Fading Channel Model

From results we can see that Neuron network of type CNN makes excellent work in predicting expected result.

Learning process looks like:

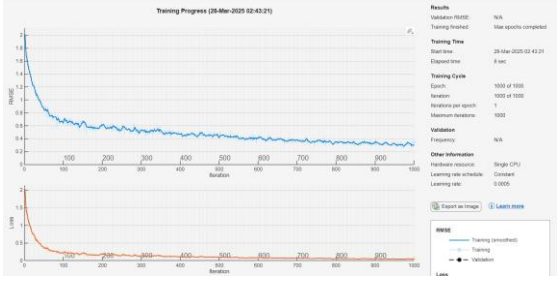


Image 6. Machine Learning process

The above demonstrates that the number of iterations required for training the neural network amounts to 1000. By employing a more sophisticated machine learning model and utilizing computational hardware superior to a standard home-use laptop, near-perfect results can be achieved. A significant advantage of this concept lies in its remarkable flexibility and ease of use. It is sufficient merely to input the distorted signal into the trained neural network model, and it will produce an output close to perfection.

5 Conclusion

We analyzed the performance of neural networks with distorted signals. The main goal was to restore the distorted signal using neural networks. The study considered three fundamental types of signal distortion: noise, attenuation, and multipath propagation.

As demonstrated in the practical part, even a simple neural network model—limited in both training time and computational power—performed exceptionally well not only in handling each type of distortion individually but also in dealing with their combined effects. Of course, this study does not cover all possible details and nuances that contribute to signal distortion. However, with a more advanced machine learning model, I am confident that the results could be further improved.

Applying this concept to real-world signal distortion problems could significantly enhance data transmission reliability and eliminate numerous challenges.

5 References and sources

[1] Pattern recognition and machine learning. Bishop, Christopher M.

[2] https://colah.github.io/posts/2015-08-Understanding-LSTMs/?spm=a2ty_o01.29997173.0.0.2177c9217xkt3M

[3] Goodfellow, I., Bengio, Y., &Courville, A. (2016). Deep Learning . MIT Press.

[4] https://commons.wikimedia.org/wiki/File:Typical_cnn.png

[5] Wireless Communications: Principles and Practice by Theodore S. Rappaport

Fotogrametria

Anastasiia Bondarenko, Andrii Smetaniuk
[1] Institute of Information and Communication Technologies, FEI STU v Bratislave

xbondarenkoa@stuba.sk xsmetaniuk@stuba.sk

Abstrakt – Táto práca sa zaoberá fotogrametriou ako modernou metódou na tvorbu 3D modelov objektov na základe fotografií z rôznych uhlov. V teoretickej časti sú vysvetlené princípy fungovania fotogrametrie, jej delenie a používaný softvér. V praktickej časti sme vytvorili 3D model miestnosti pomocou fotografií, ktoré sme spracovali v programe PostShot. Model bol následne exportovaný do Unreal Engine 5.5.4 a vizualizovaný vo virtuálnej realite cez okuliare Meta Quest 3. Cieľom práce bolo overiť presnosť fotogrametrie a demonštrovať jej využitie v prostredí VR.

1 Definícia fotogrametrie

Fotogrametria patrí medzi progresívne technológie, ktoré umožňujú rekonštrukciu trojrozmerných objektov len na základe dvojrozmerných fotografií. Vďaka rozvoju výpočtovej techniky a dostupnosti špecializovaného softvéru sa tento prístup stal široko využívaným nielen v geodézii a archeológii, ale aj v oblasti digitálnych rekonštrukcií a virtuálnej reality.

Cieľom tejto práce je ukázať celý proces vytvárania 3D modelu reálnej miestnosti pomocou fotogrametrie – od samotného fotografovania, cez spracovanie dát v softvéri PostShot, až po import a interaktívnu vizualizáciu vo VR prostredí pomocou Unreal Engine 5.5.4 a okuliarov Meta Quest 3.

1.1 Typy fotogrametrie podľa spôsobu snímania

Fotogrametria sa delí na viaceré hlavné typy podľa polohy kamery voči snímanému objektu. Každý typ má svoje špecifiká, výhody a oblasti využitia. Správna voľba typu fotogrametrie závisí od charakteru objektu, rozsahu snímania a požadovanej presnosti modelu.

Pozemná fotogrametria sa vykonáva zo zeme – buď ručne, alebo pomocou statívu. Kamera je umiestnená na úrovni objektu alebo mierne nad či pod ním. Tento typ sa často využíva na dokumentáciu architektonických prvkov, interiérov, fasád, ako aj v oblasti kriminalistiky a ochrany kultúrneho dedičstva. Je mimoriadne vhodná pri detailnej rekonštrukcii menších objektov zblízka.

Letecká fotogrametria sa realizuje pomocou dronov, lietadiel alebo vrtuľníkov a umožňuje snímať rozsiahle oblasti z výšky. Využíva sa najmä na tvorbu ortofotomáp, topografických podkladov a digitálnych modelov terénu. Vďaka dostupnosti bezpilotných systémov našla uplatnenie aj v menších projektoch v stavebníctve a poľnohospodárstve.

Satelitná fotogrametria je založená na snímkach získaných z družíc obiehajúcich okolo Zeme. Používa sa predovšetkým pri veľkoplošnom mapovaní, monitoringu krajiny, environmentálnych analýzach či geologickom prieskume. Hoci jej rozlíšenie býva nižšie ako pri pozemnej alebo leteckej fotogrametrii, poskytuje dôležité informácie v globálnom meradle.

1.2 Základné princípy fotogrametrie

Fotogrametria je metóda rekonštrukcie trojrozmernej štruktúry objektov na základe analýzy dvojrozmerných snímok, ktoré boli vytvorené z rôznych uhlov pohľadu. Vďaka geometrickým zákonom a princípu triangulácie je možné presne určiť priestorovú polohu jednotlivých bodov na povrchu objektu.

Celý proces pozostáva z viacerých krokov, ktoré na seba nadväzujú. Najprv je potrebné zhotoviť sériu fotografií z rôznych perspektív s výrazným prekrytím (zvyčajne viac ako 60 %), pričom zábery by mali pokrývať objekt zo všetkých strán – zhora, zboku aj pod uhlom.

Softvér následne automaticky identifikuje kľúčové prvky na snímkach, ako sú rohy, textúry či kontrastné detaily, a porovnáva ich medzi jednotlivými zábermi, čím určuje zhodné body. Na základe týchto bodov a známej polohy kamier sa pomocou triangulácie vypočítavajú priestorové súradnice, ktoré tvoria mračno bodov reprezentujúce tvar objektu.

Z tohto mračna sa následne vygeneruje geometrický model – teda sieť trojuholníkov alebo štvoruholníkov, ktoré spájajú identifikované body do jednej súvislej štruktúry. Poslednou fázou je nanesenie textúr, teda farebných informácií extrahovaných z fotografií, vďaka čomu výsledný 3D model pôsobí realisticky a vizuálne verne zachytáva skutočný objekt.

2 Použité vybavenie a softvér pre fotogrametriu

Na realizáciu fotogrametrie je potrebná kombinácia vhodnej techniky, softvéru a výpočtových prostriedkov. Kvalita výsledného 3D modelu do veľkej miery závisí od technického vybavenia a podmienok snímania. V rámci nášho projektu sme sa snažili využiť dostupné zariadenia a nástroje, ktoré nám umožnili prejsť celým procesom od fotografovania až po vizualizáciu vo virtuálnej realite.

Na zhotovenie kvalitných snímok sme použili digitálnu zrkadlovku Canon EOS 600D s objektívom Canon EF-S 18–55 mm f/3.5–5.6 DC III. Táto kombinácia poskytla dostatočné rozlíšenie a optickú kvalitu pre účely presnej rekonštrukcie. Aj

keď by použitie statívu prispelo k väčšej stabilite a ostrosti záberov, nemali sme ho k dispozícii, a preto sme sa snažili kameru pri snímaní držať čo najpevnejšie.

Rovnomerné osvetlenie je kľúčovým faktorom pri zachovaní konzistentnosti textúr. Ideálne je nasvietiť objekt z viacerých strán, aby sa predišlo tvrdým tieňom a prepalom. V našom prípade sme boli odkázaní na prirodzené svetlo, čo síce nebolo ideálne, ale stále umožnilo vytvoriť použiteľný vizuálny výstup.

Spracovanie veľkého množstva obrazových dát a generovanie 3D modelu si vyžaduje výkonný počítač. Použili sme notebook so samostatnou grafickou kartou, ktorá zvládla výpočtovo náročné úlohy ako generovanie mračna bodov a textúrovanie.

Z hľadiska softvérového vybavenia sme pracovali s programom Jawset PostShot, ktorý slúžil na spracovanie fotografií a tvorbu samotného 3D modelu. Pre jeho vizualizáciu vo virtuálnom priestore sme použili Unreal Engine 5.5.4, ktorý poskytuje nástroje na import, editáciu scény a interakciu vo VR prostredí.

3 Praktická realizácia

3.1 Spracovanie v programe PostShot

Po fotografovaní sme použili softvér Jawset PostShot, ktorý automaticky generuje 3D model na základe série snímok. V prvom kroku sme do prostredia načítali všetky fotografie. Program následne automaticky určil polohu kamery a vypočítal, z akých pozícií a uhlov boli zábery zhotovené. Potom identifikoval kľúčové body na jednotlivých snímkach a porovnal ich medzi sebou, čím vytvoril riedke mračno bodov. Tento základný model bol ďalej spracovaný do hustého mračna bodov, ktoré lepšie reprezentovalo detaily objektu. Zo získaných bodov sa následne vygenerovala geometrická sieť tvorená trojuholníkmi. Na záver boli na tento 3D model aplikované textúry – teda farby a povrchové detaily extrahované priamo z pôvodných fotografií.

Fotogrametria začína kvalitným fotografickým snímaním objektu. V našom prípade sme vytvorili sériu fotografií reálnej miestnosti z rôznych uhlov, pričom sme dodržali niekoľko základných pravidiel:

Aby sme dosiahli čo najvyššiu presnosť rekonštrukcie, vyhotovili sme viac ako 400 fotografií, ktoré zabezpečili dostatočné pokrytie a detailnosť pre následné spracovanie v softvéri.



Obr. 1. Fotografiía miestnosti príklad 1.



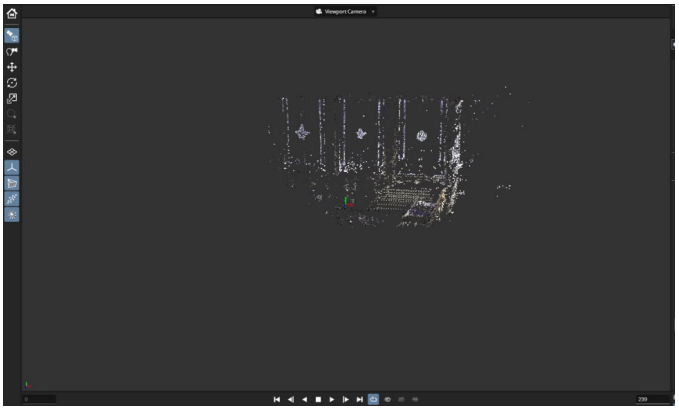
Obr. 2. Fotografiía miestnosti príklad 2.



