CRNOGORSKA AKADEMIJA NAUKA I UMJETNOSTI

UNIVERZITET CRNE GORE

162

Naučni simpozijum

NAUKA I MALE ZEMLJE: SINERGIJA DIJASPORE, MATICE I PRIJATELJA CRNE GORE

2023

CRNOGORSKA AKADEMIJA NAUKA I UMJETNOSTI
MONTENEGRIN ACADEMY OF SCIENCES AND ARTS

Naučni simpozijum

NAUKA I MALE ZEMLJE: SINERGIJA DIJASPORE, MATICE I PRIJATELJA CRNE GORE

2–3. novembar 2022, Podgorica, Crna Gora

Organizatori

Crnogorska akademija nauka i umjetnosti
Univerzitet Crne Gore

Scientific Symposium

SCIENCE AND SMALL COUNTRIES: SYNERGY OF THE DIASPORA, THE MOTHERLAND & FRIENDS OF MONTENEGRO

November 2 & 3, 2022, Podgorica, Montenegro

Organizers

Montenegrin Academy of Sciences and Arts
University of Montenegro

MONTENEGRIN ACADEMY OF SCIENCES AND ARTS
UNIVERSITY OF MONTENEGRO

Scientific Meetings
Volume 162

Scientific Symposium

SCIENCE AND SMALL COUNTRIES: SYNERGY OF THE DIASPORA, THE MOTHERLAND & FRIENDS OF MONTENEGRO

November 2–3, 2022, Podgorica, Montenegro

Editor
Academician Predrag Miranović

Podgorica
2023

CRNOGORSKA AKADEMIJA NAUKA I UMJETNOSTI
UNIVERZITET CRNE GORE

Naučni skupovi
Knjiga 162

Naučni simpozijum

NAUKA I MALE ZEMLJE: SINERGIJA DIJASPORE, MATICE I PRIJATELJA CRNE GORE

2. i 3. novembar 2022, Podgorica, Crna Gora

Urednik
Akademik Predrag Miranović

Podgorica
2023.

Prihvaćeno na sjednici Predsjedništva Crnogorske akademije nauka i umjetnosti, 22. marta 2023. godine.

Adopted at the Session of the Presidency of the Montenegrin Academy of Sciences and Arts, on 22 March, 2023.

OTVARANJE

OPENING

Akademik Dragan K. VUKČEVIĆ,
predsjednik Crnogorske akademije nauka i umjetnosti

RIJEČ NA OTVARANJU NAUČNOG SIMPOZIJUMA

Poštovani prijatelji, uvaženi rektore Univerziteta Crne Gore profesore Božoviću, poštovana ministarko nauke u Vladi Crne Gore profesorice Šćepanović, cijenjeni profesore Šehtman, kolege akademici i profesori univerziteta, dragi gosti,

Čini mi čast da priložim svoju riječ u ovom svojevrsnom sabranju umnosti i učenosti. Kasno smo kao društvo i država zakoračili na put koji vodi iz sela u grad, sa njive u fabriku i iz crkve u školu. Cijenu zakašnjele moderne plaćamo i danas. Zato je, po mom sudu, neznanje najveći problem savremene Crne Gore.

Teško shvatamo da su institucije krovovi pod kojima se oblikuje kultura jednog društva i njegove države, i teško prihvatamo da je planiranje elita oblikovanje sopstvene budućnosti. Zbog nedostatka tih institucija naše nekadašnje elite oblikovane su i stvarane van Crne Gore.

Duga je i široka rijeka onih darovitih ljudi kojima smo oticali iz sebe. Mnogi od onih, koji su danas ovdje i svojim dolaskom učinili čast i Crnogorskoj akademiji nauka i umjetnosti i Univerzitetu Crne Gore i Crnoj Gori, potomci su tih darovitih ljudi. Želim da vam izrazim iskreno i duboko poštovanje i zahvalnost što ste danas ovdje. Vjerujem da samo sa boljima od nas pored sebe možemo postati bolji od onih kakvi jesmo.

Zahvalnost dugujem i profesoru Veljku Milutinoviću, inostranom članu Crnogorske akademije nauka i umjetnosti, na ideji i na organizovanju ovog skupa. I on je sin jednog od tih darovitih ljudi, koji su svojevremeno svoje uspjehe oblikovali u institucijama van Crne Gore, ne zaboravljajući odakle su krenuli. Ali, polako i, rekao bih, nepovratno u Crnoj Gori sazrijeva svijest da je kultura posljednja odbrana države, da je profesionalizam najbolji oblik patriotizma i da je ulaganje u znanje najveća zaloga razvoju. Samo na takvom putu moguće je težiti razvijenom svijetu, biti njegov dio, a pritom ne izgubiti sebe.

Ja vjerujem da je ovaj skup značajan korak na tom putu. Crnogorska akademija nauka i umjetnosti ostaje otvorena za ovakve susrete u dolazećem vremenu.

Našem pozivu odazvao se i profesor Dan Šehtman.

Dear Friends,

It is our great pleasure to welcome in our country the Nobel Laureate in Chemistry in 2011 for the discovery of quasicrystals, Professor Dan Shechtman, from Israel.

Professor Shechtman,

You belong to the people who has paid a great price of its survival throughout history. In that perseverance through time, your people has created a lot defending itself with its own culture. Similar was the path of survival of Montenegro in history.

I hope that there will be other opportunities to meet again on the occasion of such gatherings, which represent a special celebration of knowledge.

Poštovani učesnici, vi koji ste došli na ovaj skup, možda možete bez Crne Gore, ali znajte da ona zna da je bolja sa vama i da su vam njena vrata otvorena. Hvala.

Prof. dr Vladimir BOŽOVIĆ,
rektor Univerziteta Crne Gore

RIJEČ NA OTVARANJU NAUČNOG SIMPOZIJUMA

Uvaženi profesore Šehtman, predsjedniče Crnogorske akademije nauka i umjetnosti profesore Vukčeviću, ministarko Šćepanović, uvaženi rektori, akademici, ekselencije, drage kolege i gosti,

Ova konferencija, iako joj naslov određuje okvir, nema samo naučni karakter. Ima širi kulturološki, a prije svega motivacioni značaj. Ona je, po svom sadržaju, a potpuno sam siguran i po odjecima koje će imati, zlata vrijedna poruka — da možemo. Da, možemo!

Vrijeme je da preokrenemo perspektivu pogleda na naše mogućnosti.

Naša veličina ne može biti utisnuta samo u „suvim" brojevima, već iznad svega u ideji koju nosimo. Vjerujem da Crna Gora ima potencijal koji daleko nadmašuje statističke parametre kojima se suvoparno enciklopedijski opisuje. Ovdje jeste izvor dobrog duha, britkih ideja, ali i tradicije koju moramo brižljivo čuvati u izazovnim vremenima, njegovati od najranijeg uzrasta. Samo tako se gradi pečat snažne kulture, kojom su se i neki mali narodi vinuli visoko na ljestvici opštečovječanskih dostignuća. Naučnih i svih ostalih.

Naša snaga i značaj se može mjeriti i time koga možemo da privučemo i okupimo. Dovoljno je pogledati spisak onih koji će ovdje govoriti i jasno je — u pitanju je naučni događaj elitnog svjetskog nivoa. Time svjedočimo njegoševsku poruku, misao lovćenskog genija iz nešto drugačijeg ugla. Onoliko koliko je lako iz grmena velikog izaći na svjetsku pozornicu, toliko je teško iz male zemlje skromne naučne tradicije. Zaista je teško, ali je moguće, potvrđuju glasovi nauke i znanja, okupljeni u ovoj sali, na ovoj konferenciji.

U tom duhu, a očekujući plodove i ovog današnjeg okupljanja, bila je i moja porukama mladim kolegama, koji su ove godine zakoračili na Univerzitet

Crne Gore. Zamolio sam ih da budemo odgovorni prema darovima koje imamo. Da ih ne protraćimo i ne dopustimo da budemo radna snaga za ostvarenje tuđih snova. To što smo mala zajednica, nikako ne znači da ne treba da se borimo kako bi bili najbolji u onom što radimo. Možda je nama zato teže, ali su zato plodovi uspjeha slađi i dugotrajniji.

Vrijednost nauke ne smije biti u znanju koje se hladno niže u spratovima vremena, bez topline za život društva, bez odgovornosti za čovjeka i njegove potrebe. Mi ne želimo visokotehnološku, apatičnu i hladnokrvnu civilizaciju, u kojoj je čovjek puka statistička jedinica, a negdje čak i višak. Duboko vjerujem da je nauci potreban danas povratak smislu u humanističkom okviru, što je i njeno rodno mjesto. Isto toliko vjerujem da je ova konferencija korak u tom pravcu.

Ako svake godine budemo okupljali naučnu elitu kao što smo to uradili, zahvaljujući ogromnom entuzijazmu i trudu uvaženog profesora i prijatelja Veljka Milutinovića, onda Crna Gora neće biti „samo“ divna turistička zemlja, već ćemo ovim silnim planinama dodati i neke eminentne naučne vrhove.

Dear professor Shechtman,

On behalf of the University of Montenegro, and my personal name, we owe a special appreciation to you for the time you have dedicated to show support to our scientific and academic community. Thus, in line with the title of the Symposium, you are proving that you are a true friend of this University and Montenegro in general. Thank you for coming.

Zahvaljujem svima i želim uspješan i plodan rad.

Prof. dr Biljana ŠĆEPANOVIĆ,
ministarka nauke i tehnološkog razvoja u Vladi Crne Gore

CANU — UCG, NAUČNI SIMPOZIJUM

Poštovani predsjedniče Crnogorske akademije nauka i umjetnosti profesore Vukčeviću, poštovani rektore Univerziteta Crne Gore profesore Božoviću, poštovani profesore Šehtman, uvaženi članovi Predsjedništva CANU i Rektorskog kolegijuma UCG, poštovani gosti i učesnici skupa sa raznih meridijana, uključujući crnogorski,

Hvala vam što smo se danas okupili u Crnoj Gori, povezani naukom i samom Crnom Gorom, u smislu života i rada u njoj, crnogorskog porijekla ili saradnje i jakih prijateljskih veza sa Crnom Gorom.

Jesmo mala država — teritorijom, populacijom... Ali, ta mala država ima svoje djeliće, svoje predstavnike širom svijeta, koji govore univerzalnim jezikom — jezikom nauke. I srećna je što je nijesu zaboravili, što ne samo da održavaju i jačaju svoje veze sa njom, već i pomažu da uspostavi naučnu konekciju sa drugim državama, da se crnogorske naučnoistraživačke ustanove i istraživači povežu i sarađuju sa svojim kolegama širom svijeta.

Jaka naučna dijaspora, osim izuzetnog značaja za matičnu državu, podrazumijeva i veliku odgovornost za tu državu. Usudiću se da kažem: što je država manja, to je ova odgovornost veća. Veze i saradnju ne može njegovati samo jedna strana. Kao što rekoh, mi smo u Crnoj Gori srećni što je naša dijaspora otvorena za saradnju, ali o toj saradnji moramo pažljivo brinuti i sami, kako na ličnom, tako i na institucionalnom i državnom nivou.

Stoga ideja CANU i UCG za ovakvo okupljanje, te napori za njenu realizaciju, zaslužuju izuzetno poštovanje, a ja sam počastvovana što danas imam priliku da vam se obratim u ime MNTR, koje pruža punu podršku ovakvim inicijativama. U jednu rečenicu sam smjestila 3 vodeće institucije kad je u pitanju nauka Crne Gore. Ali, ne treba da zaboravimo i zanemarimo ostale subjekte i aktere naučnoistraživačke djelatnosti u našoj državi. Kao građevinskom inženjeru, dozvolićete mi ovakvo poređenje — svi smo mi u jednoj zajedničkoj kući nauke i svi imamo svoje mjesto, svoj značaj i svoju ulogu u njoj. Za korisnika te kuće, za udobnost njegovog života, svaka prostorija u njoj je značajna, budući da imaju različite funkcije i nema smisla pitati ga koja mu je najznačajnija. S druge strane, ono o čemu korisnik ne razmišlja stalno, a od suštinskog je značaja za kvalitet njegovog života, jeste sigurnost konstrukcije te kuće, za koju je značajna uloga svih njenih konstruktivnih elemenata — krova, stubova, greda, temelja... Svaki od njih ima svoj posao i mora ga obavljati valjano da bi kuća opstajala i služila svojoj svrsi.

Nadam se da mi naučnoistraživačka zajednica neće zamjeriti ovo poređenje institucija i subjekata sa prostorijama ili elementima konstrukcije kuće. Suština je u ključnoj riječi iz naziva ovog okupljanja — sinergija — zajednički rad, udruživanje snaga, pogotovo kad su u pitanju male države, pogotovo kad je u pitanju zajednica u kojoj su naše uloge isprepletane. Sinergija nam je neophodna, kako unutar matice, tako u i odnosima sa dijasporom i prijateljima. Stoga me izuzetno raduje što smo se danas okupili u najljepšem sastavu — matica, dijaspora i prijatelji Crne Gore, da govorimo zajedničkim jezikom budućnosti — jezikom naučnog istraživanja.

U ime Crne Gore, ponosna sam što danas pozdravljam punu salu, a posebno me raduje što već planiramo ovakve skupove za naredne godine.

Još jednom bih zahvalila svim našim gostima što su svojim odzivom omogućili uspješnu organizaciju skupa, kao i organizatorima što su sve pokrenuli. Želim vam uspješan rad danas i sjutra, i dugu tradiciju ovakvih manifestacija u narednom periodu.

MEDICINSKE NAUKE

MEDICAL SCIENCES

Dušica BABOVIĆ-VUKSANOVIĆ*

GENOMICS: MODERN FORTUNE TELLER

Abstract: Genetic science has evolved over the last century and has become integral part of everyday medical practice. Instead of its initial use as a diagnostic tool to explain condition or confirm clinical suspicion, we see genomics more and more providing information to be used in predictive and preventive purposes. Genetic testing is now widely available, as the price of testing has become more affordable. Determining genetic predisposition for disease often result in preventive measures that can avoid development of disease or in screening programs that can detect disease in early stage when treatment is more effective. In addition, genetic information may guide treatment by selecting appropriate medication and dose for the individual patient. Genetics may be seen as a modern fortune teller!

Key words: *genetics, genomics, sequencing, predictive medicine, preventive medicine*

INTRODUCTION

Prevention and cure of disease, longevity and immortality have been human dream for centuries. They used culturally specific spiritual, religious theories and self-educated healers and fortune-tellers. As environmental factors, such as availability and quality of food, wars, infections and toxic exposures have been gradually eliminated as primary influencers in survival and human wellbeing, at least in the developed world, more and more attention is now directed to genetic factors as main determinants of our outcome.

The discovery of the double helix structure of deoxyribonucleic acid (DNA) by James Watson and Francis Crick in 1953 marked a milestone in the history of science. Only 50 years after their discovery, the genetic

* Dušica Babović-Vuksanović, MD, Professor of Medical Genetics and Pediatrics; Consultant, Department of Clinical Genomics, Mayo Clinic, Rochester, MN, USA; Director of Neurofibromatosis Comprehensive Program, Mayo Clinic

code was revealed by completion of the Human Genome Project in 2003. The $3 billion, 15 year endeavor described the 2.85 billion nucleotides and 20,000–25,000 protein-coding genes comprising a human genome. Since that breakthrough, the last quarter century has been marked with fast and productive development of technology, bringing genomics into medical practice and revolutionizing how we practice medicine.

MAIN TEXT

Medical practice has for long time followed the classic model of reactive medicine, where a sick patient seeks treatment for symptom relief. In this way many conditions are diagnosed too late, when treatments are not effective. Drugs have typically been given based on patient's age, weight and possible drug interactions. Now, with emerging genomic information, the concept of precision medicine and individualized approaches in diagnosis and treatment has been rapidly advanced. We are trying to predict susceptibility in order to prevent disease and develop novel, targeted treatments. We want to move away from fortune-teller strategies.

The development of rapid sequencing methods coupled with lower price and improved computing models for analysis, allowed for increasing utilization of genetic information and set the stage for personalized medicine. Now, large gene panels or whole exome or genome sequencing has become the first line diagnostic test for many phenotypes. Once the genetic cause of disease is identified in the individual patient, it may serve as tool for predictive testing for others in the family. This can facilitate early detection and more effective therapy or even preventive measures. Familial cancer syndromes, such hereditary breast and ovarian cancer, Lynch syndrome, and Li-Fraumeni syndrome are all cardinal examples.

As sequencing tests analyze more and more genes, the possibility of detecting incidental findings has increased, and the debate was opened about appropriateness, ethical and legal aspects of reporting the secondary finding. The American College of Genetics and Genomics (ACMGG) has issued recommendations for actionable conditions that should be reported to patients, which could be considered as the opportunistic screening for selected variants that could influence medical management [1]. These guidelines have been widely adopted by diagnostic laboratories.

Pharmacogenomic (PGx) applications are rapidly becoming an integral part of modern practice with many validated examples of the clinical utility. Both common and rare pharmacogenomics variants may inform individualized drug selection and dosing. This type of clinical genomic information is increasingly being generated in clinical laboratories, incorporated into electronic health records, and used to „tailor" or individualize drug therapy [2]. The inclusion of PGx results as a secondary (actionable) findings in WES testing is still not a common practice due to additional challenges related to technical problems (pseudogenes, haplotypes composed of several variants, incomplete coverage of intronic sequence), complicated „star allele" nomenclature and challenges in implementing PGx results into wide clinical practice due to lack of competencies in interpretation of these results. Also, some have argued that PGx results are not actionable until the decision is made to prescribe a drug related to the particular variant [1]. On the other hand, our experience has suggested that these incidental findings are often highly relevant. IN our series of WES results on 94 young patients with median age of 10 years, a subset of PGx variants in *CYP2C19*, *CYP2C9*, and *VKORC1* were reported. The majority of tested patients had dominant neurological pathology (71%), and was currently taking medications (90%). Of the 94 PGx-evaluated patients, 91% had at least one variant allele reported and 20% had potential immediate implications on current medication use [3]. Therefore, decision about selecting the patient-appropriate test, with or without components such as PGx or secondary results should take in account multiple factors including indication for testing and clinical scenario.

There is also a growing interest for application of genomics in predictive testing of healthy individuals and population screening, which remains controversial due to high risk of result misinterpretation especially in direct-to-consumer genetic testing, without involvement of qualified genetic This concern led to the ACMG Statement on Direct-to-Consumer Genetic Testing, highlighting that it is critical for the public to realize that genetic testing is only one part of a complex process that includes genetic risk assessment, diagnosis, and disease management [4]. Due to potential for performing DTC testing in minors, the AAP and ACMG published the statement which strongly discourage the use of direct-to-consumer and home kit genetic testing of children because of the lack of oversight on test content, accuracy, and interpretation. They also highlighted need for caution with predictive and carrier testing in children [5, 6, 7].

Despite rapid development of genomic-based medicine and widespread use of genetic information in medical practice, it cannot replace expert clinical phenotyping and use of classical medical and genetic principles. The family history analysis had been a primary tool that physicians have used to estimate genetic predisposition for illness, and detailed phenotyping has been a foundation for clinical diagnoses. These tools continue to play a fundamental role in a genetic practice. Development of new automated tools for analysis of genetic information did not eliminate them, but has even increased its value in interpretation of complex or ambiguous genetic data, which can only be understood based on correlation with clinical information and segregation of disease phenotype with genetic change [8].

Lately, we are facing spectacular headlines of human ability to clone animals (maybe including humans), correct genetic alteration in embryos and perhaps even be able to choose certain traits in our offspring by manipulating DNA. There remains constant argument between pro-futuristic and risk-taking proponents who see the genomic progress as a holly-grail of human outcomes and medical progress, and those who express concerns about consequences of premature overuse of genomic information and lack of scientific evidence to guide its application. The healthy and balanced argument is needed to crystalize directions and allow comfortable entry of genomics in radical change how we will practice medicine in the future [8].

With all of the fascinating and rapid progress in genomic science, the education of providers and the public regarding appropriate use, limitations, risks and benefits is paramount and one of most important factors in bringing genomic medicine to full utilization. Rapid pace of these advances is opening the gap between the knowledge available about the clinical relevance of genomic information and the ability of clinicians to include such information in their practices. The educational gap threatens to be rate limiting to the clinical adoption of genomic medicine. This prompted National Human Genome Research Institute to initiate development of genomic practice competencies for physicians in various medical disciplines [9].

CONCLUSION

Wise and evidence-based use of genomic tools will certainly continue to transform medical practice and take us much further away from crystal ball fortune-tellers and towards modern, safe and reliable personalized medicine.

REFERENCES

[1] Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics *Genet Med* 2017 Feb; 19(2): 249–255. doi: 10.1038/gim. 2016.190. Epub 2016 Nov 17.

[2] Weinshilboum RM and Wang L. Pharmacogenomics: Precision Medicine and Drug Response. *Mayo Clin Proc*. 2017 Nov; 92(11): 1711–1722. doi: 10.1016/j. mayocp. 2017.09.001. Epub 2017 Nov 1. Review.
[3] Cousin MA, Matey ET, Blackburn PR, Boczek NJ, McAllister TM, Kruisselbrink TM, Babovic-Vuksanovic D, Lazaridis KN, Klee EW. Pharmacogenomic findings from clinical whole exome sequencing of diagnostic odyssey patients. Mol Genet Genomic Med. 2017 Mar 19; 5(3): 269–279. doi: 10.1002/mgg3.283. eCollection 2017 May. PMID: 28546997
[4] ACMG Board of Directors: Direct-to-consumer genetic testing: a revised position statement of the American College of Medical Genetics and Genomics *Genetics in Medicine* 2016(18): 207–208.
[5] COMMITTEE ON BIOETHICS, COMMITTEE ON GENETICS, AND, THE AMERICAN COLLEGE OF MEDICAL GENETICS AND, GENOMICS SOCIAL, ETHICAL, AND LEGAL ISSUES COMMITTEE. Ethical and Policy Issues in Genetic Testing and Screening of Children, *Pediatrics,* 2013, 131 (3) 620–622.
[6] Ross LF, Saal HM, David KL et al. Technical report: Ethical and policy issues in genetic testing and screening of children. *Genet Med* 2013: 15(3): 234–245.
[7] Brothers KB, Vassy JL, Green RC. Reconciling Opportunistic and Population Screening in Clinical Genomics. Mayo Clin Proc. 2019 Jan; 94(1): 103–109. doi: 10.1016/j. mayocp. 2018.08.028. PMID: 30611438; PMCID: PMC6339566.
[8] Babovic-Vuksanovic D. Genomics as a Scientifically Based Fortune-teller. Mayo Clin Proc. 2019 Jan; 94(1): 7–9. doi: 10.1016/j. mayocp. 2018.11.008. PMID: 30611456.
[9] Korf BR, Berry AB, Limpson M et al. Framework for development of physician competencies in genomic medicine: report of the Competencies Working Group of the Inter-Society Coordinating Committee for Physician Education in Genomics. *GenetMed.* 2014 Nov; 16(11): 804–9. doi: 10.1038/gim. 2014.35. Epub 2014 Apr 24.

Dušica BABOVIĆ-VUKSANOVIĆ

GENOMIKA: MODERNI PREDSKAZIVAČ BUDUĆNOSTI

Sažetak

Genetika je doživjela veliki napredak u zadnjih 100 godina i postala integralni dio svakodnevne medicinske prakse. Umjesto upotrebe genetskih testova samo za postavljanje dijagnoze, danas se genetska informacija sve više koristi u preventivne i prediktivne svrhe. Sa smanjenjem cijene testova, genetsko testiranje je sada dostupno u svakodnevnoj praksi. Saznanje o nečijoj genetskoj predispoziciji može da rezultira u prevenciji bolesti ili u uspostavljanju dijagnoze u ranom stadijumu bolesti kada je izlječenje još moguće. Nadalje, genetska informacija može da služi za izbor lijeka i doze koja odgovara svakom individualnom pacijentu. Genetika postaje moderni predskazivač budućnosti!

Ključne riječi: *genetika, sekvenciranje, prediktivna medicina, preventivna medicina*

Antonija KRSTAČIĆ*

AMIOTROFIČNA LATERALNA SKLEROZA — „LOU GEHRIGS DISEASE“

Sažetak: Amiotrofična lateralna skleroza (ALS) ili Charcotova bolest po francuskom neurologu, koji je među prvima opširnije opisao bolesti i izdvojio je ALS od drugih mišićnih atrofija, ili Lou Gehrigova bolest dobila je ime po američkom igraču bejzbola oboljelom od ove bolesti. Lou Gehrig je na vrhuncu svoje slave obolio i preminuo od ove bolesti 1941. ALS se karakterizira progresivnom degeneracijom i gubitkom motornih neurona sa ili bez slične lezije motornih jezgri mozga uz zamjenu izgubljenih stanica gliozom. Spada u skupinu bolesti motornog neurona te zahvaća i gornji i donji motorni neuron. Može se javiti familijarno ili sporadično. Obiteljska incidencija ALS-a opisana je u rijetkim publikacijama i u literaturi je imala ograničenu pozornost sve do izvješća Kurlanda i Muldera iz 1955, koji sugerira da ALS može biti obiteljski u gotovo 10% slučajeva. Sve veći broj dokaza iz kliničkih istraživanja sugerira da ALS ima više uzroka s važnom, iako raznolikom, genetskom komponentom. Dostupni podaci sugeriraju da genetski rizik za ALS vjerojatno predstavlja kombinirane učinke jednog ili više gena koji utvrđuju ukupnu genetsku osjetljivost osobe, djelujući u sprezi s okolišnim i slučajnim učincima, koji dovode do pojave bolesti.

Ključne riječi: *amiotrofična lateralna skleroza, Lou Gehrigova bolest, bolest motornih neurona*

Amiotrofična lateralna skleroza (ALS) je neurodegenerativna bolest, koju karakterizira progresivna mišićna slabost. Opisana je još u devetnaestom stoljeću, a poznatija je postala kada je američki igrač bejzbola Lou Gehrig obolio od nje sredinom dvadesetog stoljeća. „Amiotrofična“, navodi Charcot, termin je koji se odnosi na denervacijsku atrofiju mišića zbog degeneracije neurona prednjeg roga medulle spinalis. Termin „lateralna skleroza“ odnosi se na sklerozu kortikospinalnog puta uslijed degeneracije motornih

* Antonija Krstačić, Fakultet za dentalnu medicinu i zdravstvo Osijek, Sveučilište Josipa Jurja Strossmayera u Osijeku; Medicinski fakultet Osijek, Sveučilište Josipa Jurja Strossmayera u Osijeku; KBC „Sestre milosrdnice“, Klinika za traumatologiju, Zagreb

neurona u primarnog motornog korteksa i zamjene sa glija stanicama. Spada u skupinu bolesti motornog neurona te zahvaća i gornji i donji motorni neuron. Prvi znak bolesti može biti asimetrična slabost i atrofija mišića ekstremiteta ili se najprije javljaju bulbarni simptomi. Bolest je progresivna i atrofija postepeno zahvaća sve mišiće. Disfagija, fascikulacije, slabost mišića i disartrija najčešći su simptomi, a oboljeli najčešće umiru zbog respiratorne insuficijencije. Bulbarni oblik bolesti brže progredira i nakon 2 do 3 godine dolazi do zatajivanja disanja, dok je kod spinalnog oblika bolesti preživljavanje nešto duže, 3 do 5 godina. [1, 2]

Može se javiti familijarno ili sporadično. Prosječna dob početka bolesti je 56 godina. Incidencija ALS-a u Evropi je 2–3/100.000 u općoj populaciji. ALS u 90% slučajeva nastaje sporadično (SALS), dok je 10% familijarno (FALS). Današnja saznanja govore u prilog hipotezi o složenom gensko-okolišnom međudjelovanju kao uzroku degeneraciji motornih neurona. Uzrok nastanka ALS-a nije poznat, ali smatra se da, uz genetsku mutaciju, on može biti i oksidativni stres ili toksičnost glutamata, poremećaj imunoloških procesa, djelovanje neurotrofičnih čimbenika i čimbenika okoline (teški metali, pesticidi, trauma). Epidemiološka istraživanja uočila su moguću povezanost između ALS-a i pušenja, mehaničke traume, fizičke aktivnosti, izlaganja elektromagnetnom polju, ionizirajućeg zračenja, utjecaja pesticida i herbicida, utjecaja virusnih infekcija, teških metala olova, žive, aluminija, cinka, selenija. [3–9] Primjena molekularno genetskih tehnika na ALS-u, obilježena izvješćem iz 1993. godine o povezanosti gena superoksid dismutaze 1 (SOD1) u obiteljskom ALS-u, signalizirala je sve veći fokus na genetiku u ALS-u kao sredstvo za stjecanje uvida u patogenezu bolesti, identificiranje terapijskih ciljeva i olakšavanje dijagnoze. [10] Sve veći broj dokaza iz kliničkih istraživanja sugerira da ALS ima više uzroka s važnom, iako raznolikom, genetskom komponentom. Genetski čimbenici u ALS-u kreću se od visoko penetrantnih varijanti gena povezanih s ALS-om do varijanti sekvence s naizgled ograničenim utjecajem na podložnost bolesti. Fenotipovi povezani s ovim varijantama sekvence uključuju klasični ALS, primarnu lateralnu sklerozu (PLS) i progresivnu mišićnu atrofiju (PMA). Važna „ekstramotorna" značajka, povezana s nekim varijantama gena povezanih s ALS-om, frontotemporalna je demencija (FTD), koja se može razviti uz, prije ili nakon pojave motoričkih znakova kod ALS-a, te samo kao FTD. Manje uobičajene kliničke značajke povezane s nekim varijantama gena povezanih s ALS-om uključuju ekstrapiramidne značajke i miopatiju inkluzijskih tjelešaca. Dok je obiteljski ALS uglavnom poremećaj s početkom u odraslih, nekoliko gena povezanih s ALS-om može imati fenotipove karakterizirane juvenilnim početkom. Sve se više, kao zasebni oblik ALS-a, spominje i juvenilni ALS (JALS), a koristi se za osobe oboljele prije 25. godine života.

U 20% familijarnih oblika i u 5% sporadičnih oblika ALS-a nađena je mutacija gena koji kodira enzim Zn/Cu superoksid dizmutazu (SOD1), snažni antioksidans, poznat kao ALS1. Ostali geni koji su najčešće uključeni u ovaj oblik ALS-a su C9orf72, FUS (ALS6) i TARDBP (ALS10). ALS se u 5–10% slučajeva pojavljuje familijarno (FALS) i tada je odnos oboljelih muškaraca i žena 1:1. Najčešće se nasljeđuje autosomno dominantno, no postoje autosomno recesivni i X-vezani oblici bolesti. Cilj genetičkih istraživanja je otkrivanje gena odgovornog za nastanak bolesti. U 20% pacijenata s autosomno dominantnim oblikom FALS-a i u 2% sa sporadičnim oblikom ALS-a utvrđena je mutacija gena na 21. kromosomu (21q22.1), koji kodira enzim Cu/Zn superoksid dizmutazu (SOD1). SOD1 je snažan antioskidans, koji štiti stanicu od slobodnih radikala, koji su visoko reaktivni i štetni stanični proizvodi i mogu uzrokovati preranu smrt stanice. SOD1 gen bio je prvi opisan mutirani gen u ALS-u, otkriven 1991. godine. Mutacija gena umanjuje enzimsku aktivnost SOD1 za oko 50% i nastaje sporo propadanje neurona. Istraživanja su pokazala da je dio mutacija lokaliziran u mitohondrijima u kojima tvori proteinske agregate štetne za neurone. U kralježničnoj moždini oboljelih od ALS-a nađene su biokemijske i morfološke mitohondrijske abnormalnosti. Mitohondrijska disfunkcija ne nalazi se samo u neuronima, već i u drugim tkivima, poput skeletnih mišića. Nađeno je preko 100 mutacija SOD1 gena. [11–15] „Chromosome 9 open reading frame 72“ (C9ORF72) / Ponavljanje heksanukleotida GGGGCC u prvom intronu gena koji kodira protein nepoznate funkcije na kromosomu 9, C9ORF72, najčešća je varijanta gena povezana s FALS-om, pronađena u 40% FALS-a i oko 6–8% pacijenata sa SALS-om. Povećanje ovih ponavljanja od ≤ 23 kod zdravih ljudi do 1000 kod oboljelih.

Fenotipovi „C9FTD/ALS“ uključuju klasični ALS (rijetko PMA ili PLS), ALS/FTD i FTD, kao i fenotipovi Parkinsonove i Huntingtonove bolesti, koji ne reagiraju na dopu. FTD ili manje teško frontotemporalno kognitivno oštećenje u bolesnika s C9FTD/ALS-om s ALS-om može se pojaviti prije ili nakon pojave motoričkih znakova u do 50% bolesnika. Nasljeđivanje je autosomno dominantno s nepotpunom penetracijom; srednja dob početka bolesti je 58 godina, u rasponu od 4. do 9. desetljeća.

Identifikacija varijanti TARDBP-a u bolesnika s ALS-om uslijedila je nakon otkrića iz 2006. godine da su neuronske citoplazmatske inkluzije imunoreaktivne na ubikvitin, patološki znak u većini slučajeva FALS-a i SALS-a, također su imunoreaktivne na TDP-43. Oko polovice bolesnika s patološki dokazanom frontotemporalnom lobarnom degeneracijom (FTLD, patološka osnova za klinički sindrom FTD) imaju slične TAR DNA binding protein 43 (TDP-43). TDP-43 imunoreaktivne inkluzije uspostavile su patološku vezu između ALS-a i FTD-a i dovele do koncepta da ALS, ALS-FTD i FTD predstavljaju klinički i patološki spektar, koji se naziva TDP-43 proteinopatije. Varijante gena u TARDBP-u, koji kodira 43-kD TAR DNA vezujući protein 43 (TDP-43), nalaze se u približno 4% FALS-a i 1% SALS-a, s određenim regionalnim varijacijama. TDP-43 regulira ekspresiju gena i spajanje RNA-a. Dostupni dokazi upućuju na to da se disregulacija ekspresije gena, uključujući spajanje RNA-a, pripisuje patogenim varijantama TARDBP-a, u kombinaciji s toksičnim pojačanjem funkcije mutiranog proteina TDP-43, doprinose neurodegeneraciji.

Varijante gena (FUS) povezane su s autosomno dominantnim ALS-om u oko 4% bolesnika s FALS-om i 1% pacijenata sa SALS-om. Patogene varijante FUS-a uključuju točkaste mutacije i druge strukturne defekte, a značajne su po nekoliko izvješća koja potvrđuju de novo mutacije povezane s ALS-om. Nasljeđivanje je autosomno dominantno, osim jedne obitelji s prividnim autosomno recesivnim nasljeđivanjem. Fenotipovi ALS-a uključuju ALS s početkom u odraslih, ALS/FTD i juvenilni ALS, a rijetko kao čisti FTD. Prijavljena je i jedna obitelj s FUS „ALS-plus" sindromom s očnim, autonomnim i cerebralnim značajkama. Progresija bolesti kod juvenilnog FUS ALS-a obično je brza, bez razvoja FTD. Patološka obilježja FUS ALS-a kod odraslih u post mortem mozgu i leđnoj moždini uključuju abnormalne proteinske agregate imunoreaktivne na FUS, uglavnom u citoplazmi, ali i u jezgrama neurona i glije. Juvenilni početak FUS ALS-a pokazuje karakterističnu patologiju obilježenu neuronskim bazofilnim inkluzijama imunoreaktivnim na FUS protein; slična je patologija prijavljena u FUS FTD s početkom u odraslih.

Mutacija ubikvilina 2 (gen UBQLN2) otkrivena je kao uzrok X vezanog ALS-a i ALS-a s frontotemporalnom demencijom. Mutacija za profilin 1 (PFN) gen otkrivena je kod familijarnog oblika. Mehanizmi u patofiziologiji ALS-a i mehanizmi koji vode u patogenezu, odnosno sam početak bolesti i dalje su nepoznati, vjerojatno više od navedenih mehanizama sudjeluje u patogenezi. Pretpostavlja se da abnormalan gen ili produkt gena igra ulogu u trigeriranju početka bolesti kod nasljednog oblika i vjerojatno ima ulogu u progresiji bolesti, iako sama činjenica da postoji abnormalan gen, nije dovoljna da se razvije bolest. Dokaz tome je što neki nosioci abnormalnog gena nikad ne razviju bolest. Kao što i prisutnost normalnog gena, ne mora značiti da se ne može razviti sporadični oblik bolesti. [16–22]

Ne postoji specifični biomarker pri dijagnostici ALS-a pa se nastoje prvo isključiti sve ostale dijagnoze koje mogu imati slične ili iste simptome. Dijagnoza ALS-a temelji se na kliničkoj slici, elektrodijagnostičkom testiranju i isključenju stanja koja mogu imitirati ALS. Progresija simptoma i znakova oštećenja gornjih i/ili donjih motornih neurona jedna je od najvažnijih značajki bolesti. Kako bolest sve više napreduje, pacijent razvija karakterističnu kliničku sliku kombinacije znakova oštećenja gornjeg i donjeg motornog neurona. Klinička slika okarakterizirana je progresivnim gubitkom mišićne mase i snage, fascikulacijama, spasticitetom, hiperrefleksijom i ekstenzornim plantarnim odgovorom (pozitivan znak Babinski refleksa), kao i razvojem dizartrije i disfagije. Konačno su zahvaćeni svi voljni mišići i pacijenti gube snagu i mogućnost micanja rukama, nogama i cijelim tijelom. Ne mogu samostalno stajati, hodati, ustati iz kreveta niti samostalno jesti. Poteškoće u žvakanju i gutanju povećavaju rizik od gušenja. Kada se razvije slabost dijafragme i interkostalnih mišića, pacijenti više ne mogu samostalno disati. U završnoj fazi bolesti, pacijentu je nužno ugraditi respirator kako bi se na potpomognut način omogućilo disanje, odnosno ugradnja PEG-a / sonde za hranjenje. Većina oboljelih na kraju umire zbog respiratorne insuficijencije, najčešće unutar 2 do 5 godina od početka simptoma, no 10% pacijenata preživi 10 ili više godina. Kada se bolest ranije pojavi, može se očekivati duži vijek preživljavanja. Početkom bolesti simptomi koji se javljaju su uglavnom neprimjetni, pa su često zanemareni. Bolest je okarakterizirana progresivnom mišićnom slabošću u rukama, nogama i mišićima odgovornim za disanje, govor i gutanje. Jedan od prvih znakova zahvaćenosti gornjeg motoneurona GMN-a je gubitak fine motorike, uz razvoj spasticiteta na mišićima. Spastičnost je stanje povišenog mišićnog tonusa. Javlja se fenomen džepnog nožića. Pojačani miotatski refleksi znak su zahvaćenosti GMN-a i posljedica su nedostatne inhibicije zbog disfunkcije GMN-a, tako da je potreban mali podražaj kako bi se izazvao refleks. Prisutni su patološki refleksi, poput Babinskog i Hoffmanova znaka. Kada je zahvaćen kortikobulbarni put, koji kontrolira mišiće za govor, žvakanje i gutanje, javlja se spastička bulbarna paraliza (pseudobulbarna paraliza) uz disfagiju i dizartriju. Česta je i slaba emocionalna kontrola karakterizirana spontanim napadajima smijeha i plača. Ovaj fenomen nastaje zbog gubitka inhibicije motornih neurona limbičkog sustava koji kontroliraju primitivnu vokalizaciju (smijanje i plakanje) i naziva se pseudobulbarni afekt. Ovisno o početnim simptomima kod otprilike 2/3 oboljelih javlja se spinalni oblik bolesti kod kojega su početni simptomi slabost, a potom atrofija malih mišića šaka; u 1/3 pacijenata javlja se bulbarni oblik bolesti, koji započinje smetnjama pri gutanju i govoru dok u oko 5% slučajeva bolest započinje izoliranom slabošću respiratornih mišića, koja se manifestira dispnejom, ortopnejom, nesanicom, jutarnjom glavoboljom, umorom, anoreksijom, slabom koncentracijom i čestim promjenama raspoloženja.

Zahvaćenost donjeg motoneurona DMN-a također se očituje mišićnom slabošću i atrofijama. Najčešće je izražena na mišićima šaka, nadlaktica i ramenog obruča te mišićima potkoljenica i stopala. Miotatski refleksi su oslabljeni ili ugašeni uz hipotonus mišića. Javljaju se fascikulacije u mišićima koje ponekad mogu biti teško uočljive. Perkusijom neurološkim čekićem mogu se izazvati kada su ruka ili noga opušteni. Smatra se da su posljedica denervacije. Ako se kliničkim pregledom ne mogu ustanoviti, utvrđuju se elektromiografijom. Grčevi mišića još su jedan od znakova zahvaćenosti DMN-a. Smatra se da nastaju hiperekscitabilnošću motornih aksona. Javljaju se na atipičnim mjestima poput bokova, ruku, šaka, abdomena, vrata, mandibule, pa čak i jezika. Neurološke simptome i znakove degeneracije motoneurona moždanog debla nazivamo bulbarna paraliza. Kod osoba s bulbarnim oblikom ALS-a dolazi do brže progresije simptoma. Mišiće koji kontroliraju govor, žvakanje i gutanje inerviraju VII. moždani živac (n. facialis), IX. moždani živac (n. glossopharyngeus), X. moždani živac (n. vagus) i XII. moždani živac (n. hypoglossus), čiji su motorni neuroni smješteni u jezgrama moždanog debla, te može biti zahvaćen i motorni dio V. moždanog živca (n. trigeminus), koji kontrolira mišiće za žvakanje i pokrete mandibule. Kada su zahvaćeni bulbarni mišići, vidimo slabost mišića lica. Površina jezika često je nepravilna zbog atrofije i mogu se uočiti fascikulacije. Javljaju se dizartrija, disfagija, disfonija zbog slabosti mišića uključenih u stvaranje zvuka i govora. U oboljelih se javlja otežano gutanje krute i tekuće hrane, uz mogućnost aspiracije hrane i opasnosti od razvoja aspiracijske pneumonije. Također zbog atrofije i gubitka mišića, otežanog gutanja krute i tekuće hrane, dolazi do gubitka tjelesne težine, javlja se umor i pacijenti postaju sve slabiji i sve teže pokretni. ALS je progresivna bolest i svaki izostanak pogoršanja i progresije kliničke slike dovodi u pitanje dijagnozu ALS. Dijagnostikom treba isključiti druge bolesti koje mogu uzrokovati slične ili iste simptome, pa do postavljanja konačne dijagnoze prođe i do godina i pol dana, što je vrlo poražavajuće, s obzirom na veoma progresivnu bolest i spoznaju da oboljeli u prosjeku umiru u roku 3–5 godina. Prilikom dijagnosticiranja koriste se revidirani kriteriji El Escorial. [24–26]

KRITERIJI ZA POSTAVLJANJE DIJAGNOZE ALS-a EL ESCORIAL

S ciljem da se postave smjernice u diagnosticiranju ALS-a, postavljeni su 1994. godine kriteriji El Escorial, koji su potom revidirani 1998. godine. Ti se kriteriji prvenstveno odnose na kliničku sliku i čine zlatni standard dijagnoze ALS-a. Revidirani kriteriji El Escorial daju značaj nalazu elektromiografije i neurografije u postavljanju dijagnoze ALS-a.

Revidirani kriteriji El Escorial:

Klinički siguran ALS — znakovi oštećenja GMN-a i DMN-a u tri regije.

Klinički definitivan ALS — laboratorijski podržan, znakovi GMN-a i/ili DMN-a u jednoj regiji i nalaz genetske obrade da je pacijent nositelj patološkog gena.

Klinički vjerojatan ALS — znakovi GMN-a i/ili DMN-a u dvije regije sa znakovima GMN-a rostralno od znakova DMN-a.

Klinički vjerojatan ALS — laboratorijski podržan, znakovi GMN-a u jednoj ili više regija i znakovi DMN-a u nalazu EMG-a u najmanje dvije regije.

Klinički moguć ALS — znakovi GMN-a i DMN-a u jednoj regiji ili znakovi GMN-a u dvije regije ili znakovi GMN-a i DMN-a u dvije regije bez znakova GMN-a rostralno od znakova DMN-a.

Svim pacijentima kod kojih postoji sumnja na ALS potrebno je napraviti i elektrofiziološku dijagnostiku, kako bi se dokazalo oštećenje donjih motornih neurona i isključili slični sindromi. Prve kriterije za elektrofiziološku dijagnostiku objavio je Lambert 1957. i 1969. godine. Revizija dijagnostičkih kriterija El Escorial predložila je elektrofiziološke kriterije za dijagnozu ALS-a na Awaji otoku u Japanu, 2006. godine. Elektromiografija (EMG) je neurofiziološka metoda za analizu akcijskih potencijala motornih jedinica, koju čine alfa motoneuron prednjeg roga kralježnične moždine, njegov akson i svi ogranci s pripadajućim mišićnim vlaknima. Elektromiografijom utvrđuje se disfunkcija DMN-a u barem dvije od četiri regije: moždano deblo (kranijski motorni živci), vratni, torakalni i slabinski dio (motorni neuroni prednjih rogova kralježnične moždine). Promjene koje se registriraju obuhvaćaju pojavu abnormalne spontane aktivnosti kao znak akutne denervacije i teške degeneracije DMN-a. Kod maksimalne voljne kontrakcije registrira se prorijeđeni inervacijski uzorak akcijskih potencijala visokih amplituda i produženog trajanja (gigantski potencijali). Elektroneurografskim pregledom ispitujemo brzine i amplitude motornih i senzornih živaca kako bi se isključili poremećaji perifernog živca (aksonalna motorna polineuropatija, multifokalna motorna neuropatija), koji imitiraju simptome ALS-a. U ALS-u je brzina provodljivosti motornih i osjetnih živaca najčešće normalna. [27–30] Riluzol je jedini učinkoviti lijek koji produžuje vrijeme preživljavanja za prosječno 3 mjeseca, koji je FDA 1995. godine odobrila. Riluzol je antagonist glutamata koji, smanjujući razinu glutamata, reducira oštećenje motornih neurona, te je jedini lijek za koji su klinička istraživanja pokazala statistički značajno produljenje života u pacijenata s ALS-om. Osim što produljuje trajanje života za prosječno tri mjeseca, produljuje i vrijeme potrebno za ventilacijsku potporu. Riluzol je promijenio nihilistički terapeutski stav liječnika u liječenju ALS-a. Uzima se u dozi od 100 mg na dan i sastavni je dio svih smjernica za liječenje ove bolesti. Lijek je relativno dobro podnošljiv s nuspojavama kao što su astenija, mučnina i rijetko oštećenje jetre, stoga je bitno redovito pratiti jetrene probe tijekom uzimanja terapije. [30–32] Od 2017. godine odobren je i lijek edaravon, ali isključivo u SAD nakon male randomizirane kliničke studije provedene u Japanu na bolesnicima oboljelim od ALS-a u ranom stadiju bolesti. Bolesnici su primali edaravon u razdoblju od 6 mjeseci. Edaravon (MCI-186) ALS 19 studijska grupa izvijestila je o sigurnosti i djelotvornosti edaravona u fazi 3 studije na bolesnike s ranim stadijem amiotrofične lateralne skleroze. Primarna krajnja točka bila je promjena rezultata Revidirane skale funkcionalne ocjene amiotrofične lateralne skleroze (ALSFRS-R) od početne vrijednosti do 24 tjedna nakon randomizacije. Prije randomizacije, pacijenti su ušli u razdoblje promatranja od 12 tjedana, a oni pacijenti sa smanjenjem ALSFRS-R rezultata 1–4 boda nasumično su raspoređeni da primaju edaravon ili placebo. USA 2017 FDA odobrila nakon pada 33% napredovanja invaliditeta mjernih ALS FRS R score (faza 3 CT). Mehanizam djelovanja edaravon lijeka nije potpuno razjašnjen, ali smatra se da djeluje na oksidativni stres, kao jedan od mogućih okidača za razvoj amiotrofične lateralne skleroze. Za sada lijek nije u širokoj primjeni i koristi se isključivo u ranom stadiju ALS-a. [33] Ostala terapija je simptomatska. Svrha ove terapije je ublažiti komplikacije i unaprijediti kvalitet života oboljelog. Multidisciplinarni pristup iznimno je važan za oboljelog i za sve osobe koje o njemu brinu. Stručni tim čine liječnici, farmakolozi, fizioterapeuti, radni i govorni terapeuti, psiholozi, psihijatri, nutricionisti, socijalni radnici i medicinske sestre s iskustvom palijativne skrbi. Skrb o pacijentu pred kraj života, nutritivna i ventilacijska potpora tijekom posljednjeg stadija bolesti važni su aspekti liječenja.

Optimalni tretman temelji se na liječenju simptoma i održavanju kvaliteta života, osiguranih u multidisciplinarnom okruženju. [34, 35]

LITERATURA:

[1] Gordon PH, Cheng B, Katz IB, Pinto M, Hays AP, Mitsumoto H et al. The natural history of primary lateral sclerosis. Neurology 2006; 66: 647–53.

[2] Pratt AJ, Getzoff ED, Perry JJ. Amyotrophic lateral sclerosis: update and new developments. Degener Neurol Neuromuscular Dis 2012; 2012: 1–14.

[3] Worms PM. The epidemiology of motor neuron diseases: a review of recent studies. J Neurol Sci 2001; 191: 3–9. 6. Soriani MH, Desnuelle C. Epidemiology of amiotrophic lateral sclerosis. Rev Neurol 2009; 165: 627–40.

[4] De Jong SW, Huisman MH, Sutedja NA, Van der Kooi AJ, De Visser M, Schelhaas HJ, i sur. Smoking, alcohol consumption, and the risk of amyotrophic lateral sclerosis: a population-based study. Am J Epidemiol. 2012 Aug 1. 176(3): 233–9.
[5] Armon C. An evidence-based medicine approach to the evaluation of the role of exogenous risk factors in sporadic amyotrophic lateral sclerosis. Neuroepidemiology. 2003 Jul-Aug. 22(4): 217–28.
[6] Chiò A, Benzi G, Dossena M, Mutani R, Mora G. Severely increased risk of amyotrophic lateral sclerosis among Italian professional football players. Brain. 2005 Mar. 128: 472–6.
[7] Armon C. Sports and trauma in amyotrophic lateral sclerosis revisited. J Neurol Sci. 2007 Nov 15. 262(1–2): 45–53.
[8] Belli S, Vanacore N. Sports and amyotrophic lateral sclerosis. J Neurol Sci. 2008 Jun 15. 269(1–2): 191; author reply 191–2.
[9] Peters TL, Fang F, Weibull CE, Sandler DP, Kamel F, Ye W. Severe head injury and amyotrophic lateral sclerosis. Amyotroph Lateral Scler Frontotemporal Degener. 2013 May. 14(4): 267–72.
[10] Armon C. Epidemiology of ALS/MND. Shaw P and Strong M, eds. Motor Neuron Disorders. Elsevier Sciences: 2003, 167–206.
[11] Fogh I, Rijsdijk F, Andersen PM, Sham PC, Knight J, Neale B, i sur. Age at onset in sod1-mediated amyotrophic lateral sclerosis shows familiality. Neurogenetics. 2007 Aug. 8(3): 235–6.
[12] Saeed M, Yang Y, Deng HX, et al. Age and founder effect of SOD1 A4V mutation causing ALS. Neurology. 2009 May 12. 72(19): 1634–9.
[13] Al-Chalabi A, Jones A, Troakes C, King A, Al-Sarraj S, van den Berg LH. The genetics and neuropathology of amyotrophic lateral sclerosis. Acta Neuropathol. 2012 Sep. 124(3): 339–52.
[14] Mackenzie IR, Bigio EH, Ince PG, i sur. Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. Ann Neurol. 2007 May. 61(5): 427–34.
[15] Sreedharan J, Blair IP, Tripathi VB, i sur. TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science. 2008 Mar 21. 319(5870): 1668–72. 49.
[16] Kabashi E, Valdmanis PN, Dion P, i sur. TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat Genet. 2008. 40(5): 572–4.
[17] Van Deerlin VM, Leverenz JB, Bekris LM, i sur. TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. Lancet Neurol. 2008 May. 7(5): 409–16.
[18] Arai T, Mackenzie IR, Hasegawa M, i sur. Phosphorylated TDP-43 in Alzheimer's disease and dementia with Lewy bodies. Acta Neuropathol. 2009. 117(2): 125–36.
[19] DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, i sur. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron. 2011 Oct 20. 72(2): 245–56.
[20] Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, i sur. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21- linked ALS-FTD. Neuron. 2011 Oct 20. 72(2): 257–68.

[21] Smith BN, Newhouse S, Shatunov A, Vance C, Topp S, Johnson L, i sur. The C9ORF72 expansion mutation is a common cause of ALS + /-FTD in Europe and has a single founder. Eur J Hum Genet. 2013 Jan. 21(1): 102–8.
[22] Deng HX, Chen W, Hong ST, Boycott KM, Gorrie GH, Siddique N, i sur. Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature. 2011 Aug 21. 477(7363): 211–5. 50.
[23] Wu CH, Fallini C, Ticozzi N, Keagle PJ, Sapp PC, Piotrowska K, i sur. Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis. Nature. 2012 Aug 23. 488(7412): 499–503.
[24] Wijesekera LC, Mathers S, Talman P, Galtrey C, Parkinson MH, Ganesalingam J, i sur. Natural history and clinical features of the flail arm and flail leg ALS variants. Neurology. 2009 Mar 24. 72(12): 1087–94.
[25] Kim WK, Liu X, Sandner J, Pasmantier M, Andrews J, Rowland LP, i sur. Study of 962 patients indicates progressive muscular atrophy is a form of ALS. Neurology. 2009 Nov 17. 73(20): 1686–92.
[26] Mira Bučuk, Kristina Dijan, Zoran Tomić, Ivan Sonnenschein Amiotrofična lateralna skleroza, Amyotrophic lateral sclerosis medicina fluminensis 2014, Vol. 50, No. 1, p. 7–20.
[27] Brooks BR, Miller RG, Swash M, Munsat TL. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord*. 2000; 1: 293–9.
[28] de Carvalho M, Dengler R, Eisen A, England JD, Kaji R, Kimura J, Mills K, Mitsumoto H, Nodera H, Shefner J, Swash M. Electrodiagnostic criteria for diagnosis of ALS. *Clin Neurophysiol*. 2008 Mar; 119(3): 497–503.
[29] Brooks BR. El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/ Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial „Clinical limits of amyotrophic lateral sclerosis" workshop contributors. J Neurol Sci 1994; 124(Suppl): 96–107.
[30] Ervina Bilić, Dubravka Jurenić, Marija Žagar, Davorka Vranješ. Prijedlog dijagnostičkih i terapijskih smjernica u liječenju amiotrofične lateralne skleroze (ALS) [online]. Dostupno na: http: //hdnrmb. mef. hr/sites/default/files/ALS, pdf, datum pristupa: 21. 07. 2018.
[31] Bensimon G, Lacomblez L, Meininger V. A controlled trial of riluzole in amyotrophic lateral sclerosis. ALS/Riluzole Study Group. N Engl J Med. 1994 Mar 3. 330(9): 585–91. 51.
[32] Lacomblez L, Bensimon G, Leigh PN, Guillet P, Meininger V. Dose-ranging study of riluzole in amyotrophic lateral sclerosis. Amyotrophic Lateral Sclerosis/ Riluzole Study Group II. Lancet. 1996 May 25. 347(9013): 1425–31.
[33] Writing Group for the Edaravone (MCI-186) ALS 19 Study Group — Safety and efficacy of edaravone in well defined patients with amyotrophic lateral sclerosis: a randomised, double-blind, placebo-controlled trial. Lancet Neurol. 2017; 16: 505–512.
[34] Shoesmith CL, Findlater K, Rowe A, Strong MJ. Prognosis of amyotrophical lateral sclerosis with respiratory onset. J Neurol Neurosurg Psychiatry 2007; 78: 629–31.
[35] Bilić E. Palijativna skrb za oboljele od neuromuskularnih bolesti. Neurol Croat 2011; 60: 155–63.