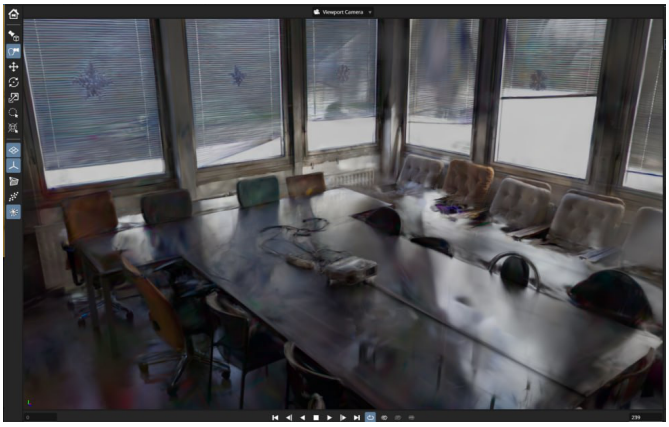
Obr. 3. Fotografiía miestnosti príklad 3.

3.2 Spracovanie v programe PostShot

Po fotografovaní sme použili softvér Jawset PostShot, ktorý automaticky generuje 3D model na základe série snímok. V prvom kroku sme do prostredia načítali všetky fotografie. Program následne automaticky určil polohu kamery a vypočítal, z akých pozícií a uhlov boli zábery zhotovené. Potom identifikoval kľúčové body na jednotlivých snímkach a porovnal ich medzi sebou, čím vytvoril riedke mračno bodov. Tento základný model bol ďalej spracovaný do hustého mračna bodov, ktoré lepšie reprezentovalo detaily objektu. Zo získaných bodov sa následne vygenerovala geometrická sieť tvorená trojuholníkmi. Na záver boli na tento 3D model aplikované textúry – teda farby a povrchové detaily extrahované priamo z pôvodných fotografií.



Obr. 4. Vytvorenie mračno bodov.



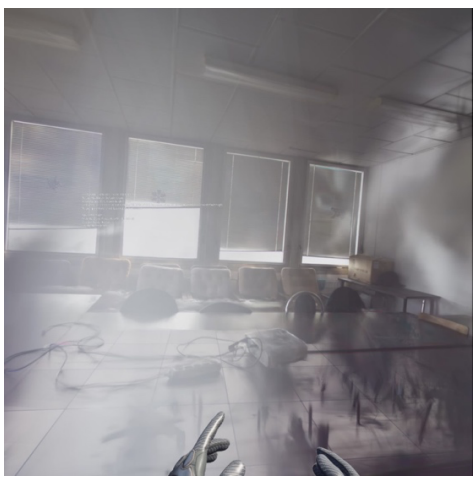
Obr. 5. Textúrovaný model.

3.3 Import do Unreal Engine a práca s VR

Po vytvorení 3D modelu sme ho pripravili na použitie vo virtuálnej realite. Model sme najskôr exportovali z programu PostShot vo formáte .ply a následne sme v prostredí Unreal Engine 5.5.4 vytvorili novú prázdnu scénu. Do tohto prostredia sme nainštalovali špeciálny plugin Jawset pre Unreal Engine, ktorý nám umožnil priamy import modelu. Po importe sme model vložili do scény a umiestnili ho na požadovanú pozíciu. Na zabezpečenie interakcie s modelom vo VR sme na notebooku vytvorili lokálny Wi-Fi hotspot a okuliare Meta Quest 3 sme pripojili k tejto sieti, čím sme umožnili bezdrôtové prepojenie so scénou.



Obr. 6. Model vo VR



Obr. 7. Pohľad cez Meta Quest 3.

4 Záver

Cieľom našej práce bolo prakticky overiť možnosti fotogrametrie pri tvorbe 3D modelu reálneho priestoru. V rámci projektu sme použili digitálnu zrkadlovku Canon EOS 600D s objektívom Canon EF-S 18–55 mm f/3.5–5.6 DC III na zhotovenie série fotografií miestnosti, ktoré sme následne spracovali v programe PostShot. Výsledný model bol importovaný do prostredia Unreal Engine a vizualizovaný vo virtuálnej realite cez okuliare Meta Quest 3.

Aj keď sa nám podarilo vytvoriť celý pracovný postup – od snímania až po interaktívne zobrazenie modelu vo VR – kvalita finálneho modelu nebola úplne ideálna. Dôvodom bol nedostatok rovnomerného osvetlenia a absencia statívu, čo spôsobilo mierne rozmazanie niektorých záberov a zníženie presnosti rekonštrukcie. Napriek týmto obmedzeniam projekt preukázal, že aj s bežne dostupným vybavením možno úspešne aplikovať fotogrametriu v praxi.

Literatúra

[1] Jawset Visual Computing. (2024). *PostShot User Manual & Documentation*. Dostupné na: https://www.jawset.com/

[2] Unreal Engine Documentation. (2024). *Using Virtual Reality with Unreal Engine 5*. Dostupné na: https://docs.unrealengine.com/

[3] Meta Quest Developer Documentation. (2024). *Streaming and Development Tools*. Dostupné na: https://developer.oculus.com/

[4] Agisoft. (2023). *Photogrammetric processing of digital images in Agisoft Metashape*. Dostupné na: https://www.agisoft.com/

[5] Wikipedia contributors. (2024). *Photogrammetry*. In Wikipedia, The Free Encyclopedia. Dostupné na: https://en.wikipedia.org/wiki/Photogrammetry

Virtuálne laboratórium pre optické vláknové medium

Kateryna Vynnyk

Ústav multimediálnych informačných a komunikačných technológií, FEI STU v Bratislave

xvynnykk@stuba.sk

Abstrakt - Cieľom tohto príspevku je prispieť k vývoju virtuálneho laboratória, ktoré umožňuje demonštrovať princípy a fungovanie metód merania prenosových parametrov technológií optických vlákien. Práca sa zameriava na charakteristiku základných princípov merania, analýzu vhodnosti softvérovej platformy Moodle a samotnú realizáciu virtuálneho laboratória s vybranými meracími metódami. Výsledkom je interaktívne prostredie, ktoré spája teoretické koncepty s praktickými návodmi, čím sa zvyšuje efektívnosť vzdelávania a odbornej prípravy v oblasti optických komunikácií.

1. Úvod

Optické vláknové technológie zohrávajú kľúčovú úlohu v moderných telekomunikačných systémoch, a preto predstavujú dôležitú tému na štúdium na univerzitách. Efektívne meranie ich prenosových parametrov je kľúčové pre zabezpečenie kvality prenosu a optimalizáciu infraštruktúry. Pri realizácii praktických cvičení so zariadeniami na meranie prenosových parametrov optického prostredia môžu mať študenti problémy s ich používaním, keďže v reálnych podmienkach nemali všetci príležitosť pracovať s týmito prístrojmi. Pochopenie ich fungovania môže preto vyžadovať značné množstvo času. Virtuálne laboratórium umožní urýchliť proces osvojenia si práce s týmito zariadeniami prostredníctvom interaktívnych výukových videí a zároveň prehĺbi vedomosti študentov tým, že demonštruje meranie prenosových parametrov v reálnych podmienkach. Táto práca sa zaoberá návrhom a implementáciou takéhoto laboratória.

2. Prostredie optického vlákna

Optické vlákna sa delia na mnohomódové step-index, mnohomódové gradient-index a jednomódové step-index. Jednomódové vlákna majú menší priemer jadra a sú vhodné na dlhé vzdialenosti s nižším tlmením, pričom jeden mód sa šíri len pri normovanej frekvencii menšej ako 2,405. Mnohomódové vlákna majú väčší priemer jadra, vyššie tlmenie a sú vhodné na kratšie vzdialenosti [1] [2].

2.1 Tlmenie

Tlmenie je spôsobené rozptylom, absorpciou a ohybom. Rayleighov rozptyl je neodstrániteľný, absorpcia závisí od vlnovej dĺžky (infračervená absorpcia spôsobuje straty nad 1600 nm) a hydroxylové ióny (OH^-) prispievajú k tlmeniu pri 950, 1240 a 1380 nm. Ohybové straty sú výraznejšie pri vyšších vlnových dĺžkach [1].

2.2 Disperzia vlákna

Disperzia sa delí na chromatickú (CD) a polarizačnú módovú (PMD) [1]. CD rozširuje impulzy a obmedzuje kapacitu prenosu, pričom ju možno kompenzovať pomocou DCF alebo EDC. PMD ovplyvňuje kvalitu signálu kvôli rôznym indexom lomu polarizačných módov. Na redukciu polarizačnej disperzie sa využívajú Braggove mriežky, autoenkodéry a hustotná matica [3] [4] [5].

2.3 Optické spoje a optické konektory

Optické konektory a spoje sú v optickom prostredí nevyhnutné na zabezpečenie rýchleho a spoľahlivého sieťového pripojenia [6].

Optické vláknové spojky sa používajú na spojenie dvoch optických vlákien, čím sa zväčšuje dĺžka cesty. Zvyčajne sa inštalujú raz a zostávajú funkčné, kým optické médium prestane fungovať. Možno ich rozdeliť na dva typy: mechanické a zvárané spojky.

3. Metódy merania prenosových parametrov optických vláknových prostredí

3.1 Meranie tlmenia priamou metodou

Priama metóda merania tlmenia optických vlákien využíva dve zariadenia – jedno na generovanie konštantného výkonu a druhé na meranie výstupného signálu [7]. Tlmenie sa vypočíta porovnaním výkonu na vstupe a výstupe pomocou vzorca:

$$\alpha = \frac{1}{L_2 - L_1} \cdot 10 \log \frac{P_{L_1}}{P_{L_2}} \; [dB/km] \qquad (1)$$

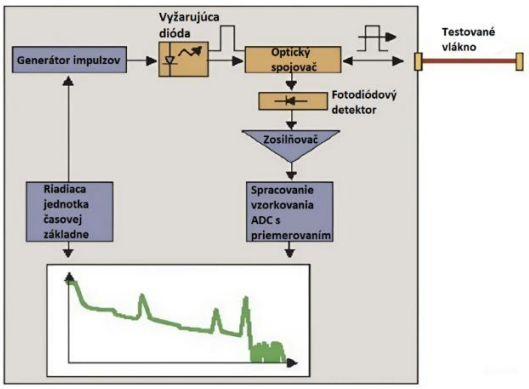
kde α - špecifické tlmenie optického vlákna, L1, L2 - zvolené dĺžky optického vlákna, PL1 a PL2 – optické výkony v [W] namerané pri zodpovedajúcich dĺžkach optického vlákna [1].

Výhodou je jednoduché vybavenie a presnosť, no metóda je vhodná len na krátke vzdialenosti, kde ohyby a kvalita spojov výrazne neovplyvňujú výsledky.

3.2 Metóda OTDR

Metóda OTDR vysiela impulz do optického vlákna a analyzuje spätný signál ovplyvnený Rayleighovým rozptylom a Fresnelovými odrazmi. Zariadenie konvertuje odrazenú energiu

na elektrický signál, ktorý sa zosilní, vzorkuje a zobrazí na obrazovke (Obr.1) [1].



Obr.1. Principiálna schéma metódy OTDR

Metóda OTDR umožňuje analyzovať optické vlákno z jedného konca a získať údaje o stratách na celej dĺžke alebo vybranom úseku. Vďaka tejto flexibilite patrí medzi najefektívnejšie metódy hodnotenia parametrov optických vlákien.

3.3 Zváranie optických vlákien

Spočiatku sa používalo mechanické spájanie, no pre vyššie rýchlosti prenosu sa rozšírilo zváranie laserom a plazmovým oblúkom, ktoré minimalizuje straty signálu a zlepšuje kvalitu spojenia.Zváranie je potrebné na bezproblémové pripojenie a minimálnu stratu signálu [8].

3.4 Videomikroskop

Optické mikroskopy odhaľujú chyby, znečistenie a poškodenie konektorov. Ponúkajú zväčšenie 200x–400x, osvetlenie a analýzu obrazu. Prenosné mikroskopy slúžia na rýchle kontroly, zatiaľ čo stolné umožňujú detailnú analýzu. Používajú sa na prevenciu strát signálu, dodržiavanie noriem a diagnostiku porúch [9] [10].

4. Praktické videá pre meranie prenosových parametrov optického prostredia

V rámci výskumnej práce bolo realizované praktické cvičenie zamerané na kontrolu stavu optických konektorov pomocou videomikroskopu. Počas cvičenia bolo zaznamenané krátke inštruktážne video (Obr.2), ktoré slúži ako praktický návod v rámci vzdelávacieho modulu Videomikroskop v e-learningovom systéme Moodle.

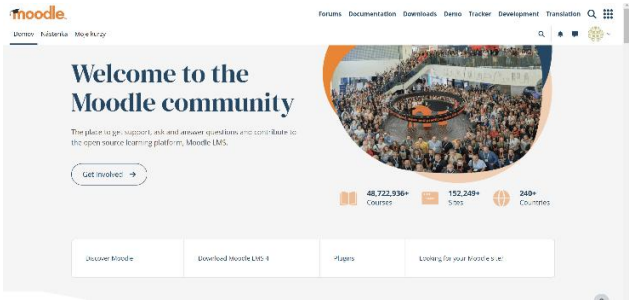
Vytvorené video je integrované do modula online kurzu s cieľom prehĺbiť porozumenie danej problematiky. Jeho využitie v prostredí digitálneho vzdelávania umožňuje študentom efektívnejšie osvojenie si teoretických aj praktických aspektov merania pomocou videomikroskopu v optických sieťach.



Obr.2. Inštruktážne video

5. Moodle

Moodle je open-source systém na správu vzdelávacieho procesu (LMS) (Obr. 3), ktorý poskytuje pedagógom nástroje na tvorbu, organizáciu a distribúciu online kurzov. Modular Object-Oriented Dynamic Learning Environment (Moodle) je navrhnutý tak, aby zabezpečoval flexibilné a prispôsobiteľné prostredie pre výučbu a samoštúdium. Vďaka svojej modulárnej štruktúre umožňuje Moodle efektívne prispôsobenie obsahu a interakcií podľa potrieb učiteľov a študentov, čím podporuje moderné metódy e-learningu [11].



Obr.3. Úvodná stránka Moodle

5.1 Funkcionalita systému Moodle

Moodle je flexibilná e-learningová platforma, ktorá umožňuje vytváranie a správu kurzov s využitím rôznych vzdelávacích nástrojov, ako sú kvízy, úlohy, fóra či multimediálny obsah. Kurzy môžu byť štruktúrované do modulov, čím sa zabezpečuje prehľadná organizácia výučby [11].

Systém disponuje pokročilými nástrojmi na hodnotenie, ktoré zahŕňajú automatizované kvízy, elektronické klasifikačné knihy a možnosť spätnej väzby k študentským prácam. Moodle podporuje aj spoluprácu prostredníctvom skupinových aktivít, diskusných fór a tímových projektov, čím sa zvyšuje interaktivita medzi študentmi a učiteľmi.

5.2 Možnosti hosťovania Moodle

Moodle môže byť nasadený rôznymi spôsobmi [12]:

- Cloudové hosťovanie – prevádzka na externých serveroch spravovaných tretími stranami.
- On-premises hosťovanie – inštalácia na vlastných serveroch organizácie s plnou kontrolou nad dátami.
- Spravované hosťovanie – hybridný model, kde poskytovateľ zabezpečuje technickú správu systému, pričom organizácia má možnosť prispôsobenia.

V rámci mojej práce bude Moodle hosťovaný na server FEI STU.

5.3 Používateľské roly a oprávnenia

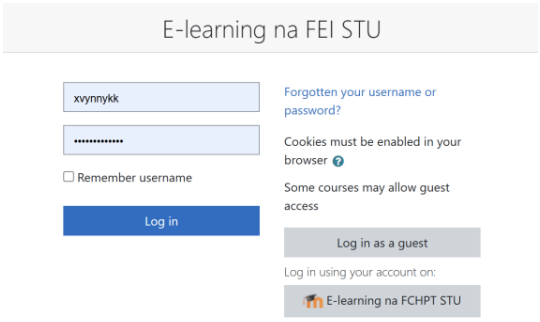
Moodle definuje viacero rolí, pričom medzi najdôležitejšie patria:

- Učiteľ – spravuje kurz, vytvára a upravuje obsah, hodnotí úlohy študentov.
- Študent – má prístup k výučbovým materiálom, odovzdáva úlohy a zapája sa do diskusií.

Pridávanie študentov môže vykonať učiteľ manuálne alebo prostredníctvom prihlasovacieho hesla.

5.4 Vytvorenie kurzu

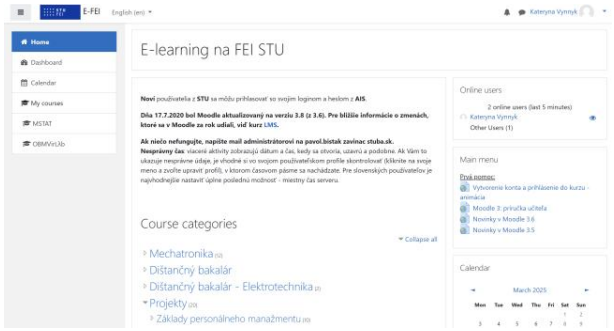
Na vytvorenie e-learningového kurzu sa najskôr musím prihlásiť na server FEI STU cez odkaz https://elearn.elf.stuba.sk/moodle/ pomocou svojho používateľského účtu (Obr.4).



Obr.4. Prihlasovacia stránka na serveri FEI STU

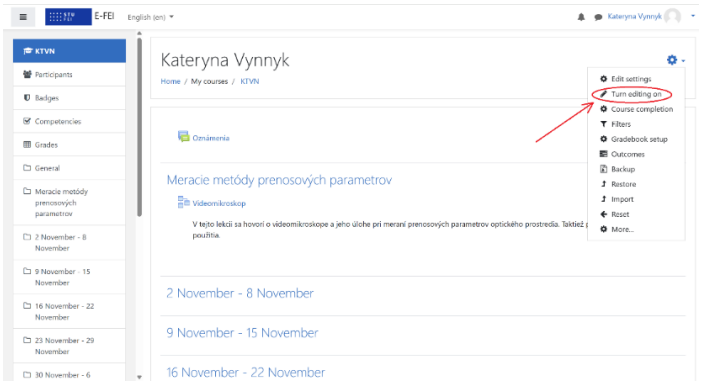
Na vytvorenie kurzu musím získať administrátorské práva v samotnom kurze od osoby, ktorá zodpovedá za Moodle na serveri FEI STU.

Po získaní administrátorských práv a prihlásení sa do svojho používateľského účtu je možné začať s tvorbou kurzu. Na úvodnej stránke Moodle na serveri FEI STU (Obr.5), kde sú zobrazené sekcie so všetkými kurzami, si vyberiem požadovaný kurz (v mojom prípade sa nachádza v sekcii Telekomunikácie), teda ten, ktorý budem napĺňať obsahom.



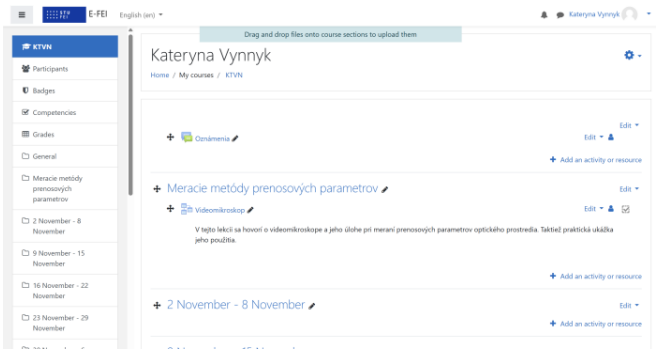
Obr.5. Úvodná stránka Moodle na serveri FEI STU

Po výbere požadovaného kurzu, na pridanie prednášok alebo cvičení je potrebné prejsť do režimu "Úprava" (Edit). Na tento účel je potrebné kliknúť na ozubené koleso v pravom hornom rohu stránky a vybrať možnosť "Turn editing on" (Obr.6).