Antonija Krstačić

AMYOTROPHIC LATERAL SCLEROSIS — THE LOU GEHRIG'S DISEASE

Summary

Amyotrophic lateral sclerosis or Charcot's disease according to the French neurologist who was among the first described the disease in more detail and distinguish ALS from other muscular atrophies, or Lou Gehrig's disease was named according the American baseball player who suffered from this disease. At the height of his fame, Lou Gehrig died in 1941. ALS is characterized by progressive degeneration and loss of motor neurons with or without similar lesions of the motor nuclei of the brain, with the replacement of lost cells by gliosis. It belongs to the group of motor neuron diseases and affects both the upper and lower motor neuron It can occur familiarly or sporadically. The familial incidence of ALS has been described in rare publications and has received limited attention in the literature since the 1955 report by Kurland and Mulder, which suggested that ALS may be familial in nearly 10% of cases. A growing body of evidence from clinical research suggests that ALS has multiple causes with an important, albeit diverse, genetic component. Available data indicate that genetic risk for ALS likely represents the combined effects of 1 or more genes that determine a person's overall genetic susceptibility, acting in conjunction with environmental and chance effects that lead to the onset of the disease.

Key words: amyotrophic lateral sclerosis, Lou Gehrig's disease, motor neuron disease

Petar M. SEFEROVIĆ[*,**], Ivan MILINKOVIĆ[*,***],
Marija POLOVINA[*,***], Giuseppe ROSANO[****]

THE ENIGMA OF THE BEST HEART FAILURE MEDICAL TREATMENT IN 2022: STILL MORE TO COME

Abstract: Over the past three decades, pharmacological treatment of heart failure (HF) with reduced ejection fraction (HFrEF) has witnessed a significant progress with the introduction of multiple disease-modifying therapies with a proven benefit on morbidity, mortality and quality of life. Recently, several novel medications (sacubitril/valsartan, sodium-glucose contransporter-2 [SGLT2] inhibitors, vericiguat and omecamtiv mecarbil) have shown to provide further improvement in outcomes in patients already receiving standard therapy for HFrEF. Available evidence suggests that sacubitril/valsartan and SGLT2 inhibitors (dapagliflozin and empagliflozin) are beneficial and well-tolerated in the majority in-patients and could be the mainstay treatment of HFrEF. Another group of medications (vericiguat and omecamtiv mecarbil) has shown promising results in reducing the risk of the composite of HF hospitalisations or cardiovascular mortality in patients with the more severe or advanced HF requiring recent hospitalisations. Therefore, these medications may be considered for the treatment of select group of patients with HFrEF and persisting or worsening symptoms despite optimal treatment. In addition, advances in pharmacological management of comorbidities frequently seen in HFrEF patients (diabetes, iron deficiency/anaemia, hyperkalaemia) provide further opportunities to improve outcomes. Given the increasing complexity of evidence-based therapies for HFrEF, there is a growing need to provide a practical perspective to their use. The purpose of this review is to summarise scientific evidence on the efficacy and safety of new and emerging medical therapies in HFrEF, with a focus on the clinical perspective of their use.

Key words: *Heart failure, Treatment, Sacubitril/valsartan, Dapagliflozin, Empagliflozin, Sotagliflozin, Vericiguat, Omecamtiv mecarbil, clinical trials, Outcomes*

[*] Faculty of Medicine, University of Belgrade, Belgrade, Serbia
[**] Serbian Academy of Sciences and Arts, Belgrade, Serbia
[***] Department of Cardiology, Clinical Centre of Serbia, Belgrade, Serbia
[****] Centre for Clinical and Basic Research, Department of Medical Sciences, IRCCS San Raffaele Pisana, Rome, Italy

INTRODUCTORY REMARKS

Medical treatment of heart failure (HF) with reduced ejection fraction (HFrEF) has significantly advanced with an introduction of disease-modifying therapies with a proven benefit on survival, morbidity and functional limitations. Most recently, an incremental improvement in outcomes has been documented with several novel medications (sacubitril/ valsartan, sodium-glucose contransporter-2 [SGLT2] inhibitors, vericiguat and omecamtiv mecarbil) in patients already receiving contemporary standard of care. Given the increasing complexity of evidence-based therapies for HFrEF, there is a growing need to provide a clinical perspective to use of the available medications, taking into the account evidence on patient eligibility and clinical characteristic, as well as the efficacy, safety and tolerability of these new medications.

THE COMPLEXITY OF HFREF TREATMENT IN THE LIGHT OF THE NEW STRATEGIES

HFrEF is a complex clinical syndrome, characterised by the compensatory activation of neurohormonal axes in response to the fall in cardiac output.[1)] In the long-run, neurohormonal activation leads to maladaptive changes in the heart, kidneys and the vasculature, causing disease progression and contribute to end-organ damage. This concept has been put to the test in the landmark clinical trials, which have conclusively demonstrated beneficial effects of neurohormonal inhibition with angiotensin-converting enzyme (ACE) inhibitors, angiotensin-II receptor blockers (ARB), beta-blockers and mineralocorticoid receptor antagonists (MRA). Despite the use of these disease- modifying and life-prolonging interventions, many of the HFrEF patients still suffer a residual risk of mortality, HF hospitalisations and disease progression. This risk could be attributed to the severity of HF, underlying aetiology, advanced age, comorbidities, frailty, limited implementation and inappropriate dosing of guideline-directed medical therapies (GDMT). However, in the recent clinical trials of novel medications in HFrEF, the residual risk remained high, despite optimal GDMT use. Following the positive results of these trials pharmacological options in HFrEF treatment has extended beyond neurohormonal blockade to include innovative therapeutic principles such as haemodynamic and metabolic benefits of SGLT2 inhibitors, improved cardiac and vascular nitric-oxide bioavailability and endothelial function with vericiguat, and increased cardiac contractility with omecamtiv mecarbil.

The complexity of clinical care for patients with HFrEF is further underlined by the high burden of comorbidities and frailty, especially among the older individuals and women. According to a recent review, the most prevalent comorbidities in clinical trials of HFrEF have been hypertension (89%), dyslipidaemia (61%), ischaemic heart disease (39%), chronic kidney disease (34%), atrial fibrillation (29%), type 2 diabetes (28%), investigator-reported depression (27%) and anaemia (12.5%), while temporal trends have shown a rise in the prevalence of several comorbidities.[2)] Clinical trial data also suggest that frailty is frequent (63%) and associated with worse symptoms and outcomes in HFrEF.[3)] Comorbidities and frailty often limit the possibilities for implementation and up-titration of GDMT due to hypotension, renal dysfunction, hyperkalaemia, poor adherence and tolerability. Nevertheless, a holistic approach to the management of HFrEF offers a possibility to improve outcomes and quality of life with the appropriate treatment of comorbidities.

NEW DRUGS, NEW MECHANISMS, ENHANCED EFFICACY, LOW SIDE EFFECTS

In the contemporary era of expanding treatment options for HFrEF, the appropriate use of multiple therapies has become more complex. New and emerging therapies have been evaluated on top of the conventional HF treatments, yet there was no suggestion of an interaction between novel drugs and background therapies in terms of efficacy. However, only a minority of patients in trials of SGLT2 inhibitors, vericiguat and omecamtiv mecarbil received sacubitril/valsartan. SGLT2 inhibitors were sporadically used in trials of vericiguat and omecamtiv mecarbil, because of an overlap in periods when the three drug classes were assessed. Also, evidence is lacking on the direct efficacy comparisons between new medications and the sequence in which they should be introduced. Despite these limitations, thorough understanding of the study eligibility criteria, patient characteristics, efficacy and safety profile of novel drugs could provide directions to the practical perspective of their use.

SACUBITRIL/VALSARTAN

Sacubitril/valsartan is a first-in-class angiotensin-receptor neprilysin inhibitor (ARNI), which combines the positive effects of neprilysin inhibition and renin-angiotensin system blockade with an ARB. Neprilysin could not be combined with an ACE inhibitor due to an excess risk of angioedema, which provides a rationale for a 36-hour washout period, when switching a patient from an ACE inhibitor to sacubitril/valsartan.

The long-term efficacy and safety of sacubitril/valsartan was evaluated in 8,399 HFrEF patients with left ventricular ejection fraction (LVEF) ≤40% in the PARADIGM-HF trial[4)] against an active comparator, enalapril. Patients with a history of angioedema, estimated glomerular filtration rate (eGFR) <30 mL/min/1.73 m^2, systolic blood pressure <100 mmHg, or decompensated HF were not included. The trial was prematurely terminated after a median of 27 months because of a significant risk reduction in the primary composite endpoint of cardiovascular death or HF hospitalization with sacubitril/valsartan (hazard ratio [HR], 0.80; 95% confidence interval [CI], 0.73–0.87; $p<0.001$, number needed to treat, NNT=21). Both components of the primary endpoint were significantly reduced, as was the risk of all-cause mortality. The beneficial effects with sacubitril/valsartan were observed early after randomisation, and no evidence of heterogeneity in efficacy was reported.

Hypotensive episodes occurred more frequently with sacubitril/valsartan (14% vs 9.2%, $p<0.001$), but deterioration of renal function and severe hyperkalaemia occurred less frequently. Of note, sub-analyses of the PARADIGM-HF trial have shown a reduction in sudden cardiac death and death due to worsening HF with sacubitril/valsartan,[5)] as well as a lower risk of worsening HF, including the requirement for therapy intensification, emergency department visits, hospitalisation in intensive care unit, all-cause and cardiovascular hospitalisations.[6)] Patients in the sacubitril/valsartan arm experienced greater reductions in haemoglobin A1c levels and less frequently required insulin initiation for diabetes treatment compared with enalapril.[7)]

The PROVE-HF trial provides evidence for reverse left-ventricular (LV) remodelling with sacubitril/valsartan in HFrEF, as demonstrated by an increase in LVEF (from 28.2% to 37.8%, difference 9.4%; 95% CI, 8.8–9.9%; $p<0.001$) and a reduction in LV volumes after 12 months of treatment, including patients with de novo HF.[8)] Further support is provided by the EVALUATE trial in which an early improvement in echocardiographic parameters occurred after 12 weeks of therapy with sacubitril/valsartan compared with enalapril,[9)] as well as by the PRIME study which suggested an improvement in functional mitral regurgitation following 12 months of sacubitril/valsartan treatment.[10)]

The PIONEER-HF study investigated the effect of an 8-week sacubitril/valsartan treatment vs. enalapril on NT-proBNP concentration in 881 patients with an LVEF ≤40%, hospitalised for acute decompensated HF [11.] The initiation of sacubitril/valsartan after haemodynamic stabilisation resulted in a greater reduction in NT-proBNP levels compared with enalapril, as early as one week after treatment initiation. The rates of worsening renal function, hyperkalaemia, symptomatic hypotension, and angioedema were comparable between the two groups. Further analysis of the PIONEER-HF trial demonstrated a significantly lower risk of an adjudicated composite outcome of death, HF rehospitalization, LV assist device implantation, and listing for cardiac transplantation in the sacubitril/valsartan arm (HR, 0.58; 95% CI, 0.40–0.85; p=0.005).[12] The TRANSITION trial assessed safety and tolerability of sacubitril/valsartan initiation at different time points after haemodynamic stabilisation in patients with acute decompensated HF and LVEF ≤40%.[13] The results point to a similar tolerability regardless whether the treatment commenced before or within two weeks after hospital discharge, whilst discontinuation rates due to adverse events remained low in both groups. Moreover, TRANSITION trial data support tolerability of sacubitril/valsartan in patients with de novo HF, previously naïve to ACE inhibitor/ARB therapy.[14] Accordingly, 29% of the trial population had de novo HF, and compared to those with prior HFrEF, they were more likely to obtain the target dose and were less likely to discontinue treatment due to adverse events.

SGLT2 INHIBITORS

The first trial to evaluate a role of an SGLT2 inhibitor in the treatment of HFrEF was DAPA-HF, which assessed the effect of dapagliflozin vs. placebo on the risk of cardiovascular mortality or worsening HF (defined as HF hospitalisation or urgent outpatient visit for HF treatment).[15]

The trial enrolled 4,744 patients with HFrEF and LVEF ≤40%, with and without diabetes. Patients with type 1 diabetes mellitus, systolic blood pressure <95 mmHg, eGFR <30 mL/min/1.73m^2 and those intolerant to SGLT2 inhibitors were excluded. The results have demonstrated a significantly lower risk of the primary endpoint (HR, 0.74; 95% CI, 0.65–0.85; p<0.001; NNT=21), as well as a significant reduction in cardiovascular mortality and worsening HF events with dapagliflozin vs placebo. The benefits occurred early after randomisation and were observed irrespective of the background GDMT, including sacubitril/valsartan (~11% of the trial patients). Importantly, the efficacy of dapagliflozin was similar in patients with and without diabetes. This was further explored in a sub-analysis which has shown consistent treatment effects across a spectrum of haemoglobin A1c.[16] There was no evidence of heterogeneity in dapagliflozin efficacy among the predefined subgroups, except possibly for a lesser efficacy in patients with the New York Heart Association functional class III–IV compared with class II. However, no heterogeneity was observed among patients with lower LVEF or higher

NT-proBNP, which suggests that dapagliflozin is similarly effective regardless of the severity of HF.[15] Dapagliflozin treatment was safe and well tolerated and no excess in serious adverse events was noted, including diabetic ketoacidosis.

The EMPEROR-Reduced trial assessed the efficacy and safety of empagliflozin vs. placebo in 3,730 patients with HFrEF and LVEF ≤40%, with and without diabetes.[17] Patients with recent acute coronary syndrome, myocardial revascularisation or stroke, acute HF, systolic blood pressure <100 mmHg, intolerance to SGLT2 inhibitors and eGFR <20 mL/min/1.73m^2 were excluded. The majority of the trial population had LVEF <30% and NT-proBNP >1000 pg/mL (73% and 79%, respectively) and almost a half of the patients had renal dysfunction (eGFR ≥20 to 60 mL/min/1.73m^2). After a median follow-up of 16 month empagliflozin treatment was associated with a lower risk of the composite primary outcome of cardiovascular mortality or HF hospitalisation (HR, 0.75; 95% CI, 0.65–0.86, p<0.001).[17] The beneficial effects of empagliflozin occurred early and were similar in patients with and without diabetes. There was no heterogeneity in treatment effects according to age, sex, and background therapy, including sacubitril/valsartan (~20% of the trial population). The effect on the primary outcome was primarily attributed to a lower risk of HF hospitalisation with empagliflozin (HR, 0.69; 95% CI, 0.59–0.81; p<0.001). There was also a reduction in the total number of HF hospitalisations (first and recurrent; HR, 0.70; 95% CI, 0.58–0.85; p<0.001) as well as a slower decline in renal function (a mean slope of change in eGFR mL/min/1.73 m^2 per year) with empagliflozin, whereas the rate of all-cause mortality was similar with placebo. Serious adverse events (hypoglycaemia, lower limb amputation, bone fracture and diabetic ketoacidosis) were rare and comparable between empagliflozin and placebo.

More recently, the SOLOIST-WHF trial has assessed the effect of sotagliflozin (combined SGTL2 and SGLT1 inhibitor) vs. placebo on the primary endpoint of the total number of cardiovascular deaths, hospitalisations and urgent visits for HF treatment (first and recurrent) in 1,222 patients with type 2 diabetes and recent hospitalisation for worsening HF.[18] Patients with end-stage HF, recent acute coronary syndrome, stroke, or myocardial revascularisation, or eGFR <30 mL/min/1.73 m^2 were excluded, The trial population mostly comprised individuals with HF and mid-range or reduced LVEF (~79%), with a median LVEF of 35%, but the SOLOIST-WHF also included diabetic patients with HF and preserved LVEF≥50% (HFpEF). The trial ended prematurely (due to the lack of funding). After a median of 9 months of follow-up, patients randomised to sotagliflozin demonstrated a significantly lower risk of the primary outcome (HR, 0.67; 96% CI, 0.52–0.85; p<0.001). This finding was consistent across several prespecified subgroups, including stratification according to the timing of the first dose of the study medications (before or 3 days after hospital discharge) and LVEF <50% or ≥50%. Also, the first secondary endpoint of the total number of HF hospitalisations was significantly reduced with sotagliflozin, whilst the rates of cardiovascular and total mortality were comparable between the study arms. These results suggest that an early introduction of sotagliflozin after stabilisation of decompensated HF may be safe, with similar benefits in patients with HFrEF and HFpEF. However, given the premature termination and a small sample size, these observations should be considered as hypothesis generating.

A meta-analysis of the DAPA-HF and EMPEROR-Reduced trials has confirmand a significant risk reduction of CV mortality or HF hospitalisation with SGLT2 inhibition (HR, 0.74; 95% CI, 0.68–0.82, p<0.0001) and a consistent effect of dapagliflozin and empagliflozin on risk reduction of cardiovascular (HR, 0.86; 95% CI, 0.76–0.98, p=0.027) and all-cause mortality (HR, 0.87; 95% CI, 0.77–0.98, p=0.018), as well as of worsening renal function (HR, 0.62; 95% CI, 0.43–0.90, p=0.013).[19] There was no excess in serious adverse events, including volume depletion, renal dysfunction, bone fractures, lower limb amputations or diabetic ketoacidosis.

VERICIGUAT

Vericiguat, a soluble guanylate cyclase stimulator acts to increase myocardial and vascular NO bioavailability, with an effect on the improvement of endothelial function, myocardial remodelling and diastolic relaxation.[20] The VICTORIA trial assessed the effects of vericiguat vs. placebo on the primary outcome of death from cardiovascular cause or first HF hospitalization in 5050 patients with HF, an LVEF <45% and a recent HF hospitalisation or urgent HF treatment.[21] Exclusion criteria were a systolic blood pressure <100 mmHg; concomitant use of long-acting nitrates, soluble guanylate cyclase stimulators, or phosphodiesterase type 5 inhibitors; and use of intravenous inotropes or implantable left ventricular assist devices. The trial population comprised high-risk patients with a mean LVEF of 29%, a median NT-proBNP of 2,816 pg/mL and a recent (<3 months) hospitalisation for HF in 67%. Over a median of 10.8 months, vericiguat treatment vs. placebo reduced the primary outcome (HR, 0.90; 95% CI, 0.83–0.98; p=0.02, NNT 24) as well as total number of HF hospitalisations and death from any cause or first HF hospitalization.

The difference favouring vericiguat became apparent after approximately 3 months from randomisation and was comparable in patients receiving (15% of the trial population) vs. those not receiving sacubitril/valsartan.

Cardiovascular and all-cause mortality rates were similar between vericiguat and placebo. An analysis of the treatment effects across prespecified subgroups indicated that the benefit of vericiguat may be attenuated in patients with severe or advanced HF, including those with the highest quartile of NT-proBNP values (>5,314 pg/mL) as well in patients older than 75 years and in those with significant renal dysfunction (eGFR <30 mL/min/1.73m^2). The overall incidence of adverse events was comparable between vericiguat and placebo, although symptomatic hypotension and syncope were numerically more frequent with vericiguat.[21)]

OMECAMTIV-MECARBIL

The effect of a selective cardiac myosin activator, omecamtiv mecarbil, on cardiovascular mortality or hospitalisations for HF was assessed in the GALACTIC-HF trial, which randomised 8256 patients with HF and an LVEF ≤35% to receive omecamtiv mecarbil or placebo, in addition to GDMT.[22)] The trial included 25% of patients currently hospitalised for acute HF (inpatients), as well as those with a hospitalisation or urgent visit for HF treatment within 1 year before screening (outpatients). Haemodynamically unstable patients, those with a systolic blood pressure <85 mmHg, an eGFR <20 mL/min/1.73 m^2, and a recent acute coronary syndrome or cardiovascular procedure were excluded. During a median of 21.8 months, patients in the omecamtiv mecarbil group experienced a reduction in the primary composite outcome of a first HF event (hospitalization or urgent visit for HF) or death from cardiovascular causes compared with the placebo group (HR, 0.92; 95% CI, 0.86–0.99; p=0.03). The treatment effects were consistent across predefined subgroups (including inpatients and outpatients), apart from a potentially greater efficacy of omecamtiv mecarbil in patients with LVEF <28%. A secondary outcome of cardiovascular mortality was not significantly reduced. Major cardiac ischaemic events and ventricular arrhythmias occurred at similar rates between the study arms. The study medications were discontinued at rates similar between the omecamtiv mecarbil and placebo arms (20.6% and 21.9%), mostly due to adverse events.[22)]

COMORBIDITIES AS A TREATMENT TARGETS IN HFREF

Type 2 diabetes

In patients with type 2 diabetes and established cardiovascular disease or with multiple risk factors, several cardiovascular outcome trials have demonstrated a consistent reduction in the risk of HF hospitalisations with

SGLT2 inhibitors (empagliflozin, canagliflozin, dapagliflozin and ertugliflozin).[23-26] Furthermore, the use of empagliflozin was associated with a reduction in cardiovascular mortality in the EMPA-REG-Outcome trial,[25] and there was a consistent effect on renal protection with SGLT2 inhibitors in patients with diabetes.[27-29] The DAPA-HF and EMPEROR-Reduced trials have shown no signal of heterogeneity in the efficacy of SGLT2 inhibition in reducing the risk of cardiovascular mortality or HF hospitalisations in patients with or without diabetes.[15][17] This is further supported by the SOLOIST-WHF trial, which has suggested that an early introduction of sotagliflozin in type 2 diabetes patients stabilised after an episode of worsening HF may be safe and effective in the prevention of recurrent hospitalisations or cardiovascular mortality.[18]

Considering other glucose-lowering medications, clinical trial data suggest a significant increase in the risk of HF hospitalisations in type 2 diabetes patients receiving thiazolidinediones (rosiglitazone and pioglitazone),[30-32] and a dipeptidyl peptidase-4 inhibitor, saxagliptin.[33] In addition, two small trials of patients with HFrEF (with and without diabetes) suggested a signal of harm with a glucagon-like peptide-1 receptor agonist, liraglutide.[34][35] Clinical trial data on the safety of metformin in patients with HFrEF are missing but accumulated clinical experience and observational data suggest that metformin is safe and associated with lesser risk of cardiovascular mortality or HF, compared with sulphonylurea agents or insulin.[36-38] Although data on the effect of insulin on the risk of mortality or worsening HF remains ambiguous,[39][40] many patients require insulin to control diabetes and it is often a part of the combined glucose-lowering regiments. Importantly, cardiovascular outcome trials have not shown an interaction between insulin treatment and the reduction of cardiovascular outcomes in patients receiving SGLT2 inhibitors.

Iron deficiency/anaemia

Iron deficiency (serum ferritin concentration <100 ng/mL or 100–299 ng/mL with transferrin saturation <20%) and anaemia (haemoglobin level <120 g/L in females and <130 g/L in males) are common in HFrEF.[41] They are independently associated with reduced exercise tolerance and a higher risk of HF hospitalisations, cardiovascular and all-cause cause mortality.[41] Several clinical trials (FAIR-HF, CONFIRM-HF and EFFECT-HF) have documented the efficacy of intravenous ferric carboxymaltose to improve symptoms, exercise capacity and quality of life in patients with HFrEF.[42-44] Prospective evaluation of the effects of parenteral iron supplementation on HF hospitalisation and mortality in HFrEF is currently underway.

HYPERKALAEMIA

Hyperkalaemia is often the reason for under-prescription, under-dosing or discontinuation of renin-angiotensin-aldosterone inhibitors, in particular MRAs, although sub-analyses of the RALES and EMPHASIS-HF trials (with spironolactone and eplerenone in HFrEF, respectively) have not documented an attenuation of therapeutic effects of MRAs in patients with serum potassium >5.5–6.0 momL/L.[45)46)] Potassium binding-agents, patiromer and zirconium cyclosilicate have shown promising results in controlling hyperkalaemia, which may allow optimisation of HFrEF treatment. The PEARL-HF trial included 105 patients with HF, an eGFR of <60 mL/min or a documented history of renin-angiotensin-aldosterone inhibitors discontinuation due to hyperkalaemia within 6 months.[47)] Following 28 days of treatment, patients randomised to patiromer had a lower serum potassium concentration, lesser incidence of hyperkalaemia and a greater proportion of patients achieved target dose of spironolactone. Significant hypokalaemia (<3.5 mmol/L) was numerically more frequent with patiromer, whilst other adverse events were mainly gastrointestinal.[47)]

Clinical trials are leading to change in clinical pharmacotherapeutic practice in HFrEF

Accumulating data suggests that sacubitril/valsartan[4)] and SGLT2 inhibitors (dapagliflozin and empagliflozin)[15)17)] are beneficial and well-tolerated in the majority of patients with HFrEF and therefore could be considered as the mainstay treatment. Sacubitril/valsartan has been proven to reduce the risk of death, worsening HF, cardiac arrhythmia, and to improve LV reverse remodelling and quality of life in HFrEF.[48)] It can be safely initiated in patients with de novo HF and during a vulnerable phase following an episode of acute HF. Therefore, this drug could be preferred to ACE inhibitors/ARBs in the majority of HFrEF patients.[49)] A caution is needed in patients with a lower systolic blood pressure, and in those hospitalised for acute HF due to a higher risk of hypotension.[11)] A lower starting dose (24/26 mg twice daily), careful up-titration and a reduction in the dose of loop diuretics should be considered in those individuals.[50)] Renal function and serum potassium should be checked within 1 to 2 weeks following an initiation of sacubitril/valsartan, and data is lacking on its safety and efficacy in severe renal impairment (eGFR <30 mL/min/1.73 m^2).

A growing body of evidence supports the role of SGLT2 inhibitors (dapagliflozin and empagliflozin) as a novel class of medications with beneficial cardiovascular and renal effects in the majority of HFrEF patients, regardless of diabetes status. These medications may be introduced early in the sequence of GDMT initiation, before obtaining target doses of other medications since clinical trial data suggest their complementary therapeutic value.[51] The use of SGLT2 inhibitors is further facilitated by a favourable safety profile, good tolerability and by the lack of requirement for up-titration. A reduction in the dose of loop diuretics may be needed due to a mild/transient diuretic/natriuretic effect of SGLT2 inhibitors. Caution is advised in patients with severe renal dysfunction, and an adjustment of antidiabetic medications may be needed in patients with diabetes.

The evidence base for an early introduction of sacubitril/valsartan (instead of an ACE inhibitor/ARB) and an SGLT2 inhibitor in patients with HFrEF is solidified by a recent cross- trial analysis in HFrEF, which has suggested that an early comprehensive disease-modifying therapy with sacubitril/valsartan, a beta-blocker, an MRA and an SGLT2 inhibitor can afford 2.7 additional years (for an 80-year-old) to 8.3 additional years (for a 55-year-old) free from cardiovascular death or first HF hospitalisation and 1.4 additional years (for an 80-year-old) to 6.3 additional years (for a 55-year-old) of survival compared with conventional therapy (ACE inhibitors/ARB and beta-blockers).[52] This concept has been endorsed by recent expert practice recommendations for the treatment of HFrEF.[49][53][54]

Vericiguat and omecamtiv mecarbil has been assessed in the more selective populations in the context of greater symptom burden, higher natriuretic peptide levels and recent hospitalization for worsening HF despite standard treatment.[20][22] These medications may not be needed in all HFrEF patients but may be considered as a treatment effective in specific populations of HFrEF patients who remain symptomatic and/or experience worsening symptoms despite optimal treatment.

Considering the treatment of comorbidities in HFrEF, recent clinical trials solidify the evidence that a holistic approach to the management of HFrEF can significantly improve outcomes. In the general population of patients with type 2 diabetes, SGLT2 inhibitors should be the first-line treatment to prevent or delay HF hospitalisation, whilst in patients with known HFrEF and diabetes, these medications should be the preferred treatment to improve clinical outcomes. Furthermore, intravenous iron supplementation in individuals with iron deficiency/anaemia provides an opportunity to improve functional status and quality of life, whilst novel therapeutic options for hyperkalaemia (potassium binders), demonstrate promising results in enabling the maintenance and up-titration of GDMT. Another practical tip from the PARADIGM-HF trial is that substitution of an ACE inhibitor with sacubitril/valsartan may be associated with a lower risk of hyperkalaemia, which may allow more space for the optimisation of GDMT in HFrEF.

CONCLUSIONS

Pharmacological management of HFrEF has witnessed major breakthroughs over the past decades and contemporary drug therapy offers a possibility to alter the natural course, prolong lives and decrease the burden of morbidity and disability in the affected patients. Accumulating evidence indicates that selecting sacubitril/valsartan (instead of an ACE inhibitor/ARB) and introducing an SGLT2 inhibitor early in the treatment pathway, along with an evidence-based beta-blocker and an MRA, is feasible and associated with an incremental prognostic benefit in the majority of patients with HFrEF. Even more, in high-risk individuals with severe or advanced HFrEF, vericiguat or omecamtiv mecarbil may be a valuable, emerging option to improve outcomes in addition to standard care. Appropriate selection of devices, surgical therapies and targeted treatment of comorbidities complete the holistic approach to the management of HFrEF. Future developments will further broaden the spectrum of emerging therapies, provide new insights into the optimal sequencing of available drugs and will lead the way to treatment pathways tailored to the individual patient's requirements.

REFERENCES

[1] Hartupee J, Mann DL. Neurohormonal activation in heart failure with reduced ejection fraction. Nat Rev Cardiol 2017; 14: 30–8.
[2] Khan MS, Samman Tahhan A, Vaduganathan M, et al. Trends in prevalence of comorbidities in heart failure clinical trials. Eur J Heart Fail 2020; 22: 1032–42.
[3] Dewan P, Jackson A, Jhund PS, et al. The prevalence and importance of frailty in heart failure with reduced ejection fraction — an analysis of PARADIGM-HF and ATMOSPHERE. Eur J Heart Fail 2020; 22: 2123–33.
[4] McMurray JJ, Packer M, Desai AS, et al. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med 2014; 371: 993–1004.
[5] Desai AS, McMurray JJ, Packer M, et al. Effect of the angiotensin-receptor-neprilysin inhibitor LCZ696 compared with enalapril on mode of death in heart failure patients. Eur Heart J 2015; 36: 1990–7.
[6] Packer M, McMurray JJ, Desai AS, et al. Angiotensin receptor neprilysin inhibition compared with enalapril on the risk of clinical progression in surviving patients with heart failure. Circulation 2015; 131: 54–61.
[7] Seferovic JP, Claggett B, Seidelmann SB, et al. Effect of sacubitril/valsartan versus enalapril on glycaemic control in patients with heart failure and diabetes: a post-hoc analysis from the PARADIGM-HF trial. Lancet Diabetes Endocrinol 2017; 5: 333–40.

[8] Januzzi JL Jr, Prescott MF, Butler J, et al. Association of change in N-Terminal pro-B-Type natriuretic peptide following initiation of sacubitril-valsartan treatment with cardiac structure and function in patients with heart failure with reduced ejection fraction. JAMA 2019; 322: 1085–95.
[9] Desai AS, Solomon SD, Shah AM, et al. Effect of sacubitril-valsartan vs enalapril on aortic stiffness in patients with heart failure and reduced ejection fraction: a randomized clinical trial. JAMA 2019; 322: 1077–84.
[10] Kang DH, Park SJ, Shin SH, et al. Angiotensin receptor neprilysin inhibitor for functional mitral regurgitation. Circulation 2019; 139: 1354–65.
[11] Velazquez EJ, Morrow DA, DeVore AD, et al. Angiotensin-neprilysin inhibition in acute decompensated heart failure. N Engl J Med 2019; 380: 539–48.
[12] Morrow DA, Velazquez EJ, DeVore AD, et al. Clinical outcomes in patients with acute decompensated heart failure randomly assigned to sacubitril/valsartan or enalapril in the PIONEER-HF trial. Circulation 2019; 139: 2285–8.
[13] Wachter R, Senni M, Belohlavek J, et al. Initiation of sacubitril/valsartan in haemodynamically stabilised heart failure patients in hospital or early after discharge: primary results of the randomised TRANSITION study. Eur J Heart Fail 2019; 21: 998–1007.
[14] Senni M, Wachter R, Witte KK, et al. Initiation of sacubitril/valsartan shortly after hospitalisation for acutely decompensated heart failure in patients with newly diagnosed (de novo) heart failure: a subgroup analysis of the TRANSITION study. Eur J Heart Fail 2020; 22: 303–12.
[15] McMurray JJ, Solomon SD, Inzucchi SE, et al. Dapagliflozin in patients with heart failure and reduced ejection fraction. N Engl J Med 2019; 381: 1995–2008.
[16] Petrie MC, Verma S, Docherty KF, et al. Effect of dapagliflozin on worsening heart failure and cardiovascular death in patients with heart failure with and without diabetes. JAMA 2020; 323: 1353–68.
[17] Packer M, Anker SD, Butler J, et al. Cardiovascular and renal outcomes with empagliflozin in heart failure. N Engl J Med 2020; 383: 1413–24.
[18] Bhatt DL, Szarek M, Steg PG, et al. Sotagliflozin in patients with diabetes and recent worsening heart failure. N Engl J Med 2021; 384: 117–28.
[19] Zannad F, Ferreira JP, Pocock SJ, et al. SGLT2 inhibitors in patients with heart failure with reduced ejection fraction: a meta-analysis of the EMPEROR-Reduced and DAPA-HF trials. Lancet 2020; 396: 819–29.
[20] Armstrong PW, Roessig L, Patel MJ, et al. A multicenter, randomized, double-blind, placebo-controlled trial of the efficacy and safety of the oral soluble guanylate cyclase stimulator: the VICTORIA trial. JACC Heart Fail 2018; 6: 96–104.
[21] Armstrong PW, Pieske B, Anstrom KJ, et al. Vericiguat in patients with heart failure and reduced ejection fraction. N Engl J Med 2020; 382: 1883–93.
[22] Teerlink JR, Diaz R, Felker GM, et al. Cardiac myosin activation with omecamtiv mecarbil in systolic heart failure. N Engl J Med 2021; 384: 105–16.
[23] Cannon CP, Pratley R, Dagogo-Jack S, et al. Cardiovascular outcomes with ertugliflozin in type 2 diabetes. N Engl J Med 2020; 383: 1425–35.
[24] Neal B, Perkovic V, Mahaffey KW, et al. Canagliflozin and cardiovascular and renal events in type 2 diabetes. N Engl J Med 2017; 377: 644–57.

[25] Zinman B, Wanner C, Lachin JM, et al. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. N Engl J Med 2015; 373: 2117–28.
[26] Wiviott SD, Raz I, Bonaca MP, et al. Dapagliflozin and cardiovascular outcomes in type 2 diabetes. N Engl J Med 2019; 380: 347–57.
[27] Mosenzon O, Wiviott SD, Cahn A, et al. Effects of dapagliflozin on development and progression of kidney disease in patients with type 2 diabetes: an analysis from the DECLARE-TIMI 58 randomised trial. Lancet Diabetes Endocrinol 2019; 7: 606–17.
[28] Wanner C, Inzucchi SE, Lachin JM, et al. Empagliflozin and progression of kidney disease in type 2 diabetes. N Engl J Med 2016; 375: 323–34.
[29] Perkovic V, Jardine MJ, Neal B, et al. Canagliflozin and renal outcomes in type 2 diabetes and nephropathy. N Engl J Med 2019; 380: 2295–306.
[30] Dormandy JA, Charbonnel B, Eckland DJ, et al. Secondary prevention of macrovascular events in patients with type 2 diabetes in the PROactive Study (PROspective pioglitAzone Clinical Trial In macroVascular Events): a randomised controlled trial. Lancet 2005; 366: 1279–89.
[31] Home PD, Pocock SJ, Beck-Nielsen H, et al. Rosiglitazone evaluated for cardiovascular outcomes in oral agent combination therapy for type 2 diabetes (RECORD): a multicentre, randomised, open-label trial. Lancet 2009; 373: 2125–35.
[32] Komajda M, McMurray JJ, Beck-Nielsen H, et al. Heart failure events with rosiglitazone in type 2 diabetes: data from the RECORD clinical trial. Eur Heart J 2010; 31: 824–31.
[33] Scirica BM, Braunwald E, Raz I, et al. Heart failure, saxagliptin, and diabetes mellitus: observations from the SAVOR-TIMI 53 randomized trial. Circulation 2014; 130: 1579–88.
[34] Jorsal A, Kistorp C, Holmager P, et al. Effect of liraglutide, a glucagon-like peptide-1 analogue, on left ventricular function in stable chronic heart failure patients with and without diabetes (LIVE)-a multicentre, double-blind, randomised, placebo-controlled trial. Eur J Heart Fail 2017; 19: 69–77.
[35] Margulies KB, Hernandez AF, Redfield MM, et al. Effects of liraglutide on clinical stability among patients with advanced heart failure and reduced ejection fraction: a randomized clinical trial. JAMA 2016; 316: 500–8.
[36] Shah DD, Fonarow GC, Horwich TB. Metformin therapy and outcomes in patients with advanced systolic heart failure and diabetes. J Card Fail 2010; 16: 200–6.
[37] Aguilar D, Chan W, Bozkurt B, Ramasubbu K, Deswal A. Metformin use and mortality in ambulatory patients with diabetes and heart failure. Circ Heart Fail 2011; 4: 53–8.
[38] Eurich DT, Weir DL, Majumdar SR, et al. Comparative safety and effectiveness of metformin in patients with diabetes mellitus and heart failure: systematic review of observational studies involving 34,000 patients. Circ Heart Fail 2013; 6: 395–402.
[39] Pocock SJ, Wang D, Pfeffer MA, et al. Predictors of mortality and morbidity in patients with chronic heart failure. Eur Heart J 2006; 27: 65–75.
[40] Gerstein HC, Miller ME, Byington RP, et al. Effects of intensive glucose lowering in type 2 diabetes. N Engl J Med 2008; 358: 2545–59.

[41] von Haehling S, Ebner N, Evertz R, Ponikowski P, Anker SD. Iron deficiency in heart failure: an overview. JACC Heart Fail 2019; 7: 36–46.
[42] Anker SD, Comin Colet J, Filippatos G, et al. Ferric carboxymaltose in patients with heart failure and iron deficiency. N Engl J Med 2009; 361: 2436–48.
[43] Ponikowski P, van Veldhuisen DJ, Comin-Colet J, et al. Beneficial effects of long-term intravenous iron therapy with ferric carboxymaltose in patients with symptomatic heart failure and iron deficiency†. Eur Heart J 2015; 36: 657–68.
[44] van Veldhuisen DJ, Ponikowski P, van der Meer P, et al. Effect of ferric carboxymaltose on exercise capacity in patients with chronic heart failure and iron deficiency. Circulation 2017; 136: 1374–83.
[45] Rossignol P, Dobre D, McMurray JJ, et al. Incidence, determinants, and prognostic significance of hyperkalemia and worsening renal function in patients with heart failure receiving the mineralocorticoid receptor antagonist eplerenone or placebo in addition to optimal medical therapy: results from the Eplerenone in Mild Patients Hospitalization and Survival Study in Heart Failure (EMPHASIS-HF). Circ Heart Fail 2014; 7: 51–8.
[46] Vardeny O, Claggett B, Anand I, et al. Incidence, predictors, and outcomes related to hypo- and hyperkalemia in patients with severe heart failure treated with a mineralocorticoid receptor antagonist. Circ Heart Fail 2014; 7: 573–9.
[47] Pitt B, Anker SD, Bushinsky DA, et al. Evaluation of the efficacy and safety of RLY5016, a polymeric potassium binder, in a double-blind, placebo-controlled study in patients with chronic heart failure (the PEARL-HF) trial. Eur Heart J 2011; 32: 820–8.
[48] Khariton Y, Fonarow GC, Arnold SV, et al. Association between sacubitril/valsartan initiation and health status outcomes in heart failure with reduced ejection fraction. JACC Heart Fail 2019; 7: 933–41.
[49] Maddox TM, Januzzi JL Jr, Allen LA, et al. 2021 update to the 2017 ACC expert consensus decision pathway for optimization of heart failure treatment: answers to 10 pivotal issues about heart failure with reduced ejection fraction: a report of the American College of Cardiology Solution Set Oversight Committee. J Am Coll Cardiol 2021; 77: 772–810.
[50] Senni M, McMurray JJ, Wachter R, et al. Initiating sacubitril/valsartan (LCZ696) in heart failure: results of TITRATION, a double-blind, randomized comparison of two uptitration regimens. Eur J Heart Fail 2016; 18: 1193–202.
[51] Docherty KF, Jhund PS, Inzucchi SE, et al. Effects of dapagliflozin in DAPA-HF according to background heart failure therapy. Eur Heart J 2020; 41: 2379–92.
[52] Vaduganathan M, Claggett BL, Jhund PS, et al. Estimating lifetime benefits of comprehensive disease- modifying pharmacological therapies in patients with heart failure with reduced ejection fraction: a comparative analysis of three randomised controlled trials. Lancet 2020; 396: 121–8.
[53] Seferovic PM, Ponikowski P, Anker SD, et al. Clinical practice update on heart failure 2019: pharmacotherapy, procedures, devices and patient management. An expert consensus meeting report of the Heart Failure Association of the European Society of Cardiology. Eur J Heart Fail 2019; 21: 1169–86.
[54] Seferović PM, Fragasso G, Petrie M, et al. Heart Failure Association of the European Society of Cardiology update on sodium-glucose co-transporter 2 inhibitors in heart failure. Eur J Heart Fail 2020; 22: 1984–6.

Danilo VOJVODIĆ*

IMUNOLOGIJA IZMEĐU DOSTIGNUĆA I PERSPEKTIVE

Sažetak: Prvo priznanje i prva procena značaja imunog sistema došli su veoma rano, u 4. veku pre nove ere, od *Thucydida*, poznatog istoričara, političara i ratnog generala. On je bio prvi koji je opisao imunu osobu kao osobu koja se oporavila od bolesti (infekcije), koja je mogla da neguje zaražene, a da ne dobije bolest drugi put. Prva terapijska imunološka procedura bila je variolizacija (XVI vek), koja je dala *Sir Edwardu Jenneru* osnovu za vakcinaciju protiv velikih boginja (XVIII vek), kao prvu ikada manipulaciju imunološkim odgovorom u cilju izazivanja zaštite. U suštini, XX vek je doneo procvat imunologije, sa značajnom transformacijom od fundamentalne nauke do aplikativne medicinske procedure. Identifikacija antiseruma, fagocita, anafilakse, antitela, komponenti komplementa, krvnih grupa, različite populacije limfocita i kasnije HLA sistema, tehnologije monoklonskih antitela, prirodnih ćelija ubica, pa čak i idiotipske — antiidiotipske mreže, napravile su nezamisliv napredak u savremenoj medicini. Sva ova dostignuća su prepoznata kao ključni kamen-temeljac. Stoga nije iznenađujuće što su brojne nobelove nagrade dodeljene za imunološki napredak u ljudskoj biologiji i medicini, od *Emila fon Beringa, Mečnikova, Erliha, Rišea, Bordea, Landštajnera, Burneta, Medavara, Edelmana, Portera, Benacerrafa, Dauseta, Snela* i mnogo više. Naravno, najuzbudljiviji uvid u imunološki mehanizam došao je iz najnovijih istraživanja u imunologiji raka. *Allison* i *Honjo*, ponovo dobitnici Nobelove nagrade, pokazali su da je ključna razmena informacija između tumorske ćelije, ćelije koja predstavlja antigen i tumor specifičnih T limfocita. Otkrili su da tumorske ćelije koriste fiziološke molekule membrane kao specifične kočnice, sposobne da indukuju inhibiciju tumora specifičnih T limfocita, čineći se zaštićenim od tekućeg imunološkog odgovora. Oni su takođe pokazali da monoklonska antitela na ovaj molekul PD-1 mogu u potpunosti da uklone inhibiciju tumora specifičnog T limfocita, što je omogućilo ogroman napredak u imunoterapiji solidnih tumora. Još jedno značajno otkriće učinilo je budućnost imunološke onkologije veoma svetlom. Korišćenjem tehnike molekularne biologije danas je moguće preneti deo antitela u T limfocit na način da T limfocit koristi delove antitela za prepoznavanje antigena kao

* Danilo Vojvodić, Medicinski fakultet, Univerzitet odbrane, Beograd; načelnik Instituta za medicinska istraživanja, Vojnomedicinska akademija (VMA), Beograd

sopstveni receptor za aktiviranje. Ovaj pristup stvara mogućnost da se bilo koje sintetičko napravljeno monoklonsko antitelo koristi kao specifični aktivirajući receptor i da se zaobiđe HLA ograničeno prepoznavanje T limfocita. Ovi T limfociti sa himernim antigenskim receptorom (CAR) su dostupna budućnost terapije raka, omogućavajući individualnu terapiju. Pored najatraktivnije tumorske imunologije, postoji impresivan napredak u oblasti imunološke dijagnostike i terapije autoinflamatornih i degenerativnih bolesti/poremećaja, sa fokusom na neuroinflamaciju i neurodegeneraciju.

Konačno, osnovna istraživanja procesa regeneracije kod beskičmenjaka ili nižih kičmenjaka pokazala su da su imunološke sile ključne u regulaciji mirovanja, aktivnosti, razmnožavanja matičnih ćelija, pa čak i u dediferencijaciji ćelija ka mlađim fazama i ponovo matičnim ćelijama. Demonstracija da je kontrolisana dediferencijacija fagocita u matične ćelije ključni korak u potpunoj regeneraciji udova guštera otvorila je novo područje. Dakle, aktuelna perspektiva imunologije je modulacija starenja i regeneracija oštećenih tkiva.

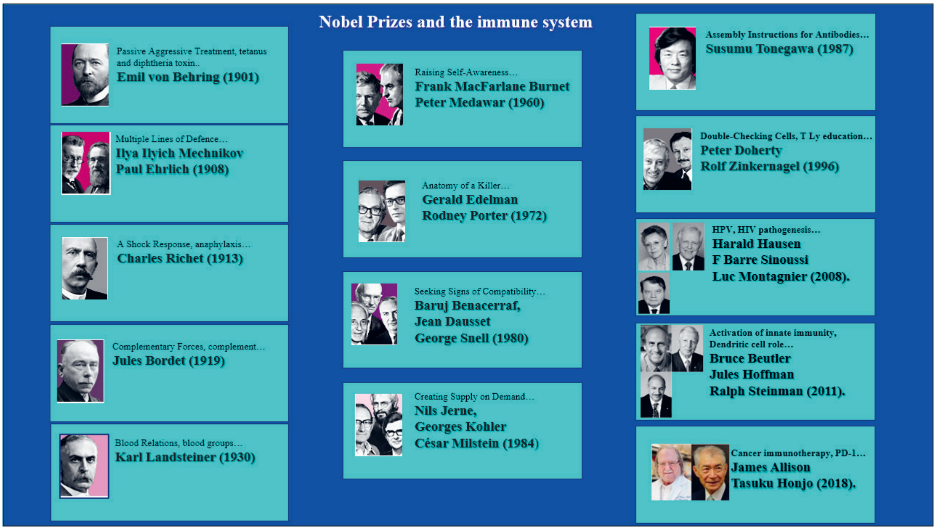
Pojam „imunosti" kao otpornosti na infekciju, prvi put se spominje negde između 500 i 400 godina p. n. e. Prvi koji je upotrebio ovaj izraz, ili preciznije definiciju, bio je *Thucydid*, atinski istoričar i general (Slika 1) (1). U vreme epidemije kuge, masovnog obolevanja i umiranja, *Thucydid* je primetio da osobe koje su prebolele infekciju mogu pomagati i negovati obolele bez opasnosti da se opet inficiraju. *Thucydid* je te osobe nazvao imunim, čime je rođen pojam kojim se uopšte opisuje otpornost na infekciju.

Imunologija je morala da čeka skoro 2000 godina do prve praktične primene. Kinezi su bili prvi koji su upotrebili infekt ili njegove delove kako bi indukovali i postigli protektivnu imunost / otpornost kod zdravih osoba izloženih toj infekciji. Pisani dokumenti iz XV veka opisuju „variolizaciju" kao

Slika 1. *Thucydid,* pojam imunosti

Slika 2. Variolacija i vakcinacija

metod zaštite od infekcije velikih boginja, putem „nazalnog udisanja" suvoga praha, koji je sadržavao viruse velikih boginja (2). Materijal je dobijan iz krasta obolelih. U XVI i XVII veku ovakav način imunizacije protiv velikih boginja postao je prilično raširen u Kini. Svakako iznenađuje standardizacija metode, koja je podrazumevala izbor pacijenata, pripremu i primenu materijala. Naime, kraste su uzimane od pacijenata obolelih od velikih boginja sa blagom kliničkom slikom jer se time sprečavala infekcija teškog oblika kod tretiranih osoba. Materijal je uziman od tri ili četiri kraste, koje su nakon sušenja mešane sa žitom i/ili mošusom i čuvane u pamučnim krpama. Neposredno pre primene ovako pripremljeni materijal bi bio pakovan u srebrne lule, kojima su ga lekari izduvavali u nozdrvu tretirane osobe. Prva primena „variolizacije" u zapadnom delu sveta zabeležena je u Londonu 1700. godine, kada su dva lekara demonstrirala ovaj metod, zahvaljujući iskustvima boravka u Kini, kao zaposlenih u *East India Company* (3). Drugačiji metod variolizacije primenjivan je na teritoriji Otomanske imperije i Africi. Metod nazvan *Dak el Jedri* koristio je tečnost iz pustuli obolelih, koja se nanosila na površinski rez (skarifikat) zdrave osobe. Engleski lekar John Fewster pokazao je 1768. godine da prethodna infekcija kravljim boginjama (*cowpox*) čini osobu otpornom na infekciju velikim boginjama, što je dalo osnova da se materijal sa životinja koristi za imunizaciju ljudi, kao mnogo bezbedniji izvor (4). Edward Jenner, engleski lekar, prvi put je primenio životinjski materijal krava obolelih od *smallpox* infekcije na skarifikatu napravljenom na nadlaktici zdravog osmogodišnjeg dečaka 1796. godine, nazvavši

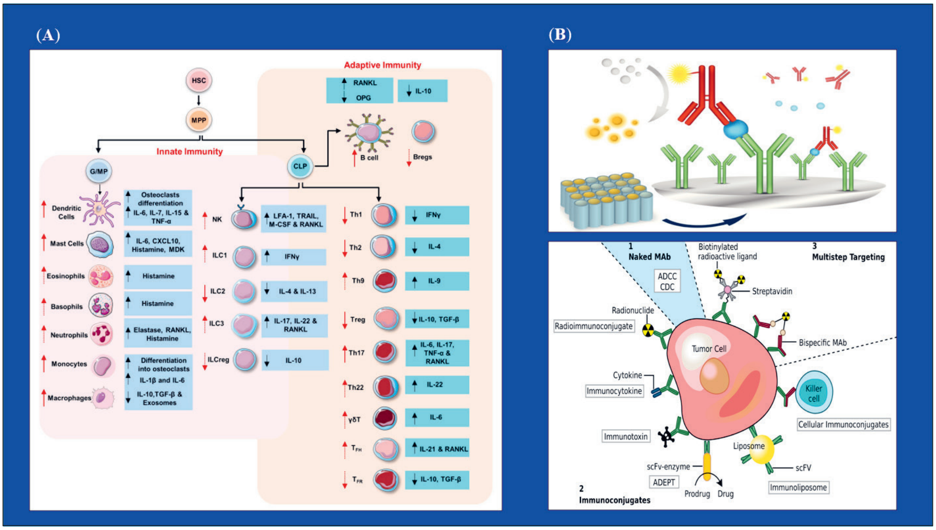


Slika 3. Nobelove nagrade u medicini dodeljene u oblasti imunologije

ovaj metod vakcinacijom (5). Nakon što je objavio opis sopstvenog metoda, vakcinacija je opšteprihvaćena (Slika 2). Uvažavajući značaj koji je vakcinacija donela u kontroli epidemije velikih boginja, pogotovo jednostavan princip same metode, praktično sve sledeće procedure imunizacije protiv difterije, malih boginja, rubele i drugih infekata nazvane su „vakcinacijom".

Pravi procvat imunologija je doživela u XX veku. Svi pronalasci, od identifikacije antitoksina, fagocitnih ćelija imunskog sistema, sistema proteina komplementa, identifikacije T i B limfocita, antitela, pa sve do prepoznavanja kritičnih molekula u komunikaciji tumorskih ćelija i tumora specifičnih limfocita, sa terapijskim mogućnostima uklanjanja inhibitornih uticaja tumor / tumor infiltrišući limfociti, predstavljala su kritična mesta u razvoju savremene biologije i medicine. Često je značaj ovih dostignuća procenjen i prepoznat kao suštinski, što je rezultovalo velikim brojem nobelovih nagrada za medicinu, od *Emila von Behringa*, *Mechnikova*, *Paula Ehrlicha*, *Richeta* pa sve do *Allisona* i *Honjoa* (7) (Slika 3).