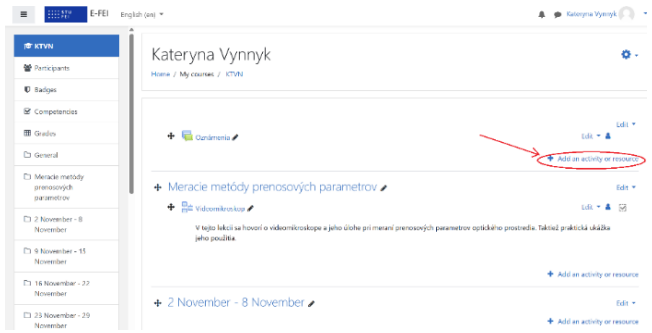


Obr.6. Tlačidlo pre režim úpravy

V režime Úprava stránka kurzu vyzerá nasledovne:



Na pridanie samostatnej prednášky alebo cvičenia je potrebné kliknúť na ikonu plus s nápisom "Add an activity or resource" (Obr.7)

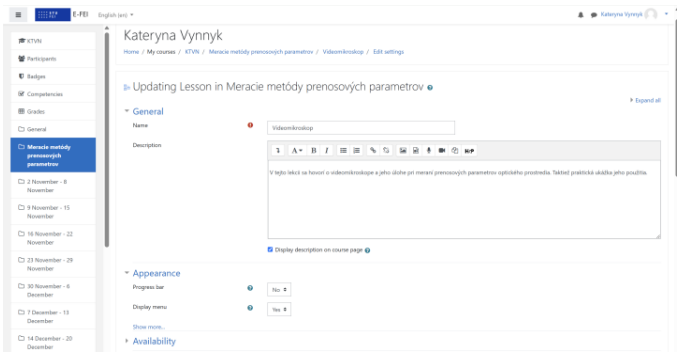


Obr.7. Tlačidlo na pridanie lekcie

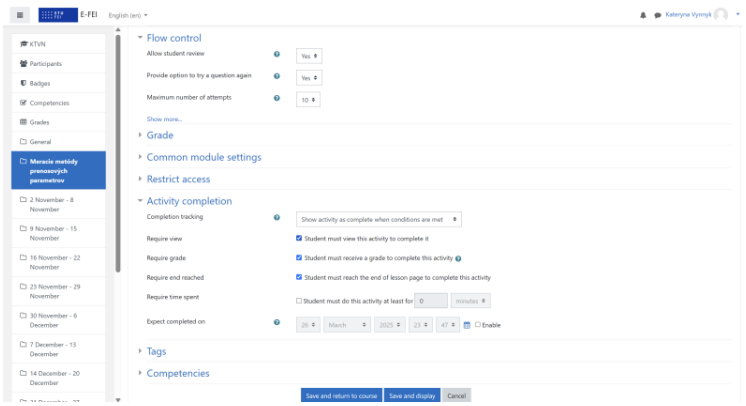
Po kliknutí sa zobrazí okno, kde budú zobrazené rôzne typy aktivít (Assignment, Attendance, Chat atď.). Na vytvorenie lekcie vyberiem aktivitu “Lesson”.

Po pridaní tohto modulu je prvé, čo treba vybrať, rôzne základné nastavenia modulu, ako napríklad názov modulu, či sa má zobrazovať pokrok počas prechádzania modulom, tlačidlo menu, či môže študent skontrolovať svoje odpovede po dokončení modulu, či môže študent opakovane odpovedať na otázky a koľko pokusov má k dispozícii, typ hodnotenia a maximálny počet bodov, aký výsledok je potrebný na postup do ďalšej lekcie a tak ďalej.

Pre modul 'Videomikroskop' som zadala názov lekcie 'Videomikroskop', pridala krátky popis lekcie; pridala tlačidlo menu v sekcii Appearance(Obr.8); povolila študentovi overiť si odpovede a tiež možnosť opakovane odpovedať na otázky, počet pokusov na opakovanú odpoveď som nastavila na 10 v sekcii Flow Control; nastavila podmienky, že modul bude označený ako dokončený, keď študent úplne prejde lekciu, získa hodnotenie a dosiahne koniec lekcie v sekcii Activity Completion. Všetky ostatné nastavenia som nezmenila (Obr.9)

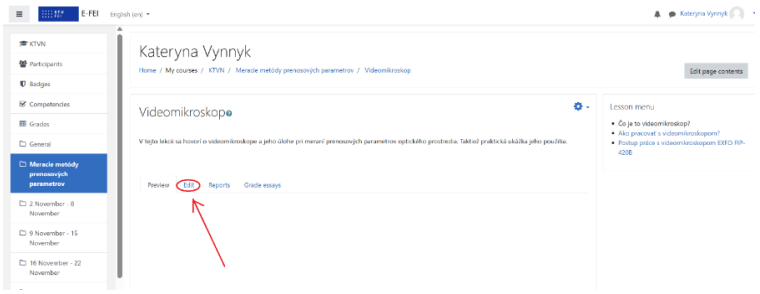


Obr.8. Základné nastavenia modulu



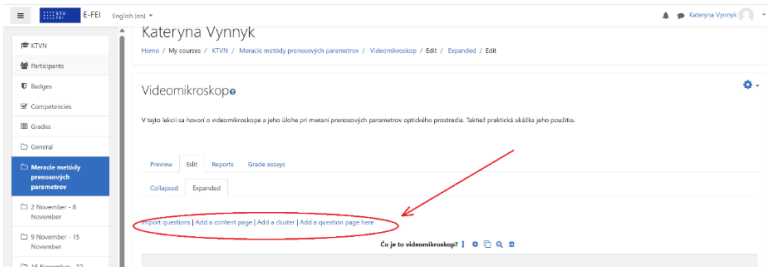
Obr.9. Ďalšie základné nastavenia modulu

Po výbere základných nastavení svojho modulu a kliknutí na tlačidlo “Save and display” (pre zobrazenie, ako bude modul vyzerať) alebo “Save and return to course” (pre návrat na stránku kurzu), som začala napĺňať modul obsahom. Na tento účel je potrebné prejsť na modul kliknutím na jeho názov a kliknúť na “Edit” (Obr.10).



Obr.10. Režim 'Edit' na úpravu obsahu modulu

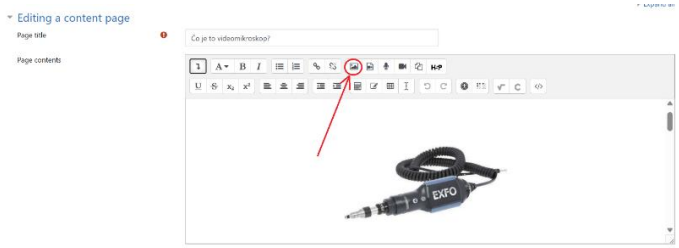
V režime 'Edit' mám možnosť pridať otázky, stránku s obsahom alebo sériu otázok, ktoré sa zobrazujú v module v náhodnom poradí pre rôznych študentov (cluster) (Obr.11).



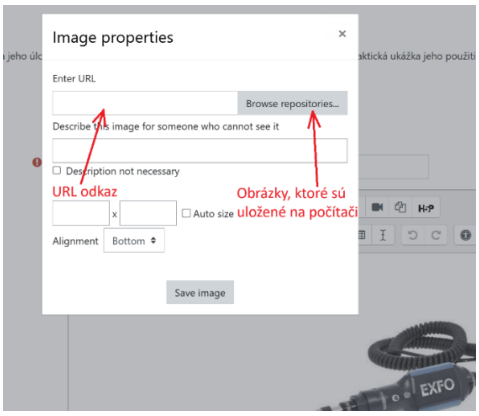
Obr.11. Možnosti na naplnenie modulu v režime Edit

Na začiatku som si vybrala možnosť pridať stránku s obsahom, kde som napísala informáciu o tom, čo je mikroskop, aké sú jeho druhy, vlastnosti a na čo je potrebný, ktorú som doplnila vizuálnymi znázorneniami. Text sa zadával v textovom poli. Obrázky boli pridané pomocou ikony s obrázkom 'Obrázok' (prvý riadok, šiesta ikona z konca) (Obr.12). Potrebné obrázky

som mala uložené na počítači, preto aby som vložila obrázok, vybrala som možnosť 'Browse repositories' a vybrala priečinok na počítači, kde sa obrázok nachádzal. Avšak existuje aj možnosť vkladať obrázky z internetu. Na to je potrebné skopírovať odkaz na obrázok a vložiť ho do riadku pre odkazy (URL) (Obr.13).

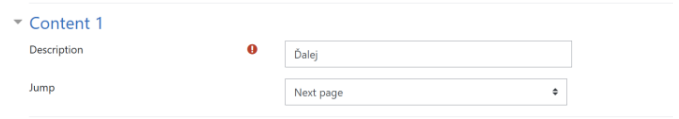


Obr.12. Ikona na vloženie obrázkov



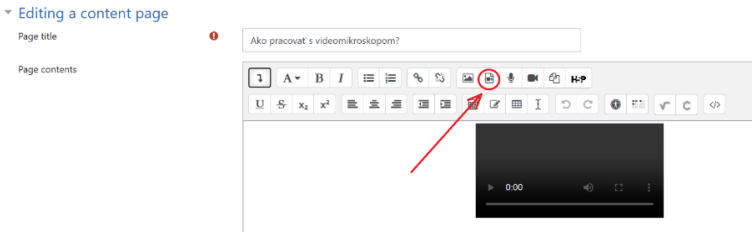
Obr.13. Dva spôsoby, ako vložiť obrázok

Potom som pridala navigačné tlačidlo 'Ďalej' v sekcii Content 1 a určila, že po kliknutí na toto tlačidlo študent prejde na ďalšiu stránku (Obr.14). A klikla som na tlačidlo 'Uložiť" (Save) na zobrazenie výsledku.



Obr.14. Tlačidlo 'Ďalej'

Nasledujúco som opäť pridala stránku s obsahom. Tentokrát som nahrala video s krátkymi inštrukciami o používaní videomikroskopu na YouTube a skopírovala odkaz. Video je možné vložiť kliknutím na piatu ikonu z konca (Obr.15), ktorá vyzerá ako súbor s kamerou, a vložením odkazu na video do poľa pre URL odkazy. A taktiež som pridala tlačidlo "Ďalej" rovnakým spôsobom ako pre predchádzajúcu stránku a tlačidlo "Späť" v sekcii Content 2, pričom som určila, že toto tlačidlo má presmerovať na predchádzajúcu stránku. A nakoniec som uložila stránku.



Obr.15. Tlačidlo na pridanie videa

Pre tretiu stránku som opäť vybrala stránku s obsahom. Tentokrát som slovami opísala, ako pracovať s videomikroskopom, a doplnila to obrázkami (rovnaké kroky ako pri vytváraní prvej stránky). Pridala som aj tlačidlá 'Späť" a 'Ďalej' rovnakým spôsobom ako pre prvú stránku. A klikla som na "Uložiť".

Po vykonaní všetkých týchto krokov som vytvorila modul s názvom "Videomikroskop".

6. Záver

V tejto práci boli analyzované prenosové parametre optického vláknového prostredia, ako aj metódy ich merania. Na základe získaných informácií a natočeného videa, v ktorom je popísaný postup práce so zariadením na meranie prenosových parametrov, bola vytvorená jedna z častí virtuálneho laboratória.

Vývoj virtuálneho laboratória a jeho implementácia ako praktickej príručky pre prácu s meracími zariadeniami môže výrazne zjednodušiť proces učenia a osvojenia si metód analýzy prenosových faktorov v optickom vláknovom prostredí.

Poďakovanie

Chcem vyjadriť svoju vďaku svojmu vedúcemu práce doc. Ing. Rastislav Róka, PhD. za podporu, ako aj za poskytnutie cenných rád a všetkých potrebných prístrojov na realizáciu tejto práce. Rovnako by som rada poďakovala ÚMIKT FEI STU za možnosť využívať potrebné zariadenia pre svoju vedeckú činnosť.

Literatúra

[1] – Prof. Ing. Ivan Baroňák, PhD.,Doc. Ing. Miloš Orgoň, PhD.,Doc. Ing. Rastislav Róka, PhD.,Ing. Erik Chromý, PhD.,Ing. Matej Kavacký, PhD., Ing. Stanislav Kľučik, PhD.,Ing. Ladislav Kočkovič, Ing. Pavol Šalík, PhD. "ŠIROKOPÁSMOVÉ TELEKOMUNIKAČNÉ SIETE.Vybrané kapitoly". SLOVENSKÁ TECHNICKÁ UNIVERZITA V BRATISLAVE.2018

[2] - FS COMMUNITY. Single-mode cabling cost vs. multimode cabling cost [online]. Dostupné na: http://app-community.fs.com/blog/single-mode-cabling-cost-vs-multimode-cabling-cost.html. [Cit.27-03-2025]

[3] - Mustafa, F.M., Zaky, S.A., Khalaf, A.A.M. et al. Chromatic dispersion compensation by cascaded FBG with duobinary modulation scheme. Opt Quant Electron 54, 819 (2022).

[4] - Li, M., & Wang, S. (2022). End-to-End Learning for Chromatic Dispersion Compensation in Optical Fiber Communication. IEEE Communications Letters, 26(8), 1829–1832.

[5] - Du, Q., Zhang, X., Guo, Y. et al. Density-matrix-formalism based scheme for polarization mode dispersion monitoring and compensation in optical fiber communication systems. Optoelectron. Lett. 19, 739–743 (2023).

[6] - Tatarko Matúš. "Spájanie optických vlákien". 2013.

[7] - DIPOL. Meranie v optických sieťach [online]. Dostupné na: https://www.dipol.sk/meranie_v_optickych_sietach_bib348.htm. [Cit.27-03-2025]

[8] - CABLERIDER. Optical fiber welding: Future connectivity [online]. Dostupné na: https://cablerider.com/cables/optical-fiber-welding-future-connectivity/#Factors_Affecting_Optical_Fiber_Welding_Quality. [Cit.27-03-2025]

[9] - FIBER SYSTEMS. Quality standards [online]. Dostupné na: https://www.fibersystems.com/pdf/whitepapers/Quality-Standards.pdf. [Cit.27-03-2025].

[10] - GAOTEK. Optical fiber microscope [online]. Dostupné na: https://gaotek.com/category/fiber-optics/advanced-fiber-testers/optical_fiber_microscope/. [Cit.27-03-2025].

[11] - LEARNINGEVEREST. Moodle LMS features and benefits [online]. Dostupné na: https://www.learningeverest.com/moodle-lms-features-and-benefits/#toc_What_is_Moodle. [Cit.27-03-2025].

[12] - OVERNETDATA. Benefits of onsite Moodle installation [online]. Dostupné na: https://www.overnetdata.com/case-study/benefits-of-onsite-moodle-installation/. [Cit.27-03-2025].

[13] - SWATKB. Roles in Moodle [online]. Dostupné na: https://swatkb.atlassian.net/wiki/spaces/MOODLE/pages/20908451/Roles+in+Moodle. [Cit.27-03-2025].

Návrh a realizácia modelu letiska M.R. Štefánika

Matúš Dúbravka

Ústav vedy a výskumu, FEI STU v Bratislave
Informačné a komunikačné technológie

xdubravkam@stuba.sk

Abstrakt – Na základe znalostí z letectva, elektrotechniky a programovania sme sa rozhodli spraviť projekt modelu letiska M. R. Štefánika. Tento nápad vznikol aj ako pomôcka pre nových pilotov, ktorí si vedia prezrieť letisko ako model a naučiť sa na ňom základné systémy používané na letiskách. Ako pilot s ambíciou stať sa inštruktorom, vidíme potenciál tohto modelu ako nástroj pri výučbe. Hlavným cieľom bolo vyhotovenie osvetlenia plôch letiska ako sú dráhy, taxi a ďalšie. V práci je tiež vyhotovené automatické nočné osvetlenie parkoviska. Pred dráhou máme spravené aj približovacie svetlá, ktoré sú vytvorené vďaka integrovanému obvodu. Ďalším stanoveným plánom bol systém na indikáciu presného priblíženia skrátene PAPI, ktorý pomáha pilotom udržiavať správnu výšku pri pristátí. Jedným z posledných cieľov bol aj starší systém, ktorý umožnil komunikáciu s vežou pri strate spojenia. Všetky tieto ciele sme splnili a na ovládanie práce sme zvolili Arduino MEGA. Práca má tiež svoju zobrazovaciu jednotku, LCD I2C displej, ktorý je ovládaný prostredníctvom vlastnej dizajnovej klávesnice. Na spustenie práce máme aj rôzne oprávnenia pre viacero užívateľov, pracujú vďaka RFID kartám. Do budúcnosti by sme určite pridali druhý snímač pre lietadlo, vyrobili navádzacie svetlo na ovládateľnom ramene, doplnili o ďalšie budovy a pridali systém na zabránenie stretu dvoch lietadiel na dráhe. Taktiež by sme vytvorili ostatné systémy na automatickú komunikáciu lietadla a veže, ktoré by mohli byť využívané v letectve na automatizáciu.

1 Úvod

Naša práca predstavuje model letiska Milana Rastislava Štefánika, známeho aj ako Bratislavské letisko. Pri jeho vytváraní sme čelili rôznym zložitým problémom, ktoré vyžadovali značný čas a sústredenie. Náš model zahŕňa reprezentáciu letiska, pričom sme sa zameriavali na vhodné úpravy rozmerov a dizajn. Každá budova bola navrhnutá s dôrazom na autentický dizajn, aby vytvorila dojem reálnosti. Letisko obsahuje viaceré budovy, ktoré boli vybavené exteriérovým dizajnom a osvetlením, čím sme dosiahli vzhľad a atmosféru. Kľúčovým cieľom bolo osvetliť všetky dráhy v súlade s ich rôznymi intenzitami osvetlenia. Okrem toho sme implementovali osvetlenie pre jednotlivé štruktúry, vrátane hlavného letiskovej haly, veže a súkromného hangáru. V rámci letiska sme integrovali rôzne zariadenia a systémy. Napríklad systém na meranie výšky lietadla nad zemou, ktorý ovládame pomocou tlačidiel pre simulovanú kontrolu. Ďalším zariadením je systém pre prípad stratenia komunikácie, ktorý sa ovláda prostredníctvom potenciometra. Pred pristávacou dráhou sme umiestnili navádzacie svetlá, ktoré pomáhajú navigácii v zhoršených podmienkach. Riadiacu jednotku tvorí Arduino MEGA, ktorá obsahuje displej s aktuálnym stavom a klávesnicu na ovládanie. Zabezpečenie práce predstavuje RFID modul, ktorý vyžaduje autentifikáciu prostredníctvom kariet a čipov pred spustením. Okrem samotného funkčného modelu sme venovali pozornosť aj dizajnu kariet a čipov, vytvárajúc jednotný vizuálny štýl. Práca bude obsahovať veľa LED diód, ktoré budeme spínať pomocou NPN tranzistorov. Táto verzia letiska je prvým krokom, avšak v budúcnosti plánujeme pridávať nové prvky a jemne upravovať existujúce komponenty.

2 NÁVRH A REALIZÁCIA MODELU LETISKA M. R. ŠTEFÁNIKA

2.1 Využitie v praxi

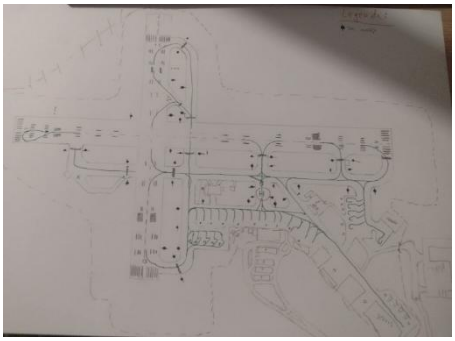
Naša práca plní hlavnú úlohu pri výučbe nových pilotov. Tí si na nej vedia pozrieť presné rozloženie letiska a dokážeme si na nej vysvetliť rôzne letiskové značenia. To pomôže pilotom lepšie sa orientovať na letisku. Taktiež dokážeme vysvetliť systémy, ktoré sa nachádzajú na letiskách. Dokážeme si vysvetliť úkony, ktoré sa musia na pozíciách vykonávať. Je to skvelá učebná pomôcka, ktorá rozšíri vedomosti a poučí nielen pilotov, ale aj širokú verejnosť.

2.2 NÁVRH PLÁNU MODELA

Elektronické zariadenia tvoria hlavnú súčasť nášho letiska, preto sme si vymysleli základné systémy, ktoré sa nachádzajú na letiskách. Pri tvorení takýchto prác treba mať taktiež pri sebe veľa výkresov, ako aj skúseností v oblasti elektrotechniky a modelovania, preto sme si najprv začali zapisovať nápady na systémy, ako sú napríklad: dráhové svetlá ovládané, osvetlenie letiska, približovacie svetlá, informácia o počasí. Prvotný náčrt obsahoval pravú mierku letiska a plôch (obr. 1.), toto však nebolo ideálne a dráhy nám pretŕčali, preto sme sa rozhodli dráhy podstatne zmenšiť, zmenšiť nepotrebné budovy a zväčšiť budovy, ktoré majú tie najdôležitejšie úlohy. Určili sme si základné požiadavky, ktoré musíme splniť a začali sme postupne robiť.

Ďalším postupom bolo spraviť si presný návrh môjho modelu. Vzhľad sme prebrali prevažne vďaka funkcii Google Earth a fotkám z okolia letiska. Ten sme začali robiť v programe od spoločnosti Adobe AutoCAD. Tá nám svojím spôsobom nepomohla, a z dôvodu chýbajúcej študentskej licencie som nakoniec skončil v 3D programe Blender. Tam sme si najprv vytvorili rozmery, ktoré by sme chceli, a následne sme upravovali výšku pre budovy. Keďže nie všetky budovy sa zmestili, museli sme pristúpiť k čisteniu ďalších budov.

Následne vďaka týmto modelom sme si vedeli spraviť základnú fyzickú predlohu nášho letiska.



Obr. 1. Náčrt plánu letiska

Po vytvorení našej predlohy sme začali pracovať na potrebných systémoch. Keďže náš projekt riadi mikrokontrolér Arduino MEGA 2560, museli sme si nájsť ideálny program na programovanie a vytvorenie simulácie, na zabránenie tvoreniu chýb. Pracovali sme v programe TinkerCAD, ktorý sa nachádza v internetovej verzii, bohužiaľ, tam neboli všetky zariadenia, ktoré sme používali, tie sme museli testovať v reálnom svete, pomocou metódy pokus a omyl.