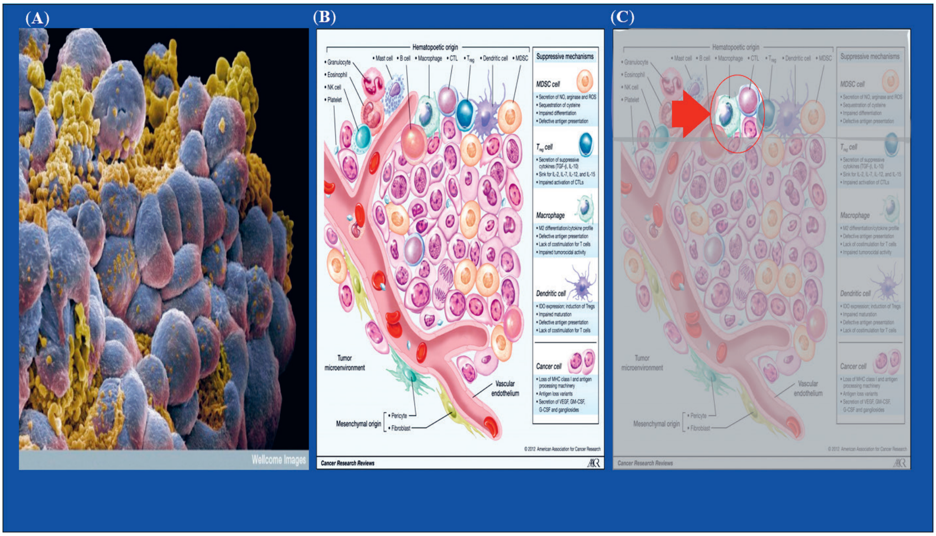
Skorašnja otkrića predstavljaju identifikaciju takozvane plastičnosti limfocita T, sopsobnosti ovih ćelija da, nakon prepoznavanja određenog patogena, ispolje veliki broj različitih funkcionalnih stanja, kao što je produkcija različitih profila citokina, medijatora koji kritično regulišu imunski odgovor (8, 9). Naime, u zavisnosti od potrebe, T limfociti specifični za isti patogen mogu biti Th1, Th2, Treg, Th17, što će uticati na kontrolisanu upalu, uklanjanje, inhibiciju aktivacije ili regeneraciju napadnutoga tkiva ili organa (Slika 4A). S druge strane, monoklonska antitela su ušla u široku dijagnostičku



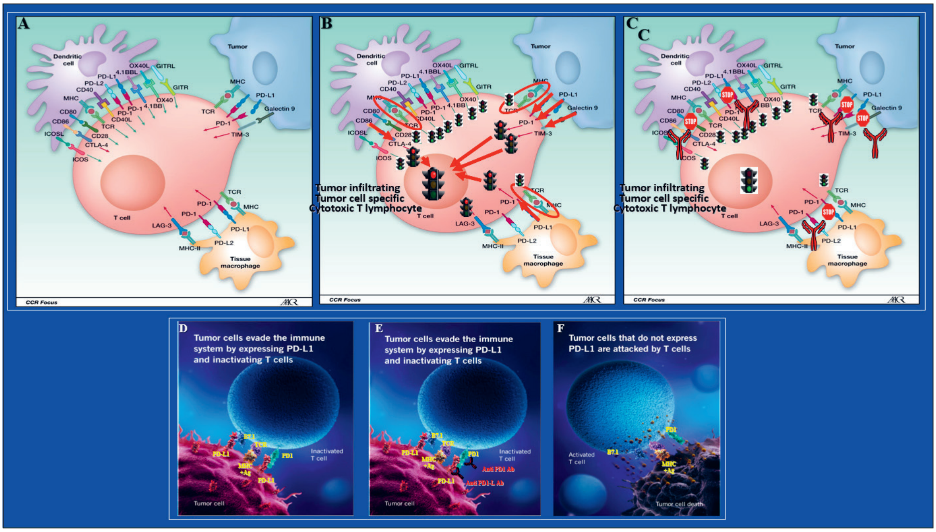
Slika 4. (A) Subpopulacije i plastičnost limfocita T; (B) Dijagnostička i terapijska upotreba monoklonskih antitela

i terapijsku primenu (Slika 4B) do te mere da je danas preko 90% laboratorijskih testova zasnovano na njihovom korišćenju (10, 11).

Među najvažnijim dostignućima moderne imunologije svakako se smatraju rasvetljenja mehanizama antitumorskog odgovora. Dugo se znalo da



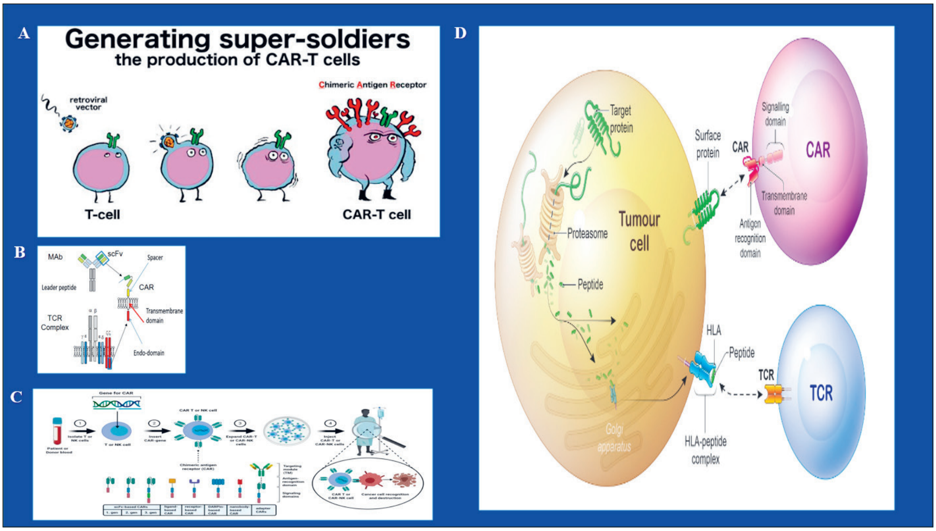
Slika 5. (A) Infiltracija tumora leukocitima; (B) Prisustvo različitih populacija leukocita u tumorskom infiltratu; (C) Interakcija tumorske ćelije, antigen prezentujuće ćelije i citotoksičnog limfocita T



Slika 6. (A) Interakcija limfocita T, antigen prezentujuće ćelije i tumorske ćelije; (B) Inhibitorni signali (crveno), koji sprečavaju eliminaciju tumorskih ćelija od citotoksičnih limfocita T; (C) Uklanjanje inhibitornih signala monoklonskim antitelima protiv PD-L1/PD-1 molekula; (D) Molekulski parovi uključeni u komunikaciju limfocita T i tumorskih ćelija; (E) Vezivanje monoklonskih antitela za PD-L1/PD-1 molekule; (F) Eliminacija tumorskih ćelija posredovana limfocitima T nakon uklonjenih inhibitornih signala

maligne ćelije ispoljavaju niz inhibitornih uticaja na zdrave limfocite T, koji infiltrišu tumorsko tkivo, mada bez preciznog razumevanja mehanizama te inhibicije (Slika 5A, tumor, infiltrišući leukociti obojeni zlatno žuto).

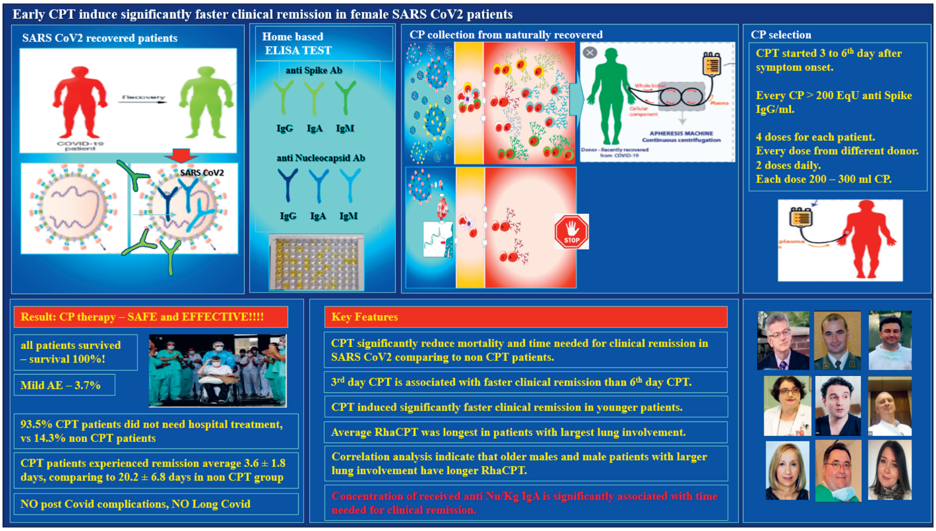
Uprkos ranijim shvatanjima da odgovor na tumor izostaje zbog nedovoljnog prisustva imunskih ćelija, tumor infiltrišu praktično svi tipovi leukocita (Slika 5B) (12). Međutim, ključna razmena informacija odigrava se na nivou komunikacije svega tri tipa ćelija: tumorske ćelije, antigen prezentujuće ćelije i citotoksičnih limfocita T (Slika 5C). Citotoksični limfocit T prepoznaje antigene determinante tumorske ćelije, tako štu su mu one prikazane u MHC molekulima na površini antigen prezentujućih ćelija (Slika 6A). Po dobijanju drugih aktivacionih signala i sledstvene aktivacije, citotoksični limfocit T bi, nakon prepoznavanja antigenih determinanti u MHC molekulama na tumorskim ćelijama, trebao da eliminiše tu tumorsku ćeliju. Međutim, eliminacija izostaje zbog inhibitornih signala isporučenih membranskim molekulima PD-L1 na tumorskim ćelijama (PD-1 na limfocitima T, Slika 6B) (12, 13, 14). Allison i Honjo su pokazali važnost ove inhibicije, kao i mogućnost da se ona prevaziđe terapijskom upotrebom monoklonskih antitela na PD-L1/PD1, čime se omogućuje fiziološka uloga



Slika 7. (A) Šematski prikaz CAR T limfocita; (B) Kreiranje CAR T receptora, deo koji specifično prepoznaje željeni molekul je kreiran od varijabilnog kraja antitela, koji je konjugovan sa transmembranskim delom konvencionalnog T ćelijskog receptora; (C) Prikaz procedure kreiranja CAR T, od uzimanja leukocita, transformacije ćelija do terapijske primene; (D) Razlika u prepoznavanju antigena između CAR T limfocita i konvencionalnog T limfocita

citotoksičnih limfocita T i posledična eliminacija tumorskih ćelija (Slika 6C, 6D, 6E, 6F).

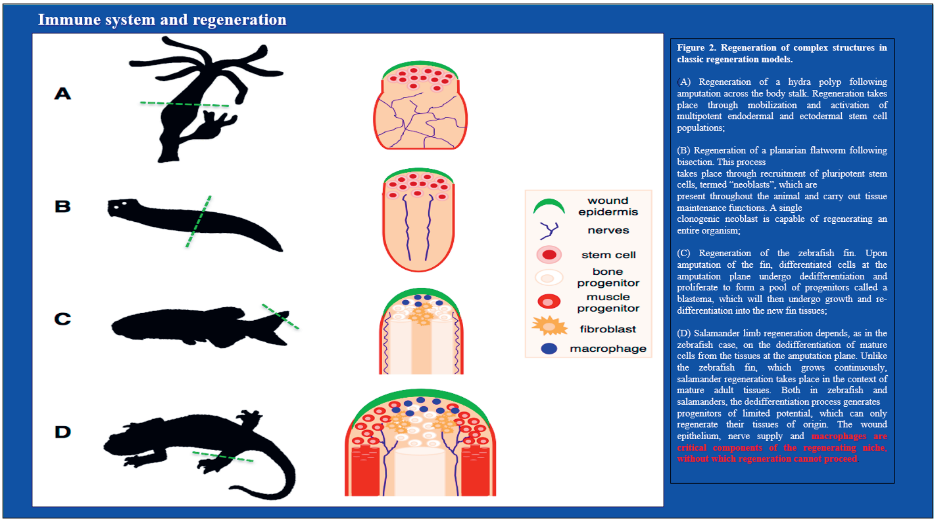
Zahvaljujući dostignućima iz oblasti molekularne biologije, omogućen je potpuno novi pristup imunološki zasnovanoj terapiji tumora, kreiranje takozvanih CAR-T limfocita, ćelija koje nose himerični antigeni receptor (Slika 7A). Himerični antigeni receptor kreiran je od delova receptora B i T limfocita, tako što deo koji obezbeđuje specifično prepoznavanje potiče od varijabilnog regiona antitela, dok je deo koji omogućuje provod signala i aktivaciju poreklom od membranskog aktivacionog dela T ćelijskog receptora (Slika 7B) (15, 16). Klinička procedura podrazumeva uzimanje leukocita (leukoferezom) od pacijenta sa ustanovljenim tumorom; genetsku modifikaciju T limfocita kako bi oni ispoljavali CAR receptor željene specifičnosti; njihovu propagaciju do dovoljnog broja i ponovno unošenje istom pacijentu (Slika 7C). Prednost terapije CAR T limfocitima leži u principu njihovog prepoznavanja tumora. Naime, za razliku od konvencionalnih limfocita T, koji moraju prepoznati antigenu determinantu samo u okviru MHC molekula, CAR T limfociti, zahvaljujući delu antitela u receptoru, prepoznaju bilo koji membranski molekul bez ikakvog ograničenja i potreba za prethodnom ćelijskom preradom (Slika 7D).



Slika 8. Primena rekonvalescentne plazme u terapiji obolelih od COVID-19

Moderna imunologija svoje uspehe ne zasniva samo na naprednoj tehnologiji i dostignućima drugih prirodnih grana nauke. U lečenju bolesnika obolelih od COVID-19, mi smo koristili princip koji je pre više od 100 godina primenio *Francesko Cenci* (17). Doktor Cenci, koji je radio u selu kraj tadašnje Peruđe, u vreme epidemije tada smrtonosnih malih boginja 1901. godine, zapazio je da osobe koje su jednom preležale infekciju, ne mogu se opet zaraziti. U vreme pre otkrića antitela, pretpostavljajući da se u krvi oporavljenih osoba formira nekakav zaštitni faktor, dr *Cenci* je uzeo 600 ml krvi od 20-godišnjeg mladića, koji je preležao male boginje, i njegovim serumom, tzv. konvalescentnim serumom, tretirao 4 dečaka stara između 4 i 8 godina. Deca koja su dobila konvalescentni serum nisu se zarazila malim boginjama, za razliku od ostalih vršnjaka. Pet godina kasnije, 1906, u drugom naletu epidemije malih boginja, dr *Cenci* je po prvi put konvalescentni serum primenio ne samo na zdravoj deci kao profilaksu, već i na deci sa teškim oblikom upale pluća, kao posledice infekcije malim boginjama. Uspeh ovakve terapije dr *Cencija*, ponavljan je u terapiji difterije, kasnije i drugih infektivnih bolesti. Naš pristup lečenju obolelih od COVID-19 zasnovan je na istom principu, sa razlikom što je u svakoj dozi date plazme precizno određena koncetracija specifičnih antitela (Slika 8).

Naime, da bi napravili selekciju idealnih donora, prvo smo kreirali test kojim smo određivali koncentraciju antitela specifičnih na kritične antigene strukture SARS CoV2 virusa, na S1 protein i proteine jedra (6 različitih antitela). Od oporavljenih sa najvećom koncentracijom antitela, konvalescentna



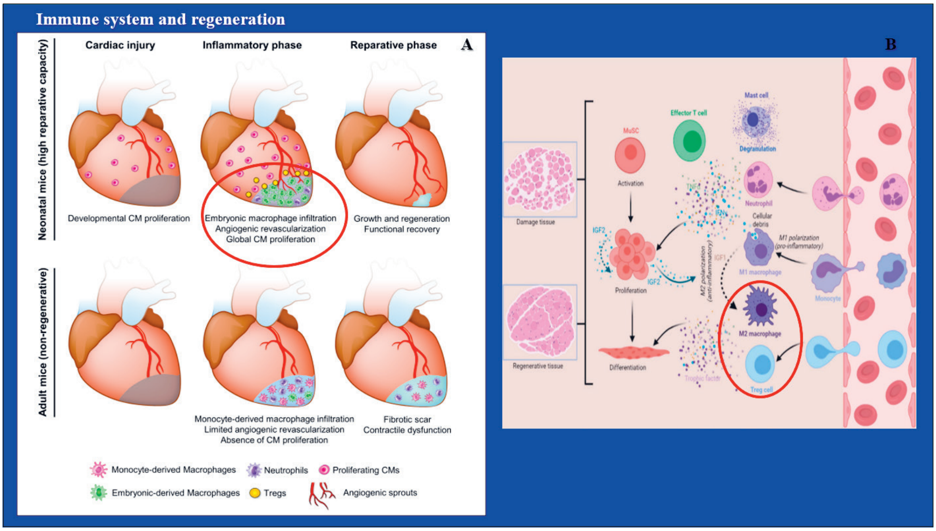
Slika 9. Mehanizam regeneracije kod pojedinih nižih i kompleksnih organizama

plazma je uzimana procesom afereze. Tretirani pacijenti su bili sa inicijalnom ili odmaklom pneumonijom. Svi pacijenti su dobili najmanje 4 doze plazme (4x 200–250 ml) od različitih davaoca, u toku prvih 4–7 dana od potvrde infekcije. Svi pacijenti (127) su preživeli, skoro 94% nije zahtevalo kiseoničnu podršku niti bolničko lečenje, nijedan pacijent nije imao komplikacije niti postkovid, a prosečno vreme do remisije iznosilo je 4 dana, naspram prosečnih 20 dana u grupi pacijenata lečenih bez konvelescentne plazme.

Dostignuća imunologije ima u praktično svakoj oblasti medicine, bazičnoj, aplikativnoj, translacionoj, dijagnostičkoj i terapijskoj oblasti. Imajući sve ovo u vidu, postavlja se pitanje koje su perspektive imunologije.

Jedno od najintrigantnijih polja u biologiji i medicini je oblast regeneracije. Neki niži organizmi (hidra, pojedini crvi), ali i neki kompleksniji organizmi (zebra, ribe, gušteri) imaju moć kompletne regeneracije pojedinih organa i tkiva, zaključno sa kapacitetom obnavljanja povređenih pluća ili srca (Slika 9) (18). Upravo kod guštera pokazano je da su kritične ćelijske komponente za regeneraciju ćelije oštećenog epitela, nervne ćelije, ali i različiti tipovi makrofaga, čije uklanjanje rezultuje neuspelim procesom regeneracije.

Novija ispitivanja na sisarima, na neonatalnim miševima (koji imaju kompletnu moć regeneracije) i adultnim miševima (nema regeneracije) pokazala su da su, nakon oštećenja srčanog mišića, neonatalni miševi sposobni da ponovo izgrade oštećeni organ. Ključna razlika u neonatalnih miševa bila je u prisustvu takozvanih embrionalnih markofaga, koji su gusto infiltrisali povređeni organ, regulišući ponovnu vaskularizaciju i matičnih ćelija (Slika 10A) (19).



Slika 10. (A) Regeneracija srčanog mišića kod neonatalnih i adultnih miševa; (B) Značaj regulatornih markofaga i regulatornih T limfocita u procesima regeneracije

Druga studija pokazala je važnost takozvanih regulatornih markofaga i regulatornih T limfocita u procesima regeneracije. Naime, pravovremena promena iz efektorskih T limfocita u regulatorne T, pravovremena promena iz upalnih procesa u procese ponovne izgradnje tkiva, predstavlja granicu između dva ishoda, procesa ožiljavanja ili procesa kompletne regeneracije, čak i kod odraslih jedinki.

ZAKLJUČAK

U poslednjih stotinu godina imunologija je prešla ogroman put, od kliničke grane skromnih terapijskih mogućnosti, do multidisciplinarne nauke, esencijalne u dijagnostici i lečenju brojnih patologija i stanja humane medicine. Posebno su uzbudljivi rezultati istraživanja i značaja imunskih mehanizama u regeneraciji tkiva i organa, što daje svetlu budućnost u rešavanju do sada neizlečivih stanja i bolesti.

BIBLIOGRAFIJA

[1] Finley J, Jr. *The Complete Writings of Thucydides: The Peloponnesian War*. Modern Library, New York; 1951.
[2] Williams, Gareth (2010). *Angel of Death*. Basingstoke: Palgrave Macmillan. ISBN 978–0230274716.
[3] Silverstein, Arthur M. (2009). *A History of Immunology* (2nd ed.). Academic Press. p. 293. ISBN 9780080919461.

[4] Pead, Patrick (2003). „Benjamin Jesty; new light in the dawn of vaccination“. *Lancet*. **362** (9401): 2104–09. doi: 10.1016/s0140–6736(03) 15111–2.

[5] L. Thurston and G. Williams (2015) „ An examination of John Fewster's role in the discovery of smallpox vaccination,“ *Journal of the Royal College of Physicians of Edinburgh*, 45: 173–179.

[6] Baxby, Derrick (1999). „Edward Jenner's Inquiry; a bicentenary analysis“. *Vaccine*. **17** (4): 301–07. doi: 10.1016/s0264–410x(98) 00207–2.

[7] Nobel Prizes and the immune system. NobelPrize. org. Nobel Prize Outreach AB 2022. Fri. 2 Dec 2022. https://www.nobelprize.org/prizes/themes/nobel-prizes-and-the-immune-system

[8] Varadé, J, Magadán, S. & González-Fernández, Á. Human immunology and immunotherapy: main achievements and challenges. *Cell Mol Immunol* 18, 805–828 (2021). https://doi.org/10.1038/s41423–020–00530–6

[9] DuPage, M, Bluestone, J. Harnessing the plasticity of CD4^{+} T cells to treat immune-mediated disease. *Nat Rev Immunol* 16, 149–163 (2016). https://doi.org/10.1038/nri.2015.18

[10] Quinteros DA, Bermúdez JM, Ravetti S, Cid A, Allemandi DA, Palma SD. Therapeutic use of monoclonal antibodies: general aspects and challenges for drug delivery. Nanostructures for Drug Delivery. 2017: 807–33. doi: 10.1016/B978–0–323–46143–6.00025–7. Epub 2017 Mar 31. PMCID: PMC7151974.

[11] Tigabu Demlie, Endale Balcha, Haben Fesseha. Monoclonal Antibody and its Diagnostic Application-Review. Biomed J Sci & Tech Res 30(4) -2020. BJSTR. MS. ID. 004997. DOI: 10.26717/BJSTR. 2020.30.004997.

[12] Sid P. Kerkar & Nicholas P. Restifo. Cellular Constituents of Immune Escape within the Tumor Microenvironment. *Cancer Res* (2012) 72 (13): 3125–3130. https://doi.org/10.1158/0008–5472.CAN-11–4094

[13] Postow MA, Callahan MK, Wolchok JD. Immune Checkpoint Blockade in Cancer Therapy. J Clin Oncol. 2015 Jun 10; 33(17): 1974–82. doi: 10.1200/JCO. 2014.59.4358. Epub 2015 Jan 20. PMID: 25605845; PMCID: PMC4980573.

[14] La-Beck NM, Jean GW, Huynh C, Alzghari SK, Lowe DB. Immune Checkpoint Inhibitors: New Insights and Current Place in Cancer Therapy. Pharmacotherapy. 2015 Oct; 35(10): 963–76. doi: 10.1002/phar. 1643.

[15] Heczey A, Louis CU. Advances in chimeric antigen receptor immunotherapy for neuroblastoma. Discov Med. 2013 Dec; 16(90): 287–94.

[16] Richards RM, Sotillo E, Majzner RG. CAR T Cell Therapy for Neuroblastoma. Front Immunol. 2018 Oct 16; 9: 2380. doi: 10.3389/fimmu. 2018.02380.

[17] Cenci F. Alcune esperienze di sieroimmunizzazione e sieroterapia nel morbillo. *Riv Clin Ped.* 1907; 5: 1017–1025.

[18] Yun MH. Changes in Regenerative Capacity through Lifespan. Int J Mol Sci. 2015 Oct 23; 16(10): 25392–432. doi: 10.3390/ijms161025392.

[19] Lai, SL, Marín-Juez, R. & Stainier, D. Y. R. Immune responses in cardiac repair and regeneration: a comparative point of view. *Cell. Mol. Life Sci.* **76**, 1365–1380 (2019). https://doi.org/10.1007/s00018–018–2995–5

[20] Fang, J, Feng, C, Chen, W. et al. Redressing the interactions between stem cells and immune system in tissue regeneration. *Biol Direct* 16, 18 (2021). https://doi.org/10.1186/s13062–021–00306–6

MATEMATIČKE NAUKE

MATHEMATICAL SCIENCES

Vladimir DRAGOVIĆ[*], Borislav GAJIĆ[**]

BRIDGING STATISTICS, GEOMETRY, AND MECHANICS

Abstract: We emphasis the importance of bridges between statistics, mechanics, and geometry. In par ticular, we developed and employed links between pencils of quadrics, moments of inertia, and linear and orthogonal regressions. For a given system of points in $\mathbb{R}^k$ representing a sample of a full rank, we recently constructed a pencil of confocal quadrics which provided a useful geometric tool to study the data.

Key words and phrases: *ellipsoid of concentration, confocal pencil of quadrics, planar moments of inertia, restricted regression, regularization and shrinkage, restricted PCA*

1. Introduction

We will talk here about bridges between statistics, mechanics, and geometry. In particular, we talk about the links we developed and employed between pencils of quadrics, moments of inertia, and linear and orthogonal regressions. This cross-fertilization between the areas appears to be beneficial for each of them. Individual quadrics have been in use in statistics from late XIX century. Recently, we constructed a new object, a confocal pencil of quadrics, associated to a given data set, see [**15, 16**]. We demonstrated in [**15, 16**] that this confocal pencil of quadrics is a natural and very useful instrument to understand the data.

1.1. Confocal pencils of quadrics. We age going to recall about confocal pencils of conics in the plane and its generalizations, the confocal pencils of quadrics in $\mathbb{R}^k$ for for any k. We will also recall the definition of the associated Jacobi elliptic coordinates.

The family of confocal conics in the plane can be given analytically:

$$\mathcal{C}_\lambda \ : \ \frac{x^2}{\alpha - \lambda} + \frac{y^2}{\beta - \lambda} = 1. \tag{1.1}$$

---

[*] Vladimir Dragović, Department of Mathematical Sciences, The University of Texas at Dallas, Richardson, TX, USA, Mathematical Institute, of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

[**] Borislav Gajić, Mathematical Institute, of the Serbian Academy of Sciences and Arts, Belgrade, Serbia

We notice that the family (1.1) contains two types of smooth conics: ellipses, when $\lambda < \beta$, and hyperbolas, when $\lambda \in (\beta, \alpha)$. Gometrically, they all share a common pair of foci. We also notice that there are two degenerated conics in the confocal family: the x-axis, for $\lambda = \beta$; and the y-axis, for $\lambda = \alpha$.

Each point in the plane, which is not a focus of the confocal family, lies on exactly two conics $\mathcal{C}_{\lambda_1}$ and $\mathcal{C}_{\lambda_2}$ from (1.1) – one ellipse and one hyperbola, which are orthogonal to each other at the intersection point.

Let an ellipsoid be given in $\mathbb{R}^k$ by:

$$\mathcal{E} \ : \ \frac{x_1^2}{\alpha_1} + \cdots + \frac{x_k^2}{\alpha_k} = 1, \quad \alpha_1 > \alpha_2 > \cdots > \alpha_k > 0. \tag{1.2}$$

The family of quadrics confocal with $\mathcal{E}$ is:

$$\mathcal{Q}_\lambda(\mathbf{x}) = \frac{x_1^2}{\alpha_1 - \lambda} + \cdots + \frac{x_k^2}{\alpha_k - \lambda} = 1, \tag{1.3}$$

where λ is a real parameter. We will say that confocal quadrics $\mathcal{Q}_\lambda, \mathcal{Q}_\mu$ are *of the same type* if there exists $j, j = 1, \dots, k-1$ such that $\alpha_j > \lambda, \mu > \alpha_{j+1}$ or $\alpha_k > \lambda, \mu$. For a point given by its Cartesian coordinates $\mathbf{x} = (x_1, \dots, x_k)$, the Jacobi elliptic coordinates $(\lambda_1, \dots, \lambda_k)$ are defined as the solutions of the equation in λ: $\mathcal{Q}_\lambda(\mathbf{x}) = 1$. The quadrics $\mathcal{Q}_{\lambda_1}, \mathcal{Q}_{\lambda_2}, \dots, \mathcal{Q}_{\lambda_k}$ which contain a given point $\mathbf{x}$ are of different types [**4**]. Jacobi introduced the Jacobi elliptic coordinates in [**31**] in 1838 when he used them to integrate the equations of geodesics on ellipsoids.

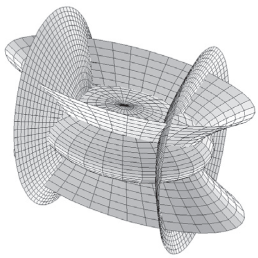


FIGURE 1. From [**17**] and [**18**]: Confocal quadrics in $\mathbb{R}^3$: one ellipsoid, one 1-sheeted hyperboloid, and one 2-sheeted hyperboloid, intersecting orthogonally at eight points.

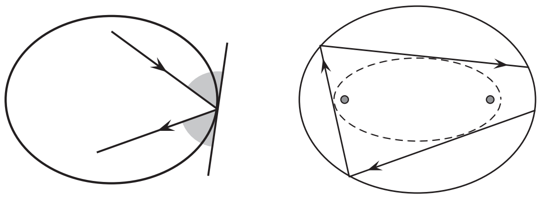
1.2. Application to billiard systems within ellipses. Similar in spirit to the use of the Jacobi elliptic coordinates in solving the equations of geodesics on ellipsoids, is the application of these coordinates to billiards within ellipsoids.

The theory of mathematical billiards is one of theoretical models for motion of a ball inside a billiard table, and its reflection off the boundary of the table. Let us suppose that a planar domain is given. *Mathematical billiard* in this domain is a dynamical system where a material point of unit mass moves under inertia without constraints and friction within the domain, and obeys the billiard reflection law off the boundary [**32**]. The billiard reflection law coincides with the law reflection of light in geometric optics. This theory provides an idealized model of the physical reality in many aspects. For example, a usual billiard ball is replaced by a material point, and the friction and spin are neglected. Nevertheless, this model has many important and natural applications, for example in geometric optics. Thus, the billiard dynamics has two different regimes: the first one is inside the billiard domain, when we assume that the material point moves under the inertia, i.e. uniformly along straight lines. The second regime concerns with the impacts off the boundary. We assume here that the impacts are *absolutely elastic*. In other words, the geometric billiard law is satisfied: the impact and reflection angles are congruent to each other and the speed remains unchanged after the impacts. Here, the trajectories of the mathematical billiards are polygonal lines with vertices at the boundary.

We focus here only on *elliptical billiards* – defined by an ellipse in the Euclidean plane as the boundary of a billiard table:

$$\mathcal{E} \ : \ \frac{x^2}{\alpha} + \frac{y^2}{\beta} = 1, \quad \alpha > \beta > 0.$$

The key property of such billiards is the existence of caustics. Each trajectory of the elliptic billiard has *a caustic*: a curve such that each segment of the trajectory lies on a line tangent to the caustic. The existence of a caustic is a geometric manifestation of the integrability of billiard systems within ellipses.



(a) (b)

FIGURE 2. 2(a): Billiard reflection law; 2(b): The caustic of a billiard trajectory.

1.3. Quadrics in Statistics. Ellipses, as two-dimensional quadrics, got incorporated in statistics in 1886 in Galton's paper [**26**]. That seminal paper introduced the law of regression.

To study the hereditary transmissions, Galton collected the data about the height of 930 adult children and 205 of their respective parentages. He introduced a "mid-parent" height, as a weighted average of the heights of the parents, and assigned it to each pair of parents. He established the average regression from mid-parent to offsprings and from offsprings to mid-parent. Using this data Galton formulated the law of regression toward mediocrity:

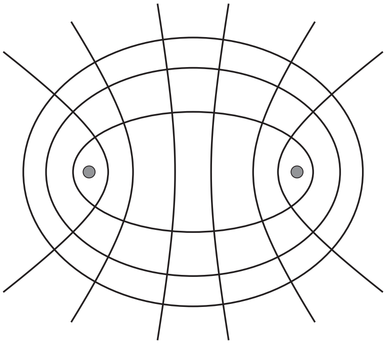


Figure 3. Confocal family of conics in plane

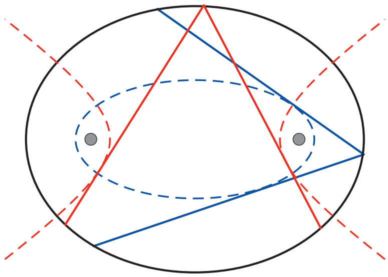


Figure 4. A billiard trajectory with an ellipse as the caustic and a billiard trajectory with a hyperbola as a caustic.

When Mid-Parents are taller than mediocrity, their Children tend to be shorter than they. When Mid-Parents are shorter than mediocrity, their Children tend to be taller than they.

Thus, the notion of *regression* got into statistics, thanks to Galton. From mathematical perspective, the background method of least squares, appeared in early 1800's, due to Gauss and Legendre.

Galton discovered an important use of ellipses in statistical analysis. Here we quote Galton:

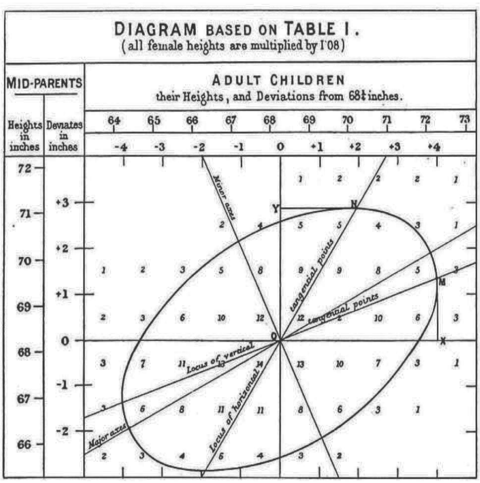


Figure 5. From [**26**].

“*...I found it hard at first to catch the full significance of the entries in the table...They came out distinctly when I ‘smoothed’ the entries by writing at each intersection of a horizontal column with a vertical one, the sum of entries of four adjacent squares... I then noticed that lines drawn through entries of the same value formed a* *series of concentric and similar ellipses. Their common center ... corresponded to* $68\frac{1}{4}$ *inches. Their axes are similarly inclined. The points where each ellipse in succession was touched by a horizontal tangent, lay in a straight line inclined to the vertical in the ratio of* $\frac{2}{3}$*; those where they were touched by a vertical tangent lay in a straight line inclined to the horizontal in the ratio of* $\frac{1}{3}$*. These ratios confirm the values of average regression already obtained by a different method, of* $\frac{2}{3}$ *from mid-parent to offspring, and of* $\frac{1}{3}$ *from offspring to mid-parent...These and other relations were evidently a subject for mathematical analysis and verification...I noted these values and phrased the problem in abstract terms such as a competent mathematician could deal with, disentangled from all references to heradity, and in that shape submitted it to Mr. Hamilton Dixson, of St. Peter's College, Cambridge...*

I may be permitted to say that I never felt such a glow of loyalty and respect towards the sovereignty and magnificent sway of mathematical analysis as when his answer reached me, confirming, by purely mathematical reasoning, my various and laborious statistical conclusions with far more minuteness than I had dared to hope... His calculations corrected my observed value of mid-parental regression from $\frac{1}{3}$ *to* $\frac{6}{17.6}$*, the relation between the major and minor axis of the ellipse was changed 3 per cent. (it should be as* $\frac{\sqrt{7}}{\sqrt{2}}$*)...*"

The notions of moments in statistics came from mechanics, where they were originally introduced in the three-dimensional space. The next Section 1.4 reviews these notions from a mechanics perspective.

1.4. Axial and planar moments of inertia in $\mathbb{R}^3$. We review the notions of the axial moment of inertia and the operator of inertia with respect to a point, as they are introduced in mechanics. Suppose that N points $M_1, ..., M_N$ with masses $m_1, m_2, ..., m_N$ are given in the space $\mathbb{R}^3$. Sum of the masses $m = \sum_{j=1}^{N} m_j$ is the total mass of the system of the points, while *the center of masses* of the N given points is the point C that satisfy

$$\sum_{j=1}^{N} m_j \overrightarrow{CM_j} = 0.$$

For a given line $u \subset \mathbb{R}^3$, *the axial moment of inertia* I_u is defined as

$$I_u = \sum_{j=1}^{N} m_j d_j^2,$$

where d_j is the distance from the point M_j to the line u, $j = 1, \ldots, N$.

Suppose that the line u contains point O and that it is defined with the unit vector $\mathbf{u}_0$.

Then the axial moment of inertia I_u for the axis u can be rewritten in the form:

$$I_u = \sum_{i=1}^{N} m_i d_i^2 = \sum_{i=1}^{N} m_i \langle \mathbf{u}_0 \times \overrightarrow{OM_i}, \mathbf{u}_0 \times \overrightarrow{OM_i} \rangle = \sum_{i=1}^{N} m_i \langle \overrightarrow{OM_i} \times (\mathbf{u}_0 \times \overrightarrow{OM_i}), \mathbf{u}_0 \rangle = \langle I_O \mathbf{u}_0, \mathbf{u}_0 \rangle.$$

If one considers u as an arbitrary line that contains fixed point O, the relation

$$I_u = \langle I_O \mathbf{u}_0, \mathbf{u}_0 \rangle$$

defines *the operator of inertia at the point O* and it is denoted by I_O. In the Cartesian coordinates $Oxyz$ the matrix of the operator I_O is symmetric and positive-definite [**4**].

Its diagonal elements I_{11}, I_{22}, I_{33} are the moments of inertia for the coordinate axes Ox, Oy, Oz respectively. For example:

$$I_{11} = \sum_{j=1}^{N} m_j(y_j^2 + z_j^2).$$

The non-diagonal elements of this matrix are called *the centrifugal moments of inertia.* For example:

$$I_{12} = -\sum_{p=1}^{N} m_p x_p y_p,$$

and similarly, one can define all other I_{ij}.

As for any symmetric matrix, one can choose an orthogonal basis in which the matrix of the operator I_O has a diagonal form:

$$I_O = \mathrm{diag}(I_1, I_2, I_3).$$

The scalars I_1, I_2, I_3 are called *the principal axial moments of inertia* while the corresponding coordinates are called *principal coordinates.*

The inertia operator I_O defines *the axial ellipsoid of inertia at the point* O:

$$\langle I_O u, u\rangle = 1.$$

In the principal coordinates, the equation of the axial ellipsoid of inertia becomes

$$I_1 x^2 + I_2 y^2 + I_3 z^2 = 1.$$

An important formula that makes a connection between the axial moments of inertia for two parallel axes, where one of them contains the center of masses, is the essence of the Huygens-Steiner Theorem (see e.g. [**4**]).

Theorem 1.1 (The Huygens-Steiner Theorem, [**4**]). *Let the axis u contain the center of masses C and let u_1 be a line parallel to u. Denote by I_{u_1} and I_u the corresponding axial moments of inertia of a given system of points with the total mass m. Then*

$$I_{u_1} = I_u + md^2, \tag{1.4}$$

where d is the distance between the parallel lines u and u_1.

As an immediate consequence, one gets an important property of the center of masses:

Corollary 1.1. *Given a system of points and one direction. Among all the lines parallel to the given direction, the least moment of inertia is attained for the line which contains the center of masses of the set of points.*

In a similar manner as for an axial moments of inertia, for a given set of points and for a given plane $\pi \subset \mathbb{R}^3$ one can introduce the planar moment of inertia. It is defined by formula:

$$J_\pi = \sum_{j=1}^{N} m_j D_j^2,$$

where D_j is the distance between the point M_j and the plane π, for $j = 1, \dots, N$. For vectors $\mathbf{n}_1$ and $\mathbf{n}_2$ define the operator

$$\langle J_O\mathbf{n_1},\mathbf{n_2}\rangle = \sum_{j=1}^{N} m_j\langle\overrightarrow{OM_j},\mathbf{n_1}\rangle\langle\overrightarrow{OM_j},\mathbf{n_2}\rangle.$$

If the plane π is determined by its point O and the normal unit vector $\mathbf{n}$, then the planar moment of inertia J_π can be rewritten in the form

$$J_\pi = \sum_{j=1}^{N} m_j\langle\overrightarrow{OM}_j,\mathbf{n}\rangle^2 = \langle J_O\mathbf{n},\mathbf{n}\rangle. \tag{1.5}$$

This is the reason to call J_O *the planar inertia operator at the point* O.

In the Cartesian coordinates $Oxyz$, the diagonal elements of the matrix of the planar inertia operator are the planar moments of inertia for the coordinate planes. For example:

$$J_{11} = \sum_{j=1}^{N} m_j x_j^2$$

is the moment of inertia for the coordinate plane Oyz. The nondiagonal elements of the planar inertia operator are also called *the centrifugal moments of inertia*. For example:

$$J_{12} = \sum_{p=1}^{N} m_p x_p y_p,$$

and similarly for other J_{ij}. The planar inertia operator is symmetric and positive-definite. Thus, one can choose a basis in which its matrix has a diagonal form:

$$J_O = \operatorname{diag}(J_1, J_2, J_3).$$

The extremal property of the center of masses is fulfilled in the planar case also. It is consequence of an analog of the Huygens-Steiner Theorem for the case of the planar moments of inertia.

Proposition 1.1. [**16**] *Given a system of points in $\mathbb{R}^3$ with the total mass m and the center of masses C. If the planes π and π_1 are parallel and π contains the center of masses C, then*

$$J_{\pi_1} = J_\pi + mD^2,$$

where D is the distance between the parallel planes π and π_1.

An immediate and important conclusion follows from the coordinate expressions for the planar and axial operators of inertia J_O and I_O of the same system of points and with respect to the same point O: both operators have a diagonal form in the same orthogonal basis, called the principal basis. Using the Pythagorean theorem, one can see that the axial moment of inertia for the axis u is the sum of the planar moments of inertia for two orthogonal planes having the line u as their intersection. For example, for the principal axes, we have:

$$I_i = J_j + J_p, \tag{1.6}$$

where (i, j, p) is a cyclic permutation of $(1, 2, 3)$.

2. The classical results of Pearson and their generalizations

Karl Pearson was one of the founding fathers of modern statistics. He investigated the question of the hyperplane which minimizes the mean square distance from a given set of points in $\mathbb{R}^k$, for any $k \geq 3$. In his own words, Pearson formulated the problem [**34**]: *"In the case we are about to deal with, we suppose the observed variables–all subject to error–to be plotted in plane, three-dimensioned or higher space, and we endeavour to take a line (or plane) which will be the 'best fit' to such a system of points. Of course the term 'best fit' is really arbitrary; but a good fit will clearly be obtained if we make the sum of the squares of the perpendiculars from the system of points upon the line or plane a minimum."*

Pearson noticed that the notion of the best fit is not uniquely determined. In the measurement error models, also known as regression with errors in variables (EIV), [**9**], a natural choice is the choice of the squares of the perpendiculars. This was indicated in Pearson's above note: "we suppose the observed variables–all subject to error". Such models are in use when the both predictors and responses are known with some error. In the usual linear regressions, the situation is different: for the predictors the exact values are known. Only responses are assumed to be known with some error. Thus, the squares of distances along one of the axes is in use in such regression models. We called that *directional regression* in a given direction. We talked more about its geometric aspects in the last section of [**16**].

Here we are going to explain in more details the classical simple linear regression model and the regression with error in variables (EIV) models, [**9**]. For a classical simple regression model, it is assumed that the values $(x^{(i)})_{i=1}^N$ are known, fixed values, as for example values set up in advance in the experiment. The values $(y^{(i)})_{i=1}^N$ are observed values of uncorrelated random variables Y_i, $i = 1, \ldots, N$ with the same variance σ^2. It is assumed that the predictors $x^{(i)}$ and responses $(y^{(i)})_{i=1}^N$ are related with a linear

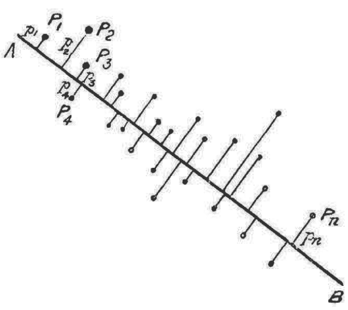


Figure 6. From [**34**].

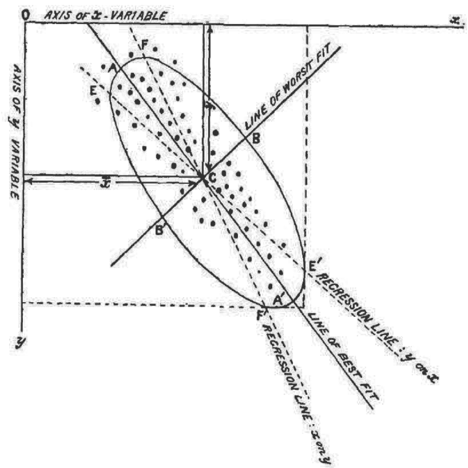


Figure 7. From [**34**].

relationship:

$$EY_i = \alpha + \beta x^{(i)}, \quad i = 1, \ldots, N.$$

This can be restated as

$$Y_i = \alpha + \beta x^{(i)} + \epsilon_i, \quad i = 1, \dots, N,$$

where ϵ_i are called *the random errors* and they are uncorrelated random variables with zero expectation and the same variance σ^2. In such models the regression is of Y on x, i.e. in the vertical direction. This model was in the background of the Galton study [**26**], mentioned above.

Important situations appear when predictors are known only up to some error. They are described by measurement error models. There the observed pairs $(x^{(i)}, y^{(i)})_{i=1}^N$ are sampled from random variables (X_i, Y_i) with means satisfying the linear relationship

$$EY_i = \alpha + \beta(EX_i), \quad i = 1, \dots, N.$$

Denoting $EX_i = \xi_i$, the errors in variables model can be defined as

$$Y_i = \alpha + \beta\xi_i + \epsilon_i, \quad X_i = \xi_i + \delta_i, \quad i = 1, \dots, N,$$

where both X_i and Y_i have error terms which belong to mean zero normal distributions, such that all ϵ_i, $i = 1, \dots, N$ have the same variance σ_ϵ^2 and all δ_i, $i = 1, \dots, N$ have the same variance σ_δ^2. Since x_i and y_i are both known with an error, the orthogonal least square method is more natural to apply here, that is to say to apply the orthogonal regression. This we are going to introduce next under the assumption that $\eta = \sigma_\epsilon^2/\sigma_\delta^2 = 1$ and in a general dimension k. One should also mention that applications of the orthogonal least square method in the models with measurement errors have limitations, based on the fact that value of η could be unknown, see e.g. [**8**], [**7**]. Here we deals with the cases when *η is known*.

The case $\eta = 1$ historically originated from [**1, 2**]. Pearson established *orthogonal regression*, using the squares of the perpendiculars that corresponds to the case $\eta = 1$. Our aim is to study geometric aspects of the orthogonal regression. We also adopted Pearson's generality full rank assumption, that the given system of points is in a general position, which means that the points do not belong to a hyper-plane.

Let N points $(x_1^{(i)}, x_2^{(i)}, \dots, x_k^{(i)})_{i=1}^N$ be given. The mean values of the coordinates define *the centroid*:

$$\bar{x}_j = \frac{1}{N}\sum_{i=1}^N x_j^{(i)}, \quad j = 1, \dots, k.$$

The variances are

$$\sigma_{x_j}^2 = \frac{1}{N}\sum_{i=1}^N (x_j^{(i)} - \bar{x}_j)^2, \quad j = 1, \dots, k.$$

Due to the generality full rank assumption, all $\sigma_{x_j}^2$, for $j = 1, \dots, k$ are non-zero. Then, *the correlations* are

$$r_{jl} = \frac{p_{jl}}{\sigma_{x_j}\sigma_{x_l}}, \quad j, l = 1, \dots, k, l \neq j,$$

where

$$p_{jl} = \frac{1}{N}\sum_{i=1}^N (x_j^{(i)} - \bar{x}_j)(x_l^{(i)} - \bar{x}_l), \quad j, l = 1, \dots, k, l \neq j,$$

are *the covariances*. *The covariance matrix* K is a $(k \times k)$ matrix with the diagonal elements

$$K_{jj} = \sigma_{x_j}^2, \quad j = 1, \dots, k,$$

and the off-diagonal elements

$$K_{jl} = p_{jl}, \quad j, l = 1, \dots, k, l \neq j.$$

The covariance matrix is always symmetric positive semidefinite. Here, due to the generality full rank assumption, K is a positive-definite matrix. In particular, it has the inverse K^{-1} and all its eigenvalues are positive. Pearson defined *the ellipsoid of residuals* by the equation

$$\sum_{j,l=1}^{k} K_{jl} x_j x_l = const.$$

Denote the eigenvalues of K as $\mu_1 \geq \cdots \geq \mu_k > 0$.

THEOREM 2.1. *[Pearson, [**34**]] The minimal mean square distance from a hyperplane to the given set of N points is equal to the minimal eigenvalue of the covariance matrix K. The best-fitting hyperplane contains the centroid and it is orthogonal to the corresponding eigenvector of K. Thus, it is the principal coordinate hyperplane of the ellipsoid of residuals which is normal to the major axis.*

Then Pearson studied the lines which best fit to the given set of points and proved

THEOREM 2.2. *[Pearson, [**34**]] The line which fits best the given system of N points contains the centroid and coincides with the minor axis of the ellipsoid of residuals.*

Pearson integrated the visualization of the linear regression with the orthogonal regression in the planar case in [**34**] in Fig. 7. The ellipse in Fig. 7 is dual to the ellipse of residuals and it coincides with the object studied by Galton.

The main results of [**16**] were to generalize the above classical results of Pearson in the following directions.

The first result: *For a given system of N points in $\mathbb{R}^k$, for any $k \geq 2$, under the full rank assumption, we consider all hyperplanes which equally fit the given system of points. In other words, for any fixed value s not less than the smallest eigenvalue μ_k of the covariance matrix K, we consider all hyperplanes for which the mean sum of square distances to the given set of points is equal to s. Starting from the ellipsoid of residuals, we are going to effectively construct a pencil of confocal quadrics with the following property: For each $s \geq \mu_k$ there exists a quadric from the confocal pencil which is the envelope of all the hyperplanes which s-fit the given system of points.*

We stress that the ellipsoid of residuals does not belong to the confocal family of quadrics. The construction of this confocal pencil of quadrics is fully effective, though quite involved. The obtained pencil of confocal quadrics have the same center as the ellipsoid of residuals and moreover, the same principal axes.

EXAMPLE 2.1. Let us recall that μ_k denotes the smallest eigenvalue of the covariance matrix K. In the case $s = \mu_k$ there is only one hyperplane which s fits the given set of N points. This is the best-fitting hyperplane described in Theorem 2.1. The envelope of this single hyperplane is this hyperplane itself. This hyperplane is going to be a degenerate quadric from our confocal pencil of quadrics.

The second result: *For a given system of N points in $\mathbb{R}^k$, for any $k \geq 2$, under the full rank assumption, find the best fitting hyperplane under the condition that it contains a selected point in $\mathbb{R}^k$. We also provide an answer to the questions of the best fitting line and more general the best fitting affine subspace of dimension ℓ, $1 \leq \ell \leq k-1$ under the condition that they contain a given point.*

A careful look at the Galton's figure (see Fig. 5) discloses an intriguing geometric fact that the line of linear regression of y on x intersects the ellipse at the points of vertical tangency, while the line of linear regression of x on y intersects the ellipse at the points of horizontal tangency. Further analysis of this phenomenon leads us to our third result.

The third result is to formulate linear regression in $\mathbb{R}^k$ in a coordinate free, i.e. in an invariant form. We answered the following question: *for a given direction and for a given system of N points, under the generality full rank assumption, what is the best fitting hyperplane in the given direction, under the condition that it contains a selected point in $\mathbb{R}^k$.*

Apparently, the second and the third result are obtained using the same confocal pencil of quadrics constructed in relation with the first result.

3. Conclusion

For a given system of points in $\mathbb{R}^k$, under the full rank assumption, we constructed in [**16**] an explicit pencil of confocal quadrics with the following properties:

(i) All the hyperplanes for which the hyperplanar moments of inertia for the given system of points are equal, are tangent to one of the quadrics from the pencil of quadrics. As an application, we developed regularization procedures for the orthogonal least square methods, analogues of lasso and ridge methods from linear regression. Another motivation for this study was the gradient descent methods in machine learning. An optimization algorithm may not be guaranteed to arrive at the minimum in a reasonable amount of time. As pointed out in e.g. [**28**] it often reaches some quite low value of the cost function equal to some value s_0 quickly enough to be useful. Here we deal with the hyperplanar moment as the cost function, in application to the orthogonal least square. From [**16**] we know that the hyperplanes which generate the hyperplanar moment equal to s_0 are all tangent to the given quadric from the confocal pencil of quadrics, where the pencil parameter is determined through the value s_0.

(ii) For any given point P among all the hyperplanes that contain it, the best fit is the tangent hyperplane to the quadric from the confocal pencil corresponding to the maximal Jacobi elliptic coordinate of the point P. The worst fit among the hyperplanes containing P is the tangent hyperplane to the ellipsoid from the confocal pencil that contains P. We also determined the best and worst fit among ℓ-dimensional planes containing P, for any $\ell : 1 \leq \ell \leq k-1$.

Both results (i) and (ii) are generalizations of the famous theorem of Pearson on orthogonal regression [**34**], or in other words on the orthogonal least square method (see e.g. [**9**]). For the original Pearson's statement, see Theorem 2.1 above. It is well known that the Pearson result also initiated the Principal Component Analysis (PCA), see e.g. [**3**]. Similarly, our results have a natural interpretation in terms of PCA. The confocal pencil of quadrics provides a universal tool to solve the Restricted Principal Component Analysis restricted at any given point which we formulated and solved in [**16**]. Our generalizations of the Pearson Theorem have natural and important applications in the statistics of the measurement error models, for which the orthogonal regression is known to provide a natural framework, see [**9**], [**24**], [**8**].

Acknowledgements. We are grateful to Maxim Arnold for indicating the reference [**27**] and for interesting discussions. We are also grateful to Pankaj Choudhary, Sam Efromovich, and Frank Konietschke for interesting discussions. This research has been partially supported by Mathematical Institute of the Serbian Academy of Sciences and Arts, the Science Fund of Serbia grant Integrability and Extremal Problems in Mechanics, Geometry and Combinatorics, MEGIC, Grant No. 7744592 and the Ministry for Education, Science, and Technological Development of Serbia and the Simons Foundation grant no. 854861.

References

[1] Adcock, R. J., *Note on the method of least squares.* Analyst 4, 183–184, (1877).

[2] Adcock. R. J., *A problem in least squares.* Analyst 5, 53–54, (1878).

[3] Andreson, T. W., *An Introduction to Multivariate Statistical Analysis* John Wiley and Sons, 1958

[4] Arnold, V. I., *Mathematical methods of classical mechanics*, Graduate Texts in Mathematics, **60**, 2nd edition, Springer-Verlag, New York, 1989, pp. xvi+508

[5] Audin, M., *Courbes algébriques et systèmes intégrables: géodesiques des quadriques*, Exposition. Math. **12**, (1994) pp. 193–226.

[6] Bialy, M., Tabachnikov, S., *Dan Reznik's identities and more*, arXiv:2001.08469, European Journal of Mathematics, , 10.1007/s40879-020-00428-7, (2020).

[7] Carroll, R. J., Ruppert, D. *The use and misuse of orthogonal regression estimation in linear errors-in-variables models.* The American Statistician, 50, 1–6, (1996).

[8] Carroll, R., Ruppert, D., Stefanski, L., Crainiceanu, C., *Measurement Error in Nonlinear Models, A Modern Perspective* Second Edition, Monographs on Statistics and Applied Probability 105, Taylor and Francis (2006).