2.3 Zakresľovanie LED diód

Keďže zariadenia a kód sme postupne robili, prešli sme k umiestňovaniu našich LED diód pre celú prácu. Ten prebiehal nasledovne, pri sebe sme mali oranžovú, červenú, zelenú, žltú a bielu fixku (Obr. 3). Začali sme navrhovať umiestnenie najprv na parkovisku, to bolo ľubovoľné a nie je zhodné s originálnym letiskom.

Pôvodne sme chceli umiestniť až 3 LED diódy do veže, ale nakoniec sme neboli schopní dokončiť dizajn spodnej časti veže, preto tam zostala len jedna.

V hlavnej letiskovej budove sa nachádzajú tri, jednu pre 1. poschodie, druhú pre prízemie a tretiu pre vyhradenú miestnosť.

Taxi svetlá boli svetlá, ktoré nemajú detailné umiestnenie, pretože v realite ich môže byť obrovské množstvo, preto sme ich dali len na naše podstatné miesta v obmedzenom počte. Následne sme robili prvú dráhu a to 04/22, ktorá je dlhšia, ale s osvetlením je rovnaká ako dráha 13/31, pri tomto sme chceli dokázať, že týchto dizajnov je viacero.

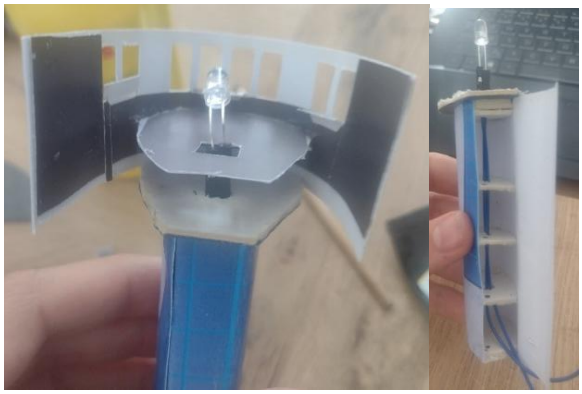
Teraz prišlo na rad spočítanie LED diód. Po spočítaní sme si všetko zapísali do výkresu a celkový počet LED diód je: 270.

Spravili sme si aj Taxi cedule, ktoré umožňujú ľahšiu orientáciu na letisku.

2.4 Modely letiska

2.4.1 Riadiaca veža

Ako prvé sme si vytlačili dizajny a pripravili narezané kúsky maslového papiera. Najprv bolo treba spraviť vežu pomocou škrupinových vrstiev. Následne sme do týchto vrstiev vyrezali diery pre kábel. Postupne sme prilepovali papiere s vytlačeným dizajnom a vytvorili tak silnú a pevnú podstavu. Následne sme hore nalepili ďalšiu, už trochu väčšiu a vytvorili vnútro tejto veže. Osadili sme LED diódu a zalepili vrch. Vyrezali sme presné rozmery na plot a osadili ho. Následne sme vytlačený kus spodnej budovy veže jednoducho nalepili a odložili hotovú vežu bokom (obr. 2.). Aktuálny prístup do veže je prísne obmedzený. Vo veži sa však nachádzajú stanoviská s obsluhovaním frekvencií: Bratislava veža, Bratislava approach, Bratislava Departure.



Obr. 2. Konštrukcia riadiacej veže

2.4.2 Hlavný Terminál

Hlavná budova bola jedna z najťažších budov modelu. Museli sme vytvoriť malý model pevných bodov maslového papiera, a následne krehkých bodov, a to fólie, ktorá dokonale simuluje sklo. K tomuto modelu sme pridali malé piliere dopredu a začali lepiť dané kúsky maslového papiera k sebe. Pilierom sme museli vytvoriť spodnú oporu a vrchnú, aby držali (obr. 3.)



Obr. 3. Konštrukcia hlavného terminálu

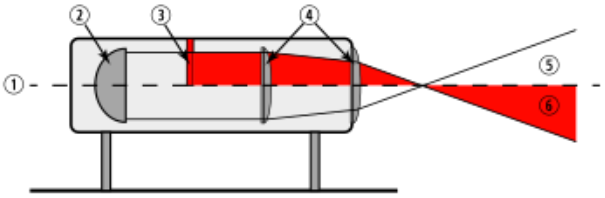
Nalepili sme sklo k ostatným voľným miestam. Strechu sme dávali až po osadení všetkých dizajnov vo vnútri. Osadili LED diódy a lepili sme druhé poschodie. Káble sme nechali trčať zo spodku. Pred samotným zatvorením sme skontrolovali funkčnosť LED diód. Strechu sme lepili postupne, najprv sme naniesli kúsok lepidla a držali strechu a budovu silno k sebe. Následne sme pokračovali až do konca, čo malo za následok prekrásnu vlnovitú strechu.

2.5 Elektrické zariadenia

Elektrické zariadenia sa nachádzajú na našej práci.

2.5.1 Systém PAPI

Tento systém pomáha pilotom udržiavať správnu výšku pri pristátí. Sú to štyri svetlá, ktoré nadobúdajú červenú a bielu farbu. Tento systém je ovládaný vďaka optickému systému (obr. 4.), ktorý môžeme vidieť na obrázku.



Obr. 4. optický systém PAPI

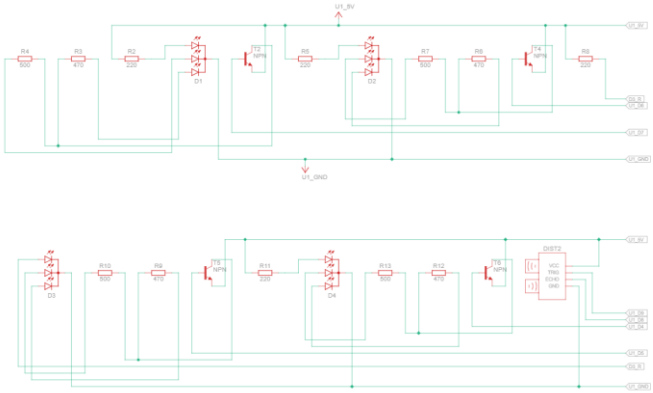
Na našej práci sme tento systém nerobili pomocou optického systému, ale pomocou snímača a riadiacej jednotky Arduina.

V práci sme chceli pôvodne využiť 2 ultrazvukové snímače, ktoré snímajú vzdialenosť snímača od objektu. Pri ich použití by sa silno ovplyvňovali a systém by nefungoval, preto sme vymysleli iba simuláciu.

Umiestnili sme jeden snímač na začiatok dráhy. RGB LED diódy sme umiestili na ich miesto, vedľa dráhy na viditeľné miesto. Vyhotovili sme tlačidlá, ktorými si nastavujeme svoju výšku. V knihe Učebnice pilota 2022 sme našli údaje o stupňoch, pri ktorých sme vysoko a pri ktorých sme nízko. To znamená, že ak pilot udržuje lietadlo na správnej trajektórii, dosiahne bezpečné pristátie na pristávacej dráhe.

Vďaka známej výške, ktorú si určíme tlačidlom a odmeranej vzdialenosti lietadla od dráhy, tak si môžeme vypočítať uhol medzi nimi, pomocou sin rovnice na výpočet uhla..

Na začiatku obvodu máme pripojený ultrazvukový snímač, následne každá RGB LED dióda má prepojené spolu mínusom a červené nožičky na (230 ohm) rezistor, ktorý je napojený na 5V (obr. 5.).



Obr. 5. Schéma zapojenia systému PAPI

Po otvorení tranzistora začne obvodom tiecť prúd, ktorý pustí prúd na ultrazvukový snímač, rozsvietia sa RGB led diódy na červeno. Následne pre zelenú nožičku má každý (150 ohm) rezistor a pre modrú (230 ohm) rezistor. Po vyslaní 5V od Digital PINU z Arduina sa tranzistor otvorí a začne tiecť prúd cez modrú a zelenú farbu RGB LED diódy. Tieto hodnoty

rezistorov sú nastavené na bielu RGB LED diódu.

Pri kódovaní sme si vytvorili nové premenné. Po zapnutí systému začali snímať správny údaj a ten sme následne vydelili danou výškou a dostali sme uhol, ten sme využili na porovnanie, či je väčší alebo menší. Následne sme zadali funkciu na zapnutie LED diódy alebo vypnutie. Potom sa program opakoval až do chvíle, keď sme stlačili niečo na klávesnici.

2.5.2 Navádzacie svetlo

Navádzacie svetlo je zastaraný systém na komunikáciu medzi vežou a pilotom. Tento systém sa taktiež používa pri strate komunikácie medzi pilotom a vežou. Podľa aktuálnych ICAO predpisov, by mala byť každá veža vybavená takýmto systémov. V skutočnosti to je svetelný lúč, ktorý vyzerá ako prenosné delo. Táto svetelná lampa môže vysielať 3 farby, a to zelenú, červenú a bielu. Zamestnanec veže vyšle signál podľa udeleného príkazu pre lietadlo, ktoré následne potvrdí signalizáciu, kývaním krídelka, alebo blikaním svetiel.

My sme spravili pre ľahšie vysvetlenie a vyhotovenie, otočné v jednej osi rotácie. Na tomto zariadení máme umiestnené 1 RGB Ledku, ktorá sa rozsvieti podľa zadaného príkazu (obr. 6.). Servo motor, je ovládaný pomocou potenciometra, ktorý sa nachádza v hlavnej riadiacej časti . Pri zapnutí tlačidla sa nám zapne LED dióda a zapne sa nám aj servo motorček

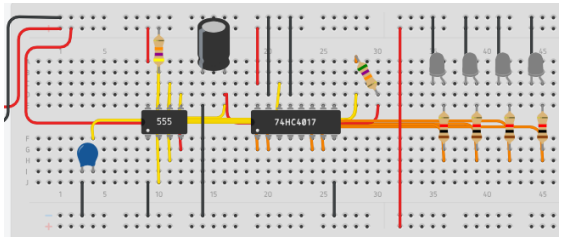
Kód je jednoduchý na prvotné nastavenie polohy serva, následne pri zaznamenaní zmeny potenciometra Arduino vyšle signál na otočenie servo motorčeka. Program sa opakuje do momentu, keď stlačíme nejaké tlačidlo na klávesnici.



Obr. 6. Ovládacie navádzacie svetlo

2.5.3 Približovacie svetlá

Je to svetelný efekt, ktorý pomáha pilotom ľahšie nájsť letiskovú dráhu. Pri nižšej viditeľnosti sú tieto svetlá veľmi užitočné. V našej práci máme pre tento svetelný efekt vyhradené 4 biele LED diódy. Obvod a jeho funkčnosť vykonáva Timer NE555 a čítač CD4017BE dekadický, ktorý nám tvorí integrovaný obvod (obr. 7.). Tento obvod je naletovaný na spodnej časti našej práce. Tento obvod sme umiestnili, len pre jednu dráhu, pre jeho vysvetlenie účelu a zámeru.



Obr. 7. Zapojenie približovacích svetiel

LED diódy máme pripojené k našej riadiacej jednotke, ktorá má viacero oddelených výstupov. Pri prijatí signálu sa zapne prvý výstup, pri prijatí druhého signálu sa vypne prvý výstup a zapne sa druhý výstup. Takto to funguje stále dookola. Rezistormi si dokážeme riadiť intenzitu tohto efektu. Na tento obvod je napojených 5V z Arduina, a následne mínus tohto

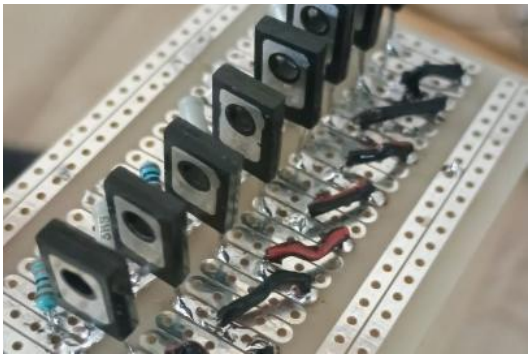
obvodu smeruje do tranzistoru, ktorý sa pri našom pokyne otvorí a obvodom začne pretekať prúd.

2.5.4 Automatické osvetlenie letiska

Na práci je spravené automatické osvetlenie parkoviska, vďaka snímaniu svetlosti na práci. Tieto svetlá nám osvetľujú parkovisko a vonkajšok letiska. Dokážeme ich vypnúť alebo zapnúť. Ako snímač sme použili najlacnejšiu a najspoľahlivejšiu súčiastku, tou je fotorezistor. Táto súčiastka sníma svetlosť a následne sa podľa nej mení odpor. Na fotorezistor privedieme 5V, ktoré následne uzemníme cez dodatočný odpor. Aby sme načítali vstupnú hodnotu, musíme ešte tento fotorezistor pripojiť na analógový vstupný port v Arduine. Pri opakovanom cykle v hlavnom kóde, pokiaľ fotorezistor zaznamená zmenu a tá prekročila hodnotu, stav sa zmení na jedna a ten otvorí tranzistor. Pokiaľ chceme vypnúť tieto svetlá, stlačíme tlačidlo na automatické svetla. Je to spravené cez viacero podmienok.

2.5.5 Celkové osvetlenie letiska

Naše osvetlenie letiska obsahuje 270 LED diód, s rôznymi farbami, ako je červená, biela, žltá, oranžová, zelená a RGB LED diódy. Každá jedna LED dióda ma svoju spotrebu. Preto sme sa rozhodli pridať externý 24V zdroj s max 2,5A. Na spínanie tohto zdroja sme použili NPN tranzistor (obr. 8.). Prúdová spotreba pre jednu LED diódu je 20mA na veľkých LED diódach a 10mA na tých malých. Napäťová spotreba závisí podľa danej farby. Najväčšiu spotrebu má biela LED dióda, ktorá má v určitých obvodoch 3V. Keďže my chceme, aby mala LED dióda odber 20mA a napätie na nej bolo 3V, tak si vypočítame, či už podľa vzorca, alebo internetu, potrebný predradný odpor. Ten následne umiestnime a z neho vstupujeme na kolektor tranzistora, ktorý je zatvorený, pretože nám neprechádza prúd. Ten začne prúdiť, až pokiaľ Arduino vyšle napätie z Digitálneho výstupového PIN 5V, ktorý prechádza rezistorom, na otvorenie tohto tranzistora. Je to z toho dôvodu, že Arduino by nerozsvietilo všetko, preto sme potrebovali pridať externý zdroj. Obvody LED diód sú navrhnuté tak, aby každá LED dióda mala príslušnú potrebnú hodnotu. Preto napríklad na dráhe je 10 zelených a 27 bielych. Týchto 10 zelených je rozdelených na 5, ktoré sú zapojené do série. Pri bielych sú do série zapojené po šiestich.



Obr. 8. NPN tranzistory na spínanie LED obvodov

2.5.6 Systém METAR

Na našom letisku sa nachádza aj systém METAR, teda aktualizovaná informácia o aktuálnom počasí na letisku. Náš METAR dokáže snímať viaceré javy, ktoré sa vyskytujú na našom letisku. Prvá vec, ktorú obsahuje naša informácia je, pre aké letisko platí, (ICAO kód LZIB) (obr. 9.). Následne vypisuje dátum 17. následne Hodina a minúta 17:31 a Zulu, ktorý označuje, že čas je v UTC. Tento čas nevypadne, pretože používame hodinový modul, RTC Tiny I2C module, ktorý obsahuje batériu, ktorá sa pri zapnutej práci nabíja. Tento modul pri vypnutom stave stále počíta Čas s presnosťou 1s na 1deň.

Ďalšia časť obsahuje teplotu a teplotu rosného bodu označenou ako 23/10, takže môžeme vidieť, že na letisku je 23°C a najnižšia teplota vlhkého vzduchu má 10°C. Aby sme ušetrili peniaze, použili sme senzor, ktorý meria teplotu a statický tlak, a ďalší senzor, na vlhkosť. Tá je potrebná spolu s teplotou na vypočítanie teploty rosného bodu.

Jedna z najdôležitejších je informácia o aktuálnom statickom tlaku. Podľa toho tlaku, následne vieme výšku nad letiskom, a tá nám dovoľuje bezpečne lietať s ostatnými lietadlami. Na obrázku môžeme vidieť Q990, teda QNH 990hPa.

Základnou vlastnosťou je aj snímanie oblačností a jej základne výšky (minimálna výška oblaku). To funguje vďaka neustálemu točiacemu sa radaru, ktorý sníma okolie letiska pomocou ultrazvukového snímača. Celý systém pracuje vďaka počítaniu každého stupňa, o ktorý sa servo motor pohne. Ak Ultrazvukový snímač zistí nejakú prekážku, pripočíta si ju k hodnote prekážka, a pritom ak je táto vzdialenosť najmenšia zatiaľ, zapíše ju ako základňu výšky oblačností. Po 360°otočení si Arduino prepočíta priemer oblačností. Podľa oficiálnych tabuliek následne zapíše, či je zamračené, málo oblačností alebo je jasná obloha. Na Správe sa preto píše FEW, pretože bola nad letiskom jemná oblačnosť, vo výške 060, teda 6cm nad zemou. Pri druhej správe, by sme mohli mať napríklad SKC, Sky clear, alebo jasná obloha.



Obr. 9. Informácia METAR

2.5.7 Klávesnica 4x4

Klávesnica je hlavná riadiaca časť na našej práci. Touto klávesnicou ovládame celú prácu. My sme chceli spraviť prácu jednoduchú, prehľadnú a peknú, preto sme sa rozhodli spraviť si aj vlastný dizajn klávesnice a vlastné zaznačenie tohto zariadenia. Táto klávesnica, ktorú sme použili, je takzvaná membránová klávesnica, tá teda obsahuje niekoľko prepínačov usporiadaných do mriežky. Vodiče, ktoré sa nachádzajú vnútri, spájajú jednotlivé stĺpce, keď chce Arduino skenovať túto klávesnicu, tak vysiela signál do každej mriežky. Pokiaľ máme stlačenú nejakú klávesu, tak sa tento riadok spojí so stĺpcom a ten 29 Arduino naskenuje a zaznamená. Preto naše klávesnica občas nereagujú a treba to postláčať párkrát. Aj napriek takto veľkému problému je táto klávesnica super na ovládanie, pretože jej cena, materiálna nenáročnosť a aj výmena dizajnu, je výborná na prácu s ňou.

Dizajn pre našu prácu sme vyhotovili v programe CANVA čo je editor fotiek s mnohými prvkami. Ako prvé sme si odmerali rozmery originálnej klávesnice. Tieto hodnoty zapísali do programu a vytvorili si svoj dizajn. Následne sme si

ho vytlačili a odlepili trochu vrchnej časti plastu (obr. 10.). Pridali sme obojstrannú pásku, nalepili vytlačený kus a zastrihli rohy podľa predlohy.



Obr. 10. Vlastný dizajn klávesnice

2.5.8 RFID systém

Ďalším systémom je prihlasovacie zariadenie. Je to RFID skener, ku ktorému priložíme kartu alebo čip pre prihlásenie. Každá karta a čip majú nejaké svoje oprávnenia, ktoré môžu na letisku vykonávať (obr. 11.). Po zapnutí práce sa nám zobrazí požiadavka od riadiacej časti na prihlásenie sa. Naše karty majú aj vlastný dizajn, pretože sme chceli mať túto prácu aj trocha krajšiu. Zariadenie je celkom ľahké na zapojenie. Stačilo nám nato iba príslušnú schému a kód. Najprv sme museli spraviť program na detekciu a výpis ID-čka kariet. Aby sme vedeli potom tieto hodnoty pripísať do kódu. Každá jedna karta má svoje ID, ktoré sme si naskenovali a následne zadali do podmienky, na začiatku sa program zastaví, dokým sa neprihlásite. Následne, keď rfid prečíta kartu, porovná ju cez podmienky. Po úspešnom prihlásení si zapamätá hodnotu vašej oprávnosti. Tú bude porovnávať pri zapnutí vykonávaní funkcií, napríklad administrátor má 1 a veža má 2. Pri zapnutí sa program opýta „je hodnota perm 1 alebo 2?“ ak áno pokračuje, ak nie vypíše, že nemáme oprávnenie.