[9] Casella, G., Berger, R. L., *Statistical Inference*, Second Edition, Duxbury Advanced Series, Duxbury

[10] Cohen, J. E., D'Eustachio, P., and Edelman, G. M., *The specific antigenbinding cell populations of individual fetal mouse spleens: Repertoire composition, size and genetic control.* J. Exp. Med. 146. 394–411, (1977)

[11] Cohen, J. E. and D'Eustachio, P., *An affine linear model for the relation between two sets of frequency counts. Response to a query.* Biometrics 34,514–516, (1978).

[12] Cramér, H., *Mathematical Methods of Statistics,* Princeton: Princeton University Press, (1946).

[13] DeGroot, M., Schervish, M., *Probability and Statistics*, Fourth Edition, Adison-Wesley, 2012.

[14] Dempster, A., *Elements of continuous multivariate analysis*, Addison-Wesley, 1969.

[15] Dragović, V. Gajić, B. (2023) Points with rotational ellipsoids of inertia, envelopes of hyperplanes which equally fit the system of points in R^k, and ellipsoidal billiards, Physica D: Nonlinear Phenomena, Volume 451, 133776

[16] Dragović, V., Gajić, *Orthogonal and Linear Regressions and Pencils of Confocal Quadrics*, preprint, arXiv: 2209.01679, 2022.

[17] Dragović, V., Radnović, M., *Hyperelliptic Jacobians as billiard algebra of pencils of quadrics: Beyond Poncelet porisms*, Advances in Mathematics, Vol. 219, 2008, pp. 1577–1607.

[18] Dragović, V., Radnović, M., *Poncelet Porisms and Beyond: Integrable Billiards, Hyperelliptic Jacobians and Pencils of Quadrics*, Frontiers in Mathematics, Basel: Springer, (2011).

[19] Fedorov, Y. N., Kozlov, V. V., *Various aspects of n-dimensional rigid body dynamics*, Translations of the American Mathematical Society-Series 2, 168, 141–172.

[20] T. B. Fomby, R. C. Hill, S. R. Johnson, *Advanced Econometric Methods*, (1984) Springer Science + Business Media.

[21] Forbes, J. D., *Further experiments and remarks on the measurement of heights by the boiling point of water*, Transactions of the Royal Society of Edinburgh, 21: 135–143, (1857).

[22] Friendly, M., Monette, G., Fox, J,. *Elliptic insights: understanding statistical methods through elliptical geometry.* Statist. Sci. 28 (2013), no. 1, 1–39.

[23] Fuller, W. A., *An affine linear model for the relation between two sets of frequency counts: Response to query.* Biometrics 34, 517–521, (1978).

[24] Fuller, W. A., *Measurement Error Models*, John Willey and Sons, 1987.

[25] Gajić, B., *The rigid body dynamics: classical and algebro-geometric integration*, Zb. Rad.(Beogr.), 16(24), (2013), 5–44.

[26] Galton, F. *Regression towards mediocrity in hereditary stature*, Journal of the Anthropological Institute, 15, (1886) pp. 246–263.

[27] Gardiner, M., *On "Confocal Quadrics of Moments of Inertia" pertaining to all Planes in Space, and Loci and Envelopes of Straight Lines whose "Moments of Inertia" are Constants*, Proceedings of the Royal Society of Victoria, 5, (1893), 200–208.
[28] Goodfellow, I., Bengio, Y., Courville, A., *Deep learning* The MIT Press, 2016.
[29] Hastie, T., Tibshirani, R., Friedman, J, *The elements of statistical learning* Second Edition, Springer
[30] Hilbert, D., Cohn-Vossen, S., *Anschauliche Geometri*, Springer, 1932. English translation: *Geometry and the imagination*. Translated by P. Neményi. Chelsea Publishing Co., New York, N. Y., 1952. ix+357 pp.
[31] Jacobi, C., *Jacobi's Lectures on Dynamics*, Springer, TRIM 51, 2009.
[32] Kozlov, V. V., Treshchev, D. V., *Billiards: A Genetic Introduction to the Dynamics of Systems with Impacts*, Translations of Mathematical Monographs, 89. Providence, RI: Amer. Math. Soc., (1991).
[33] Newton, I., *Philosophia Naturalis Principia Mathematica*, (1st edition 1687, 2nd edition 1713, 3rd edition 1726).
[34] Pearson, K,. *On Lines and Planes of Closest Fit to Systems of Points in Space*, Phil. Mag. 2 (1901), 559–572.
[35] Prasolov, V. V., Tikhomirov, V. M., *Geometry*, Translations of Mathematical Monographs, Vol. 200, AMS, (2001).
[36] Rogers, A. J., *Concentration Ellipsoids, Their Planes of Support, and the Linear Regression Model*, Econometric Reviews, 32:2, 220-243, DOI: 10.1080/07474938.2011.608055, (2013).
[37] Seber, G., Lee, A., *Linear Regression Analysis* Second Edition, John Wiley and Sons, 2003.
[38] Stachel, H., *Recalling Thread Construction of Quadrics*, J. Geom. Graph., 24, No 2, 143–157.
[39] Suslov, G.K., *Fundamentals of Anaytical Mechanics, Vol.1*, Kiev, 1900, [in Russian].
[40] Vinberg, E. B. *A Course in Algebra* Graduate Studies in Mathematics, Vol. 16, AMS.
[41] Weisberg, S., *Applied Linear Regression*, Second Edition, John Willey and Sons, 1985.

Miodrag MATELJEVIĆ[*]

NEKI ASPEKTI TEORIJE POTENCIJALA, VIZUALIZACIJA, VARIJACIONI RAČUN I PRIMENE

Sažetak. Deo teksta je preuzet iz [12]. Glavna tema rada je varijacioni račun i veza sa stacionarnim tačkama funkcionala i parcijalnim jednačinama kao i primena u drugim naukama. Parcijalne jednačine i posebno jednačine matematičke fizike opisuju mnoge pojave iz realnog sveta. Korisno je primetiti da se veći deo teksta može smestiti u zajednički okvir: rešenja eliptičkih jednačina drugog reda u zavisnosti od graničnih uslova, a kao primer navedeno je rešenje Sinaijevog problema.

Rad je podeljen na nekoliko celina:

1. Prvo je dat kratak istorijski pregled ideja i pojmova sa motivacijama iz realnog sveta. Navedene su izjave poznatih ličnosti o matematici i razmatrani pojmovi: granična vrednost, izvod, ekstremne vrednosti, stacionarne tačke, doprinos razvoju funkcionalnog razmišljanja u vezi sa konveksnošću i jednodimenzionalnim kretanjem i inercijski sistemi (jedan zadatak sa PISA testa u vezi sa pojmom konveksnosti), primene metoda vizualizacije na neke matematičke probleme . . . Modifikovan je Arhimedov pristup za izračunavanje površine koju ograničava parabola.

2. U Sekciji 2 dat je dokaz poznate nejednakosti Mihaila Petrovića o konveksnim funkcijama primenom metoda vizualizacije.

3. U trećem delu navedene su neke motivacije za matematičke teorije u realnom svetu, posebno u fizici. Razmatrane su veze između matematičke teorije potencijala i potencijala u elektrostatici: stanje ravnoteže, težnja sistema da zauzme položaj minimalne energije, izoperimetrijske nejednakosti u vezi sa kapacitetom, geometrijske nejednakosti koje uključuju Geringov problem i njihova veza sa logom Internacionalne matematičke unije, itd.

4. U četvrtom delu razmatrani su Dirihleov princip i varijacioni račun u vezi sa minimalnim površima, harmonijska preslikavanja koja su stacionarne tačke funkcionala energije i geometrijske nejednakosti.

Autor je pokušao pristupnu besedu [12](kao i ovaj rad) da prilagodi i čitaocima (slušaocima) kojima matematika nije osnovna delatnost i da pokaže da problemi kojima se bavi imaju motivaciju u realnom svetu. Takođe, treba napomenuti da su ideje uglavnom samo skicirane. [1]

1. Kratak pregled istorijskog razvoja ideja, Arhimedov pristup i pojam granične vrednosti i izvoda

1.1. **O matematici.** Možemo kao Gaus govoriti: ,,*Matematika je kraljica nauka*", i nadahnuto nastaviti ,,*Pored jezika i muzike, matematika je jedna od glavnih oblasti stvaralaštva*". Prava matematika je lepa i istinita. Najznačajniji rezultati u matematici (trijumf ljudskog uma), sadrže ne samo istinu, nego i najveću lepotu.

[*] Miodrag Mateljević, Matematički fakultet, Studentski trg 16, Beograd, Republika Srbija

[1] Izdavač je ograničio veličinu teksta.

Dobitnik Nobelove nagrade E. Vigner napisao je čuveni esej *The unreasonable effectiveness of mathematics in the natural sciences* iz koga navodimo: *Čini se da je za puno ,objašnjenje' uspeha matematike potrebno više od razumevanja jezika, psihologije, strukture mozga i njegovog delovanja, nego što se u ovom trenutku može zamisliti. Još gore, što za razvoj razumevanja možda treba, zaista mora da treba, nova vrsta i tip matematike. Ipak je važno da analiziramo obim i ograničenja matematike. Takođe, razumno je da takva analiza treba da bude neophodan deo obrazovanja i razmatranja studenta matematike. Od kakve je koristi učenik koji ne zna takve stvari?*

Postoje i druga mišljenja o matematici (možda realnija, ali takva razmatranja izlaze iz okvira ovog rada). Na primer, Albert Ajnštajn je rekao: ,,Ukoliko se zakoni matematike odnose na stvarnost, oni nisu sigurni, a ako su sigurni, oni se ne odnose na realnost".

Matematika ima višestruki značaj: važna je u unapređivanju nauke i razumevanju funkcionisanja univerzuma.

Takođe bi trebalo da budemo svesni širokog značaja matematike, i da se napreduje u primeni matematike spektakularnom brzinom. Matematika se odnosi na obrasce i strukture, ona se bavi logičkom analizom, dedukcijom (proces rezonovanja kojim se iz jedne ili više opštih izjava (premisa) dostiže određeni logički zaključak), računanjem u okviru tih obrazaca i struktura. Matematički obrasci se često koriste u veoma različitim oblastima nauke i tehnologije. Matematički obrasci se mogu koristiti da se objasne i kontrolišu prirodna dešavanja i situacije. Matematika ima prodoran uticaj na naš svakodnevni život i doprinosi bogatstvu pojedinca.

1.2. **Kratko o odnosu matematike i stvarnosti i specifičnosti matematičara.** Postoji veliki broj rasprava o odnosu matematike i stvarnosti. Kratko spomenimo neka prirodna pitanja. Da li postoji pravougli trougao u realnom svetu? Ako postoji pravougli trougao u realnom svetu, da li važi verzija Pitagorine teoreme za hiperboličku ili euklidsku geometriju? Šta je prava linija u realnom svetu? Neki fizičari smatraju da je prava linija putanja svetlosti i da svetlosni zraci skreću kada prolaze pored Zemlje. Ovakva pitanja ne izučava matematika. U matematici se definišu Rimanove mnogostrukosti i geodezijske linije (ali takvi objekti ne postoje u realnom svetu). Matematika se primenjuje na modele i aproksimacije.

Sve nauke zajedno doprinose napretku znanja o realnom svetu. Smatramo da ljudski um nikada neće razumeti realan svet potpuno.

Naše je gledište da pored drugih metoda i metod vizualizacije se može koristiti u procesu učenja i razmatranja novih problema. U celom tekstu elementi metoda vizualizacije su prisutni, a posebno u Sekciji 1 i 2 skiciramo neke probleme u kojima primenjujemo ovaj metod. Vizualizacija je sposobnost, proces i proizvod stvaranja, interpretacije, upotrebe i razmišljanja o slikama i dijagramima, u našim mislima, na papiru ili pomoću tehnoloških alata, sa svrhom prikazivanja i prenošenja informacija; razmišljanje i razvijanje ranije nepoznatih ideja i unapređenje razumevanja [14].

Postoje i posebne priče o specifičnosti matematičara. Ovde će biti navedena jedna u kojoj je autor učestvovao:

Akademik M. Marjanović postavio je na sajt članak u vezi PISA[2] testa. Sa kolegom M. Svetlikom objavio sam nekoliko članaka u vezi sa ovom temom. Jednoj

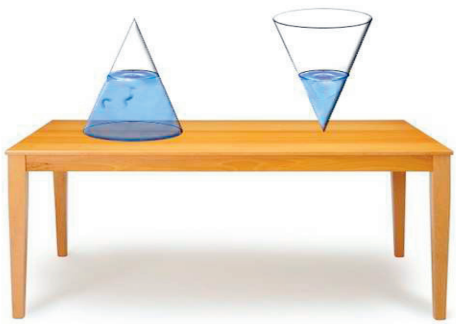
[2]Programme for International Student Assessment .

elitnoj grupi studenata matematike (treća godina studija) postavio sam zadatak sa PISA testa u kome se voda uliva u posudu u obliku kupe. Učesnici testa zaključuju da voda koja se uliva u posudu formira zarubljenu kupu. Setio sam se nekih situacija iz sopstvenog iskustva: zamislite da ste zakačili dno (vrh) posude pomoću kuke za vratilo i ulivate tečnost. Tečnost ne ostaje u posudi zbog gravitacije. Možda matematičari mogu da promene smer dejstva gravitacije. Šta znači posuda u obliku kupe?

Prava kupa nastaje obrtanjem pravouglog trougla oko ose koja sadrži jednu njegovu katetu. Površ kupe bez baze nastaje obrtanjem hipotenuze pravouglog trougla oko ose koja sadrži jednu njegovu katetu. Površ kupe sa bazom nastaje obrtanjem hipotenuze pravouglog trougla i jedne katete oko ose koja sadrži drugu njegovu katetu. Ako su BC hipotenuza, a AB i AC katete pravouglog trougla ABC i ako rotiramo oko AC, tada je C teme, a AB poluprečnik baze. Smatrao sam da je posuda u obliku kupe bez baze postavljena tako da je teme C na podu i da kupasta površ nema bazu, tj. ima otvor kroz koji se uliva tečnost. Posle odgovora studenata počeo sam da razmišljam. Možda postoje i druge mogućnosti.

Npr. probušite vrh (teme) posude u obliku kupe sa bazom, postavite bazu na pod i ulivajte tečnost. Sada voda koja se uliva u posudu formira zarubljenu kupu (Slika 1). Čudni su putevi matematike (možda postoje neke nadprirodne sile, vrhovne aksiome).

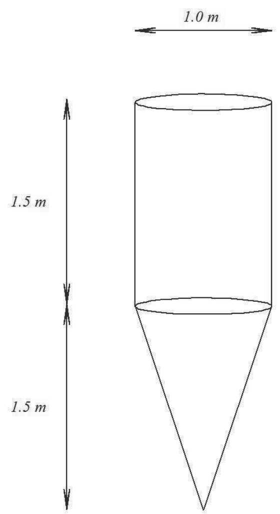
Autor je sa M. Svetlikom objavio nekoliko radova u vezi sa PISA testovima i metodikom nastave matematike.



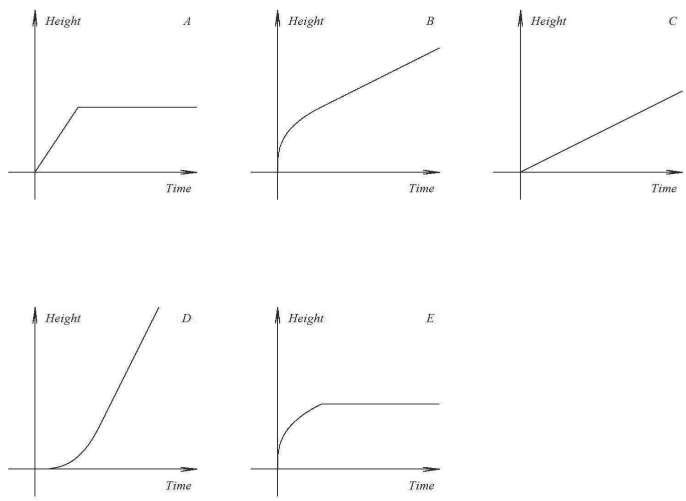
Slika 1. Ulivanje tečnosti u posudu oblika kupe

1.3. **Doprinos razvoju funkcionalnog razmišljanja u vezi sa konveksnošću i jednodimenzionalnim kretanjem i inerecijalni sistemi.** U ovoj podsekciji prikazujemo sadržaj slajda M. Mateljević and M. Svetlik [17].

Primer 1. *Rezervoar za vodu ima oblik i dimenzije kao na slici. U početku rezervoar je prazan. Zatim je napunjen vodom, brzinom punjenja od 1 litra u sekundi.*



Koji od sledećih grafika prikazuje kako se visina nivoa vode menja tokom vremena?



1.4. **Direktan problem.** Pretpostavimo da se tečnost (voda) uliva konstantnom brzinom u rezervoar oblika rotacione površi. Takođe, pretpostavimo da taj proces počinje u trenutku $t = 0$ i završava se u trenutku $t = T$, kada je rezervoar u potpunosti napunjen.

Razmotrimo zavisnost visine nivoa vode u rezervoaru od vremena. Ako sa $h(t)$ obeležimo visinu nivoa vode u trenutku t, funkciju $h : [0, T] \to \mathbf{R}$ nazivamo funkcija visine (*height filling function*).

Neka je $H > 0$ i neka je data funkcija $r : [0, H] \to \mathbf{R}$, sa sledećim osobinama:

(r1) r je neprekidna na $[0, H]$ i

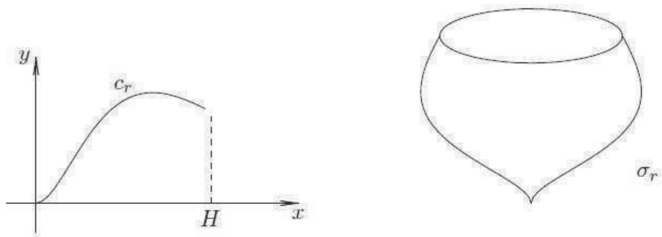
(r2) $r(x) > 0$ za $x \neq 0$.

Rotacijom krive

$$c_r = \{(0, y) : y \in [0, r(0)]\} \cup \{(x, r(x)) : x \in [0, H]\}$$

oko x-ose dobijamo površ σ_r.

Postavimo rezervoar tako da njegova osa rotacije bude u vertikalnom položaju i kroz gornji deo sipajmo tečnost konstantnom brzinom v_0 (u jednakim intervalima vremena uliva se jednaka zapremina tečnosti).



Slika 2. Generatrisa rezervoara i rezervoar

Pokazuje se da funkcija h ima sledeće osobine:

(h1) h je strogo rastuća i neprekidna na $[0, T]$;

(h2) $h(0) = 0$;

(h3a) h je neprekidno diferencijabilna na $[0, T]$ i za svako $t \in [0, T]$ važi $h'(t) > 0$ (ako je $r(0) \neq 0$)

ili

(h3b) h je neprekidno diferencijabilna na $(0, T]$, $\lim\limits_{t \to 0+} h'(t) = +\infty$ i za svako $t \in (0, T]$ važi $h'(t) > 0$ (ako je $r(0) = 0$).

Dalje važi

$$h'(t) = \frac{v_0}{\pi r^2(h(t))}$$

i

$$h''(t) = -\frac{2v_0^2}{\pi^2} \frac{r'(h(t))}{r^5(h(t))}.$$

Korišćenjem poslednje formule, koja važi ako je r diferencijabilna, dobijamo:
Ako je r rastuća, onda je $r' \geq 0$ i $h'' \leq 0$, odnosno h je konkavna na $(0,T)$.
Ako je r opadajuća, onda je $r' \leq 0$ i $h'' \geq 0$, odnosno h je konveksna na $(0,T)$.

1.5. **Obratan problem.** Razmotrimo da li važi obratno tj. ako je data funkcija h sa osobinama (h1),(h2) i (h3a) ili (h3b), da li postoji funkcija r sa osobinama (r1) i (r2) takva da je funckija h funkcija visine koja odgovara rezervoaru σ_r u koji se uliva tečnost brzinom v_0?

Odgovor:

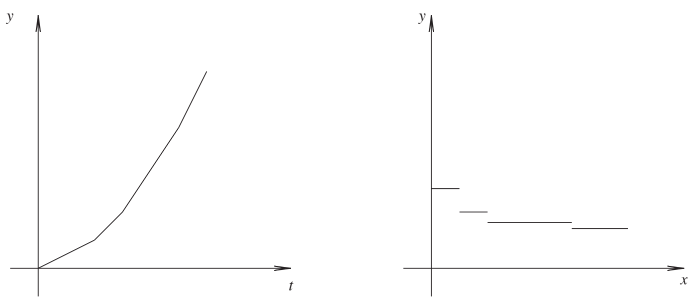
$$r(x) = \sqrt{\frac{v_0}{\pi h'(h^{-1}(x))}}, \quad x \in [0, h(T)].$$

1.6. **Uopštenje obratnog problema.** Neka je $h : [0,T] \to \mathbf{R}$ funkcija sa sledećim osobinama:

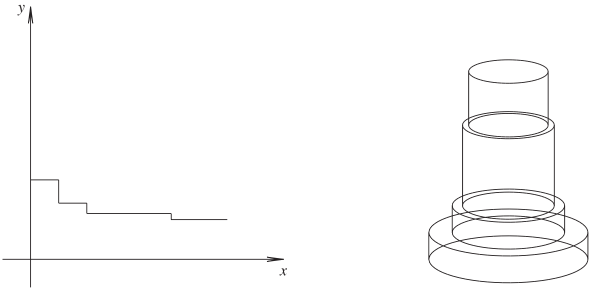
(ch1) h je neprekidna, konveksna i strogo rastuća na $[0,T]$;
(ch2) $h(0) = 0$;
(ch3) $h'_+(0) > 0$ i $h'_-(T) < +\infty$.

Da li za ovu funkciju postoji rezervoar takav da funckija h bude funkcija visine za taj rezervoar ako se u njega uliva tečnost konstantnom brzinom?

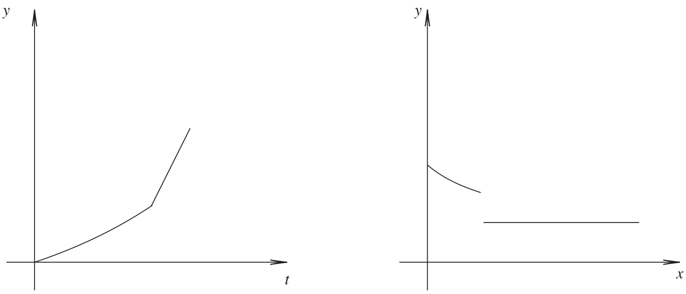
Odgovor: Da, postoji ako proširimo klasu rezervoara.



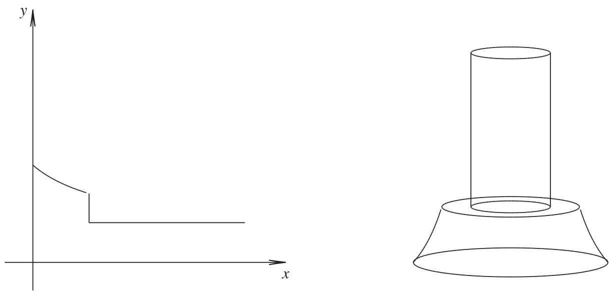
Slika 3. Funkcija visine i generatrisa rezervoara



Slika 4. Generatrisa rezervoara i rezervoar



Slika 5. Funkcija visine i generatrisa rezervoara



Slika 6. Generatrisa rezervoara i rezervoar

Primer 2. *Neka je*

$$h(t) = t^2,$$

tada je odgovarajuće

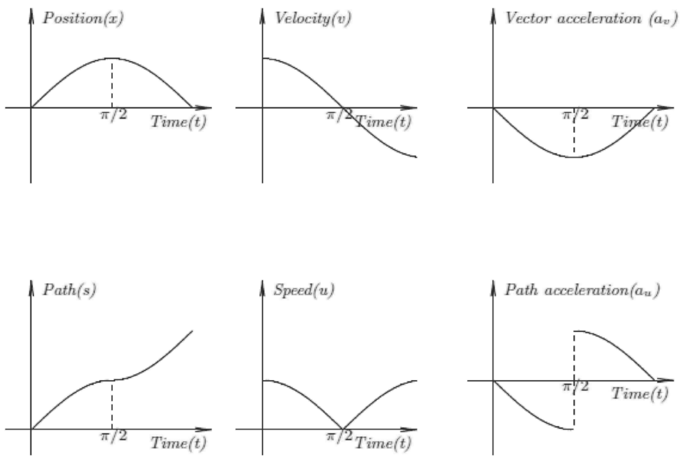
$$r(x) = \sqrt{\frac{v_0}{2\pi}}\frac{1}{\sqrt[4]{x}}.$$

Primetimo da funkcija h nezadovoljava niti uslov (h3a), niti uslov (h3b), niti uslov (ch3). Sledstveno, odgovarajući rezervoar bi imao dno beskonačne površine, ali konačnu zapreminu.

1.7. **Jednodimenzionalno kretanje.** Razmotrimo kretanje materijalne tačke duž prave linije (jednodimenzionalno kretanje).

Postoje dve mogućnosti: materijalna tačka se sve vreme kreće u jednom smeru ili menja smer kretanja.

Funkcije kojima opisujemo kretanje:
$x(t)$ pozicija u trenutku t (*position function*);
$s(t)$ pređeni put do trenutka t (*path function*);
$v(t) = x'(t)$ brzina u trenutku t (*velocity function*);
$u(t) = s'(t)$ apsolutna vrednost brzine u trenutku t (*speed function*);
$a_v(t) = v'(t)$ ubrzanje u trenutku t (*vector acceleration function*);
$a_u(t) = u'(t)$ brzina promene apsolutne vrednosti brzine u trenutku t (*path acceleration function*).



Slika 7. Funkcije kojima je opisano kretanje

Porast visine nivoa tečnosti u rezervoaru možemo videti kao jednodimenziono kretanje tečnosti duž ose rotacije rezervoara.
Pri tome važi:

$$x(t) = s(t) = h(t),$$

$$v(t) = u(t) = h'(t)$$

i

$$a_v(t) = a_u(t) = h''(t).$$

1.8. **Primeri, elementi vizuelnog pristupa.**

Primer 3 (Kretanje kuglica)**.** *Na kompjuterskom igralištu u obliku pravougaonika $ABCD$ u datom momentu $t = t_0$ počne da se kreće neki broj malih figurica (igrača) istom brzinom* 1 cm/s *paralelno sa stranicama $AB = CD = 10\,\mathrm{cm}$ u oba smera. Kada neka figurica stigne do stranica BC ili AD uklanja se iz igre. Ako se figurice sudare, onda samo promene smer kretanja.*

Da li će posle 10 s *igralište biti prazno?*

Numerišimo figurice u datom momentu $t = t_0$ i posmatrajmo njihovo kretanje na dva identična pravougaonika R i R_1.

Na pravougaoniku R simulirajmo kretanje da izbegnemo sudare, a na pravougaoniku R_1 simulirajmo kretanje tako da ako se figurice sudare razmenimo njihove brojeve. Ovaj proces vizualizacije daje jednostavno rešenje.

U svakom momentu isti brojevi zauzimaju iste pozicije na pravougaonicima; otuda igralište će biti prazno posle 10 s.

U toku mog izlaganja na naučnom simpozijumu (1–4. novembar 2022), M. Marković je postavio pitanje da li se zadatak može rešiti metodom matematičke indukcije. Interesantno je da ovaj metod daje jednostavno rešenje.

Primer 4. *Na beskonačnom putu nalaze se učenik A i B, koji ne vide jedan drugog.*

Učenik A napravi dva koraka dok učenik B napravi jedan korak. Da li učenik A može da stigne učenika B?

Up: Neka se u početnom momentu A nalazi u kordinatnom početku O i neka je rastojanje izmedju A i B jednako d koraka; neka se kreće oscilatorno i broji ukupan broj koraka. Kada se nađe u koordinatnom početku ako je do tada prešao $2n$ koraka neka predje još $4n$ koraka u istom pravcu i onda promeni smer kretanja. Za to vreme B prelazi $(2n + 4n)/2 = 3n$ koraka; za dovoljno veliko n, $4n > d + 3n$.

Neka je prešao $2n_k$ od početka do momenta kada se k put ponovo vrati u kordinatni početak; za dovoljno veliko k, $4n_k > d + 3n_k$.

Primer 5. *Na beskonačnoj kvadratnoj mreži nalaze se pauk i muva. Kreću se naizmenično tako što pauk napravi dva koraka dok muva napravi jedan korak. Korak je prelazak sa jednog čvora mreže na susedni čvor. Pauk može da vidi horizontalno i vertikalno.*

Da li pauk može da stigne muvu?

Uputstvo. Neka se u početnom momentu pauk P nalazi u koordinatnom početku O i muva m ima koordinate (m_1, m_2); zamislimo da je m_1 projekciju muve na x-osi i razmotrimo pomoćni problem: Da li pauk može da stigne projekciju muve na x-osi? Napomenimo da pauk ne vidi projekciju muve na x-osi, sem ako se muva nalazi na x-osi.

1) *Na osnovu Primera 4, posle konačno koraka P stiže projekciju muve (neka je rastojanje između njih u tom momentu d_1) i onda nastavlja da se kreće vertikalno duž pravca L_1 u smeru prema muvi.*
2) *Posle konačno koraka P stiže projekciju muve na pravavc L_1 (neka rastojanje izmedju njih u tom momentu d_2).*

Na osnovu Primera 4, posle konačno koraka P stiže projekciju muve (neka je rastojanje između njih u tom momentu d_1) i onda nastavlja da se kreće vertikalno duž pravca L_1 u smeru prema muvi. Onda nastavlja da se kreće horizontalno duž pravca L_2 prema muvi; proveriti $d_1 > d_2$, itd.

Primer 6 (Površina kupe i torusa). *Površina kupe je $P = \pi rs$, gde je s dužina izvodnice, a r poluprečnik.*

Zašto se kupa može razrezati a zatim ,razviti' u ravni u kružni isečak dužine luka $l = 2\pi r$ poluprečnika $R = s$? Ovaj proces vizualizacije treba objasniti.

U analizi 2 definišemo površinu površi npr. u 3-prostoru.

Neka je površ S zadata jednačinom $z = f(x, y)$, $(x, y) \in D$, gde je D dopustiva oblast i $f \in C^1(\overline{D})$.

Tada je

$$P = \int_D \sqrt{1 + p^2 + q^2} dxdy,$$

gde je $p = \partial_1 f$ i $q = \partial_2 f$.

Npr. rotacijom duži $z = kx$, $0 \leq x \leq r$, oko z-ose dobija se kupa $z = k\sqrt{x^2 + y^2}$, $(x, y) \in B_r$, gde je B_r krug.

Ako je h visina kupe, a s dužina izvodnice, dobija se
$h = kr$, $s^2 = r^2 + h^2$, $\sqrt{1 + k^2} = s/r$, i
$\sqrt{1 + p^2 + q^2} = \sqrt{1 + k^2} = s/r$. Otuda

$$P = \int_{B_r} \sqrt{1 + p^2 + q^2} dxdy = \pi\sqrt{1 + k^2} r^2 = \pi rs\,.$$

Neka je kupa K zadata sa $z = k\sqrt{x^2 + y^2}$ i neka je kružnica C_r presek kupe i ravni $z = r$. Da li je C_r geodezijska linija na K?

Površina zarubljene kupe je $P = \pi(r + R)s$.

Neka kupa K_1 sa temenom A ima izvodnicu dužine s_1, a R poluprečnik i neka kupa K_2 sa temenom A, koja je deo kupe K_1, ima izvodnicu dužine s_2, a r poluprečnik i neka je
$P_1 = \pi R s_1$, $P_2 = \pi r s_2$, $s_1 = s + s_2$, $s_1/s_2 = R/r; = k$;
otuda $s = (k - 1)s_2$, $Rs_1 = k^2 r s_2$ i
$P = P_1 - P_2 = \pi(k^2 - 1) r s_2 = \pi(k + 1) rs = \pi(r + R)s$.

Torus je obrtna površ koja se dobija rotacijom kružnice u trodimenzionom prostoru oko ose komplanarne sa tom kružnicom i nema zajedničkih tačaka sa njom.

Na primer, u $xz-$ravni data je kružnica K_1: $(x - R)^2 + z^2 = r^2$. Rotacijom kružnice oko z-ose dobijamo torus. Torus lako definišemo u parametarskom obliku. Neka je M tačka na torusu i v ugao između vektora RM i $x-$ose. Ako je v fiksirano tačka M opisuje kružnicu K_v poluprečnika $R' = R + r\cos v$. Projekciju K_v^z kružnice K_v na $xy-$ravan opisujemo na uobičajen način pomoću parametra u: $x = R'\cos u$ i $y = R'\sin u$. Otuda $x(u, v) = (R + r\cos v)\cos u$, $y(u, v) = (R + r\cos v)\sin u$, $z(u, v) = r\sin v$ gde su u, v parametri (u intervalu $[0, 2\pi)$), R rastojanje od središta cevi torusa do središta torusa r poluprečnik cevi torusa.

Površina torusa jednaka je

$$P = 4\pi^2 Rr = (2\pi r)\,(2\pi R)$$

Zapremina torusa jednaka je

$$V = 2\pi^2 Rr^2 = \left(\pi r^2\right)(2\pi R)\,.$$

Neka je K_2 kružnica koja se dobija rotacijom tačke $(r, 0, r)$ oko oko z-ose. K_2 se opisuje jednačinom $x^2 + y^2 = R^2, z = r$. Formula za površinu torusa navodi na pomisao da se torus može razrezati (npr. duž kružnica K_1 i K_2) a zatim ,,razviti" *u ravni u pravougaonik sa stranicama $2\pi r$ i $2\pi R$. Da li je preslikavanje $f(M) = (Ru, rv)$ izometrija?*

Primer 7. *Neka je kugla sladoleda dijametra $d = 4\,\mathrm{cm}$ i kornet u obliku kupe čija je baza dijametra $d_1 = 4\,\mathrm{cm}$ i visine $H = 10\,\mathrm{cm}$. Kada se kugla istopi koji deo zapremine korneta zauzme?*

Razmotriti slučaj $d > d_1$.

Kako je $V_0 = 4\pi R^3/3$ i $V_1 = \pi R^2/3$, dobija se $V_0/V_1 = 4R/H$. Otuda za $R = 2\,\mathrm{cm}$, nalazimo $V_0/V_1 = 4R/H = 8/10 = 0,8$. Dakle, odgovor je 80%.

Objasniti da ako se kugla stavi prirodno na kornet sladoled ne curi van korneta.

Neka je $H > R > 0$, $0 < h_1 < H$, $0 < r_1 < R$, $M = (-r_1, h_1)$, $N = (r_1, h_1)$, K kružnica sa središtem u $P = (0, H)$ i poluprečnika R, O presek tangenti l_1 i l_2 na K kroz M i N, A presek l_1 sa $x = -R$ i B presek l_2 sa $x = R$. Rotacijom trougla OAB i K oko y-ose dobija se sfera koja tangira kupu.

1.9. **Arhimedov pristup.** U geometriji se definiše površina figura kao što su trougao, pravougaonik i mnogougao. Ako imamo neku figuru koja se može rastaviti na konačan broj trouglova (takve figure nazivaju se elementarne), tada je njena površina jednaka zbiru površina trouglova (aditivnost površine).

Za pozitivnu funkciju $f : [a, b] \to \mathbb{R}^+ = \{x \geq 0\}$ definišemo ordinatni skup nad $[a, b]$ sa $\mathrm{Ord}(f) = \{(x, y) : a < x < b, 0 < y < f(x)\}$.

Prirodno je postaviti sledeće pitanje:

i) Arhimedov problem. Šta je površina P figure ordinatnog skupa funkcije $y = x^2$ nad $[0, 1]$ (površina parabole)?
ii) Ili, uopšte, šta je površina skupa $\mathrm{Ord}(f)$ pozitivne funkcije $f : [a, b] \to \mathbb{R}^+$?

Kako ordinatni skup parabole nije elementaran i ne može se rastaviti na konačan broj trouglova, potrebno je prvo definisati pojam površine za ovakve figure.

Ako smo familijarni sa integralima, možemo dati jednostavan odgovor za Arhimedov problem:

$$P = \int_0^1 x^2 dx = \frac{x^3}{3}\Big|_{x=0}^{1} = \frac{1}{3}.$$

Pokušajmo da objasnimo rhimedovu ideju za izračunavanje površine parabole. Arhimed nije, naravno, znao integrale, ali ideja koju je primenio vodi definiciji integrala. Pokušajmo da rekonstruišemo Arhimedov pristup, koristeći pojam granične vrednosti. U vreme kada je Arhimed radio, nije postojao jasan pojam limesa, ali je razmatranja koja slede obrazložio tada dostupnim metodama.

Podelimo interval $[0, 1]$ tačkama $x_k = \dfrac{k}{n}$, $0 \leq k \leq n$. Neka je P_k pravougaonik $[x_{k-1}, x_k] \times [0, f(x_k)]$ i

$$S_n = \sum_{k=1}^{n} f(x_k)(x_k - x_{k-1}). \tag{1.1}$$

Pokažimo da $S_n \to \dfrac{1}{3}$, kad $n \to \infty$, tj. da je površina figure ordinatnog skupa funkcije $y = x^2$ nad $[0, 1]$ jednaka $\dfrac{1}{3}$.

Iz (1.1) sledi

$$S_n = \sum_{k=1}^{n} \left(\frac{k}{n}\right)^2 \frac{1}{n} = T_n \frac{1}{n^3},$$

gde je $T_n = \sum_{k=1}^{n} k^2$.

Kako je

$$T_n = \frac{n(n+1)(2n+1)}{6},$$

nalazimo $S_n = \frac{a_n b_n}{6}$, gde je $a_n = \left(1 + \frac{1}{n}\right)$ i $b_n = \left(2 + \frac{1}{n}\right)$.

Kako $\frac{1}{n} \to 0$, na osnovu pravila za limes zbira i proizvoda, zaključujemo $a_n \to 1$, $b_n \to 2$ i $a_n b_n \to 2$. Dakle, S_n teži ka $\frac{1}{3}$, kada n teži ∞.

Po analogiji možemo pokušati da izračunamo površinu kruga i zapreminu lopte.

1.9.1. *Zapremina piramide.* Zapremina piramide P čija je baza B i visina h jednaka je $V = \frac{1}{3}h|B|$.

Neka je baza B piramide P u ravni $z = h$, $h > 0$, a vrh u koordinatnom početku i neka je $B(t)$ presek piramide sa ravni $z = t$. Tada je $|B(t)| = \frac{t^2}{h^2}|B|$. Koristeći integrale, nalazimo $V = \int_0^h |B(t)|dt = h^{-2} \int_0^h t^2 dt$. Otuda je $V = \frac{1}{3}h|B|$.

Pokušajmo da sledimo Arhimeda:

Podelimo interval $[0, h]$ tačkama $z_k = h\frac{k}{n}$, $0 \leq k \leq n$. Neka je $B_k = B(z_k)$, D_k projekcija B_k na $xy-$ravan, C_k prizma $D_k \times [z_{k-1}, z_k]$ i

$$V_n = \sum_{k=1}^{n} |C_k| = \sum_{k=1}^{n} |D_k|(z_k - z_{k-1}).$$

Kako je $z_k - z_{k-1} = \frac{h}{n}$ i

$$|D_k| = \frac{z_k^2}{h^2}|B| = \left(\frac{k}{n}\right)^2 |B|,$$

nalazimo

$$V_n = \sum_{k=1}^{n} |C_k| = \sum_{k=1}^{n} \left(\frac{k}{n}\right)^2 |B|\frac{h}{n} = h|B|T_n \frac{1}{n^3},$$

gde je $T_n = \sum_{k=1}^{n} k^2$. Otuda $V_n \to V_0$, gde je $V_0 = \frac{1}{3}h|B|$ zapremina piramide.

Teorija integrala, preko tablice, daje nam gotove rezultate tako da ne treba da računamo komplikovane limese nizova. Drugo je pitanje da li i koliko razumemo teoriju.

Dalji razvoj Arhimedove ideje omogućava da izračunamo površine kupe, piramide i torusa. Takođe, mogu se razmatrati pojam integrala i primitivne funkcije kao i nejednakosti vezane za obim i površinu figura (izoperimetrijski problem).

1.10. **Izvod funkcije.** Izvod funkcije jedan je od osnovnih pojmova infinitezimalnog računa i zasniva se na pojmu granične vrednosti funkcije u tački. Nastao je u 17. veku u vezi sa problemima u matematici i fizici (na primer, da se definišu tangenta, trenutna brzina, itd.), a ima značajne primene u matematičkim naukama, fizici, ekonomiji, itd.

Rasprava između Njutna i Lajbnica oko toga kome treba pripisati otkriće izvoda počinje 1699. godine, a kulminaciju dostiže 1711. godine. Njutnov argument je da je na pojmu izvoda počeo da radi još 1666. godine, međutim, svoja istraživanja nije objavio sve do 1704. godine. Kasnije postaje jasno da su Njutn i Lajbnic došli do istog rezultata, koristeći različitu notaciju i ispitujući različite probleme. Tokom godina su učenici sa obe strane iznosili nove argumente, ali rasprava o prioritetu nikada nije razrešena, pa se obojici naučnika pripisuje zasluga za konačno formiranje i povezivanje pojmova integrala i diferenciranja.

Potreba za izvodom funkcije javila se prilikom pokušaja da se pronađe univerzalni način za određivanje tangente krive u geometriji i brzine kretanja u mehanici.

Vrednost količnika promene pređenog puta i promene vremena je srednja brzina. Trenutna brzina je granična vrednost količnika promene puta i promene vremena, kada promena vremena teži nuli. Ako posmatramo kretanje na pravolinijskom putu i sa x označimo pređeni put za vreme t, u tx-koordinatnom sistemu kretanje možemo predstaviti pomoću funkcije $x = x(t)$ (koristi se i oznaka $x = s(t)$). Brzina u tački t_0 je izvod funkcije $x = x(t)$ u tački t_0 (oznaka $v = x'(t_0)$ u fizici se koristi i $\dot{x}(t_0)$), a $k = x'(t_0)$ je koeficijent pravca tangente.

Ubrzanje je promena brzine po vremenu.

Dakle, ubrzanje pokazuje koliko ,,brzo" se menja brzina posmatranog tela ili čestice. Stoga, ubrzanje a je prvi izvod brzine po vremenu, odnosno drugi izvod položaja po vremenu, tj. važi $a = \dot{v} = \ddot{s}$.

Izvod funkcije $f(x)$ u tački a se definiše kao:

$$f'(a) := f'(x)|_{x=a} = \lim_{\Delta x \to 0} \frac{f(a + \Delta x) - f(a)}{\Delta x},$$

ukoliko limes postoji. Inače, izvod u datoj tački a možemo shvatiti i kao linearni operator, koji aproksimira funkciju u okolini tačke a:

$$f(x) = f(a) + f'(a)(x - a) + o(x - a), \quad x \to a.$$

Ako je $f = f(x_1, x_2, ..., x_n)$ funkcija više promenljivih i $a = (a_1, a_2, ..., a_n)$, funkciji f pridružimo funkciju jedne promenljive $g(t) = f(t, a_2, ..., a_n)$.

Vrednost $g'(a_1)$ nazivamo parcijalni izvod funkcije f po promenljivoj x_1 u a i označavamo sa $D_1 f(a)$. Gradijent funkcije f u tački a označavamo i definišemo sa $\nabla f(a) = (D_1 f(a), ..., D_n f(a))$.

Postupak pronalaženja izvoda funkcije se naziva diferenciranjem. Diferenciranje je proces obratan u odnosu na integraciju.

Ako su vrednosti funkcije realni brojevi, tačke lokalnog minimuma i maksimuma se zajedno zovu lokalni ekstremi.

Tačka x_0 sa osobinom da je $f'(x_0) = 0$ naziva se stacionarna tačka.

1.10.1. *Brzina i ubrzanje.* Za ilustrovanje početnih veza između matematike i fizike korisno je skicirati definicije pojmova – trenutna brzina i trenutno ubrzanje pomoću pojma izvoda funkcije (koji je jedan od osnovnih pojmova infinitezimalnog računa i zasniva se na pojmu granične vrednosti funkcije).

U fizici i tehnici ponekad se navodi ,,jednostavna" (neki kažu ,,fizička") definicija da je vektor veličina koja ima intenzitet (nenegativna brojna vrednost), pravac i smer (koji se u trodimenzionom prostoru mogu pokazati prstom).

Koriste se razni opisi vektora. Na primer, u srednjoškolskim udžbenicima često piše: ,,vektor ima intenzitet, pravac i smer". Na fakultetima, u matematici, koristi se apstraktna definicija: ,,vektorski prostor (VP) je skup čije elemente nazivamo vektorima (vektor je element VP)". U realnom VP su definisane dve operacije: sabiranje vektora i množenje vektora skalarom (realnim brojem), koje imaju osobine koje odgovaraju pojmu vektora u realnom svetu.

U svakodnevnom govoru, reč brzina često označava samo vrednost (intenzitet) brzine (*speed*). Nasuprot tome, u fizici reč brzina označava vektor brzine (*velocity*).

Pretpostavimo da razmatramo pravolinijsko kretanje čestice u jednom smeru.

Vrednost količnika promene pređenog puta i promene vremena je srednja brzina. Trenutna brzina je granična vrednost količnika promene pređenog puta i promene vremena, kada promena vremena teži nuli. Dakle, ako posmatramo jednosmerno kretanje na pravolinijskom putu i sa x označimo pređeni put za vreme t (t označava vreme, računajući od nekog početnog trenutka ili položaja), u tx-koordinatnom sistemu kretanje možemo predstaviti pomoću funkcije $x = x(t)$ (koristi se i oznaka $x = s(t)$).

Kako da definišemo brzinu u određenom momentu t_0? Ako je $t > t_0$ srednja brzina na intervalu $[t_0, t]$ je

$$v_{pr} = v_{pr}(t, t_0) = \frac{x(t) - x(t_0)}{t - t_0}.$$

,,Prava" brzina (trenutna brzina) u t_0 ne zavisi od vremenskog intervala jer se dobija, kako se to popularno kaže, njegovim zamišljenim skraćivanjem na ,,beskonačno mali interval" oko određenog trenutka t_0.

Preciznije, ako postoji, granična vrednost $v_{pr}(t, t_0)$ kada t teži t_0, naziva se brzina i označava se sa $v = v(t_0)$. Dakle, brzina je izvod funkcije $x = x(t)$ u tački t_0 (oznaka $v = x'(t_0)$ u fizici se koristi i $\dot{x}(t_0)$).

Opštije, pretpostavimo da je kretanje materijalne tačke u nekom referentnom sistemu (koje ne mora biti pravolinijsko) zadato sa $\vec{r} = \vec{r}(t)$, gde t označava vreme, računajući od nekog početnog trenutka ili položaja, a $\vec{r}(t)$ vektor položaja. Ako sa s označimo funkciju $s(t)$ koja je pređeni put (dužina pređenog puta) do trenutka t, vrednost brzine je izvod puta po vremenu: $v = s'(t)$.

Ako u trenutku t tačka ima položaj $\vec{r}(t)$, ona će se nakon vremenskog intervala Δt pomeriti u položaj $\vec{r}(t + \Delta t)$. Promena položaja (ili pomak) tačke tokom tog vremenskog intervala je vektor $\Delta\vec{r} = \vec{r}(t + \Delta t) - \vec{r}(t)$, koji se dobija oduzimanjem prvog vektora položaja od drugog. Deljenjem vektora promene položaja $\Delta\vec{r}$ sa odgovarajućim vremenskim intervalom Δt dobije se prosečni vektor brzine u tome intervalu:

$$\vec{v}_{pr} = \frac{\Delta\vec{r}}{\Delta t}.$$

Da bi se dobio trenutni vektor brzine (kraće: brzina), treba koristiti izvod (kao i kod definicije intenziteta brzine): $\vec{v} = \vec{r}\,'(t)$.

Ubrzanje se može definisati slično kao brzina. Vrednost količnika promene brzine i promene vremena je srednje ubrzanje.

Trenutno ubrzanje je granična vrednost količnika promene brzine i promene vremena, kada promena vremena teži nuli. Dakle, ubrzanje pokazuje koliko ,,brzo" se menja brzina posmatranog tela ili čestice. Stoga, ubrzanje a je prvi izvod brzine po vremenu $a = \dot{v}$, odnosno drugi izvod položaja po vremenu $a = \dot{v} = \ddot{s}$.

U fizici je ubrzanje vektorska veličina, izvod brzine po vremenu (koja je isto vektor). U nekim situacijama, međutim, koristi se reč ubrzanje da označi samo intenzitet vektora ubrzanja, ako je iz konteksta jasno o čemu se radi, npr. kod pravolinijskog kretanja.

Preciznije, označimo sa $\Delta\vec{v} = \vec{v}(t_2) - \vec{v}(t_1)$ promenu brzine od trenutka t_1 do trenutka t_2, a sa $\Delta t = t_2 - t_1$ vremenski interval (proteklo vreme) između ta dva trenutka. Srednje ubrzanje na vremenskom intervalu $[t_1, t_2]$ je

$$\vec{a}_{pr} = \frac{\Delta\vec{v}}{\Delta t}.$$

Nasuprot tome, ,,pravo" ubrzanje (trenutno ubrzanje) ne zavisi od vremenskog intervala jer se dobija njegovim zamišljenim skraćivanjem na ,,beskonačno mali interval" oko određenog trenutka. Postupak se uopšte (u različitim primenama) naziva graničnim prelazom i precizno definiše pomoću pojma izvoda. Dakle, ubrzanje je izvod brzine po vremenu:

$$\vec{a} = \frac{d\vec{v}}{dt} = \lim_{\Delta t \to 0} \frac{\Delta\vec{v}}{\Delta t}.$$

Simbol $\vec{a}$ označava ubrzanje (a je prvo slovo reči akceleracija koja je latinskog porekla), simbol $\vec{v}$ označava brzinu; i jedna i druga veličina su funkcije vremena t (što se podrazumeva, pa se ne mora eksplicitno navesti). Ubrzanje opisuje kako se brzo i u kom smeru menja brzina u određenom trenutku. Budući da je brzina vektorska veličina koja može menjati i vrednost i smer, ubrzanje istovremeno opisuje i jednu i drugu promenu. No, one se mogu razdvojiti tako da se zasebno razmatraju tangencijalno i normalno ubrzanje.

Analiza kretanja u dinamici često polazi od Drugog Njutnovog zakona, koji (u nerelativističkoj aproksimaciji) glasi ,,suma sila jednaka je proizvodu mase i ubrzanja" (tj. $\sum \vec{F} = m\vec{a}$). Odatle se, iz sila koje deluju na materijalnu česticu, jasno dobija ubrzanje čestice. Potom se iz ubrzanja, brzina čestice računa integracijom:

$$\vec{v} = \vec{v}_0 + \int_{t_0}^{t} \vec{a}\,\mathrm{d}t$$

gde je $\vec{v}_0$ brzina u trenutku t_0 (tzv. početna brzina). Iz brzine se može odrediti jednačina putanje ili pređeni put. Ako se umesto materijalne tačke razmatra telo, navedena razmatranja odnose se na kretanje njegovog centra masa.

2. Vizualizacija i modeliranje matematike

U ovoj sekciji je skiciran dokaz nejednakosti Mihaila Petrovića korišćenjem vizualizacije konveksnih funkcija. Pristup je zasnovan na radu [10] u kome je dat doprinos razvoju funkcionalnog razmišljanja u vezi pojma konveksnosti. Interesantno je da je klica ovog rada jedan zadatak višestrukog izbora iz PISA testa. Uopštavajući ovaj zadatak razmatra se veza između oblika rezervoara za vodu (posude) u koji se voda uliva konstantnom brzinom i grafika funkcije koja pokazuje kako se visina vode u rezervoaru za vodu menja kao funkcija vremena. Ako u momentu t označimo sa $h(t)$ visinu vodenog stuba onda h nazivamo funkcijom visine (ulivanja).

Neka je posuda P_r nastala rotacijom grafika funkcije $x = r(z)$, $0 \leq z \leq H$ oko z-ose i pretpostavimo da je negativna orijentacije z-ose u smeru teže i da je dno posude u obliku kruga poluprečnika $r_0 = r(0)$.

U [10] je dokazano da se konveksna funkcija na $[0, T]$ može interpretirati kao funkcija visine (ulivanja) u posudu koja ima oblik rotacione površi definisane pomoću nerastuće funkcije $r : [0, h] \to \mathbb{R}$.

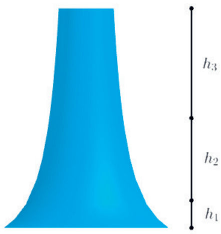
2.0.1. *Petrovićeva nejednakost za konveksne funkcije.*

Teorema 1 (Petrovićeva nejednakost za konveksne funkcije)**.**
Neka je $h : [0, \infty) \to \mathbb{R}$ *konveksna funkcija i neka su* $t_1, t_2, \ldots, t_n$ *pozitivni brojevi. Tada važi sledeća nejednakost*

$$h(t_1) + \ldots + h(t_n) \leq h(t_1 + \ldots + t_n) + (n-1)h(0).$$

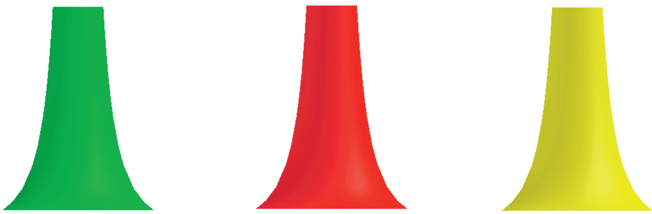
Ponovimo, funkcija h se može interpretirati kao funkcija visine punjenja posude V, koja ima oblik rotacione površi P_r, gde je $r : [0, h] \to \mathbb{R}$ nerastuća funkcija.

Korak 1. Neka je $h_1 = h(t_1), \ldots, h_n = h(t_n)$ i $H = h_1 + \ldots + h_n$. Formirajmo posudu V koja odgovara funkciji h i čija je visina H (Slika 8).

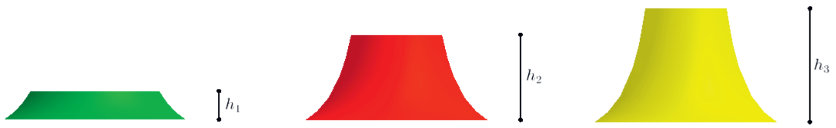


Slika 8. Posuda V koja odgovara funkciji h

Korak 2. Uzmimo n kopija posude V (videti Sliku 9 za $n = 3$) i isecimo ih na visinama $h_1, \ldots, h_n$. Time dobijamo posude $V_1, \ldots, V_n$ (Slika 10).

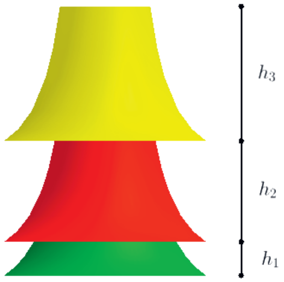


Slika 9. Tri kopije posude V



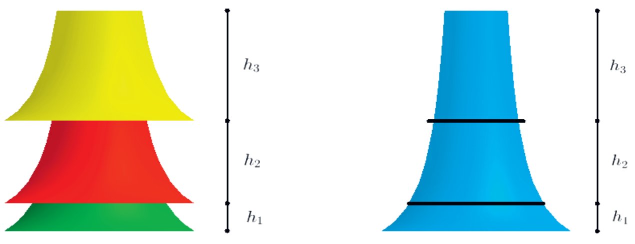
Slika 10. Posude V_1, V_2 i V_3

Korak 3. Isecanjem delova baza ovih posuda, dobijamo novu posudu V' (Slika 11).



Slika 11. Posuda V'

Korak 4. Pretpostavimo da u istom trenutku počnemo da punimo obe posude V i V' (Slika 12). Kako je $r : [0, h] \to \mathbb{R}$ nerastuća funkcija, zapremine odgovarajućih delova posude V manje su od zapremina odgovarajućih delova posude V'. Stoga, posuda V će se prva napuniti i otuda sledi Petrovićeva nejednakost (ostavljamo čitaocu da popuni detalje).

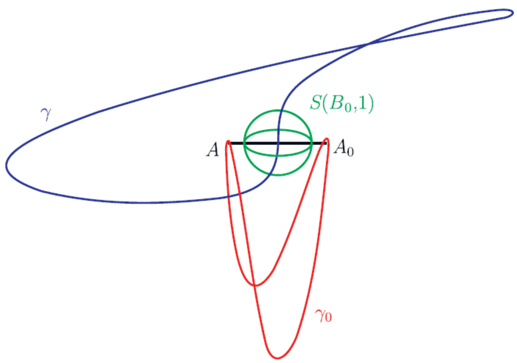


Slika 12. Posude V' i V

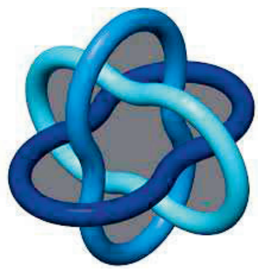
2.0.2. *Geringov problem.* Navedimo samo jedan specijalni izoperimetrijski problem (Geringov problem). Ovaj problem autor je rešio i objavio 1975. godine (Uopštenje je objavio 2015. godine.):

Teorema 2. *Ako su γ i γ_0 ulančane krive u $\mathbb{R}^3$ na rastojanju 1, dokazati da je dužina svake od ovih krivih najmanje 2π.*

Rešenje ovog problema ima veze sa logom IMU [3] (videti slike 13 i 14; takođe, videti i Borominov prsten).



Slika 13. Rešenje Geringovog problema



Slika 14. IMU logo

3. Njutnov zakon i Kulonov zakon

3.1. **Njutnov zakon.** Drugi Njutnov zakon glasi: „Intenzitet sile koja pokreće telo jednak je proizvodu mase tela i ubrzanja koje telo dobija delovanjem te sile”, tj. $F = ma$, gde je F sila koja pokreće telo i daje mu ubrzanje, m masa tog tela i a ubrzanje koje telo dobija delovanjem te sile. Ubrzanje tela srazmerno je spoljašnjoj sili koja deluje na telo i obrnuto srazmerno njegovoj masi. Smer ubrzanja je u smeru ukupne sile koja deluje na telo.

[3] Internacionalna matematička unija.

Njutnov zakon o opštoj univerzalnoj gravitaciji ima sledeću formulaciju: Svako materijalno telo u Vasioni privlači drugo telo silom koja deluje u pravcu prave koja spaja tela (preciznije centre masa), intenziteta srazmernog proizvodu masa tela, a obrnuto srazmernog kvadratu rastojanja. Dakle, intenzitet sile je

$$F_g = G\frac{m_1 m_2}{r^2},$$

gde je G univerzalna gravitaciona konstanta (po Gausu).

U nameri da damo precizniju formulaciju pretpostavimo da su x i y respektivno središta prvog i drugog tela, $\overrightarrow{r} = \overrightarrow{xy}$, $r = |\overrightarrow{r}|$ međusobna udaljenost između tela (m), i $\mathbf{r}_0 = \overrightarrow{r}/|\overrightarrow{r}|$. Označimo sa $\overrightarrow{F}_1$ silu kojom drugo telo dejstvuje na prvo i sa $\overrightarrow{F}_2$ silu kojom prvo telo dejstvuje na drugo . Tada je

$$\overrightarrow{F}_1 = -\overrightarrow{F}_2 = G \cdot \frac{m_1 \cdot m_2}{r^2}\mathbf{r}_0,$$

gde je G univerzalna gravitaciona konstanta koja otprilike iznosi $6,67428 \cdot 10^{-11}\mathrm{Nm}^2\mathrm{kg}^{-2}$, m_1 masa prvog tela (kg), m_2 masa drugog tela (kg).

U literaturi se ponekad sila označava bez strelice i tvrdi da važi $F = F_1 = F_2$, gde je F uzajamna sila privlačenja između dva tela. Preciznije je reći da su intenziteti uzajamnih sila privlačenja između dva tela jednaki.

Tako, Danilo P. Rašković u svom udžbeniku Dinamika, iz 1948. godine, piše da je ovaj Njutnov zakon gravitacije uneo mnogo poleta u naučna dokazivanja istina o kretanju nebeskih tela i citira da se o njegovoj važnosti najlepše izrazio velikan naše nauke Milutin Milanković: *Tako se Njutnov zakon najveličanstveniji što ga je ikad smrtni čovek mogao da dokuči, pokazao kao opšti zakon prirode kome se pokorava cela Vasiona.* I Veljko A. Vujičić u svojoj monografiji Preprincipi mehanike, iz 1998. godine, takođe citira našeg genijalnog naučnika Milutina Milankovića sledećim rečenicama: *Njutnovim zakonom posta odgonetnuta hiljadugodišnja zagonetka planetskog kretanja i nova saznanja sledovaše, sama od sebe iz njega. Sve nejednakosti kretanja planeta i Meseca ispoljiše se kao prirodna posledica toga zakona, kao i jasni izražaj međusobnog privlačnog dejstva tih nebeskih tela. Ne samo da je tim priroda tih nejednakosti postala rastumačena: one su se mogle izračunavati i pratiti i u prošlost i u budućnost. Pokazalo se i za komete, ubrzo iza postavljenja Njutnovog zakona, da on važi za sva nebeska tela bez izuzetka, dakle i izvan Sunčevog sistema. Precesija ravnodnevnica, koju je, kao što smo čuli, prvi konstatovao Hiparhos, našla je Njutnovim zakonom svoje potpuno razjašnjenje, a isto tako kasnije opažena rotacija Zemljine ose. I oblik naše Zemlje, a naročito njena spljoštenost usled rotacije dobi, u svim pojedinostima, mehaničko i geometrijsko obrazloženje. To isto važi i za prastaro pitanje o postanku morske plime, koja se pokazala kao neposredna posledica privlačnog dejstva Sunca i Meseca. Zahvaljujući ovom Njutnovom zakonu iznikla je jedna nova nauka - Nebeska mehanika.*

Postoje razlike u metodama provere neke teorije u matematici i naukama koje se bave realnim pojavama. Sa matematičke tačke gledišta može se postaviti pitanje da li se masa definiše pomoću sile i obrnuto (pomoću Njutnovog zakona) i da li se može postići napredak u vezi ovog pitanja.

Osvrnimo se kratko na razvoj nauke i nove teorije koje su se pojavile posle Njutna.

a) Danas se smatra da Njutnovi zakoni važe samo u klasičnoj mehanici, gde je brzina kretanja tela relativno mala u odnosu na brzine svetlosti, a masa tela znatno veća od mase atomskih delova (elektrona, protona i neutrona). U slučaju izuzetno velikih brzina, uporedivih sa brzinom svetlosti, ili izuzetno malih masa, uporedivih sa masom atoma, pojavljuju se drugi efekti koji se precizno opisuju zakonima kvantne mehanike i relativističke fizike.

b) Teorija relativnosti koju je razvio Albert Ajnštajn (1905. godine) ne prihvata pojam apsolutnog vremena, koji je na prvi pogled potpuno logičan. Da bi se uporedilo vreme između dva događaja, koji mere dva posmatrača u različitim referentnim sistemima, potrebno je koristiti neki signal. Jedini mogući način u realnom svetu je upotreba svetlosnog signala. No kako je brzina svetlosti konstantna i nezavisna od referentnog sistema, Ajnštajn je pokazao da vreme mora zavisiti od referentnog sistema. Vreme je, a prema tome i pojam istovremenosti dva događaja, relativno. Tretirajući vreme kao promenljivu ekvivalentnu prostornim promenljivim, Ajnštajn je, sledeći put na koji je ukazao Herman Minkovski, izgradio pojam četvorodimenzionog prostora – prostorno vremenskog kontinuuma. Geometriju takvog prostora određuje materija, a gravitacija je samo posledica geometrije fizičkog prostora. Drugim rečima, gravitacija je samo posledica činjenice da kontinuum prostor-vreme nije ravan, nego zakrivljen. Čestica materije ubačena u svemir ne bi se kretala po pravoj liniji, kako sledi iz Njutnove jednačine, nego po tzv. geodezijskoj liniji, koja fizički predstavlja (vremenski) najkraći put između bilo koje dve tačke u svemiru. Svetlosni zraci takođe se ne šire pravolinijski, nego se i oni savijaju u gravitacionom polju (gravitacijske leće).

v) U savremenoj fizici kvantna mehanika se primenjuje na mikronivou. Teorija relativiteta i kvantna mehanika su dve fundamentalno različite teorije sa potpuno različitim formulacijama i prema sadašnjem razumevanju nekompatibilne teorije realnosti. Za razliku od teorije relativiteta, stvarnost na koju se odnosi kvantna mehanika sačinjena je od kvantnih skokova. Ishodi događaja nisu određeni i opisuju se pomoću verovatnoće. Kretanje subatomskih čestica nije neprekidno.

g) Teorija struna je prvi pokušaj da se ujedine ove dve teorije. Pogodno je u fizici koristiti modele u kojima se najsitnije, elementarne čestice razmatraju kao bezdimenzione tačke u prostoru i predstavljaju pomoću koordinata. Obično zamišljamo, npr. elektron kao tačku bez unutrašnje strukture. Osnovna ideja teorije struna je: sve te različite ,,osnovne" čestice u standardnom modelu su samo različite manifestacije jednog osnovnog objekta strune. Supersitne strune su dužine od oko 10^{-35} metara (npr. proton je 100 milijardi puta veći od tih struna). Po teoriji struna, koja je nastala šezdesetih godina, najsitnije, elementarne čestice (elektroni, kvarkovi, hadroni, bozoni, fermioni...) imaju prirodu jednodimenzionih struna. Na primer, prema teoriji struna elektron je sićušna petlja strune koja može da osciluje na različite načine i da se ponaša kao foton ili kvark ili graviton. Univerzum je sačinjen od elemenata koji vibriraju, čije ponašanje je slično ponašanju zategute žice u prostoru. Oscilovanje elementarnim česticama daje naelektrisanje, masu i spin.

Mnogi teorijski fizičari veruju da je teorija struna korak napred u shvatanju fundamentalnog opisa prirode, mada je eksperimentalno izuzetno teško proveravati predviđanja ove teorije. Istraživači u oblasti teorijske fizike proučavaju matematiku na kojoj se bazira teorija struna i kvantna teorija polja. To uključuje primenu najnovijih matematičkih disciplina, kao što je nekomutativna geometrija, ali takođe uključuje razvoj i proučavanje nove matematike koja je potrebna za dalji razvoj teorije.

d) O suštini Faradejevih ideja Maksvel piše: ,,Faradej je svojim misaonim okom video linije sile kako prožimaju ceo prostor tame gde su matematičari videli centre sila koji deluju iz daleka; Faradej je video medijum tamo gde oni nisu videli ništa sem odstojanja; Faradej je tražio sedište fenomena u realnim delovanjima koja se prostiru kroz medijum, dok su se oni zadovoljavali da ga nađu u dejstvu na daljinu kojem su podvrgnuti električni fluidi”.

3.1.1. *Kulonov zakon.* Elektrostatika je oblast fizike koja proučava statički elektricitet ili naelektrisanje koje miruje. Postoji pozitivno i negativno naelektrisanje. Nosioci pozitivnog naelektrisanja su protoni, a negativnog elektroni. Protoni i elektroni su nosioci najmanje količine naelektrisanja u prirodi i nazivaju se *elementarna naelektrisanja.* Naelektrisanje protona obeležava se sa e (ili p^+), a naelektrisanje elektrona obeležava se sa $-e$ (ili e^-).

Privlačne električne sile se javljaju između pozitivno i negativno naelektrisanih tela. Odbojne električne sile se javljaju između istoimenih naelektrisanja. Na atomskoj skali, zbog principa neodređenosti kvantne mehanike, naelektrisana čestica nema preciznu poziciju, ali je predstavljena raspodelom verovatnoće, tako da se naelektrisanje pojedinačne čestice ne koncentriše u jednoj tački, već se ,,razmazuje” u prostoru i ponaša se kao prava kontinuirana distribucija naelektrisanja.

Za telo kažemo da je u ravnoteži (equilibrium) ako ne menja svoj položaj u odnosu na referentno telo, tj. ne kreće se u odnosu na njega. U matematičkoj analizi, grani matematike, izvod je mera kako (koliko brzo) funkcija menja svoje vrednosti kada joj se ulazne vrednosti menjaju. Izvod krive u nekoj tački predstavlja koeficijent pravca tangente date krive u toj tački.

Osnovna jednačina elektrostatike je Kulonov zakon, koji opisuje silu između dve tačke (naelektrisanih tela). Veličina elektrostatičke sile između dva naelektrisanja je direktno proporcionalna proizvodu veličina naelektrisanja i obrnuto proporcionalana kvadratu udaljenosti između tačaka koje su naelektrisane opterećenjima Q_1 i Q_2:

$$F = \frac{Q_1\, Q_2}{4\pi\varepsilon_0 r^2},$$

gde je ε_0 konstanta koja se naziva permitivnost vakuuma (dielektrična konstanta vakuuma).

U elektrostatici električno polje je naročito fizičko stanje u okolini naelektrisanih tela i električnih opterećenja, koje se vidno manifestuje u pojavi mehaničke sile, koja deluje na probno električno opterećenje uneto u polje. Ako se u prostor između naelektrisanih tela unese punktualno električno opterećenje q, koje je toliko malo da se njegov uticaj na raspodelu električnih opterećenja koja stvaraju polje može zanemariti, na njega će delovati mehanička sila čiji su intenzitet, pravac i smer potpuno određeni u svakoj tački prostora. U jednoj tački polja intenzitet sile je srazmeran probnom opterećenju dok pravac i smer ostaju nepromenjeni.

Pomenutu osnovnu manifestaciju električnog polja, da dejstvuje mehaničkom silom na uneseno probno opterećenje, iskoristićemo za defininiciju kvantativne karakteristike polja, koju nazivamo jačina električnog polja. Jačina polja je vektor, koji obeležavamo simbolom $\mathbf{E}$ i definišemo kao količnik mehaničke sile, kojom polje dejstvuje na pozitivnu probnu količinu elektriciteta, i same te količine elektriciteta: $\mathbf{E} = \mathbf{F}/q$. Vektor jačine polja ima pravac i smer sile koja deluje na pozitivno probno opterećenje, a intenzitet mu je brojno jednak sili koja deluje na jedinično opterećenje.

Podvucimo da postoji sličnost između Njutnovog zakona i Kulonovog zakona. Formule su analogne, masa igra ulogu naelektrisanja, a matematika koja se odnosi na teoriju potencijala je ista i to može da ima veze sa univerzalnim principima.

4. Varijacioni račun

Pronalaženje ekstrema funkcionala je slično pronalaženju maksimuma i minimuma funkcija. Maksimumi i minimumi funkcije mogu biti locirani pronalaženjem tačaka gde se izvod anulira (tj. jednak je nuli). Ekstremi funkcionala mogu se dobiti pronalaženjem funkcija gde je funkcionalni izvod jednak nuli. To dovodi do rešavanja pridružene Ojler - Lagranžove jednačine. Navedimo jedan poznat primer iz istorije matematike. Plato je eksperimentisao sa sapunicom kako bi pokazao postojanje minimalne površi sa datom granicom. Pogodno je formulisati odgovarajući problem u sledećoj formi.

4.0.1. *Platov problem.* Odrediti funkciju $z = f(x, y)$ zadatu na domenu D u ravni, čiji grafik minimizira površinu, uz pretpostavku zadatih vrednosti na granici domena D. Koristeći standardni argument u računu varijacija, može se pokazati da rešenje Platovog problema (koje nazivamo minimalni graf $z = f(x, y)$) zadovoljava nelinearnu parcijalnu diferencijalnu jednačinu

$$(1 + f_y^2)f_{xx} - 2f_xf_yf_{xy} + (1 + f_x^2)f_{yy} = 0.$$

Neka je G domen u ravni okružen Žordanovom krivom C. Pretpostavlja se da je Lagranžijan $L(x, y, u, p, q)$ dovoljno glatka funkcija svojih pet argumenata i neka je $u : G \to \mathbb{R}$. Mi koristimo notaciju $p = u_x$ i $q = u_y$. Razmotrimo

$$J[u] = \iint_G L(x, y, u, u_x, u_y)dxdy.$$

Za $t \in \mathbb{R}$ i $v : \overline{G} \to \mathbb{R}$, uvedimo $h(t) = J[u + tv]$. Ako je $v \in C^1(\overline{G})$ i $v = 0$ na bG, tada je

$$h'(0) = \iint_G (vL_u + v_xL_p + v_yL_q)dxdy.$$

Pomoću parcijalne integracije dobijamo

$$h'(0) = \iint_G v\big(L_u - (L_p)_x - (L_q)_y\big)dxdy.$$

Otuda je

$$L_u - (L_p)_x - (L_q)_y = 0 \tag{4.1}$$

i

$$L_u - L_{xp} - L_{yq} - u_xL_{up} - u_yL_{uq} - u_{xx}L_{pp} - 2u_{xy}L_{pq} - u_{yy}L_{qq} = 0. \tag{4.2}$$

Izvedimo Ojler - Lagranžovu jednačinu za minimalnu površ

$$L = \sqrt{1 + p^2 + q^2}.$$

Neka je $f : G \to \mathbb{R}$. Sa $\Gamma(f) = \{(x, y, f(x, y)) : (x, y) \in G\}$ označavamo grafik preslikavanja f. Razmotrimo 1-parametarsku familiju $f^t := f + tg$ i definišimo

$A(\Gamma(f)) = \int_G dA$ i $A(t) = A(\Gamma(f^t))$, gde je $dA = \underline{A}dxdy$ i
$\underline{A} = \sqrt{1 + |\nabla f|^2}$. Dalje nalazimo

$$A'(0) = -\int_G g \operatorname{div}\left(\frac{\nabla f}{\underline{A}}\right) dxdy$$

i stoga $\operatorname{div}\left(\frac{\nabla f}{\underline{A}}\right) = 0$.

Otuda, kako je $\underline{A}^2 = 1 + f_x^2 + f_y^2$ mi prvo nalazimo $\underline{A}\,\underline{A}_x = f_x f_{xx} + f_y f_{xy}$ i stoga

$$(1 + f_y^2) f_{xx} + (1 + f_x^2) f_{yy} - 2 f_x f_y f_{xy} = 0.$$

Površina tečnosti se ponaša kao elastična gumena membrana (opruga). Generalno, sistem pod uticajem sila kreće se ka ravnotežnoj konfiguraciji koja odgovara minimalnoj potencijalnoj energiji. Od svih tela zadate zapremine lopta ima najmanju površinu. Od svih zatvorenih površi zadate površine sfera sadrži najveći volumen i minimalnu površinsku potencijalnu energiju.

Nema kubnih kišnih kapi. Sila F_T kojom površina tečnosti deluje na objekt koji je u kontaktu s njom je direktno proporcionalna dužini L linije kontakta (ili površini A područja kontakta ako je predmet površ). Konstanta proporcionalnosti γ naziva se koeficijent površinskog napona tečnosti. Dakle, važi $F_T = \gamma L$. Stoga se koeficijent površinskog napona može izraziti u formi: $\gamma = F_T/L$. Čvrstoća membrane varira za različite tečnosti, npr. mnogo je manje za sapunastu vodu nego za čistu vodu. Na primer, ako $\varphi(x, y)$ označava pomeranje membrane iznad domena D u ravni xy, tada je njena potencijalna energija proporcionalna njegovoj površini:

$$U[\varphi] = \iint_D \sqrt{1 + \nabla\varphi \cdot \nabla\varphi} dx\, dy.$$

4.0.2. *Dirihleov princip.* Za male $|\nabla\varphi|$, $\sqrt{1 + \nabla\varphi \cdot \nabla\varphi} = \sqrt{1 + |\nabla\varphi|^2}$ se aproksimira sa $1 + |\nabla\varphi|^2/2$.

Često je dovoljno razmotriti samo male pomake membrane, čija energetska razlika od početnog položaja se aproksimira integralom energije

$$E[\varphi] = E_D[\varphi] = \frac{1}{2} \iint_D \nabla\varphi \cdot \nabla\varphi \, dx\, dy. \tag{4.3}$$

Ovde je $L = p^2 + q^2$. Kako je $L_u = L_{xp} = L_{yq} = L_{up} = L_{pq} = 0$ jednačina (4.2) se svodi na $L_u - u_{xx} L_{pp} - -u_{yy} L_{qq} = 0$. Dalje, $L_p = 2p$, $L_q = 2q$ i stoga $L_{pp} = L_{qq} = 2$. Otuda nalazimo $\Delta u := u_{xx} + u_{yy} = 0$. Funkcije koje zadovoljavaju poslednju jednačinu nazivaju se harmonijske.

Može se koristiti i formula (4.1). Tada treba $L_p = 2p$ razmatrati kao $L_p = 2p = 2u_x$ i $L_q = 2q = 2u_y$. Otuda $(L_p)_x = 2u_{xx}$ i $(L_q)_y = 2u_{yy}$.

Primetimo da se u fizici koristi notacija $\nabla^2 = \nabla \cdot \nabla$ za Laplasov operator Δ (Laplasijan).

Ako je u glatka funkcija na domemu $D \subset \mathbb{R}^n$, motivisani prethonim razmatranjima i formulom (4.3) definišemo

$$E[u] = E_D[u] = \int_D |\nabla u|^2 dx$$

i nazivamo integral energije preslikavanja u.

Razmatranja u ovoj sekciji prirodno dovode do sledećeg problema: minimizirati funkcionale U i E među svim probnim funkcijama f koje imaju zadate vrednosti na granici bD.

Prvo ćemo navesti Dirihlov princip za harmonijske funkcije.

Teorema 3 (Dirihleov princip). *Pretpostavimo da je D domen u $\mathbb{R}^n$ i:*

(a) *g je neprekidna funkcija na $\overline{D}$;*
(b) *g ima prve parcijalne izvode koji su neprekidni na D;*
(v) *integral energije funkcije g je konačan.*

Ako je u neprekidna funkcija na granici $\overline{D}$, harmonijska na D i ako je $u = g$ na granici $\overline{D}$ onda je

$$E(g) \geq E(u),$$

pri čemu se jednakost dostiže ako i samo ako je $u = g$ na bD.

5. Teorija potencijala i potencijal u elektrostatici

Podvucimo da se razmatranja u ovoj sekciji odnose na 3D-trodimenzioni prostor.

O osnovama elektrostatike i elektromagnetike videti na primer u [15]. Neka je e pozitivno naelektrisana čestica locirana u tački y. Ako se naelektrisana čestica s jediničnim naelektrisanjem nalazi u tački x, po Kulonovom zakonu naelektrisanje e deluje na jedinično naelektrisanje u x sa silom $\overrightarrow{F}(x) = \dfrac{e\overrightarrow{r}}{r^3}$, gde je $\overrightarrow{r} = \overrightarrow{yx}$ i $r = |x - y|$.

Polje koje potiče od naelektrisanja e je potencijalno polje, tj. postoji funkcija

$$\varphi = -\frac{e}{|x-y|}$$

takva da je $F = \nabla\varphi$. Funkcija φ se naziva potencijal. Ponovimo da važi $\Delta = \mathrm{div}\nabla = \nabla \cdot \nabla$.

Dajemo nekoliko posledica Kulonovog zakona. Pošto je

$$\Delta\frac{1}{r} = 0,$$

imamo presudnu činjenicu da je $\dfrac{1}{r}$ harmonijska funkcija u $\mathbb{R}^3 \setminus \{0\}$ i stoga $\mathrm{div}F = 0$.

Interesantno da u smislu distribucija važi $\Delta\dfrac{1}{r} = -4\pi\delta$, pri čemu je δ Dirakova distribucija.

Otuda sledi da je elektrostatičko polje solenoidno. U vektorskom računu solenoidno vektorsko polje (poznato i kao nekompresibilno vektorsko polje, vektorsko polje bez divergencije, ili poprečno vektorsko polje) je vektorsko polje v sa divergencijom nula u svim tačkama u polju:

$$\nabla \cdot \mathbf{v} = 0.$$

Sada razmotrimo sistem naelektrisanja $e_1, \ldots, e_n$ koja se nalaze u tačkama $x_1, \ldots, x_n$. Ovom sistemu možemo pridružiti polje sila (unošenjem probnog naelektrisanja), koje se može smatrati Rimanovom sumom,

$$\mathbf{F}(x) = \sum_{k=1}^{n} \frac{e_k}{|x-x_k|^3}(x - x_k)$$

sa potencijalom

$$\varphi(x) = -\sum_{k=1}^{n} \frac{e_k}{|x - x_k|}. \tag{5.1}$$

Ako se dva provodnika proizvoljnog oblika opterete jednakim količinama elektriciteta suprotnog znaka,

$$q_1 = -q_2 = Q,$$

u prostoru oko provodnika se formira električno polje, a između provodnika potencijalna razlika

$$U = \varphi_1 - \varphi_2 = \int_1^2 E dl.$$

Linijski integral se može uzeti po proizvoljnoj konturi koja spaja površi provodnika koje su ekvipotencijalne.

Između električnog opterećenja kondenzatora, Q, i potencijalne razlike U postoji linearna zavisnost, što je direktna posledica linearnih odnosa koji vladaju između električnih opterećenja i potencijala. Ako se ukupna količina elektriciteta na nekom provodniku poveća x puta, relativna raspodela opterećenja na njegovoj površini se neće izmeniti, već će se samo površinska gustina u svim tačkama povećati x puta. U istoj razmeri će se povećati i potencijal. Odnos između opterećenja i potencijalne razlike između elektroda kondenzatora je funkcija oblika, dimenzija i međusobnog položaja elektroda i naziva se kapacitivnost: $C = Q/U$.

Neka je G naelektrisano provodno telo u $\mathbb{R}^3$ u kom je ukupno naelektrisanje Q. Ako je sistem u stanju ravnoteže, potencijal V je konstantan u G.

Veličina $C(G) = \operatorname{cap}(G) = \dfrac{Q}{V}$ naziva se kapacitet provodnog tela G. U nastavku teksta je data matematička definicija kapaciteta oblasti u $\mathbb{R}^3$.

Totalni rad kada se sva naelektrisanja premeste u beskonačnost je

$$W = \sum_{k=1}^{n} W_k = \sum_{k<l} \frac{e_k e_l}{|x_l - x_k|} = \frac{1}{2} \sum_{k \neq l} \frac{e_k e_l}{|x_l - x_k|}. \tag{5.2}$$

Koristeći rezultate za potencijal naelektrisanja, možemo dati odgovarajuće matematičke definicije.

Smatraćemo da je gustina naelektrisanja neprekidna i da konačne sume teže integralu. Umesto raspodele naelektrisanja posmatraćemo mere.

Ako je površ S (odnosno oblast G) naelektrisana i A podskup na površi S (odnosno oblasti G) onda A sadrži neku količinu naelektrisanja $q = q(A)$. Dakle svakom podskupu A dodeljujemo $q = q(A)$, količinu naelektrisanja koju sadrži. Opštije, umesto naelektrisanja smatraćemo da imamo meru μ koja ,,meri" podskupove od S; $\mu(A) \geq 0$.

Sa $D^c = \mathbb{R}^n \setminus D$ označavamo komplement skupa D.

Motivisani formulama (5.1) i (5.2) definišemo:

i) Njutnov potencjial V^μ (predlažemo i oznake $N[\mu]$ i $\Gamma[\mu]$) mere μ sa nosačem u D kao konvoluciju:

$$V^\mu(x) = \Gamma[\mu](x) = \int_D \frac{1}{|x - y|} \mu(y)\, dy;$$

ii) energiju $I(\mu)$ mere μ sa:

$$I(\mu) = \int_D V^\mu(x)d\mu(x) = \int_{D\times D} \frac{1}{|x-y|} d\mu(x)d\mu(y).$$

Ako je $E(x) = \dfrac{1}{|x|}$, tada je $V^\mu = E * \mu$ konvolucija E i μ. Važi da je V^μ je harmonijska funkcija na $\overline{D}^c$.

Kada je sistem (telo G opterećeno naelektrisanjem Q) u ravnoteži, odgovarajući potencijal u_0 se može normalizovati tako da je u_0 jednako 1 na bG. Tada je $u_0 = u_G$ on G^c, gde je u_G harmonijska funkcija na $\overline{G}^c$, neprekidna na G^c, koja je jednaka 1 na bG.

Interesantno je da je kapacitet skupa G povezan sa energijom harmonijske funkcije u_G:

$$4\pi \mathrm{cap}(G) = E_G[u_0]. \tag{5.3}$$

Pre nego što skiciramo dokaz ove jednakosti, navodimo jedan dovoljan uslov za rešenje spoljašnjeg Dirihleovog problema.

Ako je G ograničena oblast sa glatkom granicom (preciznije Helderove klase $C^{1,\alpha}$) postoji rešenje spoljašnjeg Dirihleovog problema za neprekidnu realno vrednosnu funkciju f na bG, tj. funkcija $u = u_G$ koja je neprekidna na $\mathbb{R}^n \setminus G$, harmonijska na $\overline{G}^c = \mathbb{R}^n \setminus \overline{G}$, $u = f$ na bG i $u(x)$ teži c kada x teži ∞.

Sada možemo dati matematičku definiciju kapaciteta pomoću varijacinog problema: Preciznije $C(G)$ je infimum $E(\mu)$ preko familije svih pozitivnih mera μ za koje $N[\mu] \geq 1$ na G. Podvucimo kako je $V^\mu = -4\pi\mu$ i kako V^μ ima kompaktan nosač na osnovu parcijalne integracije, dobijamo

$$E(V^\mu) = -\int_{\mathbb{R}^3} V^\mu \Delta V^\mu dx = 4\pi E(\mu). \tag{5.4}$$

Ako označimo potencijal V^μ sa $u(x)$ i primetimo da funkcija $u(x)$ ima kompaktan nosač, dolazimo do sledeće definicije:

$C(G) = \mathrm{cap}(G)$ je infimum $\dfrac{E[u]}{4\pi}$ preko familije svih funkcija $u \in C_0^1(\mathbb{R}^n)$ (ili $u \in C_0^\infty(\mathbb{R}^n)$) za koje je $u(x) \geq 1$ na G.

Otuda, jednakost (5.3) sledi na osnovu Dirihleovog pricipa.

Primetimo da formula $E[u] = 4\pi E(\mu)$ daje još jedno objašnjenje zašto $E[u]$ nazivamo integral energije.

5.1. **Tomsonova teorema.** Na osnovu elektrostatičkih razmatranja zaključujemo da elektrostatička ravnoteža znači da ne postoji neto protok električnog naelektrisanja ili da nema struje. Ovde ćemo proučavati svojstva provodnika u elektrostatičkoj ravnoteži.

1. Električno polje unutar provodnika je nula u elektrostatičkoj ravnoteži.
2. Bilo koja neto naelektrisanja na izolovanom provodniku nalaze se na njegovoj površi.
3. Svaka tačka na površini provodnika u elektrostatičkoj ravnoteži je na istom potencijalu – površ je ekvipotencijalna.

Želeli bismo da razmotrimo Tomsonovu teoremu sa matematičke tačke gledišta.

Postoji analogija između raspodele naelektrisanja, elektrostatičkog potencijala, elektrostatičke energije električnog polja, respektivno i mere, Njutnovog potencijala, energija mera itd.

Teorema 4 (Tomsonova teorema)**.** *Za raspodelu naelektrisanja, za koju su obe površi provodnika ekvipotencijalne, elektrostatička energija električnog polja je minimalna.*

Umesto dva provodnika možemo razmatrati n-provodnika koji su ograničeni površima S_i opterećeno sa datim naelektrisanjima q_i, $i \in \{1, 2, \dots, n\}$.

U ovoj situaciji Tomsonova teorema tvrdi: *Za raspodelu naelektrisanja, za koju su sve površi provodnika ekvipotencijalne, elektrostatička energija električnog polja je minimalna.*

Na seminaru za analizu pokazali smo da je ovaj rezultat posledica Dirihleovog principa i dobili dalje rezultate.

5.2. **Karlemanov rezultat.** Neka je $F \subset D$ i $A = D \setminus F$ topološki prsten u $\mathbb{C}$. Ako je $S_1 = \text{area}(D)$ i $S_0 = \text{area}(F)$ tada je

$$\text{cap}^{-1}(D, F) \leq \ln \frac{S_1}{S_0}\,.$$

Ovaj rezultat se može iskazati i u sledećoj formi:

Teorema 5. *Među svim topološki prstenastim domenima u ravni sa datom površinom ,,rupa", domen ograničenog sa dva koncentrična kruga daje najmanju vrednost za* $\text{cap}(D, F)$.

Za više detalja o rezultatima prikazanim u ovoj podsekciji videti [8]. Na osnovu Žordanove teoreme znamo da prosta zatvorena Žordanova kriva γ u ravni deli ravan na dve oblasti. Označimo sa $\text{Int}(\gamma)$ ograničenu komponentu.

Neka su γ_0 i γ_1 dve proste zatvorene Žordanove krive u ravni i neka je $D_0 \subset D_1$ i $A = D_1 \setminus D_0$, gde je $D_0 := \text{Int}(\gamma_0)$ i $D_1 := \text{Int}(\gamma_1)$.

Interesantno je da su modul i kapacitet prstena jednaki, tj. važi

$$\text{cap}(A) = M(A) = d_A(\gamma_0, \gamma_1),$$

gde je $M(A)$ modul A i $d_A(\gamma_0, \gamma_1)$ je ekstremalno rastojanje između γ_0 i γ_1 u A.

Mekmulen je dokazao verziju Karlemanovog rezultata sa $M(A)$ (Milnor [13] koristi naziv Mekmulenova nejednakost) i primenio je u kompleksnoj dinamici.

Autor je dokazao razne verzije i uopštenja ovog rezulta u [8] (videti teoreme 5.3, 5.4 i 6.1).

Poenkare i Sege dobili su sledeći rezultat: ako S ograničava domen D u 3D-prostoru zapremine $V = \frac{4}{3}\pi R^3$, tada je

$$C(S) \geq C(S(0; R)) = \left(\frac{3}{4\,\pi}\right)^{1/3} V^{1/3}.$$

Drugim rečima, između svih domena date zapremine lopta ima najmanji kapacitet.

5.3. **Sferna simetrizacija.** Definišimo p-integral energije

$$E_p[u] = \int |\nabla u|^p dx.$$

Neka je $A \subset \mathbb{R}^n$ otvoren skup i $C \subset A$ kompaktan skup. Par $E = (A, C)$ naziva se kondenzator, a $p - \operatorname{cap}(A, C)$ je infimum

$$E_p[u] = \int_A |\nabla u|^p dx$$

preko familije svih funkcija $u \in C_0^1(A)$ (ili preko svih nenegativnih funkcija ACL^p sa kompaktnim nosačem u A) za koje je $u(x) \geq 1$ na C.

Opštije ako je G podskup u $\mathbb{R}^n$ definišimo kapacitet skupa G:
$C(G) = \operatorname{cap}(G)$ je infimum

$$E_p[u] = \int_{\mathbb{R}^n} |\nabla u|^p dx$$

preko familije svih funkcija $u \in C_0^1(\mathbb{R}^n)$ (ili $u \in C_0^\infty(\mathbb{R}^n)$) za koje je $u(x) \geq 1$ na G.

Zimer je dokazao $p - \operatorname{cap}(A, C) = M_p(\Delta(C, bA; A))$, gde M_p označava p-modul familije krivih $\Delta(C, bA; A)$ koje spajaju C i bA u A.

Definicija 1. *Neka je L zrak iz x_0 do ∞ i $E \subset \overline{\mathbb{R}}^n$ kompaktan skup.*
Sferna simetrizacija E^* skupa E u odnosu L definiše se na sledeći način:

1) *$x_0 \in E^*$ ako i samo ako $x_0 \in E$;*
2) *$\infty \in E^*$ ako i samo ako $\infty \in E$;*
3) *za $r \in (0, \infty)$ skup $E^* \cap S^{n-1}(x_0, r)$ je zatvorena sferna kalota centrirana na L sa istom $(n-1)$-dimenzionom merom kao i $E(x_0, r) := E \cap S^{n-1}(x_0, r)$ za $E \cap S^{n-1}(x_0, r) \neq \emptyset$ i $\emptyset$ ako je $(n-1)$-dimenziona mera skupa $E(x_0, r)$ nula.*

Gering je dokazao da ako je (A, C) kondenzator, tada za $p \geq 1$ važi

$$p - \operatorname{cap}(A, C) \geq p - \operatorname{cap}(A^*, C^*).$$

6. Harmonijska preslikavanja i geometrijske nejednakosti

6.1. **Harmonijska preslikavanja.** Podsetimo da smo u prethodnim sekcijama razmatrali varijacioni račun i pridruženu Ojler–Lagranžovu jednačinu i elektrostatičku teoriju potencijala, i specijalno Dirihleov princip i Platov problem, koji se odnosi na pronalaženje funkcije $z = f(x, y)$ zadate na domenu D u ravni, čiji grafik minimizira površinu, uz pretpostavku zadatih vrednosti na granici D, rešenja se nazivaju minimalnim površima. Ponovimo ova razmatranja mogu se smestiti u zajednički okvir: rešenja eliptičkih jednačina drugog reda u zavisnosti od graničnih uslova. U tom pravcu u ovoj sekciji dajemo uslove pod kojima je gradijent rešenja ograničen, a kao primer navodimo i rešenje Sinaijevog problema.

Napomenimo da se teorija potencijala može razviti i za rešenja eliptičkih jednačina drugog reda. Dodatno razmatramo i neke geometrijske nejednakosti.

Ako je M podsup $\mathbb{R}^n$ sa bM označavamo granicu M.

Holomorfne funkcije su kompleksne funkcije definisane na otvorenom podskupu kompleksne ravni koje su diferencijabilne. Kompleksna diferencijabilnost ima mnogo jače posledice nego uobičajena (realna) diferencijabilnost. Na primer, holomorfne funkcije su beskonačno puta diferencijabilne, što nikako ne važi za realno diferencijabilne funkcije. I više od toga holomorfne funkcije su predstavljive stepenim redovima po kompleksnoj promenljivoj z. Većina elementarnih funkcija, uključujući eksponencijalnu funkciju, trigonometrijske funkcije, i sve polinomijalne funkcije, holomorfne su.

Razmatranja u vezi sledećeg problema su važna u geometrijskoj teoriji funkcija: ako su D i G dva domena u n-dimenzionom prostoru, odrediti funkcije koje minimiziraju Dirihleov integral

$$\int_D |\nabla f|^2 dx$$

po klasi funkcijama $f : \overline{D} \to \overline{G}$ koje preslikavaju bD u bG i ,,ne degenerišu se na granici".

U ravni stacionarne tačke su holomorfne funkcije, a u prostoru harmonijske funkcije. Na primer, u prostoru ne postoji biholomorfno preslikavanje između lopte $\mathbb{B}_2$ i polidiska $\mathbb{U}\times\mathbb{U}$, a u ravni postoji između domena istog topološkog tipa pri dodatnim uslovima.

Za jedinični vektor $\mathbf{h}$ koji ima početak u tački x definišimo pravolinijski put $C(t) = x+th$. Ako t interpretiramo kao vreme, preslikavanje f definiše put $C_{\mathbf{h}}(t) = f(x + th)$. Brzina kretanja čestice duž ovog puta je istezanje funkcije f u pravcu (izvod u pravcu) vektora $\mathbf{h}$.

Za realno vrednosne funkcije $\dot{C}_{\mathbf{h}}(t) = (\nabla f, \mathbf{h})$ važi da je $|\nabla f(x)|$ najveće istezanje.

U ravni holomorfne funkcije istežu isto u svim pravcima. U prostoru kažemo da je preslikavanje $f : D \to G$, $K-$ kvazikonformno ako je količnik između najvećeg i najmanjeg istezanja uniformno ograničen sa K u svakoj tački $x \in D$.

U [4] razmatramo

$$|\Delta u| \le a|\nabla u|^2 + b \tag{6.1}$$

i pokazujemo da su kvazikonformna preslikavanja između glatkih domena koja zadovoljavaju (6.1) Lipšicova.

Za dalja uopštenja videti Apendiks 3 i teoremu 6.9 u [7].

Teorema 6 ([7])**.** *Neka su D i Ω glatki domeni u $\mathbb{R}^m$. Ako je $u : D \to \Omega$ kvazikonformno (g, g')-harmonijsko preslikavanje (ili zadovoljava (6.1)), tada je preslikavanje u Lipšicovo na D.*

Ovaj rezultat je razmatran na Seminaru za analizu u Beogradu i na konferenciji u Čenaju, Indija (Workshop on Harmonic Mappings and Hyperbolic Metrics, Chennai, India, Dec. 10–19, 2009).
Videti i https://sites.google.com/site/iwhmhm09/course-materials.

Na ovoj konferenciji autor je održao nekoliko predavanja. Ovde samo skiciramo ideju dokaza.

Pretpostavljamo da su domeni D i Ω glatki, tako da možemo primeniti lemu 6.18 (i teoremu 6.19) iz [3]. Neka je $x_0 \in bD$, i $y_0 = u(x_0)$. Tada $y_0 \in b\Omega$ i postoji lokalna karta ψ definisana i glatka na $\overline{B}$ gde je $B = B(y_0, r_0)$ lopta, koja preslikava $D \cap B$ u hiperravan $y^m = 0$ u $\mathbb{R}^m$.

Definišimo $\tilde{u} = \psi \circ u$ i neka je $\tilde{u}^m$ m-ta kordinata $\tilde{u}$. Podvucimo da je $\tilde{u}^m$ 0 na $u^{-1}(B \cap \partial\Omega)$. Definišemo metriku g'' na $\overline{\psi(B)}$, $\overline{\psi(B)} \subset \mathbb{R}^m$ tako da je ψ izometrija u odnosu na euklidsku metriku na B. Primenom leme 6.18 [3] može se pokazati da je $\tilde{u}$ Lipšicovo preslikavanje.

V. Božin i autor su pokazali da su domeni Ljapunova kvazikrugovi i da se mogu aproksimirati sa konveksnim domenima. Koristeći specijalnu tehniku, rešili su jedan otvoren problem (star oko dvadesetak godina, koji je verovatno prvi formulisao D. Kalaj):

Teorema 7 ([2]). *Kvazikonformna harmonijska preslikavanja između domena Ljapunova su bi-Lipšicova.*

Napomenimo da u [7] ukazujemo na neke razlike između teorije u ravni i u prostoru.

6.1.1. *Pitanje Jakoba Sinaija.* Autor je imao priliku da 2016. godine na Prinstonu razgovara sa Jakobom Sinaijem, dobitnikom Abelove nagrade, akademikom RAN i profesorom na Prinstonu. U tom razgovoru Sinai je izložio neke probleme vezane za dinamičke sisteme.

Takođe, tokom komunikacije pojavilo se i sledeće pitanje: neka je G ograničen domen u ravni klase C^2, osim u konačno mnogo tačaka na granici, gde ima singularitete ,,tipa uglova". Šta možemo reći o regularnosti Dirihleovih sopstvenih funkcija na G blizu granice? (O ovom pitanju govorimo kao o Sinaijevom pitanju.)

Kažemo da je f Dirihleova sopstvena funkcije na domenu G ako postoji skalar λ tako da je $\Delta f = \lambda f$ na G i $f = 0$ na bG. U klasičnom smislu pretpostavljamo da je f neprekidna na zatvorenju G i da ima parcijalne izvode drugog reda na G. Opštije možemo govoriti o Dirihleovim sopstvenim funkcijama u prostorima Soboljeva. Funkcija f pripada prostoru Soboljeva $W_0^{1,2}(G)$ ako je istezanje integrabilno i trag $\mathrm{tr}(f)$ nula na granici. Ako je f neprekidna na zatvorenju G tada se $\mathrm{tr}(f)$ svodi na restrikciju f na bG.

Teorema 8. *Pretpostavimo da je G ograničen konveksan domen u ravni. Ako je $w \in W_0^{1,2}(G)$ Dirihleova sopstvena funkcija, onda je gradijent funkcije w ograničen.*

6.2. **Geometrijske nejednakosti.**

6.2.1. *Švarcova lema.* Neka $b \in \mathbb{S} = \{w : -1 < \mathrm{Re} w < 1\}$ i ϕ_b konformni izomorfizam $\mathbb{U} = \{z : |z| < 1\}$ na $\mathbb{S}$, $\phi_b(0) = b$ i $\phi_b'(0) > 0$. Preslikavanje ϕ_b ima sledeća svojstva:

i) ϕ_b je rastuća na $(-1, 1)$ i preslikava $(-1, 1)$ na sebe;

ii) Za $r \in [0, 1)$ važi $\phi_b([-r, r]) = [m_b(r), M_b(r)]$, gde je $m_b(r) = \phi_b(-r) = \dfrac{4}{\pi} \arctan \dfrac{a - r}{1 - ar}$, $M_b(r) = \phi_b(r) = \dfrac{4}{\pi} \arctan \dfrac{a + r}{1 + ar}$ i $a = \tan \dfrac{b\pi}{4}$;

iii) Ako je $z = re^{i\alpha} \in \mathbb{U}$, i R_α^+ i R_α^- rotacije definisane množenjem sa $e^{-i\alpha}$ i $-e^{-i\alpha}$ respektivno, $u^+ = \mathrm{Re}\phi_b \circ R_\alpha^+$ i $u^- = \mathrm{Re}\phi_b \circ R_\alpha^-$, tada je $u^+(z) = M_b(r)$ i $u^-(z) = m_b(r)$.

Za razmatranja u vezi Švarcove leme videti [9]. Ovde navedimo samo jedan rezultat. U radu [11] autor i M. Svetlik su dokazali:

Teorema 9 ([11]). *Neka je $u : \mathbb{U} \to (-1, 1)$ harmonijska funkcija i $u(0) = b$. Tada je*

$$m_b(|z|) \le u(z) \le M_b(|z|), \quad \text{za svako} \quad z \in \mathbb{U}, \tag{6.2}$$

i nejednakost je optimalna za svako $z \in \mathbb{U}$ u smislu iii).

6.2.2. *Lokalna distorzija kvazikonformnih preslikavanja.* Neka Δ označava jedinični krug i $f : \Delta \to \mathbb{C}$ kvazikonformno preslikavanje. Definišimo
$\delta = \mathrm{dist}(0, \partial f(\Delta))$, $mu^+(z) = \sup\{|\mu(\zeta)| : |\zeta| = |z|\}$, gde je μ kompleksna dilatacija i pretpostavimo da integral

$$\tau = \int_0^1 \frac{2\mu^+(t)}{1 + \mu^+(t)} \frac{dt}{t}$$

konvergira.

Teorema 10 ([6]). *Neka je f kvazikonformno preslikavanje na Δ, koje je konformno u 0 i neka je $a_1 = f'(0)$. Pretpostavimo gornje oznake. Tada je*

$$|f'(0)|e^{-\tau} \leq 4\delta. \tag{6.3}$$

Ako je f konformno preslikavanje, onda je $\tau = 0$, i otuda kao posledicu dobijamo čuvenu Kebeovu teoremu.

7. Dodatni komentari

Kratka biografija, prikaz i odjek naučnih rezultata autora dobijenih posle izbora za dopisnog člana SANU dat su u [16]. Ovde navodimo samo da je Kristofer Bishop, pozvani predavač na Međunarodnom kongresu matematičara 2018. godine, u Math Reviews dao izuzetno pozitivan prikaz rezultata Mateljevića o harmonijskim preslikavanjima: [a41], MR2796877 (2012i: 30046) ,,Ovo je dobro napisan i vrlo zanimljiv rad"; [a40], MR2791734 (2012i: 30047) ,,Ovaj rad je zanimljiv doprinos teoriji harmonijskih kvazikonformalnih preslikavanja."; [a35], MR2337479 (2008d: 30037) ,,Ovaj rad je fundamentalan za geometrijsku teoriju funkcija i kompleksnu dinamiku jer postavlja stroga ograničenja na veličinu lokalne distorzije koju kvazikonformno preslikavanje može imati." Radovi [a35], [a40] i [a41] navedeni su u [16]; rad [a35] citiran je u ovom članku kao [6].

7.0.1. *Koautori.* Autor je sarađivao i objavio radove sa M. Arsenović, V. Božin, D. Kalaj, M. Knežević, V. Marković, M. Marković, M. Pavlović, M. Jevtić, D. Šarić, I. Anić, N. Lakić, M. Svetlik, M. Albijanić, N. Mutavdžić, M. Vuorinen, S. Chen, S. Ponnusamy, ... Autorovi doktorandi (V. Marković, D. Kalaj...) imaju značajne matematičke doprinose u svojim oblastima. Na primer, V. Marković je izabran za Fellow of the Royal Society (ekvivalentno akademik u Engleskoj), u 2014. godini dobio je niz prestižnih nagrada i bio pozvani predavač na Međunarodnom kongresu matematičara u Seulu, Južna Koreja).

Zahvaljujem akademiku A. Đorđeviću i kolegama M. Arsenoviću, O. Mihić i M. Svetliku na korisnim sugestijama, i posebno kolegi M. Svetliku koji je izradio slike i tehnički obradio tekst.

U vezi vizualizacije videti takođe članke:

(1) Piotr Zarzycki, *From visualizing to proving*, September 2004 Teaching Mathematics and its Applications 23(3) DOI: 10.1093/teamat/23.3.108, SourceOAI.
(2) James Robert Brown, University of Toronto, https://pirsa.org/04110002, Proofs and Pictures: The Role of Visualization in Mathematical and Scientific Reasoning

(3) Antonia Makina, *The role of visualisation in developing critical thinking in mathematics*, Perspectives in Education 28(1), March 2010 : 24–33, Project: Leadership for e-learning.
(4) Yilmaz, R. & Argun, Z. *Role of visualization in mathematical abstraction: The case of congruence concept.* International Journal of Education in Mathematics, Science and Technology (IJEMST), 6(1), 2018, 41–57. DOI:10.18404/ijemst.328337 2006. In Novotná, J., Moraová, H., Krátká, M. & Stehlková, N. (Eds.). Proceedings 30th Conference of the International Group for the Psychology of Mathematics Education, Vol. 4, pp. 457–464. Prague: PME. 4–457
(5) Bettina Rösken Katrin Rolka, *A picture is worth a 1000 words-the role of visualization in mathematics learning*, University of Duisburg-Essen, Germany , https://www.emis.de/proceedings/PME30/4/457.pdf

The paper [14] describes some important aspects concerning the role of visualization in mathematics learning. We consider an example from integral calculus which focuses on visual interpretations. The empirical study is based on four problems related to the integral concept that highlight various facets of visualization. In particular, we are interested in the visual images that students use for working on specific problems and how they deal with given visualizations. The findings show the importance as well as the difficulties of visualization for the students. Visualization is the ability, the process and the product of creation, interpretation, use of and reflection upon pictures, images, diagrams, in our minds, on paper or with technological tools, with the purpose of depicting and communicating information, thinking about and developing previously unknown ideas and advancing understandings. Problem 1: Draw a figure to illustrate the geometric definition of the integral.

Literatura

[1] Nives Baranović, *Razvoj vizualno-prostornih vještina i geometrijskog mišljenja studenata učiteljskog fakulteta zasnovan na metodi usmjerenog opažanja i teoriji Van Hila*, (doktorska disertacija), septembar, 2022.
[2] V. Božin, M. Mateljević, *Quasiconformal and HQC Mappings Between Lyapunov Jordan Domains*, http://arxiv.org/abs/1805.04313, prihvaćen za štampu u The Annali della Scuola Normale Superiore di Pisa, Classe di Scienze.
[3] D. Gilbarg, N. Trudinger, *Elliptic partial Differential Equation of Second Order*, Second Edition, 1983.
[4] D. Kalaj, M. Mateljević, *Inner estimate and quasiconformal harmonic maps between smooth domains*, Journal d'Analise Math. 2006. Vol. 100, pp. 117–132.
[5] M. Knežević, M. Mateljević, On the quasi-isometries of harmonic quasi-conformal mappings, J. Math. Anal. Appl, 2007, Vol 334, No 1, 404–413.
[6] M. Mateljević, *Quasiconformal and quasiregular harmonic analogues of Koebe's theorem and applications*, Ann. Acad. Sci. Fenn. Math. **32** (2007), 301–315.
[7] M. Mateljević, *The Lower Bound for the Modulus of the Derivatives and Jacobian of Harmonic Injective Mappings*, Filomat 29:2 (2015), 221–244.
[8] M. Mateljević, *Isoperimetric Inequality, F. Gehring's Problem on Linked Curves and Capacity*, Filomat 29:3 (2015), 629–650.
[9] M. Mateljević, *Schwarz lemma and Kobayashi Metrics for harmonic and holomorphic functions*, J. Math. Anal. Appl. 464 (2018) 78–100.
[10] M. Mateljević, M.Svetlik, *A contribution to the development of functional thinking related to convexity and one-dimensional motion*, The Teaching of Mathematics, XIV 2 (2011), 87–96.
[11] M. Mateljević, M. Svetlik, *Hyperbolic metric on the strip and the Schwarz lemma for HQR mappings*, Appl. Anal. Discrete Math. 14 (2020), pp. 150-168.

[11] M. Mateljević, M. Svetlik, *Hyperbolic metric on the strip and the Schwarz lemma for HQR mappings*, Submitted on 20 Aug 2018, arXiv:1808.06647v1 [math.CV], http://arxiv.org/abs/1808.06647

[12] M. Mateljević, Neki aspekti teorije potencijala, vizualizacija, varijacioni račun i primene, Akademske besede knj. 2, 2019, redovinih članova SANU koji su izabrani 8. novembra 2018 god. U izdanju Srpske akademije nauka i umetnosti, 73-110, ISSN 2466-5134

[13] J. Milnor, *Dynamics in One Complex Variable*, Ann. Math. Studies **160**, Princenton U. Press, Princenton, 2006.