Funkcia	Administrátor	Riadiaca veža	Polícia	Satelit
Svetlá Dráhy 13/31	Oprávnený ☑	Oprávnený ☑	Zamietnutý ✗	Zamietnutý ✗
Svetlá Dráhy 22/04	Oprávnený ☑	Oprávnený ☑	Zamietnutý ✗	Zamietnutý ✗
Taxi svetlá	Oprávnený ☑	Oprávnený ☑	Zamietnutý ✗	Zamietnutý ✗
Riadiaca veža	Oprávnený ☑	Oprávnený ☑	Zamietnutý ✗	Zamietnutý ✗
Hlavná Budova	Oprávnený ☑	Zamietnutý ✗	Oprávnený ☑	Zamietnutý ✗
Hangár	Oprávnený ☑	Zamietnutý ✗	Oprávnený ☑	Zamietnutý ✗
Letiskové plochy	Oprávnený ☑	Oprávnený ☑	Oprávnený ☑	Zamietnutý ✗
Vonkajšie svetlá	Oprávnený ☑	Zamietnutý ✗	Oprávnený ☑	Zamietnutý ✗
Systém PAPI	Oprávnený ☑	Oprávnený ☑	Zamietnutý ✗	Zamietnutý ✗
Približovacie svetlá	Oprávnený ☑	Oprávnený ☑	Zamietnutý ✗	Zamietnutý ✗
Navádzacie svetlá	Oprávnený ☑	Oprávnený ☑	Zamietnutý ✗	Oprávnený ☑

Obr. 11. Tabuľka oprávnení podľa kariet a čipov

2.5.9 LCD I2C Display

Zobrazuje nám aktuálny stav práce. My používame I2C zbernicu, tak namiesto 10 potrebných PIN a súčiastok, máme 1 vstup a 1 výstup, čiže SCL a SDA. Zariadenie môže zobraziť 16 znakov v 2 riadkoch. Tento modul funguje ako predchádzajúci systém RFID systém. Chceli sme ho neskôr nahradiť buď jeho väčšou verziou 20x4, alebo dotykovím NEXTON displejom, v ktorom je zabudovaný samostatný procesor. Keďže naša práca nepotrebuje pokročilú zobrazovaciu jednotku, zostali sme pri tejto

2.5.10 Bzučiak

Pri zapínaní, preklikávaní alebo ovládaní sme umiestnili bzučiak, ktorý bzučí tichšie a na nižšej frekvencii. Tento dodatočný zvuk je skvelý pri nastavovaní, pretože hneď vieme, či práca funguje alebo nie. Pre toto zvukové zariadenie sme museli dodatočne vypáliť na krabičke dieru, aby sme ho pekne a bez odrazu počuli.

2.5.11 RTC modul

Hodinový modul, ktorý si uchováva svoj nastavený čas. Ako prvé sme museli na tomto module nahrať čas cez počítač, následne si uloží tieto hodnoty a z batérie bude brať energiu na pripočítavanie času, keď je tento modul odpojený. Pri zavolaní funkcie v kóde tento modul pošle čas do zobrazovacej jednotky. Ten je zobrazený v hlavnom menu, kde sa zapínajú rôzne funkcie. Toto zariadenie, vždy keď je Arduino pripojené do počítača, si svoj čas aktualizuje na miestny čas a začne počítať znova. Tým, že toto zariadenie potrebuje napájanie, máme tam uloženú mincovú baterku. Prvým a najhorším možným problémom bolo nabíjanie mincovej baterky, a taktiež aj nefunkčnosť kúpeného modulu. RTC modul automaticky dobíjal mincovú batériu. Prvotná batéria nebola nabíjateľná, a musela byť okamžite vymenená. Bohužiaľ autor toto nikde neuviedol a zázrakom sa našej práci nič nestalo. Zistili sme to až pri kontrolovaní modulu prečo nefunguje. Následne po meraní sme zistili, že neboli spojené všetky zeme. Bola to výrobná chyba v danom modely. Aktuálne už modul funguje a zobrazuje správny čas. Tento modul sa volá RTC Tiny I2C module. Modul pracuje iba pri prevzatí signálu z Arduina, vtedy vyšle signál o aktuálnom čase.

2.5.12 Zmenšenie letiska a Hlavná riadiaca časť

V dôsledku veľkosti práce sme museli pár nepodstatných budov odobrať. Teraz sme potrebovali vymyslieť dizajn pre hlavnú riadiacu časť. Tú sme spravili v našich farbách a vymysleli ideálne osadenie týchto komponentov. Na vrch sme umiestnili zobrazovaciu jednotku, pod to vykonávaciu jednotku a ešte nižšie doplnkové funkcie ovládania. Celý dizajn letiska sme vyhotovovali v CANVA, odmerali sme si rozmery našej práce a začali sme si tvoriť dizajn. Po vytvorení dizajnu, sme si dorobili samostatne parkovisko, ktoré sme následne vytlačili zvlášť. Letisko sme tlačili ako A3 a spájali, nakoniec sme nalepili parkovisko, ktoré sme museli trocha upravovať, kvôli 5mm chybám v presnosti. Následne sme si odstrihli nepodstatné zelené plochy, ktoré neskôr nahradíme trávou

Hlavná riadiaca časť (obr. 15.)sa nachádza v krabičke, v nej sa nachádzajú potrebné zariadenia, ktoré riadia celú prácu. Hlavnou riadiacou časťou našej práce je mikrokontrolér Arduino MEGA, ktorý je rozšírenou verziou jeho základnej verzie Arduino UNO. V tejto krabici taktiež môžeme nájsť, tlačidlá, potenciometer, LCD displej, RFID scanner, klávesnicu, bzučiak, RTC model. Ako prvé sme umiestnili Arduino a potom ostatné zariadenia.



Obr. 12. Hlavná riadiaca časť

2.5.13 Vŕtanie a spájkovanie LED diód

Zoberieme si hlavnú veľkú nakreslenú schémy, kde máme zafarbené a načrtnuté miesta pre LEDky. Budeme potrebovať vŕtačku a vrták s prierezom M1,5. Do maslového papiera si môžeme vytvoriť 2 diery, ktoré budú slúžiť ako predloha LEDky. Následne na určité miesto priložíme pomôcku a vyvŕtame dieru. Pokiaľ sú diery vidieť, dáme tam príslušnú diódu.

Inak si zoberieme batériu a presvietime to cez tieto diery. Vo výsledku budeme vidieť, kde máme diódu osadiť. Spájkovanie LEDiek, je dlhá a komplikovaná práca.

Keďže máme 270 LED diód, treba spraviť len jednoduchšiu metódu, buď si spravíme v strede dráh cestičky, po ktorých budeme osádzať káble a vystupovať k daným LED diódam, alebo pre úsporu času a káblov si nameriame kábel, ktorý pripojíme k LED dióde, pôjdeme na doske k ďalšej LED dióde, treba si označovať káble, inak by sme sa v tom stratili. Kladný pól zo zdroja smeruje na tieto segmenty LED diód, segmenty máme spravené podľa príslušnej farby.

2.5.14 Zapájanie daných zariadení

Keďže všetko máme osadené, dané zariadenia si zapojíme podľa schémy. Všetky tieto systémy sme spojili iba do jedného. Spojili sme kód a vytvorili tak postupnosť v našom kóde. Pri vytváraní tohto zapojenia nám vzniklo mnoho chyb, ktoré sme museli premerať vymeniť a prerobiť. Najhorším problémom bol nefunkčný zdroj a vymenenie polarity z diód na tranzistory. Na zdroji sme nakoniec vymenili NTC termistor, čo je súčiastka, ktorá mení svoj odpor podľa teploty

2.5.15 Výber správnych zariadení a súčiastok

Súčiastky a zariadenia sme vyberali podľa potrebných funkcií a podľa ceny. Preto niektoré zariadenia nie sú dokonalé, nám to však stačí a pri druhej verzií tejto práce by sme určite postupovali iným spôsobom. Aj cez to u nás snímače a ostatné zariadenia majú celkom rýchly reakčný čas, okrem klávesnice, ktorá z dôvodu vykonávania viacero funkcií zaniká. Dôležitým parametrom pre výber bolo určite určenie potrebných rezistorov pre LEDky. Tu sme počítali takzvaný predradný odpor LED diódy. Pre niektoré obvody bol nutný 1,5kohm, pre niektoré iba 150ohm. Všetko záležalo od množstva LED diód v sérií danej farby.

2.5.16 Rozpočet a použitý materiál

Prácu bolo náročné a zdĺhavé priviesť do stavu funkčnosti. Keďže sme išli od základov, na práci sme museli vytvoriť potrebné schémy a dizajny pre budovy. Problémy bolo treba odstrániť a pozmeniť viacero vecí. Maximálny rozpočet na prácu bol do 400€. Naša práca má cenu bez práce s DPH 372,90€. Celkový čas strávený na práci sa rovná 683 hodín. Najviac času sme strávili pri zhotovení, bolo to presne 307 hodín. Ďalším dlhým procesom bolo napísanie kódu, jeho upravenie, návrh zariadení a ich následne spojenie do jedného veľkého celku, to nám zabralo cez 250 hodín. Pri vytváraní práce sme mali aj menšie prestávky, tie sme však odrátali a vždy sme si zapisovali správny čas na práci. Ďalšie hodiny sa aktualizujú pri vylepšení zariadení optimalizácií hlavného riadiaceho kódu. Taktiež aj pri vytváraní nových systémov.

3 Záver

V dôsledku kombinácie znalostí z letectva, elektrotechniky a programovania sme sa rozhodli vytvoriť komplexný model letiska Milana Rastislava Štefánika, známeho aj ako Bratislavské letisko. Tento projekt nevznikol len ako technická výzva, ale aj ako potenciálna pomôcka pre nových pilotov, ktorí by mohli využívať model na vzdelávanie a precvičovanie základných systémov používaných na letiskách. Ako pilot s ambíciou stať sa inštruktorom vidíme v tomto modeli potenciálne nástroj na efektívnu výučbu.

Naším hlavným cieľom bolo nielen vytvoriť funkčný model, ale aj prispieť k autenticite projektu prostredníctvom detailného dizajnu každej štruktúry a implementácie osvetlenia pre dráhy a rôzne budovy, vrátane hangárov a veže. Ovládanie modelu sme úspešne zabezpečili pomocou Arduino MEGA, pričom jeho funkcie sú prehľadne zobrazené na LCD displeji s vlastnou klávesnicou.

Významným prínosom tejto práce je aj zabezpečenie bezpečnosti prostredníctvom RFID modulu pre autentifikáciu používateľov. V rámci budúcich plánov máme v úmysle pridanie nových prvkov, rozšírenie funkcií a jemné úpravy existujúcich komponentov. Táto verzia letiska predstavuje náš prvý krok smerom k celkovej simulácii, pričom sme pripravení na ďalší vývoj a zdokonaľovanie nášho projektu

Poďakovanie

Týmto sa chcem poďakovať svojim sponzorom: Progresus

Attack AIR, Senec Červený majer letisko; Mstar s.r.o. Elektrárenská 4, 831 04 Nové Mesto; Schneider Electric Slovakia: Slovensko. Tí mi pomáhali pri práci, či už finančne, materiálne fotografiami, ale aj vysvetľovaním a poradenstvom. Taktiež sa chcem poďakovať aj svojim rodičom, príbuzným a kamarátom, ktorí tiež podali svoju pomocnú ruku pri problémoch a návrhoch.

Literatúra

[1] MOUCHA, O. 2022. Učebnica Pilota 2022. In: NEJEZCHLEBA, J., SOUČEK, L., DOUDA, O. Praha 402. publikace. ISBN 978-80-7573-101-2.
[2] Proaviate 2019, Krajná 29, Bratislava, vyučovacie materiály.
[3] Progresus Attack Air, 2023. študijné materiály. In: progressus attack air, Senec
[4] Arduinogetstarted.com copyright 2018 – 2024, kódy, schémy a materiál.