[14] Bettina Rösken, Katrin Rolka, A picture is worth a 1000 words-the role of visualization in mathematics learning, https://www.emis.de/proceedings/PME30/4/457.pdf.

[15] J. Surutka, *Elektromagnetika*, Akademska Misao, Beograd 2006.

[16] Akademici A. Ivić, S. Pilipović i G. Milovanović, *Predlog za izbor Miodraga Mateljevića za redovnog člana SANU*, Avgust 2018.

[17] Miodrag Mateljevic and Marek Svetlik, A Contribution to the Development of Functional Thinking Related to Convexity and One-dimensional Motion and Inertial Frames, slide.

Vladimir FILIPOVIĆ*

METAHEURISTIC OPTIMIZATIONS IN BIOINFORMATICS AND MACHINE LEARNING

Abstract: This paper deals with the usage of metaheuristic optimization methods aimed at solving problems in bioinformatics and machine learning. Two metaheuristic optimization methods have been selected and explained in detail: Electromagnetism-like Metaheuristics and Variable Neighborhood Search. Results obtained by applying those two metaheuristic optimization methods on various problems in bioinformatics and machine learning are described. More precisely, the following problems have been solved: dimensionality reduction, support vector machine parameter selection, maximum betweeness problem and k-plex partitioning problem. At the end, algebraic topology concepts aimed at enhancing Electromagnetism-like Metaheuristics and Variable Neighborhood Search are proposed and their design is described.

Key words: *Metaheuristics, Optimization, Electromagnetism-like Metaheuristics, Variable Neighborhood Search, Algebraic topology*

1. INTRODUCTION

In recent years, there has been a tremendous growth of interest in applications of optimization in bioinformatics and machine learning. Optimization is frequently used for designing and modeling complex systems, which are essential in biomedical and biological research, e. g. in solving the maximum betweeness problem [1], in partitioning complex networks to k-plex structures [2], in identifying functionally related protein groups in weighted PPI networks [3], etc.

The main purpose of this paper is to address the usage of metaheuristic optimization methods in bioinformatics and in machine learning. In order to do so, we will concentrate on three problems from these domains (Dimensionality Reduction Problem, Maximum Betweeness Problem and Maximum Edge k-plex Partitioning Problem) and on two metaheuristic

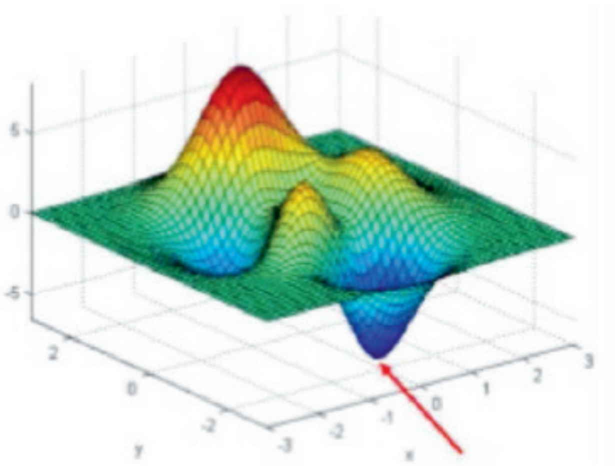
* Vladimir Filipović, University of Belgrade, Faculty of Mathematics, Department of Computer Science, Belgrade, Serbia

optimization methods (Electromagnetism-like Metaheuristics and Variable Neighborhood Search).

2. PROBLEMS

Firstly, let us define the optimization problems. Since minimization and maximization optimization problems have basically the same structure, the minimization optimization is defined.

Following elements are known: search space S; solution space X, $X \subseteq S$; objective function f, $f: S \longrightarrow R$, which maps elements of S to real numbers. In minimization optimization problems, the goal is to calculate $x^* \in X$, such that $f(x^*) = min\{f(x) | x \in X\}$. Picture 1 illustrates one example of the minimization optimization problem.



Picture 1 — Minimization optimization problem illustration

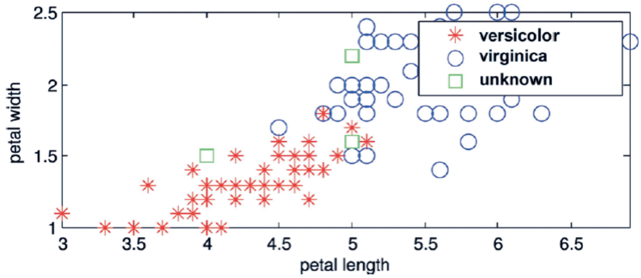
There are many important optimization problems in the fields of machine learning and bioinformatics, the most relevant are enlisted in [4] and [5].

This paper will address three machine learning and bioinformatics combinatorial optimization problems: Dimensionality Reduction Problem, Maximum Betweeness Problem and Edge-weight k-plex Partition Problem.

2.1. DIMENSIONALITY REDUCTION PROBLEM

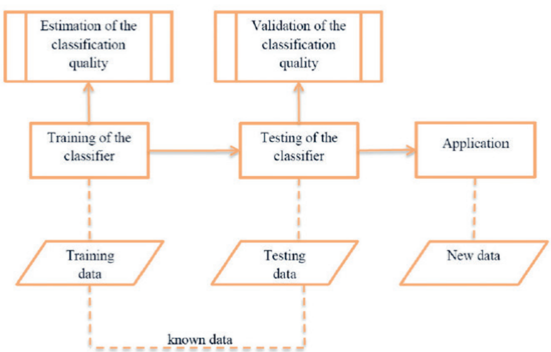
Data mining is one of the most popular and exciting discipline of applied informatics. It allows researchers to discover complex and hidden patterns in data, which can potentially lead to completely new conclusions in different disciplines, where sometimes even experts in the disciplines cannot do better. Nowadays, there is an extremely rapid growth in the volume of data stored in biological databases, with increased complexity of data and a very high dimensionality.

Data mining includes classification, which predicts a certain outcome based on a given input. An illustrative example of a classification task is given in Picture 2.



Picture 2 — Basic concept of classification

In order to learn how to predict outcome, the algorithm uses a set of training records containing a set of attributes and the respective outcome. The classification algorithm then, in the so-called training phase, tries to discover relationships between the attributes that would make it possible to predict an outcome. After this step, the algorithm is given a dataset not seen before, called a set of testing records, which contains the same set of attributes, except for the prediction attribute that is not yet known. The algorithm analyses the input and produces a prediction — this is the testing phase. The prediction accuracy defines the quality of the classification algorithm. After the testing phase, classifier is used in real-life conditions. The classification process is described by the flowchart in Picture 3.



Picture 3 — Flowchart of the classification process

There are two types of benefits for applying dimensionality reduction by feature selection for the classification process: firstly, by eliminating unnecessary features, it is possible to eliminate dataset noise that degrades the quality of the classification model; secondly, the problem dimension is decreased and the efficiency is increased.

The criteria for dimensionality reduction can vary, and in terms of classification problems (which are one focus of this paper) are usually referred to the classification accuracy, model efficiency, level of dimension reduction, or composition of former criteria. Dimensionality reduction algorithms have been utilized as a support for solving various and often real problems. There are three general types of dimensionality reduction algorithms [6]:

1) wrapper methods that use a classification algorithm as a black box for the evaluation of feature subsets;

2) filter methods that form feature subsets in the preprocessing phase, and do not depend on the employed classification algorithm;

3) embedded methods that form a feature subset in the training process and are specific to a given classification algorithm.

Various optimization techniques are used for dimensionality reduction by feature selection: genetic programming, fractional 0–1 programming, neural-genetic algorithms, particle swarm optimization, etc. In-depth study of Dimensionality Reduction Problem is given in [6], where the wrapper method is considered for dimensionality reduction, where the 1-nearest-neighbor classifier (1-NN) and support vector machine (SVM) are used as underlying classification mechanisms. Here, specific Electromagnetism-like Metaheuristics is proposed, various computational experiments are executed and obtained results are compared to results reported from other state-of-the-art methods for the considered problem.

2.2. MAXIMUM BETWEENESS PROBLEM

The betweeness problem is a well-known optimization problem. For a given finite set S of n objects $S = \{x_1, x_2, \ldots, x_n\}$ and a given set C of triples $\left(x_i, x_j, x_k\right) \in S \times S \times S$, the betweeness problem is a problem of determination of the total ordering of the elements from S, such that triples from C satisfy the „betweeness constraint", i. e. the element x_j is between the elements x_i and x_k. The problem presented in the paper [1], called the Maximum Betweeness Problem (MBP), deals with finding the total ordering that maximizes the number of satisfied constraints.

The MBP, as well as other betweeness problems, belongs to a class of discrete optimization problems. Those problems have important applications in various fields, including bioinformatics. For example, the MBP is used for solving some physical mapping problems in molecular biology. During the radiation hybrid experiments, X-rays are used to fragment chromosomes. If the markers on chromosomes are more distant, the probability that the given dose of an X-ray will break a chromosome is greater. In this way, markers are placed on two separate chromosomal fragments. By estimating the frequency of the breaking points, and thus the distances between markers, it is possible to determine their order within a chromosome in a manner analogous to meiotic mapping. In this context, improvement of the radiation experiment can be achieved by finding the total ordering of the markers that maximizes the number of satisfied constraints.

2.3. PROBLEM OF PARTITIONING OF COMPLEX BIOLOGICAL NETWORKS (K-PLEX PARTITION PROBLEM)

Partitioning networks into high density subnetworks, especially cliques, has already been proven as a useful technique for obtaining new information in understanding complicated relations between biological elements. For example, partitioning in protein threading analysis can be reduced on maximum edge weight clique problem, the protein side chain packing problem is transformed into a problem of finding a maximum weight clique.

Finding cliques is also one of the methods for identification of the clusters that are later divided into protein complexes and dynamic functional modules. By analyzing the multibody structure of the network of protein–protein interactions (PPI), molecular modules that are densely connected within themselves, but sparsely connected with the rest of the network, are discovered. Cliques have a similar use in modular decomposition of PPI networks. This decomposition allows to combine proteins into the actual functional complexes by identifying groups of proteins acting as a single unit.

On the other hand, a number of biological networks classes contain only sparse networks. Dealing with such networks, partitioning into cliques can be too restrictive method, so many potentially useful information about the interference of biological objects can be neglected. Therefore, clique relaxation approaches could be even more useful.

In the approach presented here, partitioning is followed by the principle that the objects in each partition are still highly connected in a particular way, but not so restrictively to form a clique. By relaxing cliques to sparse graphs, biological objects become connected in semantically or functionally logical groups which we call k-plexes, having in mind that the total sum of weights in all partitions should be as large as possible.

Here, we deal with the partitioning of the edge-weighted networks into k-plex components, where a subset of some $n$vertices in a network is a k-plex if the degree of each vertex in the subnetwork induced by this subset is at least $n-k$. The aim of the Maximum Edge-weight k-plex Partitioning Problem is to find the k-plex partitioning with the maximal total weight of edges.

More formal definition of this problem is given in [2]: Let a network be denoted as $G=(V,E)$, where $V=\{1,2,\ldots,n\}$ is the set of nodes and $E\subseteq V\times V$ is the set of edges. With uv we simply denote the edge $\{u,v\}\in E$. With real numbers $w_{uv}>0$ we denote the weight of the edge connecting nodes u and v. We call u and v the end-vertices of the edge uv.

Let $k\geq 1$be an integer. A set of nodes S is called k-plex if the degree of each node in the subnetwork induced by S is at least $n-k$. The weight of a k-plex is the sum of all its edge weights.

The weight of the whole partition is the sum of the weights of all its k-plex components. The Maximum Edge-weight k-plex Partitioning Problem is then defined as finding such a partition of V which is of the maximum total weight and each component is a k-plex. If $k=1$, the k-plex is a clique and the maximum edge-weight k-plex partitioning problem is brought down to the Maximum Edge-weight Clique Partitioning Problem.

3. METAHEURISTIC OPTIMIZATION METHODS

Metaheuristic methods are generalized computational intelligence methods that can be successfully adopted to various problem domains. They are trying to obtain the optimal solution, or the solution that is close to optimal one. Metaheuristic algorithms are characterized with approximation and non-determinism.

Basic metaheuristics concepts are abstractly represented. They should be adapted to problem domain, otherwise they should won't obtain enough good solution.

As the contrast to exact methods, which produces the exact solution (but have issues with time resources, with memory resources and sometimes solution cannot be obtained at all), metaheuristics methods produce approximate solution, which optimality is not guaranteed (but it works with limited computational resources, solution is always produced, and it usually has good quality).

Metaheuristic methods can be population-based (like Evolutionary algorithms, Particle Swarm Optimization, Electromagnetism-based Metaheuristics, etc.) or single-solution (like Taboo Search, Simulated Annealing, Variable Neighborhood Search etc.).

In the following subsections we will focus on two methods, one population based and one single-solution, namely to Electromagnetism-like Metaheuristic (EM) and Variable Neighborhood Search (VNS).

3.1. ELECTROMAGNETISM-LIKE METAHEURISTICS (EM)

Electromagnetism-like Metaheuristic (EM), proposed in [7], represents a population-based optimization technique inspired by mechanisms of interaction among electrically charged particles (called EM points). The method employs a proficient search process governed by EM points, where each of them represents single candidate solution of the underlying problem. EM points that represent better solutions are awarded with higher charge. This is crucial for leading the search process towards promising solution regions, because EM points with higher charge attract other points more strongly. The exact attraction-repulsion relationship is given in formula analogues to Coulomb's Law.

Electromagnetism-like algorithms turn out to be successful in solving many problems with practical and theoretical background: in [8] EM technique is adopted to solve feature selection problem, EM method for uncapacitated multiple allocation hub location problem is proposed in [9], EM method for the SVM parameter tuning [10], etc.

Overall structure of the EM algorithm is described in the flowchart in Picture 4.

```
input: N_it, M
p = createInitialPoints(M);
for iter ← 1 to N_it do
    for i ← 1 to M do
        objFunction(p_i);
    end
    charges(p);
    forces(p);
    relocate(p);
end
printSolution();
```

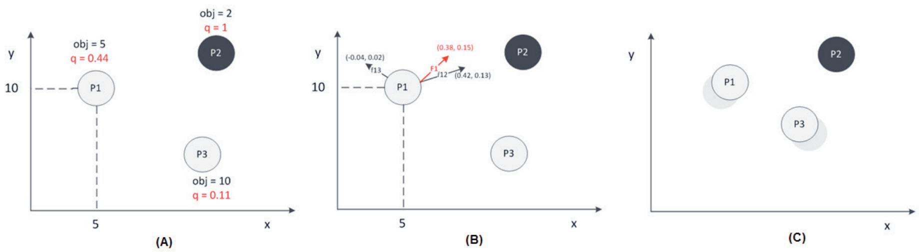
Picture 4 — Outline of the EM method

EM requires only two control parameters: $N_{it}$is the number of the main loop iterations and M represents the number of EM points. The points are first assigned with initial solutions and after that, the algorithm enters into the main loop. The main loop iterates N_{it} times and within iteration every EM point $\boldsymbol{p_i}$, $i = 1,\ldots M$ is subjected to the objective value calculation, i. e. measuring the quality of solution represented by that point.

The next step is the calculation of the EM point's charges. The charge of a fixed EM point will depend on its solution quality, according to formula:

$$q_i = e^{-N \frac{obj(\boldsymbol{p_{best}}) - obj(\boldsymbol{p_i})}{\sum_{k=1}^{M} obj(\boldsymbol{p_{best}}) - obj(\boldsymbol{p_k})}}$$

In previous formula, N is dimensionality on EM point space, obj is an objective function and $\boldsymbol{p_{best}}$ denotes EM point with the highest objective value. Illustrative example with EM points and its calculated charges is shown in Picture 5 (A).



Picture 5 — EM — (A) calculation of the charges, (B) calculation of the forces, (C) movements of the EM points

After all charges are calculated, the total impact on each point is calculated by superpositioning particle pairwise interaction forces, which are calculated by following formula:

$$\boldsymbol{F_i} = \begin{cases} \sum_{j=1, j\neq i}^{M} \left(\boldsymbol{p_j} - \boldsymbol{p_i}\right) \frac{q_j \times q_i}{\left\|\boldsymbol{p_j} - \boldsymbol{p_i}\right\|^2}, & p_j^{obj} < p_i^{obj} \\ \sum_{j=1, j\neq i}^{M} (\boldsymbol{p_i} - \boldsymbol{p_j}) \frac{q_j \times q_i}{\left\|\boldsymbol{p_j} - \boldsymbol{p_i}\right\|^2}, & p_j^{obj} \geq p_i^{obj} \end{cases}.$$

It should be noticed that forces are calculated similar to the Coulomb's Law, in a sense that the force between every two particles is proportional to the product of their charges and inversely proportional to their distances. Picture 5 (B) shows calculated force vectors for every EM point given in previous illustrative example.

Upon calculation of all the forces $\boldsymbol{F_i} = \left(F_i^1, F_i^2, \ldots, F_i^N\right)$ $i = 1,\ldots M,$ the movement procedure is applied. Movement of each EM point $\boldsymbol{p_i} = \left(p_i^1, p_i^2, \ldots, p_i^N\right)$ is guided by direction and magnitude of corresponding force vector $\boldsymbol{F_i}$. The following formula determines movements of EM points:

$$p_i^k = \begin{cases} p_i^k + \lambda \frac{F_i^k}{\|\boldsymbol{F_i}\|}\left(1 - p_i^k\right), & F_i^k > 0 \\ p_i^k + \lambda \frac{F_i^k}{\|\boldsymbol{F_i}\|} p_i^k, & F_i^k \le 0 \end{cases}$$

Illustrative example that describes movements of EM points whose charges and forces are calculated is shown in Picture 5 (C).

Overall structure of the EM within any problem domain is like previously described. However, EMs aimed at solving different problems differs in EM point representation, in calculating objective function for EM points, and in values of the parameters used during the execution of the method.

3.2. VARIABLE NEIGHBORHOOD SEARCH (VNS)

Variable Neighborhood Search (VNS) method is a robust metaheuristic introduced by Mladenović and Hansen [11]. The main searching principle of a VNS is based on the empirical evidences: (a) multiple local optima are correlated in some sense (usually close to each other) and (b) a local optimum found in one neighborhood structure is not necessarily a local optimum for some other neighborhood structure.

The overall structure of the VNS algorithm [1] is shown on the Picture 6.

The input of the VNS algorithm, consists of:

- n_{min} and n_{max} — minimal and maximal VNS neighborhood structure size;
- it_{max}, $itrep_{max}$, t_{max} — maximal number of total iterations, maximal number of iterations without improvement, and maximal execution time in seconds, respectively;
- $prob$ — probability to move to the other solution of the same quality;
- k- size of the neighborhood (integer value);

```
input: n_min, n_max, it_max, itrep_max, t_max, prob, k
output: x

x ← initializeSolution();
n ← n_min;
it ← 1;
while it < it_max ∧ (it − it_lastimpr) < itrep_max ∧ t_run < t_max do
    x′ ← shaking(x,n);
    x″ ← localSearch(x′, k);
    move ← shouldMove(x, x″, prob);
    if move then
        x ← x″ ;
    else if n < n_max then
        n ← n + 1;
    else
        n ← n_min;
    it ← it + 1;
end
return x;
```

Picture 6 — Outline of the VNS method

The main loop VNS algorithm usually imposes three main procedures: shaking, local search (LS) and neighborhood change.

o Shaking — in order to escape local suboptimal solutions, a new solution within a parametrized neighborhood of the current best solution is generated.

o Local search — starting from the new solution obtained in the previous step, other possible solutions within local neighborhood are systematically examined with the aim of finding the local optimum.

o Neighborhood change — depending on the success of the previous two procedures, the current neighborhood size is adjusted. More precisely, when the current best solution is changed, neighborhood size is reduced to minimal, otherwise it is cyclically increased by 1 (cycle ends at maximal neighborhood size).

In the algorithm presented on Picture 6, previously described procedures are iteratively called, until no further improvements of the best solution can be made inside the current neighborhood. When that situation appears, the algorithm steps into the next neighborhood. When the last neighborhood n_{max} is explored, the search restarts at the first neighborhood n_{min}.

The execution of the VNS is stopped when either of the following conditions becomes satisfied: a maximum number of iterations is reached, a maximum number of iterations without any improvement of the current best solution is reached, or a maximum allowed execution time is reached.

Design of the VNS for optimizing specific problem requires that solution representation, shaking procedure and LS procedure should be defined in the way that is the most efficient for that specific problem. Of course, parameters that governs VNS search process differs from problem to problem. However, overall structure of the VNS method is the one described above.

4. SOLUTIONS AND OBTAINED RESULTS

4.1. EM FOR DIMENSIONALITY REDUCTION PROBLEM

In order to achieve desired results, proposed EM metaheuristic method have the custom representation, custom shaking and custom LS. Detailed elaboration of the proposed method, computational experiments that are designed and executed, and the analysis of the obtained results (which indicate that proposed method outperforms other method on the well-known problem-instance set) is given in [6]. The same data are publicly available in the author's GitHub repository [12] in the supplemental data section.

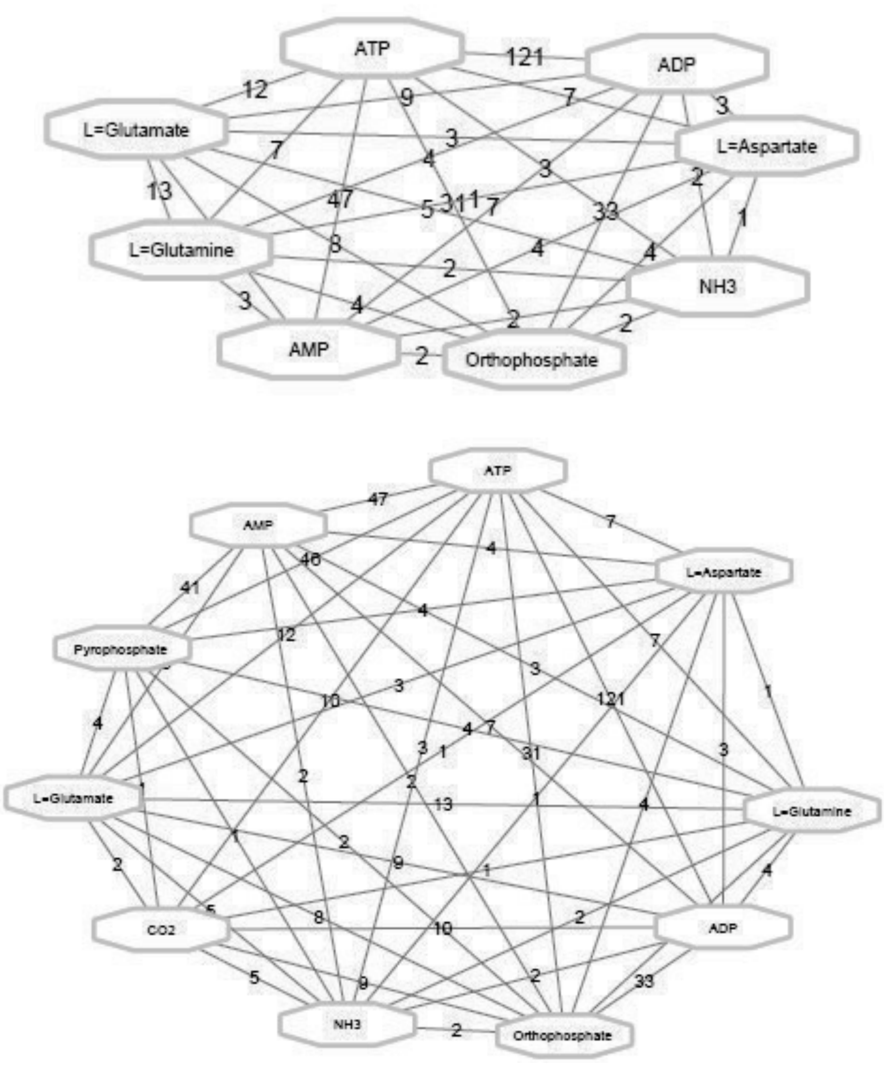
4.2. EM FOR MAXIMUM BETWEENESS PROBLEM

EM metaheuristic designed to solve Maximum Betweeness Problem have specific the EM representation and LS designed specifically for that problem. Detailed elaboration of the designed method, executed computational experiments and obtained results is given in [1]. Obtained results that indicate superiority of the designed method to other state-of-the art counterparts are publicly available as supplemental data in GitHub repository of the author [12].

4.3. VNS FOR MAXIMUM EDGE-WEIGHT K-PLEX PARTITION PROBLEM

Detailed elaboration of the proposed method specification, executed computational experiments and obtained results is given in [2]. Results from [2] clearly indicate that proposing of the newly designed method to this problem is justified.

Moreover, the relaxation of the clustering requirements lead to more useful information from biological point of view. Among many metabolic



Picture 7 — Largest clusters obtained for k=1 and k=2

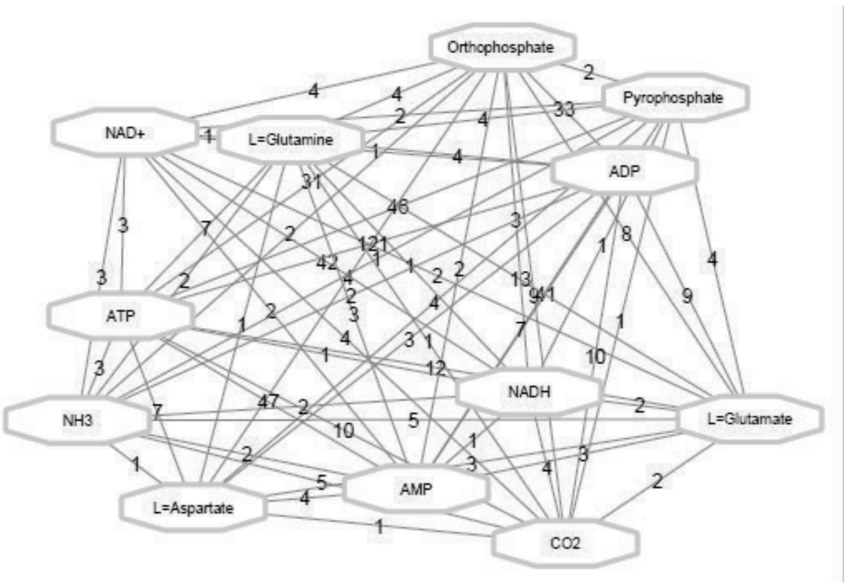
processes that appeared in various k-plexes obtained by the proposed VNS algorithm, in paper [2], the following processes are discussed: amino acid degradation process, fatty acids synthesis, vitamin B6 synthesis, oxidation of the succinate to the fumarate and formaldehyde oxidation. In this section, we will set focus to amino acid degradation process — which is one of the most important processes in metabolism.

In order to confirm the reliability of the obtained results, particular information of the biochemical pathways of considered organism Saccharomyces cerevisiae are checked and confirmed with the data presented in Yeast Pathways Database.

Ammonia presented in the organism is used as a resource of nitrogen for amino-acid synthesis and if it released in larger quantity, it must be removed because of its toxicity. In the considered organism Saccharomyces cerevisiae, ammonia can be incorporated into the amino group of glutamate, by two pathways: the reductive amination of 2-ketoglutarate, catalyzed by glutamate dehydrogenase where NADPH serves as the source of electrons, or by the ATP-dependent synthesis of glutamine from glutamate and ammonia catalyzed by glutamine synthetase. Picture 7 visualize the results obtained by proposed method, for $k \in \{1,2\}$.

The first cluster in the left side of Picture 7 is a clique with 8 vertices and contains the main intermediates which figure in ammonia synthesis from glutamic and aspartic acids. On the right side of Picture 7 is shown the largest cluster obtained for $k = 2$. A wider set of intermediates is now shown, also including additional reactions.

More detailed graphical interpretation is shown in Picture 8, obtained by the proposed method for $k = 3$.



Picture 8 — Largest clusters obtained for k=3

Since the condition for forming clusters is now more relaxed, more intermediates figures in the cluster. In addition to the previous ones, in the cluster shown in Picture 8 we see the process of the oxidative deamination.

By this process, two toxic products are synthesized: hydrogen peroxide and ammonia. In Picture 8 we see that the method grouped all these intermediates in one cluster, which was not the case with the more strict conditions (cases $k = 1$ and $k = 2$).

Last, but not the least, data and results are the publicly available in the author's GitHub repository [12], in the supplemental data section.

5. TOPOLOGY SENSITIVE METAHEURISTIC OPTIMIZATION METHODS

The main motivation for integrating topology and metaheuristics comes from the notion that metaheuristics might use the topological regularities inside the solution space to better maneuver through it. This can become especially useful when the solution space becomes extremely large. In such situation, classical metaheuristics might use too much resources in order to search the solution space. Although this sounds like it could lead to premature convergence to local optima, we stress that our conceptual design essentially generalizes and encompasses the classical metaheuristic algorithms.

This means that the proposed metaheuristics, during its execution, gradually converge to its classical variants. Also, by imposing adequate parameters, these topologically sensitive metaheuristics can be used as a classical through its whole execution.

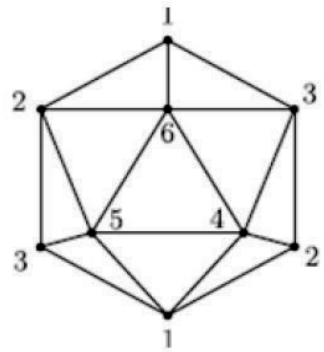
Execution of proposed topologically sensitive metaheuristics will resemble execution of any other metaheuristic: each execution creates a path in fitness landscape in order to reach global optima and avoids local optima. Therefore, fitness landscape analysis, which includes analysis of local optima positions, is very important for design of such metaheuristics. In other words, if some topological regularity in fitness landscape is detected, that regularity can be exploited and used for designing metaheuristic that will perform better than the alternatives. Topology-based models and techniques already achieved good results in revealing hidden structures and detecting new regularities [13], so it can be expected that it will be helpful in this domain.

In-depth discussion of the concepts described in this section are given in the paper [14].

The most important topology (more precisely, algebraic topology) concepts in this domain are simplicial complexes, homology groups and persistent homologies.

Simplicial complexes are combinatorial objects (abstract schemes) used from the early days of algebraic (combinatorial) topology as a bookkeeping device for triangulations of geometric objects. Conversely, they are used in the opposite direction for geometric presentation (visualization, geometric analysis and quantification) of the information (databases, point clouds) of any kind (not necessarily of geometric origin). The importance and versatility of simplicial complexes is illustrated by the fact that they appear under different names and in disguise in different areas of science, in and outside of mathematics. In cooperative game theory they are known as „simple games“ (after John von Neumann and Oskar Morgenstern). A similar use is in social choice theory (reliability theory). We meet them as threshold complexes (of „short sets“), both in weighted voting games and in the geometry of configuration spaces of polygonal linkages (protein chains). Closely related concepts are monotone hypergraphs, monotone Boolean functions, finite partially ordered sets, etc.

For illustration, the hemi-icosahedron on Picture 9, triangulates the real projective plane and can be used for a combinatorial analysis of this object (homology calculation, non-embeddability in the 3-space, etc.). On the other hand, it provides an important example of a cooperative voting scheme (simple game) for six persons (parties), with 10 winning and 10 losing, 3-element coalitions, which is not realizable as a weighted voting scheme.



Picture 9 — hemi-icosahedron

Simplicial complexes provide historically the first foundation for the theory of homology groups, which capture the idea of higher (dis)connectivity (voids, holes) in geometric object. For example, the edge path 1–6–4–1 surrounds an essential 1-hole in the hemi-icosahedron, while if we traverse this edge-path twice, and perturb it to the edge-path 1–6-4–2-5–

1 (in the same homology class), we obtain a trivial 1-cycle (illustrating the torsion phenomenon in homology groups).

The so called persistent homology (initiated by A. Zamorodian, H. Edelsbrunner, G. Carlsson, and others) opened a new chapter of applications of homological methods and created a large part of what is today known as applied and computational algebraic topology. The importance of persistent homology is in its applicability to dynamical simplicial complexes (complexes depending on a parameter) which arise in the analysis of large databases (large finite metric spaces, point clouds, sparse matrices). Information collected from biological systems is typically less structured, and can be organized, via a concept of clustering into a dynamical (filtered) simplicial complex.

In order to show that topological enhancement can be done in general case (in top-down manner), two complementary techniques are selected: the first (VNS) is single-solution based and discrete coded metaheuristic, while the second (EM) is population-based and real coded metaheuristic. In paper [14], the conceptual design of the two topologically sensitive metaheuristics is presented: Topologically Sensitive Electromagnetism metaheuristics (TEM) and Topologically Sensitive Variable Neighborhood Search (TVNS).

5.1. TOPOLOGICALLY SENSITIVE ELECTROMAGNETISM-LIKE METAHEURISTICS

TEM is designed as a generalization of EM that builds on m-simplex data, where for special case $m = 0$ that algorithm becomes a classical VNS. The main difference in TEM, in comparison to classical EM, is in the movement step. For each solution point, within TEM, we try to find new solution position inside the solution space that will form a m-simplex with other m solution points from the current population (two variants: with or without that solution considered as a candidate). This should be done in such a way that average or maximal distance among 0-simplices is minimized.

More precisely, process starts with the most restrictive m-simplex, i. e. $m = m_{max}$. As mentioned in the description of the EM algorithm, the movement is controlled partially by forces that affect the solution point and partially by the randomness. In TEM, the movement is also controlled by forces, but now the randomness is restricted with respect to parameter m. This means that for $m > 0$, the set of possible positions from which the new position is randomly chosen now becomes smaller, i. e. it becomes the subset of the set of possible positions where $m = 0$ (classical EM).

If, for a given solution point and current simplex size m, the movement is not possible, the simplex size is reduced by 1. Finally, if m reaches 0, that basically means that algorithm fell back to its classical version where every movement within distance threshold is allowed.

Note that this process of reducing the simplex size is done for each solution point separately. Also, note that the distance function threshold for deciding whether two 0-simplices are connected can be implicitly set by an algorithm to a sufficiently large value which will always allow a movement in the classical fallback EM.

The overall effect that we expect TEM will have on the search process in comparison to EM is increased preservation of the same or similar topological regularity through time (if this regularity exists). We also believe that the expansion of already existing simplices, especially large ones, is well motivated. This is because the existence of regular formation of local optima itself is an indicator that more new local optima may be found around that formation. Another important observation is that since TEM falls back to classical EM, we can expect that TEM will be generally applicable, i. e. if the topological regularity is low and cannot be exploited, TEM should work at least as good as classical EM (though performance might get deteriorated). Similar observation can be made for TVNS algorithm as well.

5.2. TOPOLOGICALLY SENSITIVE VARIABLE NEIGHBORHOOD SEARCH

TVNS is essentially conceived as a generalization of VNS that builds on m-simplex data (with special case $m = 0$ being a classical VNS). We will also sometimes refer to $(m + 1)$-simplex neighborhood which corresponds to collection of all valid simplices that can be formed by adding 0 -simplex to observed m-simplex. Therefore, 1-simplex neighborhood correspond to classical VNS, while m-simplex neighborhoods where $m > 0$ refer to its topological generalizations.

The main loop of TVNS should be made in such a way that the sequence of neighborhood structures, that are now parametrized by m and k, starts with the most restrictive neighborhood and after that proceeds with the sequence of more relaxed ones. Therefore, the neighborhoods will start with smallest neighborhood size $k = k_{min}$ and the most restrictive simplex structure $m = m_{max}$, and further proceed with reduction of m by 1. When

m reaches 0, it basically means that classical VNS algorithm is to be performed. After that, the k is increased by 1 and m is reset to m_{max}. The full cycle through neighborhoods is done when k reaches k_{max} and m reaches 0. If, at some moment, current solution is improved, both k and m are reset to its initial values.

6. CONCLUSION

Two metaheuristic optimization methods aimed at solving specific problems in bioinformatics and machine learning are described and the obtained results are analyzed. Topological enhancement for those methods are proposed.

Further research will be focused on theoretical characteristics of the proposed enchantments, on design and execution of computational experiments.

BIBLIOGRAPHY

[1] V. Filipović, A. Kartelj and D. Matić, „An electromagnetism metaheuristic for solving the Maximum Betweenness Problem," *Applied Soft Computing,* vol. 13, no. 2, pp. 1303–1313, 2013.

[2] M. Grbić, A. Kartelj, S. Janković, D. Matić and V. Filipović, „ Variable Neighborhood Search for Partitioning Sparse Biological Networks into the Maximum Edge-Weighted k-Plexes," *IEEE/ACM Transactions on Computational Biology and Bioinformatics,* vol. 17, no. 5, pp. 1822–1831, 2019.

[3] M. Grbić, D. Matić, A. Kartelj, S. Vračević and V. Filipović, „A three-phase method for identifying functionally related protein groups in weighted PPI networks," *Computational Biology and Chemistry,* vol. 86, 2020.

[4] P. Pardalos and E. Romeijn, Handbook of Optimization in Medicine, New York: Springer, 2009.

[5] V. Filipović, „Optimization, classification and dimensionality reduction in biomedicine and bioinformatics," *Biologia Serbica,* vol. 39, no. 1, pp. 83–98, 2017.

[6] A. Kartelj, „An Improved Electromagnetism-like Method for Feature Selection," *Journal of Multiple-Valued Logic and Soft Computing,* vol. 25, no. 2, pp. 169–187, 2015.

[7] I. Birbil and S. Fang, „An Electromagnetism-like Mechanism for Global Optimization," *Journal of Global Optimization,* vol. 25, pp. 263–282, 2003.

[8] C. Su and H. Lin, „Applying electromagnetism-like mechanism for feature selection," *Information Sciences,* vol. 181, no. 5, pp. 972–986, 2011.

[9] V. Filipovic, „An Electromagnetism Metaheuristic for the Uncapacitated Multiple Allocation Hub Location Problem," *Serdica Journal of Computing,* vol. 5, no. 3, pp. 261–272, 2011.

[10] A. Kartelj, N. Mitić, V. Filipović and D. Tošić, „Electromagnetism-like Algorithm for Support Vector Machine Parameter Tuning," *Soft Computing,* vol. 18, no. 10, pp. 1985–1998, 2013.
[11] N. Mladenović and P. Hansen, „Variable neighbourhood search," *Computers & Operational Research,* vol. 24, pp. 1097–1100, 1997.
[12] V. Filipović, „Suplemental info," 22 9 2022. [Online]. Available: https: // github. com/vladofilipovic/documents-science-public/tree/main/conferences/canu-2022/suplemental. [Accessed 22 9 2022].
[13] M. Reimann, M. Nolte, M. Scolamiero, K. Turner, R. Perin, G. Chindemi, P. Dotko, R. Levi, K. Hess and H. Makram, „Cliques of Neurons Bound into Cavities Provide a Missing Link between Structure and Function," *Frontiers in Computational Neuroscience,* vol. 11, 2017.
[14] A. Kartelj, V. Filipović, S. Vrećica and R. Živaljević, „Topologically sensitive metaheuristics," arhiv, 2020.

PRIRODNE NAUKE

NATURAL SCIENCES

Ivan BOŽOVIĆ*

ALCHEMY OF XXI CENTURY: DIGITAL SYNTHESIS OF QUANTUM MATERIALS

Abstract: Atomic-layer-by-layer molecular beam epitaxy (ALL-MBE) is a new technique developed in the last few decades to synthesize functional quantum materials, including high-temperature superconductors, other complex oxides, and two-dimensional materials such as graphene and borophene. It even enables one to synthesize novel metastable materials that cannot be produced by standard methods. Several examples are presented of ALL-MBE alchemy — the creation of artificial materials with novel and unique electronic properties. The ability to engineer the materials atasingle-atomicmonolayerlevelenabled important discoveries, further illustrating the power of ALL-MBE.

Key words: *Atomic layer-by-layer Molecular beam epitaxy, Digital synthesis, Artificial metastable materials*

1. INTRODUCTION

Ancient alchemists attempted to purify and perfect certain materials, and I will use the word „alchemy" in this sense. (The attempts to transmute elements, discover the panacea, and the philosopher's stone will be ignored.) The materials of our focal interest will be high-temperature superconductors (HTS) cuprates, but I will also mention other complex oxides and two-dimensional (2D) quantum materials such as graphene and borophene. As for the synthesis, purification, and perfection technique, the focus will be on Atomic-layer-by-layer molecular beam epitaxy (ALL-MBE), which I had the privilege of developing, in parallel with several other colleagues and collaborators, over the last three decades. In some cases, ALL-MBE has even

* Ivan Božović, Brookhaven National Laboratory, Condensed Matter Physics and Materials Science Division, Upton, New York 11973–5000, USA; Yale University, Department of Chemistry, New Haven CT 06520, USA

enabled the creation of metastable novel artificial materials that the standard thermodynamic synthesis methods cannot obtain.

The paper is organized as follows; In Section 2, I describe our hardware — three ALL-MBE systems, each unique and custom-designed for a different purpose. Section 3 briefly describes the know-how — the technique of using ALL-MBE for materials synthesis, heterostructure engineering, and device fabrication. Section 4 is central to our topic, presenting concrete examples of ALL-MBE alchemy — the creation of several new, artificial metastable materials with novel and unique electronic properties. Section 5 briefly mentions several important discoveries enabled by ALL-MBE alchemy and notably by its ability to custom-engineer the materials and samples for the needs of a particular physics experiment. Conclusions are given in Section 6.

2. THE HARDWARE: ATOMIC LAYER-BY-LAYER MOLECULAR BEAM EPITAXY (ALL-MBE)

MBE-1 system at BNL. Atomic layer-by-layer (or „digital") synthesis of novel quantum materials requires specialized and sophisticated hardware. An ALL-MBE system at Brookhaven National Laboratory, illustrated in Figures 1 and 2, is a prime example. [1] It is a cluster of several interconnected ultra-high vacuum (UHV) chambers distributed between two neighboring laboratories. The main growth chamber (Figure 1) is equipped with the following tools: (i) a 6-degree-of-freedom sample manipulator, (ii) a diode-laser heater, (iii) 16 metal atom sources, mostly thermal effusion (Knudsen) cells, but also including several rod-fed electron-beam evaporators used for depositing refractory (high-melting point) metals, (iv) a source of pure ozone gas, (v) a scanning quartz-crystal monitor for measuring deposition rates as a function of position, before and after the film growth, (vi) a 16-channel atomic absorption spectroscopy system for measuring the deposition rates in real-time, (vii) a 20-channel scanning reflection high-energy electron diffraction (RHEED) system, and (viii) a time-of-flight ion scattering and recoil spectroscopy (TOF-ISARS) system for measuring the surface chemical composition.

A pure ozone beam provides sufficient oxidation under high vacuum conditions, which permits in-situ monitoring of the surface structure by RHEED and TOF-ISARS.

The main growth chamber is highly modular. It is equipped with 16 spools ('arms'), each of which houses a metal atom source. Each arm has a gate valve and autonomous pumping using a dedicated small (80 1/s) turbo-molecular pump and a roughing line. It is thus possible to valve off each

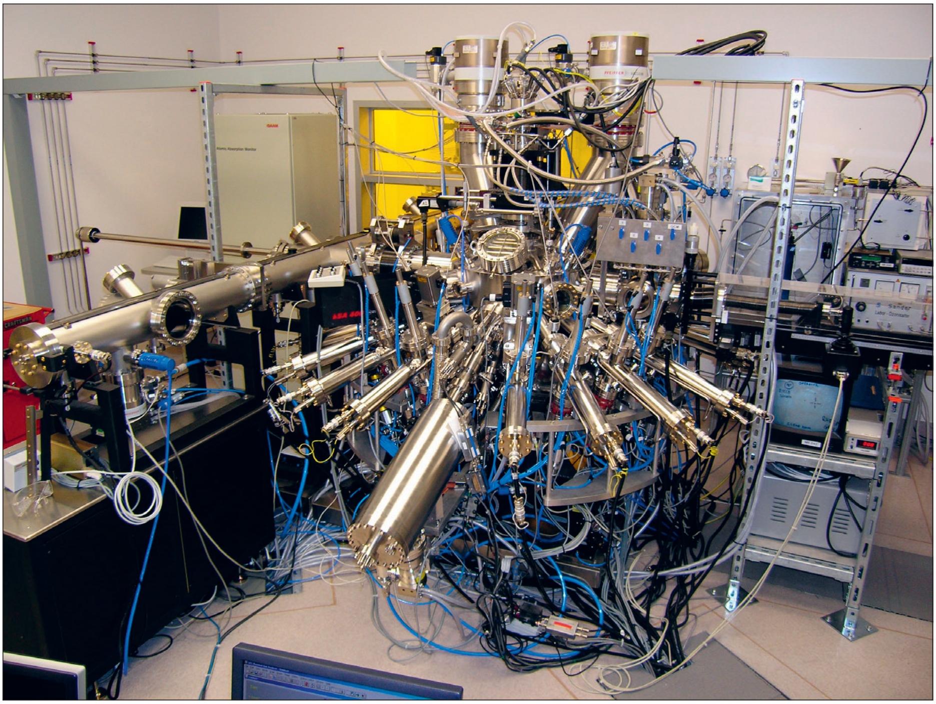


Figure 1. BNL oxide-MBE system: the synthesis chamber

arm and to recharge, service, interchange, replace, or outgas each source, within minutes, without breaking the vacuum in the main chamber, even while a film is being grown.

Various metal sources are used: low-temperature, high-temperature, and double-filament thermal effusion cells, as well as rod-fed electron-beam sources, for refractory metals. All the sources are easily interchangeable. Altogether, we can access most elemental metals in the periodic system with this range of sources.

The controlling software allows us to track the temperature of all the sources, the ozone pressure, the substrate position and orientation, the deposition rates from each source, etc., and to control or modify the stoichiometry within each monolayer.

The processing chamber (Fig. 2) is installed in the clean room. This way, the substrates can be prepared in a class-100 clean environment and loaded into the system without surface contamination. This chamber contains several tools for critical lithographic operations: a 5-pocket electron-beam evaporator used to deposit metallization and insulation layers, an oxygen plasma source for burning photoresist residue and cleaning the surface, a large-diameter (2”) ion-beam source for ion-milling, and a secondary-ion

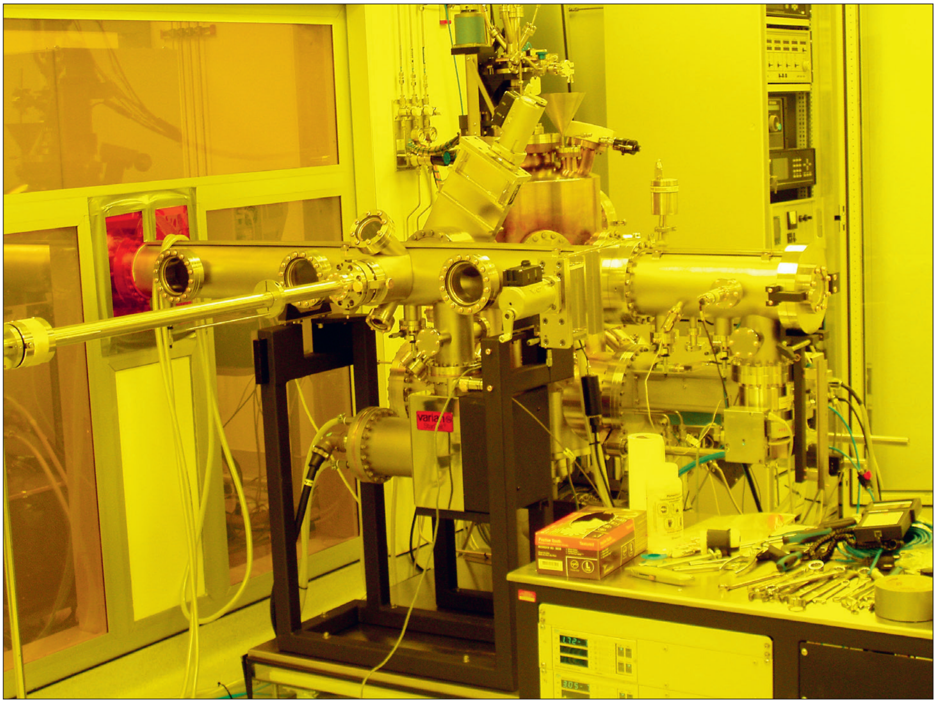


Figure 2. BNL oxide-MBE system: the chamber for lithographic processing

mass spectrometer (SIMS) enabling atomically precise control of the depth of ion-milling. The processing chamber allows the key micro-fabrication steps (metallization, insulation, surface cleaning and oxidation, and ion-milling) to be performed without breaking the vacuum.

The growth and processing chambers are connected via a 20' long UHV transfer chamber. Thus, the samples can be shuttled between the two chambers for multiple process steps without breaking the vacuum. This capability is critical, e. g., for the fabrication of sandwich junctions with superconducting contacts deposited to the top.

MBE2: OASIS system at BNL. Our second ALL-MBE system is a part of the OASIS facility at BNL. [2, 3] This new multi-module system consists of an Oxide molecular beam epitaxy module, Angle-resolved photoemission spectroscopy (ARPES) module, and a Spectroscopic Imaging Scanning tunneling microscopy (SI-STM) module. All these modules are interconnected under UHV; we can thus grow a film and transfer it to the ARPES or the STM module without breaking the vacuum and contaminating the surface.

This OASIS MBE module is conceptually akin to the ALL-MBE system described above, but there are some variations. The OASIS system is smaller and more compact, with just eight elemental sources (plus a source

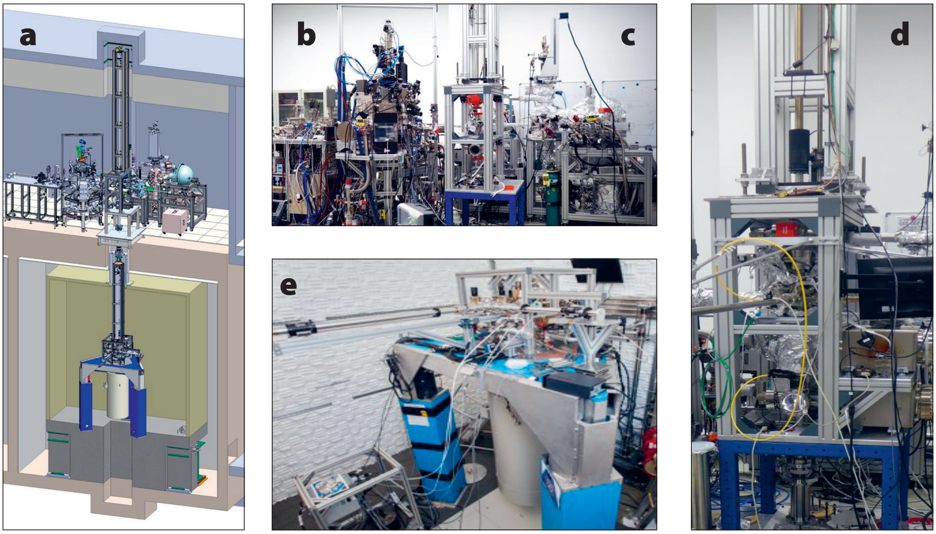


Figure 3. a, Schematics of the OASIS system at Brookhaven National Laboratory. b, the Oxide MBE module. c, the ARPES module. d, the horizontal and vertical transfer station. e, the SI-STM module.

of pure ozone gas). The sample holder and the transfer systems are very different, custom-designed to accommodate the conflicting demands of oxide MBE (high temperature, low thermal contact to the chamber, high oxidation power, horizontal transfer), STM (rigid contact to reduce vibrations, vertical transfer, cryogenic temperature), ARPES (precise sample rotation, cryogenic temperature, sample electrically grounded). To maintain UHV, the sample heater is an infrared source placed outside the MBE chamber, and a 100 cm long quartz crystal rod is used as a waveguide.

Traditionally, APRES and STM studies of HTS cuprates and other complex oxides were limited to materials that can be readily cleaved under UHV. The OASIS system is a solution for this problem, greatly expanding the list of materials amenable to ARPES and STM studies to include any material that can be synthesized by oxide MBE. Moreover, one can control and choose the surface termination layer and use heterostructures to study the effects of the epitaxial strain, various proximity effects, etc.

MBE3: The LEEM/MBE/multiprobe system at Yale. To synthesize and characterize novel quantum materials, including borophene, we have put together at Yale a unique system with a small MBE chamber attached to two additional modules for sample characterization. One is a low-energy electron microscope (LEEM), providing real-space microscopic images of the surface, with a resolution better than 10 nm, in real-time and at a

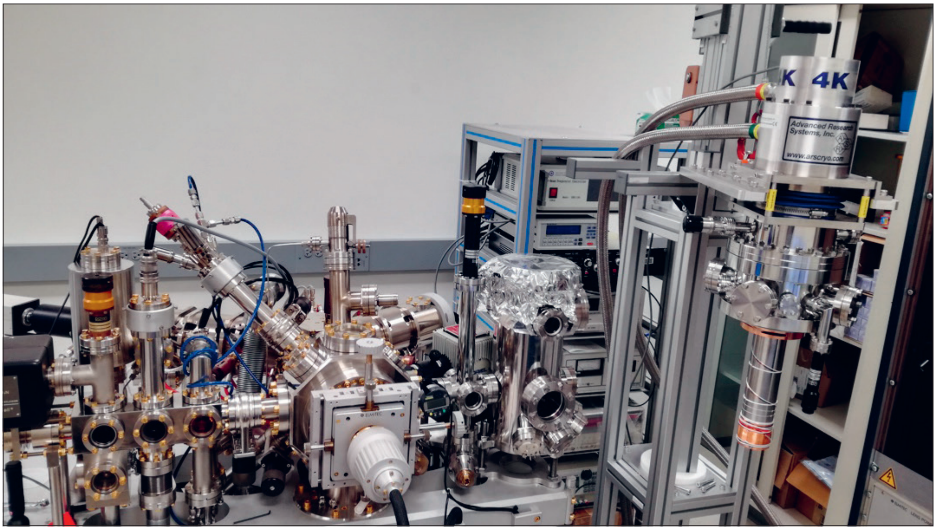


Figure 4. A system at Yale University that integrates MBE, LEEM/LEED/PEEM, and a multiprobe chamber for electrical transport and magnetic susceptibility measurements.

video rate. LEEM enables one to observe, study, and quantify the film nucleation and growth and identify the growth mechanism(s), such as island nucleation or step-flow growth. The same instrument can also be operated in the electron diffraction (LEED) mode, providing inverse-space information about the crystalline structure and order. In the photo-emission electron microscopy (PEEM) mode, the incoming electron beam is switched off and replaced by an ultraviolet light source, which is also implemented in our system. [4, 5]

The third module is a UHV chamber for in-situ cryogenic measurements of resistivity and magnetic susceptibility. All three modules are interconnected under UHV, so we can synthesize reactive and air-sensitive materials and study their fundamental electronic properties of interest without breaking the vacuum.

3. THE TECHNOLOGY: SYNTHESIS, HETEROSTRUCTURE ENGINEERING, AND DEVICE FABRICATION

Our advances in the MBE apparatus and technique enabled us to develop a technology for reproducible fabrication of atomically smooth films of cuprate superconductors and other related complex oxides. Over two decades, we synthesized well over 3000 thin films. These included single-crystal films of $La_{2-x}Sr_xCuO_4$ (with $0 \le x \le 0.45$), $Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+2}$ (with n

= 1, 2, 3, 4,...), $DyBa_2Cu_3O_{7-x}$, $BaBiO_3$, $La_{2-x}Sr_xNiO_4$, $LaAlO_3$, $LaSrAlO_4$, $La_{2-x}Sr_xZnO_4$, etc., as well as a large number of heterostructures, multilayers, and superlattices. [6–17] Some contain layers that are just one unit cell (1 UC) thick and have virtually atomically perfect interfaces with little or no cation interdiffusion. [18–20]

Each film was characterized in real-time by RHEED and ex-situ by atomic force microscopy (AFM) and magnetic susceptibility measurements. RHEED oscillations provide a digital count of atomic layers and real-time film quality control. Selected films were also characterized *in situ* by time-of-flight ion scattering and recoil spectroscopy and ex-situ by X-ray diffraction (XRD), transport measurements, and Rutherford backscattering. Additional characterization was done by our collaborators using atomic-resolution scanning transmission electron microscopy and electron energy loss spectroscopy, resonant elastic and inelastic synchrotron X-ray scattering, synchrotron-based X-ray phase-retrieval techniques such as coherent Bragg rod analysis, muon spin rotation, ultrafast electron diffraction, ultrafast optical and THz pump-probe techniques, second-harmonic generation spectroscopy, shot-noise measurements, scanning SQUID magnetometry, etc.

Thin films of LSCO synthesized by ALL-MBE are single crystals without any secondary phase precipitates and feature atomically smooth surfaces and interfaces; we have seen the RMS surface roughness as low as 1 Å. The films can be made ultrathin, down to a single unit cell thick, without any pinholes; this is advantageous for transmission measurements. Much thicker films can also be grown for reflectance measurements.

To alleviate the problem of oxygen non-uniformity, we have performed over a thousand experiments that involved annealing in ozone, oxygen, or vacuum, spanning 13 orders of magnitude in pressure, from $10-^{8}$ Torr to 200 atm. Before and after each annealing step, the films were characterized by AFM, transport, and XRD. We have developed recipes that involve multiple annealing steps at different temperatures and pressures, yielding the most homogeneous films with the sharpest superconducting transitions based on these extensive studies.

Altogether, we have developed a reliable and robust technology for synthesizing thin films of high-T_c superconductors and other complex oxides, including various heterostructures and superlattices. We have also developed a large spectrum of device designs, lithographic masks, and well-tested recipes for the fabrication of electronic devices, including SNS and SIS tunnel junctions, SQUIDS, nanowires, nano-rings, sunbeam patterns of Hall bars for angle-resolved transport measurements, and many others. [16, 21–23]

4. THE ALL-MBE ALCHEMY

Tunnel junctions, modulation doping. ALL-MBE synthesis provides excellent flexibility in switching from one compound to another and tailoring surfaces and interfaces. This opens new avenues in the fabrication of tunnel junctions. [16, 21, 22] Traditionally, the barrier layer is inserted between the two naturally terminated layers of the host material (the bottom and the top electrode). ALL-MBE makes it possible to insert a layer of a foreign compound in-between two „inner" planes of a host compound, which are not natural termination planes. The first HTS superconductor-insulator-superconductor (SIS) junctions were fabricated in this way. [21] Next, one can deposit just a fraction of a monolayer of a given atomic species and then complete the monolayer with another element. In this way, if there is no bulk diffusion, one can select the monolayers to be doped at will. The dopant may be picked to have a different valence, which enables one to control the local charge carrier density. Modulation doping in HTS materials has been accomplished in this manner. [19, 21]

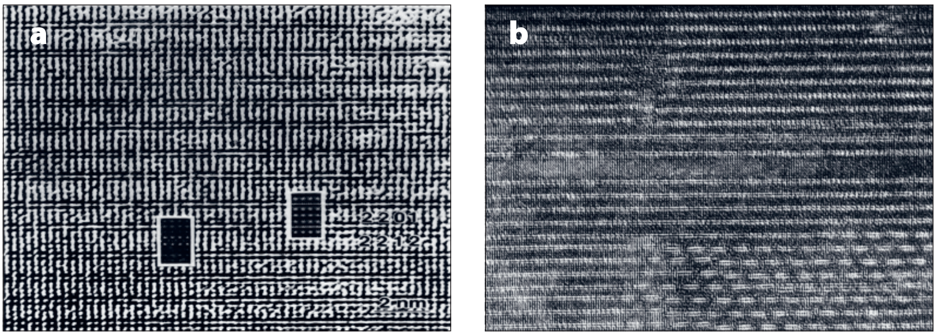


Figure5.a,$Bi_2Sr_2CuO_6$: $Bi_2Sr_2CaCu_2O_8$ superlattice. b, $Bi_2Sr_2Ca_7Cu_8O_{20}$: a novel artificial HTS meta-material. (Top and bottom electrodes: $Bi_2Sr_2CaCu_2O_8$)

Superlattices. ALL-MBE enabled synthesis of precise $[(\text{Bi-2201})_n + \text{Bi-2212}]_m$ superlattices, with $n = 1, 2, ..., 10$. [24] The cross-section TEM image in Fig. 5a shows essentially perfect Bi-2201 / Bi-2212 interfaces. In-plane transport measurements showed that in such superlattices, T_c was equal to that in thick single-phase Bi-2212 films, independent of the number of the spacer Bi-2201 layers inserted between one-half unit cell (15.4 Å) thick Bi-2212 monolayers. This indicated that Bi-2212 monolayers were well connected and continuous (in-plane) on a macroscopic length scale. The same observation demonstrated that the HTS phenomenon is essentially quasi-2D in nature. We have also seen undiminished T_c in superlattices with one-half unit cell (6.6 Å) thick LSCO layers spaced away by thicker, undoped, insulating LCO layers. [25]

Metastable phases. In layered oxides, one frequently finds several energetically nearly degenerate phases. A well-known example is provided by 2201, 2212, 2223, 2234, etc., phases in the $Bi_2Sr_2Ca_{n-1}Cu_nO_{2n+4}$ family. In general, by cooling the Bi-Sr-Ca-Cu-O melt, one obtains a mixture of all these phases for entropy reasons. Layer-by-layer growth enables one to discriminate among such phases, i. e., to selectively grow the desired one only, including the phases that do not exist as bulk equilibrium compounds. Unless the temperature is raised too much, the topological barrier protects the phase grown because phase separation would require extensive bulk diffusion. In this way, it has been possible to assemble metastable compounds such as Bi-2234, Bi-2245, and Bi-2278 phases (see Fig. 5b). These films were superconducting, with the highest T_c = 60 K in single-phase Bi-2278 films and T_c = 75 K in Bi-1278. These were the first 'engineered' high-T_c materials. [26]

Artificial compounds. By our digital synthesis technique, we can stack various atomic layers trying to create new, synthetic, metastable compounds. One example is $Ba_3LaBi_2O_9$, a layered bismuthate related to the perovskite $BaBiO_3$ (BBO). The latter is a charge-density wave (CDW) insulator, but upon doping with potassium, it becomes metallic and superconducts with T_c = 30 K. This is very high for a cubic material; T_c higher than that only occurs in 2D materials. The question is what T_c one could get by forcing BKBO to be 2D. Our attempts at d-doping with K failed because it is volatile; at the lowest substrate temperature sufficient to make the film crystalline, K migrates away and out of the film.

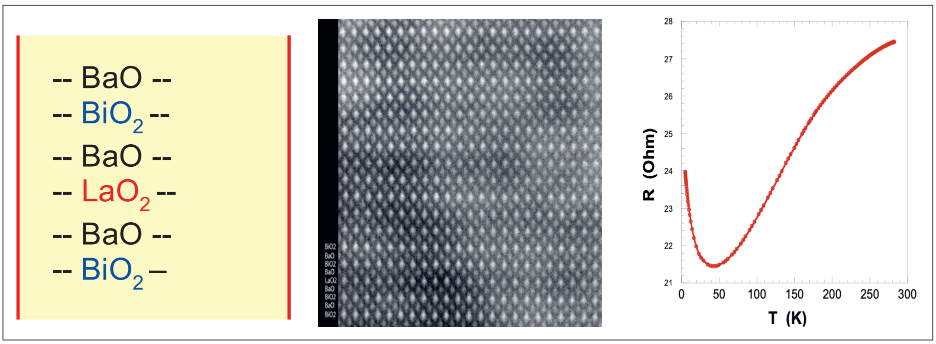


Figure 6. a., Schematic of $Ba_3LaBi_2O_9$ (BLBO). b, A cross-section image of BLBO film obtained by transmission electron microscopy. c, The high-temperature dependence of BLBO is metallic, but it develops a CDW at low temperatures.

For this reason, we tried d-doping with La3 +, the only cation with adequate size and charge. Replacing every third BiO_2 plane with a LaO_2 plane should provide one hole per two BiO_2 planes, making them metallic. The experiment was spectacularly successful insofar that we could synthesize this new, metastable compound that standard thermodynamic techniques cannot make. The films were atomically perfect and smooth. They were metallic but regrettably not superconducting so far because of the competing CDW instability. [27] In the future, we will try to suppress this CDW, and hopefully, superconductivity will emerge.

Borophene. Another spectacular example of new quantum alchemy is the discovery, after two centuries, of several new allotropes of boron, employing MBE synthesis. Generally, discoveries of new allotropes of elements are rare and hailed as milestone events in Chemistry. Take carbon as an example. Amorphous carbon has been known since prehistory. Diamond was known since 2000 BCE in China, although it was identified as carbon only in 1772 by Lavoisier. Seven years later, graphite was also identified as carbon by Schelle. Two centuries later, in 1985, fullerenes (buckyballs and carbon nanotubes) were discovered to be recognized with the Nobel Prize in 1996. Graphene was discovered next in 2004, earning another Nobel Prize in 2010.

Boron is the neighbor of carbon in the periodic system. While the borax glaze was used for pottery in China since at least the 3rd century, boron was identified as an element in 1808 by Humphrey Davy. Boron also forms several allotropes; one amorphous and four crystalline forms have been known for a long time. Two groups recently reported the first synthesis of a new allotrope, borophene — an atomically-thin, covalent-bonded boron sheet analogous to graphene. [28, 29] As the substrate, they used silver crystal that was polished to expose the (111) surface with hexagonal symmetry, like graphene. The drawback was that only nanometer-sized crystalline flakes of borophene were produced this way, precluding device fabrication and the measurements of interesting physical properties.

Using MBE synthesis empowered by real-time LEEM and LEED monitoring, we were able to grow, on copper substrates, monocrystal borophene islands of an area exceeding 100 square micrometers. [4, 5] More importantly, we discovered two new borophene structures, depending on whether we used the hexagonal Cu(111) or tetragonal Cu(100) facets. These new allotropes of boron captured much interest since they host electronic spectra and states with unusual topological properties and feature exotic electronic quasiparticles. Borophene bilayers were also fabricated recently. [30]

5. NEW PHYSICS ENABLED BY ALL-MBE

One distinguishing feature of our signature MBE design is complete system modularity. [1] Every system component — the main chamber, each atom source, RHEED, QCM, TOF-ISARS — is placed in a separate UHV compartment equipped with a gate valve and a turbo-molecular pump. The MBE-1 synthesis module thus hosts 24 turbo pumps. The great advantage of this design is that one component can be serviced without breaking the vacuum in the others; for example, an atomic source can be cooled down, vented, recharged, pumped down, and outgassed even while a film is being grown. This ensures a very high system up-time, essentially 100%, and excellent productivity. We have been routinely able to complete over 200 synthesis experiments per year.

This, in turn, enabled a focused and comprehensive study that spanned almost two decades and over three thousand HTS cuprate samples, probably without precedence in Condensed Matter Physics. We have measured the fundamental physical parameters of the normal and superconducting states and established their precise dependence on doping, temperature, strain, and external fields. This enormous database — several tens of thousands of data files — contains a wealth of information that has enabled critical new insights into the physics of cuprates. Our ability to engineer the material at the single-atomic monolayer level for the specific needs of a particular experiment has also played a crucial role. A brief list of some key discoveries is as follows.

An extraordinary insulating state. Undoped cuprates (the „parent“ compounds) are antiferromagnetic insulators. Upon doping, they become metallic and superconducting. For example, in $La_{2-x}Sr_xCuO_4$ this quantum phase transition occurs at $x = 0.06$. For $x < 0.06$, the material is insulating in the sense that its resistivity increases as the temperature is lowered. However, it does not behave as a standard semiconductor or insulator. The transport becomes erratic, hysteretic at very low temperatures, and shows memory effects. [31,32] These are signatures of an unusual quantum charge-cluster glass state. Moreover, the charge is quantized in the units of 2e (two electron charges) — electrons in this insulator are paired. [33]

An extraordinary metallic state. For $x > 0.06$, LSCO is metallic; its resistivity decreases as the temperature is lowered. But this is a „strange metal“; the dependences of the resistivity on the temperature and magnetic field are incompatible with the standard textbook description of metals as a gas of weakly-interacting quasiparticles. [15] Even more strange, the rotational symmetry of the electron fluid in the normal metallic state above T_c is always spontaneously broken — the so-called „electronic nematicity“. [14]

Moreover, magnetic excitations (paramagnons) indicate the vestiges of short-ranged, short-lived antiferromagnetic correlations at all doping levels. [34]

An extraordinary superconducting state. In standard superconductors, electrons pair and condense at some sharply-defined critical temperature T_c. In HTS cuprates, including both BSCCO and LSCO, we detected electron pairing and vestiges of superconductivity (superconducting fluctuations) way above T_c and at energies way above the superconducting gap. [16, 17, 36] Next, when the doping level is increased above $x = 0.16$, the superfluid density decreases and drops to zero at $x = 0.26$ instead of increasing; this strange behavior is unexpected and, so far, unexplained. [13]

Extraordinary proximity effects. Heterostructures also brought in some big surprises. For example, in $La_{2-x}Sr_xCuO_4/La_2CuO_4$ multilayers, for $x < 0.16$, there is no charge transfer; unexpectedly, the HTS state and the antiferromagnetic state separate sharply on the Ångstrom scale. In stark contrast, in $La_{1.84}Sr_{0.16}CuO_4/La_{2-x}Sr_xCuO_4$ multilayers, for $x > 0.06$, supercurrent runs through non-superconducting barrier layers that are 100 thicker than what is conceivable based on the standard theory — the „Giant proximity effect", also unexplained so far. [37] Finally, in $La_{2-x}Sr_xCuO_4/La_2CuO_4$ multilayers, for $x > 0.16$, interface superconductivity is seen, with $T_c \approx 40$ K independent on the doping level x and confined to a one-atom-thick CuO_2 layer — the thinnest superconductor known. [18–20]

6. CONCLUSIONS

ALL-MBE of HTS cuprates and other complex oxides has gone a long way from its humble origins, over three decades ago, at the margins of industrial R&D. By now, it has taken center stage in the fundamental Condensed Matter Physics and Materials Science, enabling a series of important experiments and discoveries, a few of which were listed above. By the strength of this fact, coupling ARPES or STM with an MBE in one complex UHV system has become the norm in the community today, while just a decade ago, this was generally considered excessive if not technically impossible.

Apart from the fabrication of state-of-the-art single-crystal HTS films and heterostructures, ALL-MBE has enabled engineering materials at a single-atomic monolayer level. By stacking atomic layers at will, new, artificial, metastable quantum materials have been created. They can be fine-tuned by various tricks — epitaxial strain, epitaxial stabilization, heterostructure engineering, interface effects, proximity effects — to tailor the desired electronic properties, including some unprecedented and very exotic ones. New alchemy has been developed, and we have just touched the tip of the iceberg. We can say one thing about the future: it will surpass our imagination.

ACKNOWLEDGEMENT

The research at BNL was supported by the U. S. Department of Energy, Basic Energy Sciences, Materials Sciences and Engineering Division. The research at Yale was supported by the Gordon and Betty Moore Foundation's EPiQS Initiative through Grant GBMF9074.

REFERENCES

[1] I. Božović: „Atomic-Layer Engineering of Superconducting Oxides: Yesterday, Today, Tomorrow", IEEE Trans. Appl. Superconduct. 11, 2686–2695 (2001).
[2] Z. Wu, D. Putzky, A. K. Kundu, H. Li, S. Yang, Z. Du, S. H. Joo, J. Lee, Y. Zhu, G. Logvenov, B. Keimer, K. Fujita, T. Valla, I. Božović, and I. K. Drozdov: „A homogeneous superconducting gap in $DyBa_2Cu_3O_{7-\delta}$ synthesized by oxide molecular beam epitaxy", Phys. Rev. Materials 4, 124801 (2020).
[3] C. K. Kim, I. Drozdov, K. Fujita, J. C. Davis, I. Božović, and T. Valla: „In-situ angle-resolved photoemission spectroscopy of copper-oxide thin films synthesized by molecular beam epitaxy", Journal of Electron Spectroscopy and Related Phenomena 257, 146775 (2022).
[4] R. Wu, I. K. Drozdov, S. Eltinge, P. Zahl, S. Ismail-Beigi, I. Božović, and A. Gozar: „Large-area single-crystal sheets of borophene on Cu(111) surfaces", Nature Nanotechnology 14, 44–49 (2019).
[5] R. Wu, S. Eltinge, I. K. Drozdov,A. Gozar,P. Zahl,J.T.Sadowski,S. Ismail-Beigi,andI. Božović: „Borophene on Cu(100) surface", Nature Chemistry 14, 377 (2022).
[6] P. Abbamonte, L. Venema, A. Rusydi, G. A. Sawatzky, G. Logvenov, and I. Božović: „A Structural Probe of the Doped Holes in Superconducting Cuprates", Science 297, 581–584 (2002).
[7] N. Gedik, D.-S. Yang, G. Logvenov, I. Božović, and A. Zewail: „Non-equilibrium Phase Transitions in Cuprates Observed by Ultrafast Electron Crystallography." Science 316, 425 (2007).
[8] M. P. M. Dean, R. S. Springell, C. Monney, K. J. Zhou, I. Božović, J. Pereiro, B. Dalla Piazza, H. M. Ronow, J. van den Brink, T. Schmitt, and J. P. Hill: „Spin Excitations in a single La_2CuO_4 layer", Nature Materials 11, 850–4 (2012).
[9] D. H. Torchinsky, F. Mahmood, A. T. Bollinger, I. Božović, and N. Gedik: „Fluctuating charge density waves in a cuprate superconductor", Nature Materials 12, 387–391 (2013).
[10] I. Božović and C. Ahn: „A new frontier for superconductivity", Nature Physics 10, 892–895 (2014).
[11] P. D. C. King, H. I. Wei, Y. F. Nie, M. Uchida, C. Adamo, S. Zhu, X. He, I. Božović, D. G. Schlom, and K. M. Shen: „Atomic-scale control of competing electronic phases in ultrathin $LaNiO_3$", Nature Nanotechnology 9, 443–447 (2014).

[12] I. Božović: „High-temperature superconductivity: a conventional conundrum", Nature Physics 12, 22–24 (2016).
[13] I. Božović, J. Wu, X. He, and A. T. Bollinger: „Dependence of critical temperature in overdoped copper oxides on superfluid density", Nature 536, 309–311 (2016).
[14] J. Wu, A. T. Bollinger, X. He, and I. Božović: „Spontaneous breaking of rotational symmetry in copper oxide superconductors", Nature 547, 432–436 (2017).
[15] P. Giraldo-Gallo, J. A. Galvis, Z. Stegen, K. A. Modic, F. F Balakirev, J. B. Betts, X.-J. Lian, C. Moir, S. C. Riggs, J. Wu, A. T. Bollinger, X. He, I. Božović, B. J. Ramshaw, R. D. McDonald, G. S. Boebinger, and A. Shekhter: „Scale invariant magnetoresistance in the strange metal phase of a cuprate superconductor", Science 361, 479–481 (2018).
[16] P. Zhou, L. Chen, Y. Liu, I. Sochnikov, A. T. Bollinger, M.-G. Han, Y. Zhu, X. He, I. Božović, and D. Natelson: „Electron pairing in the pseudogap state revealed by shot noise in copper-oxide junctions", Nature 572, 493–496 (2019).
[17] I. Božović and J. Levy: „Pre-Formed Pairs in Copper Oxides and $LaAlO_3/SrTiO_3$", Nature Physics 16, 712–717 (2020).
[18] A. Gozar, G. Logvenov, L. Fitting Kourkoutis, A. T. Bollinger, L. A. Giannuzzi, D. A. Muller, and I. Božović: „Interface superconductivity between a metal and a Mott insulator", Nature 455, 782–785 (2008).
[19] G. Logvenov, A. Gozar, and I. Božović: „High-temperature superconductivity in a single copper-oxygen plane", Science 326, 699–702 (2009).
[20] J. Wu, O. Pelleg, G. Logvenov, A. T. Bollinger, Y. Sun, G. S. Boebinger, M. Vanević, Z. Radović, and I. Božović: „Anomalous independence of interface superconductivity on carrier density", Nature Materials 12, 877–881 (2013).
[21] I. Božović and J. N. Eckstein: „Atomic-layer fabrication of HTS tunnel junctions", J. of Alloys & Compounds 251, 201–205 (1997).
[22] I. Božović, G. Logvenov, M. A. J. Verhoeven, P. Caputo, E. Goldobin, and T. H. Geballe: „No mixing of superconductivity and anti-ferromagnetism in a high-critical-temperature superconductor", Nature 422, 873–875 (2003).
[23] I. Sochnikov, A. Shaulov, Y. Yeshurun, G. Logvenov, and I. Božović: „Large oscillations of the magnetoresistance in nanopatterned high-temperature superconducting films", Nature Nanotechnology 5, 516–519 (2010).
[24] I. Božović, J. N. Eckstein, M. E. Klausmeier-Brown, and G. F. Virshup: „Superconductivity in Epitaxial $Bi_2Sr_2CuO_6$ / $Bi_2Sr_2CaCu_2O_8$ Superlattices: The Superconducting Behavior of Ultrathin Cuprate Slabs", J. Supercond. 5, 19–23 (1992).
[25] S. Smadici, J. C. T. Lee, S. Wang, P. Abbamonte, A. Gozar, G. Logvenov, C. Deville Cavellin, and I. Božović: „Superconducting Transition at 38 K in Insulating-Overdoped La_2CuO_4-$La_{1.64}Sr_{0.36}CuO_4$ Superlattices: Evidence for Interface Electronic Redistribution from Resonant Soft X-Ray Scattering", Phys. Rev. Letters 102, 107004 (2009).
[26] I. Božović, J. N. Eckstein and G. F. Virshup: „Superconducting Oxide Multilayers and Superlattices: Physics, Chemistry, and Nano-engineering", Physica C 235–240, 178–181 (1994).
[27] X. He, A. T. Bollinger, M.-G. Han, Y. Zhu, and I. Božović, „Search for superconductivity in $BaBiO_3$-based heterostructures": J. Supercond. Nov. Magn. 33, 241–248 (2020).

[28] A. J. Mannix, X.-F. Zhou, B. Kiraly, J. D. Wood, D. Alducin, B. D. Myers, X. Liu, B. L. Fisher, U. Santiago, J. R. Guest, M. J. Yacaman, A. Ponce, A. R. Oganov, M. C. Hersam, and N. P. Guisinger: „Synthesis of borophenes: Anisotropic, two-dimensional boron polymorphs“, Science 350, 1513–1516 (2015). \
[29] B. Feng, J. Zhang, Q. Zhong, W. Li, S. Li, P. Cheng, S. Meng, L. Chen, and K. Wu: „Experimental realization of two-dimensional boron sheets“, Nature Chemistry 8, 563 (2016).
[30] I. Božović: „Doubling-down on borophene electronics“, Nature Materials 21, 11–12 (2022).
[31] X. Shi, D. Popović, C. Panagopoulos, G. Logvenov, A. T. Bollinger, and I. Božović: „Emergence of superconductivity from the dynamically heterogeneous insulating state in $La_{2-x}Sr_xCuO_4$“, Nature Materials 12, 47–51 (2013).
[32] J. Wu, A. T. Bollinger, Y.-J. Sun, and I. Božović: „Hall Effect in quantum critical charge-cluster glass“, Proc. Natl. Acad. Sci. USA (PNAS) 113, 4284 (2016).
[33] A. T. Bollinger, G. Dubuis, J. Yoon, D. Pavuna, J. Misewich, and I. Božović: „Superconductor–insulator transition in $La_{2-x}Sr_xCuO_4$ at the pair quantum resistance“, Nature 472, 458–460 (2011).
[34] M. P. M. Dean, G. Dellea, R. S. Springell, F. Yakhou-Harris, K. Kummer, N. B. Brookes, X. Liu, Y.-J. Sun, J. Strle, T. Schmitt, L. Braicovich, G. Ghiringhelli, I. Božović, and J. P. Hill: „Persistence of magnetic excitations in $La_{2-x}Sr_xCuO_4$ from the undoped insulator to the heavily overdoped non-superconducting metal, Nature Materials 12, 1019–23 (2013).
[35] L. S. Bilbro, R. Valdés Aguilar, G. Logvenov, O. Pelleg, I. Božović, and N. P. Armitage: „Temporal correlations of superconductivity above the transition temperature in $La_{2-x}Sr_xCuO_4$ probed by terahertz spectroscopy“, Nature Physics 7, 298–302 (2011).
[36] J. Corson, R. Mallozzi, J. Orenstein, J. N. Eckstein, and I. Božović: „Vanishing of phase coherence in underdoped $Bi_2Sr_2CaCu_2O_{8+d}$“, Nature 398, 221–223 (1999).
[37] I. Božović, G. Logvenov, M. A. J. Verhoeven, P. Caputo, E. Goldobin, and M. R. Beasley: „Giant Proximity Effect in cuprate superconductors“, Phys. Rev. Letters 93, 157002 (2004).
[38] E. Morenzoni, B. M. Wojek, A. Suter, T. Prokscha, G. Logvenov, and I. Božović: „The Meissner effect in a strongly underdoped cuprate above its critical temperature“, Nature Communications 2, 272 (2011).

Ivan BOŽOVIĆ

ALHEMIJA XXI VEKA: DIGITALNA SINTEZA KVANTNIH MATERIJALA

Sažetak

Epitaksija molekularnim zracima, atomski sloj po sloj (ALL-MBE), nova je tehnologija, razvijena u poslednjih nekoliko decenija, za sintezu funkcionalnih kvantnih materijala, uključujući visokotemperaturne superprovodnike, druge kompleksne okside, i dvodimenzionalne materijale kao grafen i borofen. Ova tehnika omogućava sintezu novih metastabilnih materijala, koji se ne mogu proizvesti standardnim metodama. ALL-MBE alhemija — kreiranje sintetskih materijala sa novim i jedinstvenim elektronskim svojstvima je ilustrovana sa nekoliko primera. Mogućnost da se materijal oblikuje na nivou pojedinačnih atomskih monoslojeva je omogućila važna otkrića, koja ukazuju na prednosti ALL-MBE tehnologije.

Fedor MESINGER*

THE ETA MODEL IN WEATHER AND CLIMATE: BACKGROUND, AND LESSONS LEARNED

Abstract: The idea of writing a limited area atmospheric model came following my several years stay at the University of California in Los Angeles, UCLA, at the end of the sixties. Exposed at UCLA to what I refer to as the Akio Arakawa approach, encouraged by my having an idea for a scheme that seemed an improvement to what Arakawa was using, and aware of the importance of topography for weather of the country I returned to, led to my writing in 1973 a limited area code that eventually became the Eta model. Refinements introduced in subsequent years including those of a collaborator I acquired, Zaviša Janjić, led to the code that when installed in 1984 at the then U. S. National Meteorological Center, attracted attention. It is the model's performance when compared to two competing candidate models that eventually led to it becoming in 1993 the primary U. S. operational regional weather prediction model. Some of the key events that enabled this to happen are recalled. Although the model after about a decade + eventually was replaced in this role, it continued to be used, such as for forecasts by the Brazilian National Institute for Space Sciences (INPE), as a Regional Climate Model (RCM) in numerous climate change studies, and as a tool for the North American Regional Reanalysis (NARR), run in near-real time by the U. S. National Centers for Environmental Prediction. Model refinements in this later period are summarized, including the introduction of the so-called cut-cell discretization of its representation of topography. More recently, using about the same resolution, the model showed ensemble skill mid-tropospheric jet-stream forecast accuracy superior to its highly acclaimed driver European Centre for Medium Range Forecasts (ECMWF) model.

Key words: *Eta model, cut-cell schemes, finite-volume schemes, topography representation*

* Fedor Mesinger, Serbian Academy of Sciences and Arts, Belgrade, Serbia

1. THE BEGINNINGS

The reason I enrolled to study meteorology was my reluctance to abandon either one of the basic sciences taught in high school, chemistry, mathematics, physics. Also, because it offered understanding of why events we see take place. It turned out however that no chemistry was taught to meteorology students, and that mathematics was what I found the most enjoyable. However, methods used for weather prediction, clearly the main practical purpose of meteorology, at the time were analyses of weather maps, something I was not at all comfortable with. But weather prediction using mathematics, or, to use more modest words, using computers, seemed on the horizon, and to me the only way to make progress that looked like real science.

Much of an encouragement I received from a book „Numerical Weather Analysis and Prediction“ (Thompson 1961) then I got hold of, and still more from my attendance of the „International symposium on numerical weather forecasting“, in Oslo, March 1963. Just about every one of the leading pioneers of the emerging numerical forecasting field was present, including Akio Arakawa, Jule Charney, Arnt Eliassen, Cecil (Chuck) Leith, Edward Lorenz, Aksel Wiin-Nielsen, more (Platzman 1963). But most of all, it included a report by Arakawa on his finite-difference horizontal advection scheme that via conservation of total kinetic energy and vorticity squared prevented the so-called nonlinear instability that Norman Phillips discovered some years earlier (Phillips 1959), and that looked like making a successful longer range numerical weather prediction (NWP) impossible. The NWP future looked bright!

At the time however there were no electronic computers in Belgrade, and in addition my education in the area needed improvement. I managed to spend some time at the National Center for Atmospheric Sciences (NCAR), Boulder, CO, some at the University of California at Los Angeles (UCLA), then some in Belgrade without a real job, and eventually at the end of the sixties at the best place possible for exposure to the at that time emerging developments in atmospheric numerical modeling, again at UCLA. This because of the ability to listen to numerical methods course of Akio Arakawa.

This was the time when people at the forefront of NWP efforts understood that for real progress one needed to move from integrating the vorticity equation to more complete so called „primitive“ equations, Navier-Stokes equations with hydrostatic approximation. These are equations which have the vertical velocity calculated using the two horizontal components. A new set of problems had to be given attention, and these were not problems one could learn about by looking at a textbook on numerical methods for solving differential equations, such as the highly respected textbook of the time by Richtmyer and Morton (1967).

There are at least two basic reasons for this. While we know the equations of motion we want to solve, they having been known now for more than 200 years, and while we also know they *are* initial-value equations, we do not really know the initial values. They are obtained via a variety of measurements and can only be approximate. And we are working with a system that does not have a solution in form of some analytic function. The solution is the atmospheric state that happens.

One of these new problems with no mathematical guidance is the distribution of variables in space. Arakawa, working with his graduate student Frank Winninghoff, had four possible square arrangements of variables of two-dimensional primitive equations analyzed (Winninghoff 1968). They are displayed in Fig. 1. The way this was done was to look at what happens when using simplest centered differences for space derivatives of linearized gravity and inertia terms when assuming wave solutions of these equations (e. g., Mesinger and Arakawa 1976, Ch. IV, Sec. 6; Arakawa and Lamb 1977, Ch. III, Sec. A). Examining the effect of the resulting space discretization error on the frequency Arakawa and Lamb (1977) conclude that except for some rare situations the fully staggered C grid gives the best result, this being important for the so-called geostrophic adjustment process. This is the establishment of approximate balance between the horizontal pressure gradient and the Coriolis force once this balance is perturbed, something that is nowadays understood to be constantly taking place in the atmosphere.

2. NWP PROGRESS, THE ETA AT NMC AND NCEP

This address of the impact of the choice of the horizontal grid via its adequacy for the numerical representation of the gravity-inertia terms at the end of the sixties might well mark the beginning of the soon to follow fast progress in the skill of the actual NWP models. Namely, the skill of operational NWP following its inception several years after the famous Charney, Fjørtoft, and von Neumann (1950) accomplishment was not impressive for quite some time. Change to more general primitive equations was one requirement, and the availability of more powerful computers they needed another. Both were gradually taking place only in early to mid-seventies.

Having returned to the then Yugoslavia in 1970, after several years dedicated mostly to preparing the courses taught, I wrote the original code of what eventually was going to become the Eta model. I decided to use the E-grid because this seemed to me best for the definition of lateral boundary conditions. Soon I was joined in this effort by Zaviša Janjić, senior year

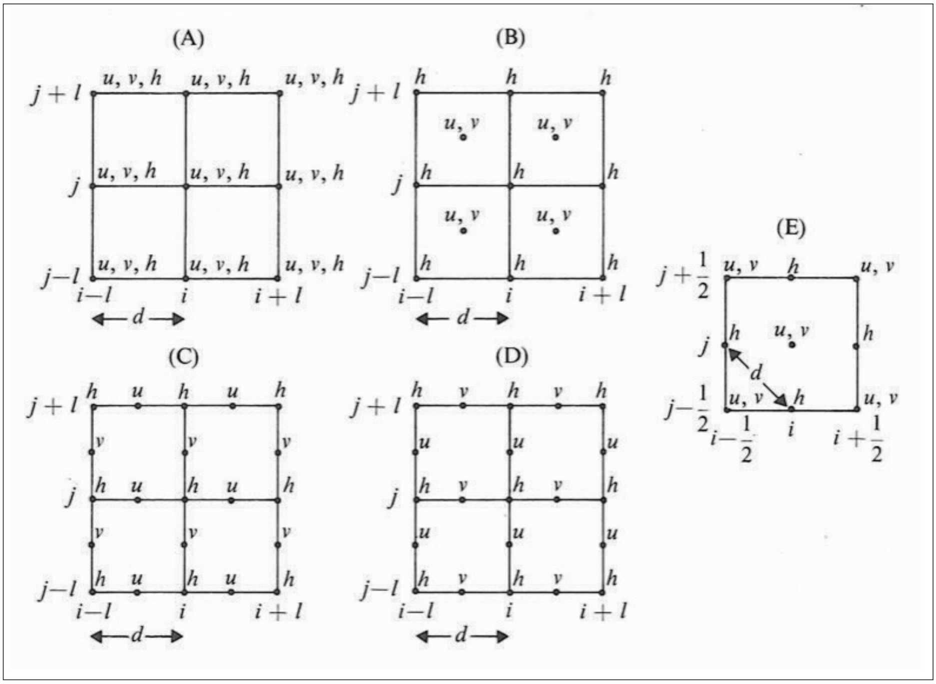


Fig. 1. Four possible distributions of horizontal velocity components *u, v* and height *h* for the so-called shallow water model, defined to have velocity components independent of height. For primitive equation models, surface pressure and temperatures are defined at the *h* points. Note that E is the same as B, except for being rotated by 45°. Thus, B and E grids at times are referred to as the B/E grid. From Mesinger and Arakawa (1976).

student when I returned to Belgrade, who made important contributions of his own. In the very first code there was one scheme directly taken from Arakawa, the vertical advection scheme, but various our contributions followed (e. g., Janjić 1977, Mesinger 1977). But those of most impact were surely these of the early eighties, with (Mesinger 1984) changing the vertical coordinate from the ubiquitous terrain-following to the terrain-intersecting step-topography eta (Fig. 2), and (Janjić 1984) discovering a way to convert the Arakawa and Lamb (1077) C-grid horizontal advection scheme to use the velocity components of the model's E-grid.

Having re-written the model code during my half a year 1984 visit to GFDL to use these upgrades I have brought the code to the then National Meteorological Center (NMC) in Camp Springs, MD, for another half a year visit. Dennis Deaven of NMC helped me install the code and run an experiment additional to those already done at GFDL (Mesinger et al. 1988).

A most distinguished person working at NMC during my 1984 visit was Norman Phillips, mentioned above for his discovery of the nonlinear

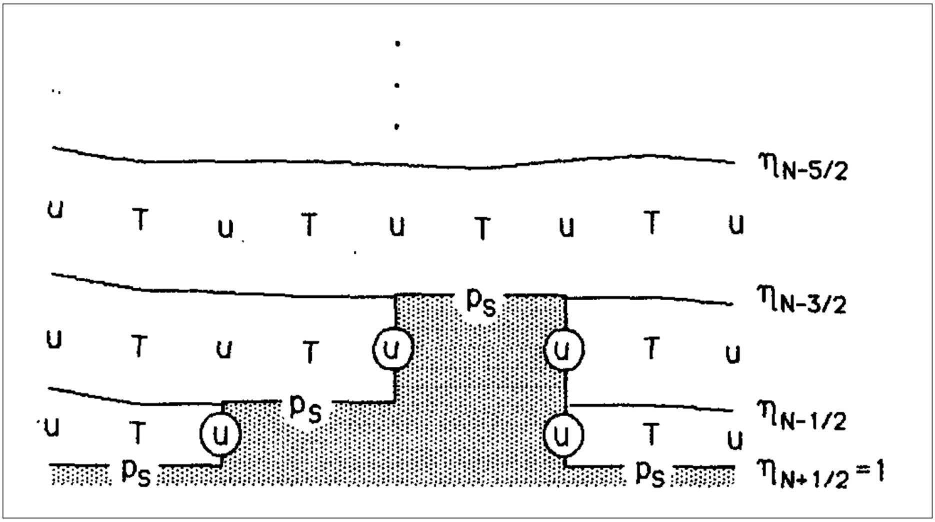


Fig. 2. Example of a 2D representation of topography using the eta vertical coordinate. *u* represents x-components of velocity, *T* temperature, and N the number of model layers. Quasi-horizontal lines are model layer interfaces with prescribed η values. From Janjić (1990).

instability. Phillips, one of the „pantheon of gods" of atmospheric numerical modeling one might say (note, e. g., Fig. 1 in Mesinger et al. 2018) left in 1974 his position of the head of the Department of Meteorology at MIT (Massachusetts Institute of Technology) to pursue numerical prediction at NMC. With the assistance of Jim Hoke, he developed the so-called Nested Grid Model (NGM), that clearly NMC expected a lot from. But when I asked him in 1984 what horizontal grid he used in the NGM, A, B, C etc., he replied he did not know! Instead, he handed me a technical note on NGM, and told me something like „Here is our publication, so you can find out". Of course, I looked into it, and found out he used the D grid, according to Arakawa the worst choice. Was it a different time with information not easily spread. Or?

The National Weather Service Director of the time, Ronald (Ron) McPherson saw to it that the model code (note yet referred to as Eta) is maintained, tested more (Black 1988), and further developed, see Tom Black's words in the Acknowledgements of Mesinger (2004). A crucial step in this further development was Zaviša Janjić's visit to NMC in 1987, during which he upgraded the „minimum physics" package of the model primarily by adding the Mellor-Yamada level 2.5 turbulence, level 2 in the lowest layer, and the Betts-Miller convection scheme, both with various modifications (Janjić 1990).

Janjić's visit was followed by mine, with the attention divided between additional model developments and verification efforts, all of these with continued assistance of Tom Black. Model upgrades included addition of a viscous sublayer over water to the surface layer scheme, so-called piecewise linear vertical advection of moisture (Mesinger and Jovic 2002), and, in cooperation with Alan Betts, refinements of the Betts-Miller convection scheme. With the Eta, as spontaneously already referred to at that time, often showing results superior to those of the operational NGM, a three-way comparison of precipitation accuracy was done of the Eta vs. the NGM as run operationally, and the NGM using the Eta's Betts-Miller convection scheme. This by objective skill scores confirmed the Eta's notable improvement over the NGM in forecasting all heavier precipitation categories during the period used, regardless of the convection scheme used in the NGM (Mesinger et al. 1990).

Another Janjić's visit followed, with a considerable upgrade of the convection scheme, subsequently referred to as the Betts-Miller-Janjic (BMJ) scheme, as well as that of the viscous sublayer scheme, now different above land and water, and minor Mellor-Yamada (MY) changes (Janjić 1994). It should be noted however that the Janjić (1994) paper covers essentially only the material prior to the end of his just referred to visit, some time in 1990, with only brief mentions of the developments after 1990.

When in early 1991 I rejoined what used to be the NMC's Development Division, reorganized so that it became the National Centers for Environmental Prediction/ Environmental Modeling Center (NCEP/EMC), its Director Eugenia Kalnay told me about a problem of the then semi-operational Eta. During the past winter it exhibited a „mixed performance over warm water", occasional overdeepening of lows and widespread rain that did not verify. Looking into a sensitive case of a real development that did not verify, eventually I invented what I referred to as the „*l* bulk" scheme, replacing the finite-difference type specification of the length scale of the model's MY-2 (Mellor-Yamada level 2) lowest layer scheme. It very much removed the overdeepening problem, and performed even better than a standard Monin-Obukhov formulation (Mesinger and Lobocki 1991, summarized in Mesinger 2010).

Another problem was identified by chance (serendipity?). Adrian Marroquin running the Eta outside NCEP, complained about having no turbulence kinetic energy (TKE) above the planetary boundary layer (PBL). I checked at EMC and noticed reasonable values in the forecast I looked at. The difference had to be in the code: Marroquin was using an older version which did not have the MY master length scale values I specified throughout the model atmosphere and thus also above the PBL, advised by Mark Helfand to do so, these specifications addressing also other issues (Mesinger 1993b). But having a look at the right place, I noticed a major oversight on the part of apparently all previous users of the MY-2.5, that made it next to impossible for TKE to be generated above the PBL (Mesinger 1993a).

Eventually the Eta became officially the U. S. operational regional model on 8 June 1993. The NGM at the time was „frozen" since 1991, but continued to be run, thus, inter alia, serving as a benchmark that could be used to assess the progress made. Another model, so called Regional Spectral Model (RSM), developed by Henry Juang, was also run on a regular basis, and it the eyes of some certainly was a contender. Note the statement of all the Development Division managers of the same year as the official implementation of the Eta „A comparison with Regional Spectral Model (RSM) will determine possible replacement by the RSM" (Kalnay et al. 1993). It continued to be run until late 1997, with resolution about the same as that of the Eta, when in almost a 2-year comparison of more than a thousand forecasts the Eta demonstrated higher precipitation accuracy across all the intensity thresholds. This despite the RSM being run later with a 12-h more recent lateral boundary conditions (Mesinger 2004, Fig. 2).

Following a variety of enhancements and refinements made, in 1998 the operational Eta achieved precipitation accuracy scores of its 24–48 h forecasts across all intensity thresholds higher than the NGM's scores of the 00–24 h forecasts.

The Eta accuracy over its U. S. domain was higher than that of the U. S. global model as well, so in 1997 a proposal was made for a „regional reanalysis" project, by the Advisory Committee of an existing global reanalysis project. The idea of such reanalyses, at the time one already having been performed, is to use a fixed model system and process all the data available as of the beginning of satellite measurements, to obtain analyses of climate change not affected by the model and analysis changes. This proposal to perform regional reanalysis using the Eta has eventually been accepted, funded, and done for its initial 25-yr retrospective period, and continues to be performed in near-real time as we speak. Paper describing the method and reporting on these first 25 years of data, Mesinger et al. (2006), is frequently cited, so that in Google Scholar as this is being written has more than 3700 citations.

However, a problem with the Eta has been noticed at the end of the nineties. An experimental 10-km Eta failed to forecast an intense downslope windstorm in the lee of the Wasatch Mountain, while a sigma system MM5 model did well (McDonald et al. 1998). Following that, Gallus and Klemp (2000) reported on a nonhydrostatic eta code in simulating flow over a 2D bell-shaped topography exhibiting a flow separation off its top instead of a descent down the lee slope. This had a rather negative impact on the Eta reputation as it was by many expected to become more detrimental with higher resolutions. Consequently, when replacing the referred to 10-km Eta by an 8-km Nonhydrostatic Mesoscale Model (NMM) using terrain-following coordinate, an announcement was made that „This choice [of the vertical coordinate] will avoid the problems … with strong downslope winds and will improve placement of precipitation in mountainous terrain“ (Geoff DiMego, personal communication, 19 July 2002).

Accordingly, upgrades of the Eta at NCEP were discontinued, and in addition a more advanced data assimilation system has been developed for the NMM. Eventually in 2006 a so-called „parallel“ was run, of four + months of forecasts of the operational Eta system, and the newly developed NMM system, at the same domains and resolution. The Eta performed clearly better with precipitation skill scores, the more so the longer-range forecasts were looked at (e. g., Mesinger and Veljovic 2017, Fig. 4). There was little else in terms of objective scores that would be favoring the NMM (DiMego 2006). Over higher topography western contiguous United States, the Eta remained more accurate in both 2-m temperatures and 10-m winds (Mesinger 2022). Nevertheless, perhaps understandably, decision was made to have the NMM be the next U. S. operational regional model.

3. THE ETA OUTSIDE NCEP

With the Eta demonstrating accuracy as summarized no wonder it was used in various ways in some countries outside the United States. In 1994–1998 it was the primary experimental tool of international summer schools organized in Krivaja, Bačka Topola, by the Federal Hydrometeorological Institute, Belgrade, sponsored and partly supported by WMO (World Meteorological Organization). It was used in three two-week workshops/conferences on regional modeling organized by the International Center for Theoretical Physics (ICTP) near Trieste, 2002, 2005, and 2008.

The most extensive use of the Eta outside NCEP was at CPTEC (Center for Weather Prediction and Climate Studies), Cachoeira Paulista, SP, Brazil. Two CPTEC people visited NMC in 1996 and have taken the Eta code to Brazil. One of them was Sin Chan Chou (pron. Shou) who continued using it and organized a dedicated group using the Eta for short range forecast and as a regional climate model (RCM). As this is written Chou had organized for the 7th time an Eta model workshop with participants being taught to run the model and make experiments. Twice it was the model of choice of an U. N. organization, CEPAL, covering Latin America and the Caribbean, for training workshops on climate change experiments, hosted by CPTEC.

CPTEC, an agency of INPE (National Institute for Space Sciences), as this is written only INPE, following NCEP's 2006 decision has become kind of a „primary home“ of the Eta, not only for hosting events but also in taking part in contributions to the further development of the model. Thus, the paper summarizing the design of the model as of 2011, Mesinger et al. (2012), of 11 coauthors includes 5 coauthors who are or were resident in Brazil.

But it is fitting to continue this presentation with summaries of some of the major model advancements following 2006. One resulted from another event due to chance. To address the „Gallus-Klemp problem“ of the Eta, I had a long-standing idea, as of 2002, on my „to do list“ what needs to be done. Since Gallus and Klemp ascribed the problem to the existence of step corners of the step topography Eta, as illustrated in the schematic of Fig. 2, I was going to change the discretization so that the velocity cells just above the corners have sloping bottoms, as illustrated in Fig. 3.

Only much later I realized that this in fact is a simple version of what originally was referred to a shaved cells method (Adcroft et al. 1997), and nowadays, perhaps as of Steppeler et al. (2011), a cut-cell scheme.

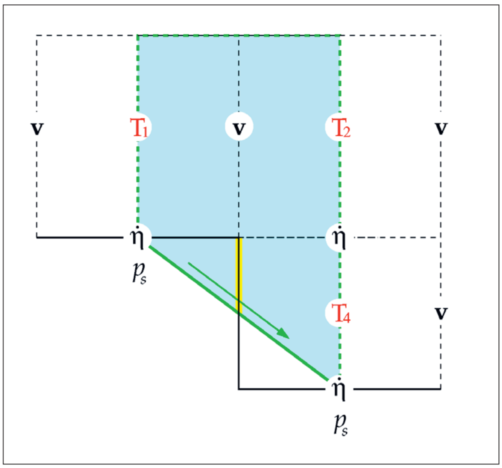


Fig. 3. 2D schematic illustrating the „sloping steps“ Eta discretization. The vertical topography side below the velocity (v) cell in the middle is replaced by a sloping side going from the left to the right surface pressure (pS) points. Symbols T denote temperature cells, η denotes the eta coordinate, with dot ($\dot{}$) above a symbol standing for the time derivative. From Mesinger et al. (2012).

Once a version of the code using „sloping steps“ was put together a test was done using an 8-km/60-layer code over a domain of the western United States and running a 48-h forecast. Obtained temperatures of the lowest layer T cells are shown in Fig. 4. Two or three spots all in mountain basins are seen responsible for the choice of the NCAR graphics routine of the plotting interval of 10°K, one of them in southwestern Montana, with the temperatures below 180°K. Another in western Alberta, not much warmer than that.

Understanding what happened was obviously needed. It was not hard to get the idea that the finite-difference vertical advection scheme, used for slantwise temperature advection such as between cells T1 and T4 in Fig. 3, is a good candidate to be responsible.

The centered finite-difference scheme used for slantwise advection in calculating temperatures of Fig. 4 was

$$\frac{dT}{dt} = \cdots - \dot{\eta}\,\overline{\frac{\Delta T}{\Delta \eta}}^{\eta} \tag{1}$$

d/dt being the time derivative, the eta vertical velocities, $\dot{\eta}$, being defined at cell interfaces, the overbar standing for two-point averaging in the direction indicated, and Δ denoting the finite difference in the direction of the η.

Suppose the left half of the schematic of Fig. 2 were the exit region of a basin with predominant flow to the right and somewhat upward. According to (1) the temperature change due to vertical advection is the average of contributions from the top and the bottom sides of T cells. If a temperature inversion were to develop between the two leftmost and lower T cells, then the vertical advection contribution from their mutual interface would cool both cells, but for the lower of them would be the only contribution, thus tending to increase the inversion, amplifying its cooling, feeding on itself. An instability like mechanism would be established, for a physically wrong reason.

It was easy to avoid the problem. As we *know* the velocity across the half of what was a vertical step, yellow in the schematic of Fig. 3, we also know the mass of air moving in the time step from cell T1 to T4, so that it is straightforward to calculate the temperature changes due to the slantwise advection in a Lagrangian way. This was coded instead of the use of (1), and a realistic lowest cell temperature forecast was obtained instead of that in Fig. 4.

However, the scheme (1) was used not only for slantwise advection, but for the vertical advection of the main prognostic variables, v and *T*, as well. Being now aware of the scheme's problem at the lowest cells, I have changed the vertical advection of v and *T* to the finite-volume scheme that the model

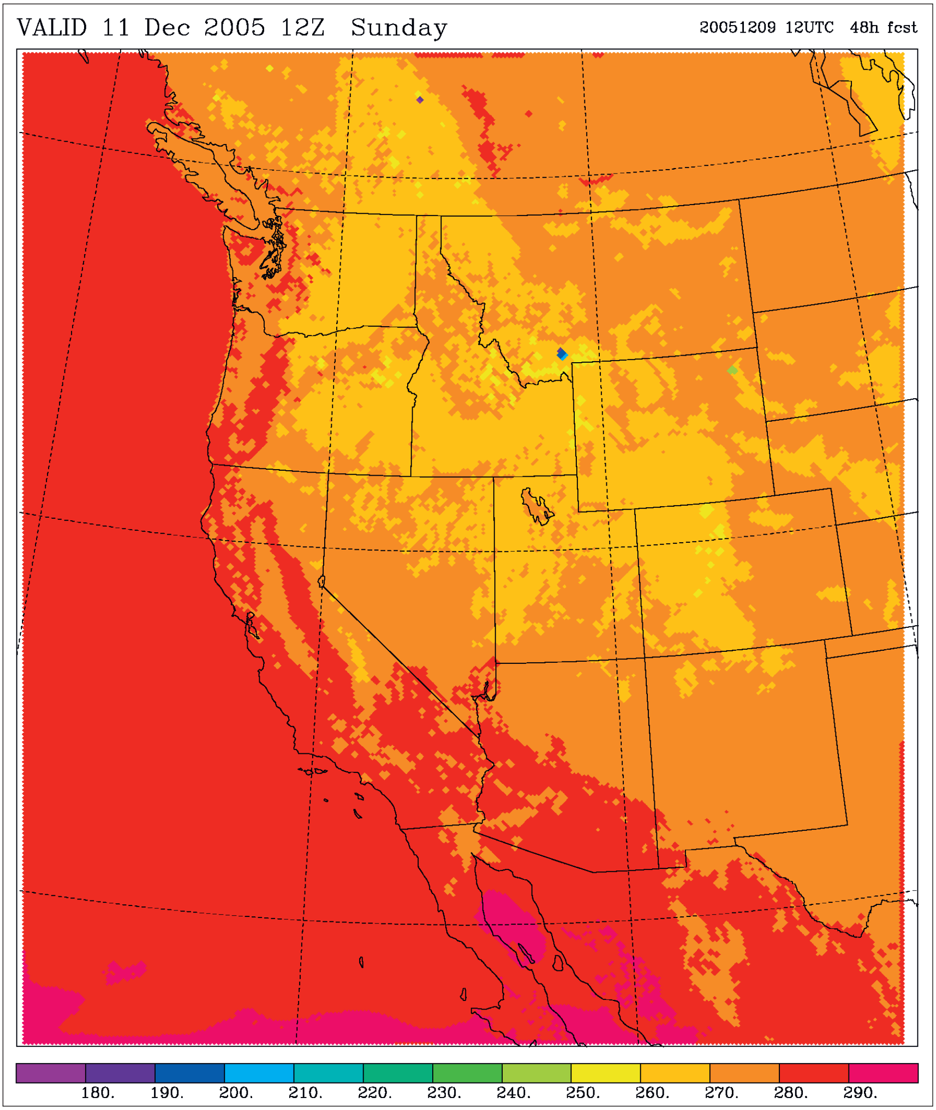


Fig. 4. Temperatures of the lowest temperature cells of a 48-h forecast with the „sloping steps“ Eta, when using a finite-difference „Lorenz-Arakawa“ centered slantwise temperature advection scheme.

was already using for moisture (Mesinger and Jovic 2002). It is a scheme that respecting the finite-volume meaning of prognostic values as cell averages, adjusts slopes inside cells toward values at the edges of neighboring cells, but without creating new minima or maxima and keeping them linear inside cells. Thus, the term „piecewise linear“ is often used. Advection can then be performed using velocities at cell boundaries.

An unexpected reward came from that change, in addressing the model's difficulty with downslope windstorms. When the change was made experiments were in progress with cases of intense downslope windstorms in the lee of the Andes. Two sections of synoptic maps illustrating one of these cases are shown in Fig. 5. The case is the same as that discussed in Section 9 of Mesinger et al. (2012). Warming of 24°C is seen at the station San Juan, about the middle of the maps. Warmings of that type are known in the Alpine region under the name „foehn", in the lee of the Andes their name is „zonda". The „reward" just referred to is illustrated by the two plots of Fig. 6. The one on the left shows the result of the forecast using (1) for the slantwise as well as for the vertical advection, while the one on the right shows the result with (1) replaced for both by the finite-volume advections. Warming in the zonda region is thereby seen increased by more than 4°K and occurs in one might say „the right place". But to get back to the total warming, compared to the 24 h forecast (not shown) the warming of Fig. 6 at San Juan is seen to be more than 20°K. Thus, notwithstanding the difference of 33 vs. 30 h, this zonda effort can be declared a success.

Be that as it may it is the flow separation issue pointed out by Gallus and Klemp (2000) that was quoted the most as the weakness warning people to stay away from the eta system, e. g., five citations listed in Mesinger (2004). Thus, it required attention, and it was on 9 March 2002 that I worked out a plan how to address it. I could look up this date because I recall I was then travelling on a ship from a meeting on „Awaji" island to Osaka. My

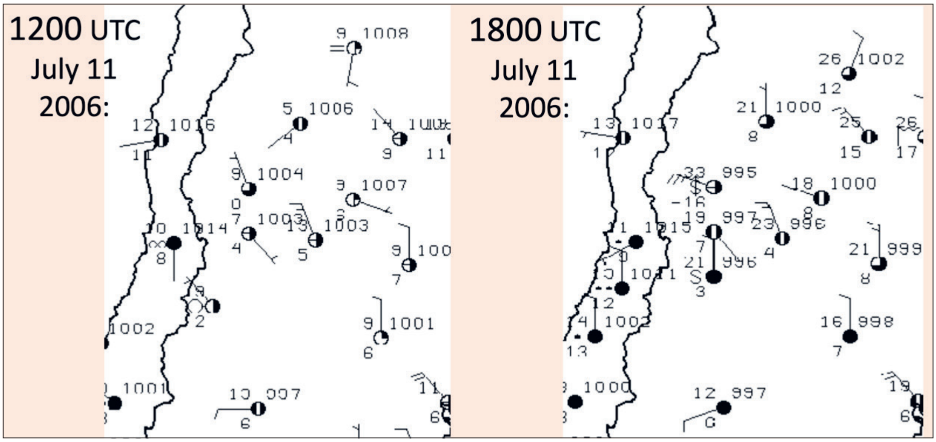


Fig. 5. Sections of surface synoptic maps illustrating a case of an intense „zonda" windstorm in the lee of the Andes. Warming from 9 to 33°C in 6 h is seen at the station San Juan, 630 m above sea level, close to the middle of the sections. Valid times are displayed in the top left corners of the maps.

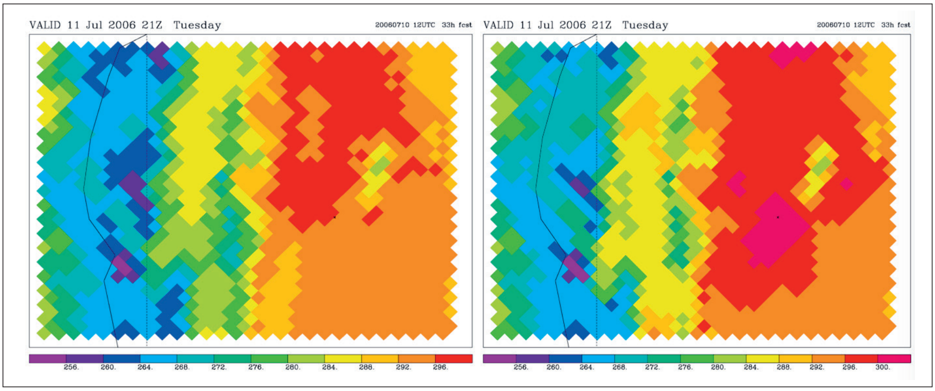


Fig. 6. Forecast lowest cell temperatures at 33 h of the case discussed in Section 9 of Mesinger et al. (2012). The left-hand plot shows the result obtained using (1) for both the slantwise and the vertical advection, while the right-hand plot shows the result with these advections replaced by the finite-volume versions. The roughly vertical line on the left sides of the plots is the Chile-Argentina border, while the straight line is the 70°W meridian. The dot to the right of the centers of plots shows the place of the San Juan station.

plan involved defining slopes at the bottoms of v cells, using the topography values of four surrounding *zS* points. Thereby the step corners of the eta topography according to Gallus and Klemp (2000) responsible for the flow separation would be eliminated, and presumably the Gallus-Klemp Eta problem as well.

Implementing the plan was not straightforward, assistance was needed in handling the code I wrote, and obtained from the Sin Chan Chou's group, primarily Jorge Gomes. My code bug produced in copying hand-written code lines to code statements was eventually discovered by Ivan Ristić of the „Weather2" Belgrade company, and the code then seemed to work fine. But the flow separation in the Gallus-Klemp experiment of 2D bell-shaped topography, Mesinger et al. (2012, Fig. 3), while visibly improved, was not completely removed.

An unusual help came in 2013 from Sandra Morelli, of the University of Modena. Morelli informed me of noticing „something strange" in the code of the so-called horizontal diffusion, code modelers use either to avoid unwanted noisiness of fields, or to simulate the impact of unresolved eddies, the latter in the case of the Eta. „Something strange" was leftover code from a previous sigma version that was with the eta coordinate not active, but a look at the right place was enough to see a problem. Horizontal diffusion code was not made aware of the „sloping steps" and was thus responsible for the remaining flow separation. This being addressed, flow over the 2D bell-shaped topography was obtained as in the right-hand plot of Fig. 7. Its left-hand plot is (c) of Fig. 6 of Gallus and Klemp (2000), which they obtained using artificial modification of the code to remove the impact of the step corners they found responsible for the flow separation problem.

The routine used for our right-hand plot of Fig. 7 did not allow for slopes, so they are not visible in the topography of the right-hand plot Fig. 7. Compared to the linear solution one could be concerned with the maximum on top of the mountain, but we are confident this could be improved with slopes over more than one cell if this were felt to be an issue of sufficient priority.

4. ETA AS RCM AND ITS LARGE-SCALE SKILL

While the Eta has been used as a global model on a cubed-sphere grid (Latinović et al. 2018), almost exclusively its use was as a limited area model (LAM), covered in two preceding sections. As stated earlier, as RCM it enjoyed extensive use for a variety of purposes mostly over South American domains (e. g., Chou et al. 2020, with references to many others).

One point can be stressed here. It is almost universally believed that the nested model should improve on smaller scales, while it should accept „large scales“ as they are in its driver global model. Consequently, so called Davies' relaxation lateral boundary conditions are applied, forcing variables in some rows around the boundary to conform to the driver model values, completely at the boundary, and less and less toward the inside of the domain. Very often investigators also apply the so-called large scale or spectral nudging inside the domain, forcing the integration variables not to depart much from those of the driver model.

It is hard to see a scientific basis for these practices. While the global driver model might be equipped with components and feedbacks missing in a LAM or RCM, if we consider just the atmospheric motion, impact of these missing components can be received via the lateral boundary conditions but large scales inside the LAM domain could still be improved if the LAM has some advantage over its driver model. The advantage could be higher resolution but can also be a dynamical core better in some ways, or both, but hardly better parameterizations because global models require very considerable parameterization efforts.

As an example, in Fig. 8 top right, average wind speed at 250 hPa of an Eta 21-member ensemble is shown, driven by a European Centre for Medium-Range Weather Forecasts (ECMWF, or EC) ensemble, with its average wind speed at the same height at top left, both at 4.5 days lead time. Space resolution of the Eta ensemble was about the same as that of the driver EC ensemble. EC analysis valid at the same time is shown at the bottom left. Mesinger and Veljovic (2020) contains more detail.

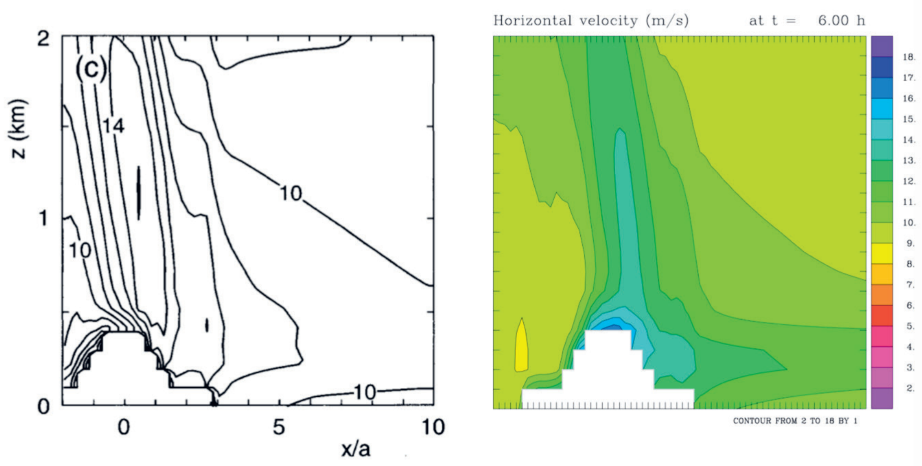
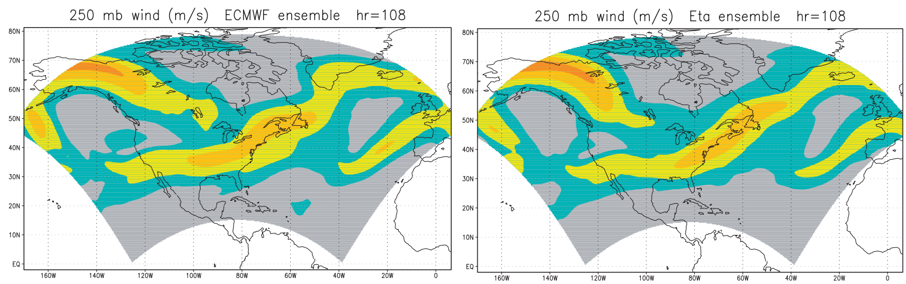


Fig. 7. Simulation of the Gallus-Klemp experiment with the Eta code, plot (c) of Fig. 6 of Gallus and Klemp (2000), left, using the sloping steps Eta code allowing for velocities at slopes in the horizontal diffusion scheme, right. From Mesinger and Veljovic (2017).

While visual impressions as those pointed out in the caption of Fig. 8 are crucial, results of models are normally assessed using some skill numbers, or scores. Three such verification scores have been used to assess the Eta skill vs. that of its driver EC model, and with each of them at 4.5-day lead time all 21 Eta members had better skill scores of 250 hPa wind speeds



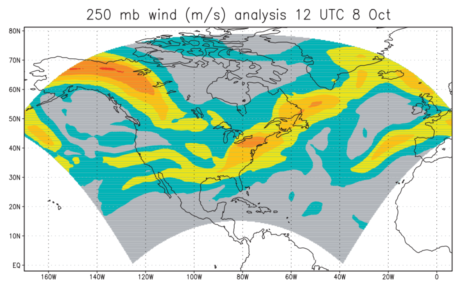


Fig. 8. Averages of 4.5 day 250 hPa wind speeds of the 21 member Eta ensemble, top right, of the EC driver ensemble member, top left, and EC verification analysis left. Note the more accurate Eta's northeastward extension of the New England jet streak of > 45 m/s, as well as the more accurate south-eastward extension of the one over the Canadian Rockies.

than their driver EC members, and with two of them even more frequently (Mesinger and Veljovic 2017). Surprisingly, when switched to use the sigma coordinate the Eta also achieved better scores, although not that much better as the Eta.

5. CONCLUDING COMMENTS

Weather and climate models require many components and are nowadays never developed by a single person. But if a model that in terms of its dynamical core features was developed primarily by only two people achieves results at least comparable with a model of a major international institution, one might wonder why this has happened.

In this text two features have been listed that contributed to the Eta skill but are absent in the EC model. One is the Eta vertical coordinate resulting in quasi-horizontal coordinate surfaces, eliminating thus large pressure gradient force errors.

Another is the finite-volume slantwise and vertical advection. They not only avoided the false advection from below ground of the finite-difference scheme used previously and made a crucial contribution to the simulation of the zonda windstorm, but they also achieved consistency with the finite-volume property of the Eta Arakawa horizontal advection. This I believe has been essential for the successful performance of the Eta model as shown, including the skill of the Eta vs. the Eta using sigma.

ACKNOWLEDGEMENTS

Very many people in numerous ways enabled the today's existence and skill of the Eta model, its code and its results. To list only a few, Zaviša Janjić, Lazar Lazić, Tom Black, Eugenia Kalnay, Dušan Jović, Sin Chan Chou, Jorge Gomes, Sandra Morelli, Ivan Ristić, Katarina Veljović, too many more to mention individually, and to all of them goes my appreciation and heartfelt gratitude.

REFERENCES

[1] Adcroft A, Hill C, Marshall J (1997) Representation of topography by shaved cells in a height coordinate ocean model. Mon Weather Rev 125: 2293–2315.

[2] Arakawa A, Lamb VR (1977) Computational design of the basic dynamical process of the UCLA general circulation model. In: J. Chang, (ed), Methods in Computational Physics, Academic Press. 17: 173–265.

[3] Black TL (1988) The step-mountain eta coordinate regional model: a documentation. NOAA/NWS National Meteorological Center, April 1988, with revisions later. http://ftp1.cptec.inpe.br/pesquisa/grpeta/petamdl/Publications/Black.1988+_EtaDocumentation.pdf
[4] Charney JG, Fjørtoft R, von Neumann J (1950) Numerical integration of the barotropic vorticity equation. Tellus 2: 237–254.
[5] Chou SC, Dereczynski C, Gomes JL, Pesquero JF, de Avila AMH, Resende NC, Alves LF, Ruiz-Cárdenas R, de Souza CR, Bustamante JFF (2020) Ten-year seasonal climate reforecasts over South America using the Eta regional climate model. Ann of the Braz Acad Sci 92: 3 https://www.scielo.br/j/aabc/a/vS95QNbfvcsmn57F8XnGMbJ/?format=pdf&lang=en
[6] DiMego G (2006) WRF-NMM & GSI Analysis to replace Eta Model & 3DVar in NAM Decision Brief. National Centers for Environmental Prediction 115 pp https://www.emc.ncep.noaa.gov/WRFinNAM/Update_WRF-NMM_replacing_Eta_in_NAM2.pdf
[7] Janjić ZI (1977) Pressure gradient force and advection scheme used for forecasting with steep and small scale topography. Contrib Atmos Phys 50: 186–199.
[8] Janjić ZI (1984) Nonlinear advection schemes and energy cascade on semi-staggered grids. Mon Weather Rev 112(6): 1234–1245.
[9] Janjić ZI (1990) The step-mountain coordinate: physical package. Mon Weather Rev 118(7): 1429–1443.
[10] Kalnay E, Baker W, Kanamitsu M, Petersen R, Rao, DB Leetmaa A (1993) Modeling plans at NMC for 1993–1997. In: 13th Conf on Weather Analysis and Forecasting, Vienna, VA. Am Meteorol Soc, pp 340–343.
Kalnay E, Baker W, Kanamitsu M, Petersen R, Rao, DB Leetmaa A (1993) Modeling plans at NMC for 1993–1997. In: 13th Conf on Weather Analysis and Forecasting, Vienna, VA. Am Meteorol Soc, pp 340–343.
[11] Latinović D, Chou SC, Rančić M, Medeiros GS, Lyra AA (2018) Seasonal climate and the onset of the rainy season in western-central Brazil simulated by Global Eta Framework model. Int J Climatol 1–17. https://doi.org/10.1002/ joc.5892
[12] McDonald BE, Horel JD, Stiff CJ, Steenburgh WJ (1998) Observations and simulations of three downslope wind events over the northern Wasatch Mountains. In: 16th Conference on Weather Analysis and Forecasting, Phoenix, AZ. Am Meteorol Soc, pp 62–64.
[13] Mesinger F (1977) Forward-backward scheme and its use in a limited area model. Contrib Atmos Phys 50: 200–210.
[14] Mesinger F (1984) A blocking technique for representation of mountains in atmospheric models. Riv Meteorol Aeronautica 44: 195–202.
[15] Mesinger F (1993a) Forecasting upper tropospheric turbulence within the framework of the Mellor-Yamada 2.5 closure. Res Act Atmos Ocean Model, WMO, Geneva, CAS/JSC WGNE Rep. No. 18, 4.28–4.29.
[16] Mesinger F (1993b) Sensitivity of the definition of a cold front to the parameterization of turbulent fluxes in the NMC's Eta Model. Res Act Atmos Ocean Model, WMO, Geneva, CAS/JSC WGNE Rep. No. 18, 4.36–4.38.
[17] Mesinger F (2004) The Eta model: design, history, performance, what lessons have we learned? In: Symp. on the 50th Anniversary of Operational

Numerical Weather Prediction. Univ of Maryland, College Park, MD. 14–17 June 2004. 20 pp https://www.ncep.noaa.gov/nwp50/Agenda/Tuesday/
[18] Mesinger F (2010) Several PBL parameterization lessons arrived at running an NWP model. In: International conference on planetary boundary layer and climate change. IOP Publishing. IOP Conference Series: Earth and Environmental Science 13 https://iopscience.iop.org/article/10.1088/1755–1315/13/1/012005
[19] Mesinger F (2022) Vertical resolution of the surface layer versus finite-volume and topography issues. Boundary-Layer Meteorol. In press. doi: 10.1007/s10546–022–00745–2
[20] Mesinger F, Arakawa A (1976) Numerical methods used in atmospheric models. WMO/ICSU Joint Organizing Committee, Geneva, Switzerland. GARP Publications Series 17, Vol. 1, vi + 64 pp.
[21] Mesinger F, Jovic D (2002) The Eta slope adjustment: Contender for an optimal steepening in a piecewise-linear advection scheme? Comparison tests. NOAA/NCEP Office Note 439 Washington, DC: US Department of Commerce
[22] Mesinger F, Lobocki L (1991) Sensitivity to the parameterization of surface fluxes in NMC's eta model. In: 9th Conf on Numerical Weather Prediction, Denver, CO. Am Meteorol Soc, pp. 213–216.
[23] Mesinger F, Veljovic K (2017) Eta vs. sigma: Review of past results, Gallus-Klemp test, and large-scale wind skill in ensemble experiments. Meteorol Atmos Phys 129: 573–593.
[24] Mesinger F, K. Veljovic K (2020) Topography in weather and climate models: Lessons from cut-cell Eta vs. European Centre for Medium-Range Weather Forecasts experiments. J Meteorol Soc Japan 98(5): 881–900. https://doi.org/10.2151/jmsj.2020–050
[25] Mesinger F, Janjic ZI, Nickovic S, Gavrilov D, Deaven DG (1988) The step-mountain coordinate: model description, and performance for cases of Alpine lee cyclogenesis and for a case of an Appalachian redevelopment. Mon Weather Rev 116(7): 1493–1518.
[26] Mesinger F, DiMego G, Kalnay E, Mitchell K, Shafran PC, Ebisuzaki W, Jovic D, Woollen J, Rogers E, Berbery EH, Ek MB, Fan Y, Grumbine R, Higgins W, Li H, Lin Y, Manikin G, Parrish D, Shi W (2006) North American Regional Reanalysis. Bull Am Meteorol Soc 87: 343–360.
[27] Mesinger F, Chou SC, Gomes JL, Jovic D, Bastos P, Bustamante JF, Lazic L, Lyra AA, Morelli S, Ristic I, Veljovic K (2012) An upgraded version of the Eta model. Meteorol Atmos Phys 116: 63–79.
[28] Mesinger F, Rančić M, Purser RJ (2018) Numerical Methods in Atmospheric Models. In: Oxford Research Encyclopedia of Climate Science. 51 pp. https://doi.org/10.1093/acrefore/9780190228620.013.617.
[29] Phillips NA (1959) An example of non-linear computational instability. In: The Atmosphere and the Sea in Motion (The Rossby Memorial Volume), 501–504. Rockefeller Institute Press, New York. pp. 509.
[30] Platzman GW (1963) International symposium on numerical weather forecasting, Oslo, March 11–16, 1963. Tellus 15(3): 284–286.
[31] Richtmyer RD, Morton KW (1967) Difference Methods for Initial-Value Problems. Interscience Publishers, New York.

[32] Steppeler J, Park SH, Dobler A. (2011) A 5-day hindcast experiment using a cut cell z-coordinate model. Atmospheric Science Letters 12: 340–344.
[33] Thompson PD (1961) Numerical Weather Analysis and Prediction. Macmillan, New York, pp. 170.
[34] Winninghoff FJ (1968) On the adjustment toward a geostrophic balance in a simple primitive equation model with application to the problems of initialization and objective analysis. PhD Thesis, Dept Meteor Univ California, Los Angeles, CA 90024.

Velimir R. RADMILOVIĆ*

ATOMISTIČKI FENOMENI U ZNO-IN_2O_3 NANOŽICAMA

Sažetak: Da bi se postigao sklad između proizvodnje energije i zaštite životne sredine, neophodan je potpun prelaz na obnovljive izvore energije i prestanak fosilnih goriva, čija upotreba ugrožava ekosistem naše planete. Da bi se riješio ovaj veoma težak problem, istraživanja idu u mnogim pravcima. Jedan od tih pravaca predstavljaju termoelektrični nanomaterijali, koji pokazuju sposobnost konverzije temperaturnog gradijenta u električnu struju, koji imaju potencijalnu primjenu u izradi senzora, potrošačkoj elektronici, medicinskim uređajima, itd. Termoelektrične $M_2O_3(ZnO)_n$ nanožice, gdje M mogu da budu In, Ga, Fe..., sintetizovane su, koristeći difuzione procese u čvrstom stanju, koji omogućavaju kontrolu defekata u strukturi na atomskom nivou i stvaranje struktura superrešetki, u kojima se stvaraju kompleksne oksidne granične površine između ZnO vurcitne kristalne rešetke i monoatomskih slojeva na bazi indijuma (In). Formiraju se dvije vrste defekata: ravanski, paralelni bazalnim ravnima vurcitne rešetke, i cik-cak defekti paralelni piramidalnim ravnima, koji omogućavaju nezavisnu kontrolu termičkih i električnih osobina. Oba ova faktora olakšavaju postizanje značajnih vrijednosti parametra valjanosti, zT, koji se koristi za rangiranje termoelektričnih materijala i uređaja. Skenirajuća transmisiona elektronska mikroskopija atomske rezolucije potvrdila je prisustvo monoatomskih planarnih defekata na bazi indijuma, normalnih na [0001] pravac rasta ZnO nanožica. Ovi defekti razdvajaju vurcitne $GaZn_nO_{(n+1)}{}^+$ slojeve različitih debljina na nanoskali, u kojima se stvaraju uslovi za kvantnu stiješnjenost, a istovremeno služe kao prepreke za transport fonona. Nepotpune monoatomske slojeve na bazi indijuma karakteriše prisustvo ivičnih dislokacija, koje omogućavaju putanje ubrzane difuzije za velike atome indijuma. Jasno je da se ove dvije alatke kvantne i talasne mehanike, kvantna stiješnjenost i fononsko rasipanje, mogu koristiti za kontrolu termoelektričnih osobina, što bi moglo da bude ključno u razvoju sljedeće generacije nanostrukturnih materijala.

---

* Velimir R. Radmilović, Srpska akademija nauka i umetnosti, Knez Mihailova 35, 11000, Beograd, Srbija; Univerzitet u Beogradu, Tehnološko-metalurški fakultet, Karnegijeva 4, 11120, Beograd, Srbija

UVODNA RAZMATRANJA

Već polovinom prošlog vijeka postalo je jasno da bi nastavak korišćenja fosilnih goriva za proizvodnju energije, sa učešćem od približno 80% u energetskom bilansu naše planete, mogao da dovede do pogubnih posljedica zbog emisije gasova staklene bašte u atmosferu i nepovoljnih klimatskih promjena. Da bi se izbjegle neželjene posljedice, ulaže se veliki napor da se izgrade alternativni energetski kapaciteti, koji se baziraju na obnovljivim izvorima, poput vjetroparkova, solarnih farmi, postrojenja za biomasu, geotermalnih postrojenja i slično. U isto vrijeme, preduzimaju se značajni istraživački napori da se razviju nove i unaprijede postojeće energetske tehnologije, koje omogućavaju prikupljanje toplotne energije, bilo solarne, ili one koja se gubi u primarnoj proizvodnji, kroz razvoj novih termoelektričnih materijala, koji koriste nanostrukture, poput nanožica, nanočestica i nanokompozitnih struktura i koji posjeduju jedinstvenu osobinu da vrše konverziju temperaturnog gradijenta u električnu energiju. Tipičan primjer, na kome je radila naša istraživačka grupa, predstavljaju cink-oksidne nanožice, dopirane elementima kao što su indijum (In), galijum (Ga), željezo (Fe)... koji mogu da se na kontrolisan način uvode u vurcitnu kristalnu strukturu i stvaraju defekte, koji igraju važnu ulogu u kontroli termičke i električne provodljivosti.

Već iz ovih nekoliko uvodnih rečenica jasno je da je ova istraživačka oblast vezana za energiju i energetske tehnologije, bilo da se radi o proizvodnji energije, ili o sakupljanju energije koja se gubi u primarnoj proizvodnji, što je ekvivalentno proizvodnji od oko 30% globalnih instalisanih kapaciteta. Nažalost, glavno ograničenje za dosadašnji skroman učinak u ovom pravcu predstavlja mala efikasnost termoelektričnih (TE) uređaja.

Pojačana zabrinutost za životnu sredinu, zbog proizvodnje energije sagorijevanjem fosilnih goriva, podstakla je interes istraživača širom svijeta, ali i agencija koje finansiraju istraživanja za unapređenje postojećih i pronalaženje novih tehnologija, čiji bi rezultat bio proizvodnja čiste energije (u popularnim tekstovima se često naziva zelena energija) iz obnovljivih izvora, ekonomski dostupne svim stanovnicima naše planete. Dok se istraživački napor ulaže u mnogim uzbudljivim pravcima za proizvodnju zelene energije, jedna od istraživačkih avenija, koja izaziva sve više pažnje naučne javnosti u posljednjih dvadesetak godina, odnosi se na termoelektrične materijale, koji omogućavaju da se sačuva dio toplote izgubljene u primarnoj proizvodnji energije, koja, kao što je već rečeno, trenutno iznosi oko 6,5 TW, a koja odgovara količini energije koju bi proizvelo 6500 nuklearnih elektrana, prosječne instalisane snage od 1000 MW.

ATOMISTIČKI FENOMENI U TERMOELEKTRIČNIM NANOŽICAMA

Tomas Zebek (Thomas Johann Seebeck, 1770–1831) objavio je 1821. godine rezultate koji su pokazali da se u spoju raznorodnih metalnih žica, u prisustvu temperaturnog gradijenta, spontano stvara električna struja. Ovu pojavu nazvao je termomagnetnim efektom. Odnos električnog potencijala, V, i temperaturnog gradijenta, ΔT, naziva se Zebekov koeficijent, $S=V/\Delta T$.

Žan Peltije (Jean Charles Athanase Peltier, 1785–1845) 1834. godine je pokazao da u prisustvu električnog potencijala, spoj odgovarajućih električnih provodnika izaziva efekat hlađenja. Odnos protoka toplote, Q, i električne struje, I, naziva se Peltijeov koeficijent, $\Pi=Q/I$. Vilijam Tomson, Lord Kelvin (William Thomson, Lord Kelvin, 1785–1845) 1951. godine objavio je i pokazao da su Zebekov i Peltijeov efekat isti fenomeni i formulisao koncept reverzibilnosti, odnosno da, pod određenim uslovima, termoelektrični par može da proizvodi struju ili izaziva efekat hlađenja. Kelvinov koeficijent predstavlja proizvod Zebekovog koeficijenta, S, i apsolutne temperature, T, $\Pi=S \cdot T$.

Hans Ersted (Hans Christian Ørsted, 1777–1851) i Žan Furije (Jean-Baptiste Joseph Fourier, 1768–1830) zainteresovali su se za ova otkrića i poslije izvjesnog vremena objavili da ove pojave nijesu elektromagnetni, nego termoelektrični fenomeni. Ersted je prvi upotrijebio naziv termoelektrični efekat.

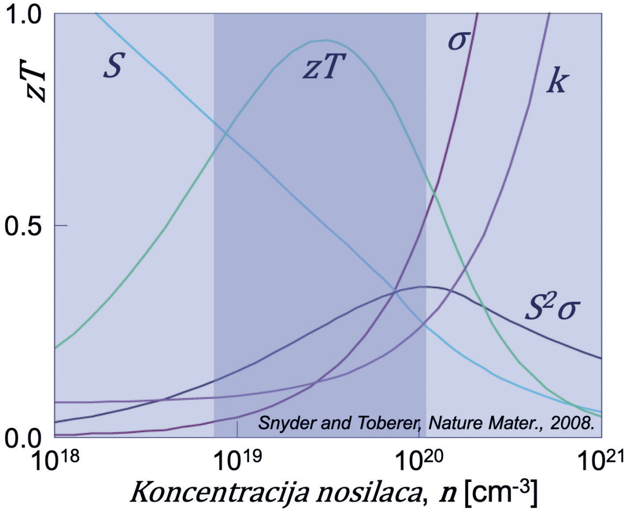
Interesantno je zapaziti da od otkrića termoelektričnog efekta 1821. godine, sljedećih oko 150 godina u ovoj oblasti malo se toga dešavalo, sve do sredine druge polovine XX vijeka [1, 2, 3].

Objašnjenje zašto se čekalo toliko dugo na prodore u ovoj oblasti može da se objasni, koristeći dijagram Snajdera i Toberera [4], prikazan na Slici 1. Sa Slike 1 se zapaža da se sa povećanjem udjela dopirajućeg elementa, kojim se kontroliše električna provodljivost povećava, takođe, i termička provodljivost, a smanjuje vrijednost Zebekovog koeficijenta, što proizlazi iz formule za parametar valjanosti, čija vrijednost ukazuje na kvalitet termoelektričnog materijala:

$$zT = \frac{S^2 \cdot \sigma \cdot T}{k}$$

gdje je $S^2 \cdot \sigma$ faktor snage (PF), čije povećanje ukazuje na veću sposobnost vršenja rada, S — Zebekov koeficijent, σ — koeficijent električne, a k — koeficijent toplotne provodljivosti. Iz ove formule lako je zaključiti da dobar termoelektrični materijal karakteriše velika vrijednost Zebekovog koeficijena, S, veliki koeficijent električne provodljivosti, a mala vrijednost koeficijenta termičke provodljivosti, k. Postizanje ovakvih osobina cilj je svakog istraživanja u oblasti poboljšanja postojećih i razvoja novih termoelektričnih materijala. Nažalost, ova tri parametra, kao što se to može vidjeti sa Slike 1, međusobno su nepovoljno zavisna, pa ih je samim tim teško optimizovati istovremeno.

Koeficijent toplotne provodljivosti, k, u jednačini za parametar valjanosti, zT, sastoji se iz dva člana, $k = k_e + k_l$, gdje k_e označava doprinos elektronske toplotne provodljivosti, a k_l: — toplotne provodljivosti rešetke.



Slika 1. Međuzavisnost parametara S — Zebekovog koeficijenta, σ — koeficijenta električne i k — koeficijenta toplotne provodljivosti u izrazu za parametar valjanosti, zT [4]

Prema Videman–Francovom (Gustav Heinrich Wiedemann, 1826–1899; Johann Carl Rudolf Franz, 1826–1902) zakonu, doprinos elektronske toplotne provodljivosti proporcionalan je, za datu temperaturu, električnoj provodljivosti, σ , i Lorencovoj konstanti, L.

$$k_e = L \cdot \sigma \cdot T \;\; ; \text{за } \sigma = \mu \cdot n \cdot e \;\; ; k_e = L \cdot \mu \cdot n \cdot e \cdot T$$

gdje su: σ — koeficijent električne provodljivosti, L — Lorencov broj (Ludvig Valentin Lorenz, 1829–1891), μ — pokretljivost nosilaca naelektrisanja (hemijski potencijal), n — koncentracija nosilaca naelektrisanja (broj valentnih elektrona u dopirajućim elemenatima). Lorencov broj može da se izračuna na osnovu empirijskog Videman–Francovog zakona, koji uspostavlja odnos između koeficijenata toplotne i električne provodljivosti, gdje je L — konstanta proporcionalnosti:

$$\frac{k}{\sigma} = L \cdot T = \frac{\pi^2}{3}\left(\frac{k_B}{e}\right)^2 \cdot T$$

koja za većinu metala ima istu vrijednost i iznosi $2.44 \times 10^{-8}\ J^2 \cdot K^{-2} \cdot C^2$ (odnosno, $2.44 \times 10^{-8}\ V^2 \cdot K^{-2}$). Kasnije se, međutim, ispostavilo da postoje odstupanja od ovog pravila, odnosno da Lorencov broj nije konstanta, već odstupa od navedene vrijednosti, posebno u poluprovodničkim materijalima [5], ali i u nekim staklastim materijalima i oksidnim strukturama. Razlog ovome je što je doprinos elektronske toplotne provodljivosti, k_e, ukupnom koeficijentu provodljivosti, k, manji od doprinosa toplotne provodljivosti rešetke, k_ℓ.

Kada su u pitanju poluprovodnički materijali, za doprinos toplotne provodljivosti rešetke, k_ℓ, koeficijentu toplotne provodljivosti, k, odgovorni su *fononi**, odnosno njihov transport kroz kristalnu rešetku fonona:

$$k_\ell = k_f = \frac{1}{3} C_V \cdot v \cdot l_{MFP}$$

Za slučaj da rasipanje fonona kontroliše *N* različitih mehanizama (fonon-fonon, granične i slobodne površine, defekti, čestice sekundarnih faza...), srednje slobodno rastojanje l_{MFP} izračunava se na osnovu formule:

$$\frac{1}{l_{MFP}} = \sum_{i=1}^{N} \frac{1}{l_i}$$

Ukupno vrijeme relaksacije fonona, τ_T, izračunava se prema izrazu [6]:

$$\frac{1}{\tau_T} = \sum_{i=1}^{N} \frac{1}{\tau_i}$$

* Toplota se prostire kroz materijal elastičnim talasima, koji nastaju usljed usklađene kolektivne oscilacije (vibracije) atoma u kristalnoj rešetki. Slično drugim vrstama elektromagnetnog zračenja, energija ovih vibracionih talasa je kvantovana. Pojedinačni kvanti vibracione toplotne energije rešetke nazivaju se *fononi*. Izraz *fonon* koristi se često u literaturi kao sinonim za vibracione talase.

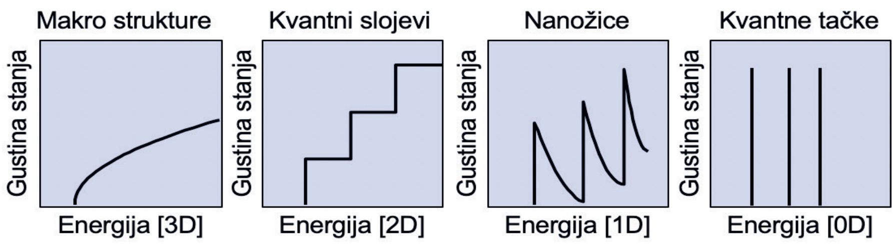
Ovdje su: e — naelektrisanje elektrona; n — koncentracija nosilaca naelektrisanja; μ — mobilnost nosilaca naelektrisanja (hemijski potencijal); k_B _ Bolcmanova konstanta; ν — brzina prostiranja fonona (srednja brzina prostiranja zvuka); τ — vrijeme relaksacije fonona (vrijeme između dva uzastopna događaja rasipanja fonona); l_{MFP} — srednje slobodno rastojanje fonona; C_V — toplotni kapacitet (specifična toplota kristalne rešetke).

U pogledu optimizacije termoelektričnih uređaja postoje dvije strategije [4, 7]:

Prva mogućnost odnosi se na promjenu elektronske strukture, povećanje gustine stanja u blizini Fermijevog nivoa, korišćenjem *kvantne stiješnjenosti*, čime se može uticati na vrijednost Zebekovog koeficijenta u nanostrukturama, odnosno na vrijednost *termosnage*, $\sigma \cdot S^2$. Ovo može da se postigne smanjenjem jedne ili više dimenzija TEG uređaja, kao što je to shematski prikazano na Slici 2. Ukoliko je rasipanje elektrona nezavisno od energije ($\mu E = \mu$), tada je električna provodljivost direktno proporcionalna *gustini stanja* (engleski naziv: *density of states*, DOS). Mot–Džonsova jednačina omogućava nam da izračunamo Zebekov koeficijent (ona važi samo za metale i visokodegenerisane poluprovodnike) [7, 8]:

$$S = \frac{8\pi^2 k_B^2}{3eh^2} m^* T \left(\frac{\pi}{3n}\right)^{\frac{2}{3}}$$

gdje je h — Plankova ($6{,}63 \times 10^{-34}$J·s), a k_B — Bolcmanova konstanta ($1{,}38 \times 10^{-23}$J/atom·K; $8{,}62 \times 10^{-5}$eV/atom·K), e — naelektrisanje elektrona ($1{,}6 \times 10^{-19}$C), m^* – efektivna masa nosilaca naelektrisanja, n – koncentracija nosilaca naelektrisanja.



Slika 2. Shematski prikaz promjene gustine stanja u funkciji energije, za 3D, 2D, 1D i 0D strukture [7]

Odavde slijedi da će povećanje Zebekovog koeficijenta uticati na povećanje termosnage povećanjem nagiba *gustini stanja* na Fermijevom nivou, E_F, kroz povećanje *efektivne mase* nosioca naelektrisanja jer su sve ostale veličine konstante, odnosno [5]:

$$S \sim \left(\frac{\partial[DOS(E)]}{\partial E}\right)_{E=E_F}; \quad DOS(E)dE = \frac{4\pi\sqrt{(2m*)^3 \cdot E}}{h^3}dE$$

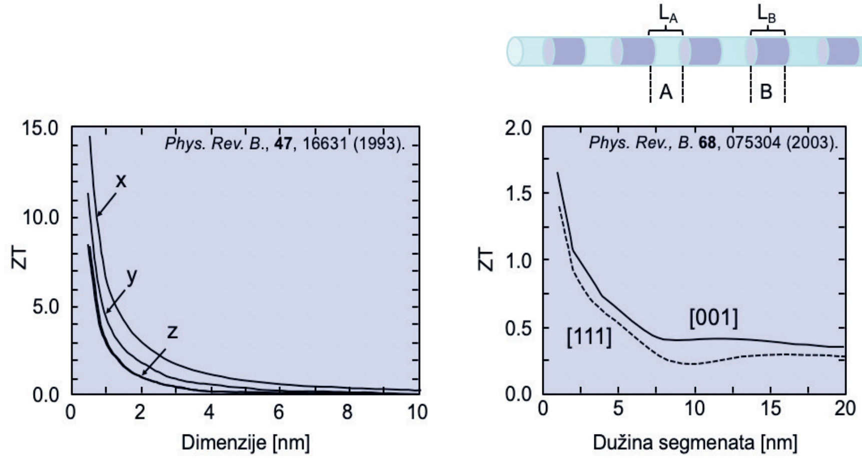
Povećanje termosnage, odnosno efektivne mase, m^*, uticaće na povećanje parametra valjanosti, ZT, a samim tim na poboljšanje efikasnosti termoelektričnog materijala i uređaja.

S obzirom da se pri tome ne mijenja koeficijent toplotne provodljivosti, može da se zaključi da će smanjenje dimenzija termoelektričnog uređaja kao krajnji rezultat imati povećanje *parametra valjanosti*, ZT. Ovo je prvi put pretpostavljeno na osnovu rezultata modeliranja i simulacije ovih fenomena u grupi M. Dreselhaus (Mildred S. Dresselhaus) na Masačusetskom institutu za tehnologiju, kao što je to prikazano na Slici 3 (lijevo) [8], a zatim eksperimentalno potvrđeni u brojnim sistemima: kvantnim tačkama, superešetkama, tankim filmovima i nanokompozitnim strukturama.

Druga mogućnost je povećanje broja različitih strukturnih elemenata (defekata), kao što su: greške u redosljedu pakovanja gusto složenih atomskih ravni, dvojnici, granice inverznih domena, rastvoreni atomi, dislokacije, uključci, granice i slobodne granične površine, čestice sekundarnih faza, itd. Prisustvo ovih defekata može veoma značajno da utiče na rasipanje fonona, a u konačnom na povećanje parametra valjanosti*. Rezultat jedne takve teorijske analize uticaja dužine segmenata u superstrukturi nanožica na parametar valjanosti, *ZT*, prikazan je na Slici 3 (desno). Rasipanje visokoenergetskih fonona, talasne dužine istog reda veličine kao defekti u kristalnoj rešetki, posebno onih koji uzrokuju lokalni gubitak translacione simetrije, može da se koristi kao dodatna „kvantna alatka" ili „alatka talasne mehanike" u optimizaciji osobina *TEG* uređaja, kojim se smanjuje toplotna provodljivost, dok ostala dva parametra, Zebekov koeficijent i električna provodljivost ostaju nepromijenjeni.

---

* U našoj literaturi često se za ZT pogrešno koristi termin „parametar pouzdanosti". Valjanost nekog uređaja predstavlja granicu do koje pokazuje ono što se očekuje da proizvede. Pouzdanost se odnosi na konzistentnost proizvedenih rezultata.



Slika 3. Promjena vrijednosti *parametra valjanosti*, *ZT,* u funkciji dimenzija *TEG* uređaja (lijevo) [12] i dužine PbS i PbSe segmenata nanokompozitne strukture nanožice debljine 10 nm, za [001] i [111] kristalografske pravce kubne *Fm3m* kristalne superrešetke (desno) [8, 9]

Kada je u pitanju korišćenje nanokompozitnih struktura u kontroli rasipanja fonona, važno je imati na umu da fononski spektar odgovoran za prenos toplote karakteriše prisustvo različitih talasnih dužina, od nekoliko nanometra do nekoliko mikrona. Za rasipanje fonona poželjno je da strukturni konstituenti ne budu, po veličini i međusobnom rastojanju, na istoj skali, jer se tim obezbjeđuje da centri za rasipanje budu efektivni u blokiranju prostiranja fonona različitih talasnih dužina. Pri tome ne bi trebalo da se utiče na srednje slobodno rastojanje za kretanje elektrona.

Interesantno je pomenuti da su promjena elektronske strukture usljed kvantne stiješnjenosti [8, 9] i pojačano rasipanje fonona [10, 11] već bili pojedinačno korišćeni u optimizaciji parametra pouzdanosti [12, 13, 14, 15], ali nijesu bili implementirani istovremeno u nekom nanostrukturnom termoelektričnom materijalu.

Naša ideja je bila da implementiramo oba koncepta, odnosno da upotrijebimo obje *kvantne alatke*, *kvantnu stiješnjenost*, za povećanje gustine stanja u blizini Fermijevog energetskog nivoa da bi mogli da povećamo *faktor snage*, i da kontrolisano unesemo *prepreke fononskom transportu u obliku monoatomskih graničnih površina na bazi indijum-oksida*, i tako značajno smanjili dopinos rešetke toplotnoj provodljivosti. Na taj način je bilo za očekivati značajno povećanje vrijednosti parametra valjanosti, *ZT*, a na taj način i povećanje efikasnosti konverzije, odnosno koeficijenta iskorišćenja, η, toplotnog gradijenta u električnu struju; time bi termoelektrični generatori (TEG) mogli da postanu konkurentni drugim toplotnim mašinama.

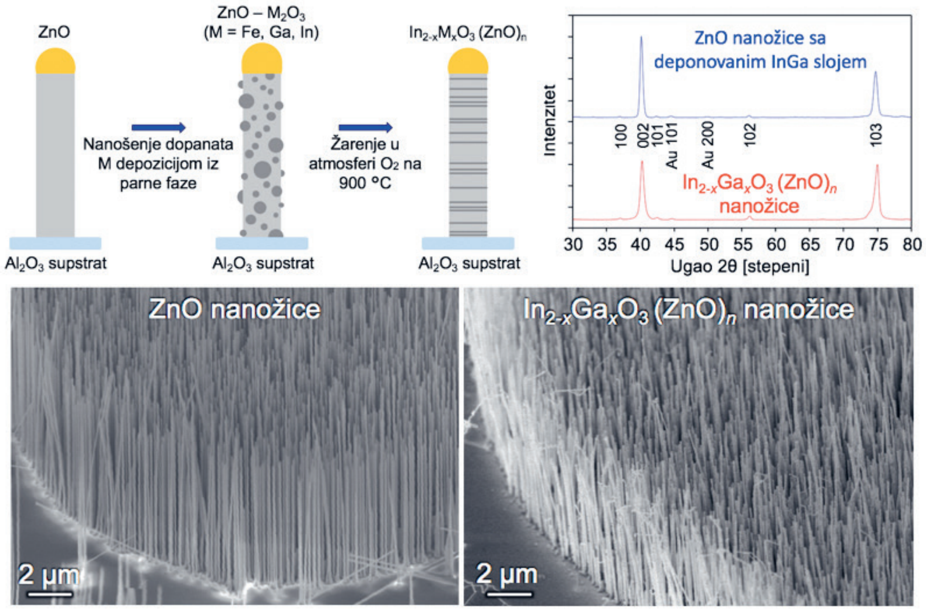
Kao model sistem za ispitivanje termoelektričnih fenomena u funkciji Zebekovog koeficijenta, električne i toplotne provodljivosti, odabrane su ZnO nanožice, dopirane, pojedinačno ili u kombinaciji elemenata, indijumom, galijumom i željezom. ZnO pripada II–VI grupi poluprovodničkih materijala sa energijskim procijepom od 3,4 eV, koji može da ima potencijalno široku primjenu u elektronici i optoelektronici [16]. Dopiranje ZnO nanožica prečnika oko 100 nm, recimo, implantiranjem galijuma (Ga), smanjuje toplotnu provodljivost za red veličine, sa 12,7 $Wm^{-1}K^{-1}$ na 1,22 $Wm^{-1}K^{-1}$, zahvaljujući rasipanju fonona na atomima galijuma, a povećava električnu provodljivost, zahvaljujući tome što Ga kao dopant povećava broj doniranih elektrona za red veličine, sa $1{,}01 \times 10^3\ \Omega^{-1}m^{-1}$ na $1{,}46 \times 10^4\ \Omega^{-1}m^{-1}$ [17].

Električna mjerenja vršena su na individualnim ZnO nanožicama dopiranim sa In/Ga sa odnosom naparenih slojeva 5/5 nm (5 nm In i 5 nm Ga), koje su nanešene na silicijumske čipove sa oksidnim slojem na površini, debljine 200 nm. Toplotni transport je mjeren uobičajenim metodama opisanim u literaturi [18].

Koristeći trostepeno procesiranje ZnO nanožica, difuzijom u čvrstom stanju, razvijena je nova metoda koja omogućava sintezu $M_2O_3(ZnO)_n$ legiranih nanostruktura sa mogućnošću kontrole sastava, strukture i električnih/elektronskih osobina, koje mogu da budu od interesa za elektronsku i optoelektronsku primjenu.

Na Slici 4 prikazan je shematski proces sinteze, po završetku rasta polaznih ZnO nanožica, a koji uključuje dopiranje i stvaranje politajpoidne nanostrukture, skenirajuće elektronske mikrofotografije ZnO i 40/40 nm In/Ga metala (M) $M_2O_3(ZnO)_n$ nanožica, kao i XRD difraktogrami ovih struktura dobijeni korišćenjem Cu-K_α zračenja.

Analizom SEM mikrofotografija zapaža se da je morfologija nanožica očuvana poslije procesa dopiranja i žarenja na 900°C, odnosno legiranja sa kombinacijom metala In/Ga, što ukazuje na potencijal ove metode za integraciju. Jaka (002) linija na XRD difraktogramu ukazuje da nanožice zadržavaju istu orijentaciju, odnosno vertikalnost uspostavljenu u toku rasta čistih ZnO nanožica. Važno je, takođe, uočiti da na difraktogramu nijesu prisutni pikovi koji pripadaju In_2O_3 ili Ga_2O_3. U 40/40 nanožici IGZO uočeni su pikovi za koje smo, mjerenjem međuravanskih rastojanja, odnosno parametara rešetke, ustanovili da iznose 8,6 Å i 11,2 Å. Ove vrijednosti su u vrlo dobroj saglasnosti sa proračunatim konstantama rešetke za $M_2O_3(ZnO)_n$ nanožice sa vrijednostima indeksa $n = 1$ i $n = 2$. Prisustvo In, Ga i Zn potvrđeno je energo-disperzionom (EDS) mikrohemijskom analizom.



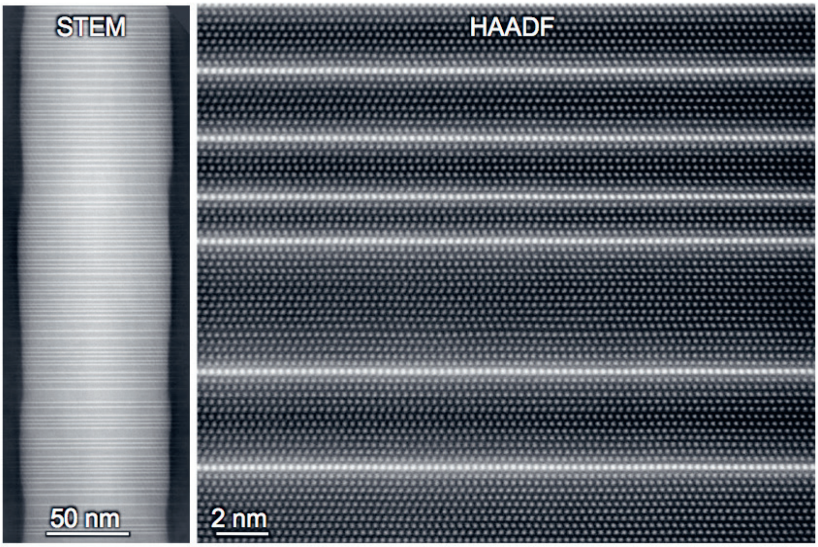
Slika 4. Shematski prikaz procesa sinteze, po završetku rasta polaznih ZnO nanožica, koji pokazuje dopiranje i stvaranje politajpoidne nanostrukture, skenirajuće elektronske mikrofotografije (SEM) ZnO nanožica, prije i poslije dopiranja sa 40/40 nm In/Ga metala (M), $M_2O_3(ZnO)_n$ nanožica, kao i XRD difraktograme ovih struktura, dobijene Cu-K_α zračenjem [19]

Za visokorezolucionu analizu atomske strukture korišćena je skenirajuća transmisiona elektronska mikroskopija Z-kontrasta (kontrast atomskog broja; kada se za formiranje slike koriste elektroni rasuti pri velikim uglovima, preko 50 miliradijana; ova metoda se naziva HAADF (skraćenica engleskih riječi: High Angle Annular Dark Field)), na TEAM mikroskopu pri radnom naponu od 300 kV, sa dvostruko korigovanom sfernom aberacijom (CS), i sa atomskom rezolucijom do nivoa od oko 50 pikometara.

Na Slici 5 prikazane su Z-kontrast transmisione elektronske mikrofotografije, pri malom povećanju (lijeva) i slika atomske rezolucije (desna) nanožice 10/10 IGZO. S obzirom da je intenzitet na Z-kontrast slikama proporcionalan kvadratu atomskog broja:

$$I = A \cdot Z^{\beta}$$

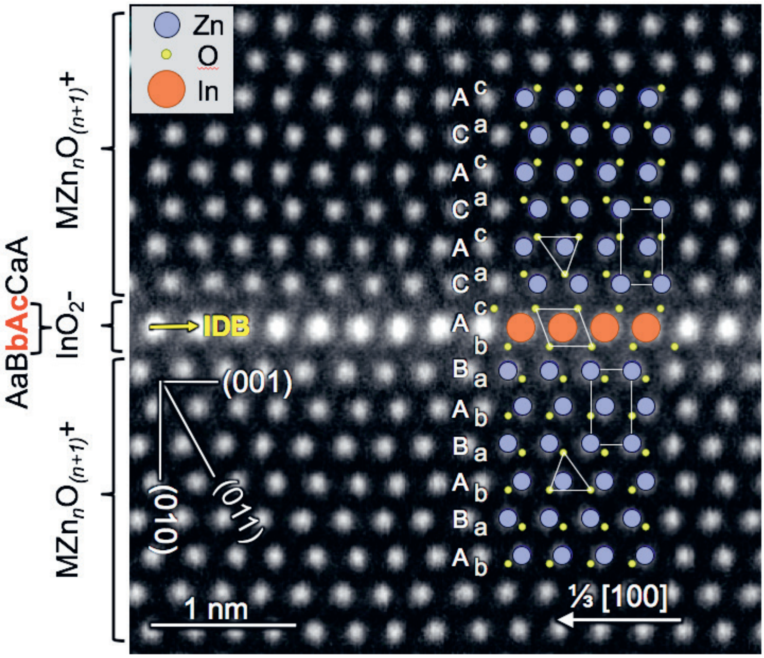
gdje β — obično ima vrijednosti između 1,2 i 2, A je konstanta, a Z — atomski broj hemijskog elementa. S obzirom da se u ZnO slabovima jasno vide atomske kolone Zn, to njihov intenzitet može da posluži kao interna kalibracija za precizno određivanje koeficijenta β, pa je lako utvrditi da su svijetle monoatomske ravni sastavljene od atoma indijuma. Atome kiseonika nije moguće vidjeti istovremeno sa atomima cinka i indijuma zbog male vrijednosti atomskog broja kiseonika, u odnosu na atomske brojeve cinka i indijuma.



Slika 5. Z-kontrast skenirajuće transmisione elektronske mikrofotografije (kvantne mape) ispitivane strukture nanožice 10/10 IGZO, u blizini [100]ZnO, pri malom povećanju (lijeva) i slika atomske rezolucije (desna) [19]

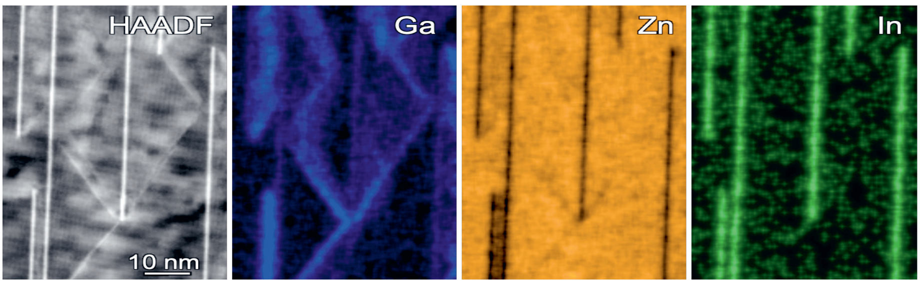
Visoko rezoluciona STEM (HAADF) slika jasno pokazuje da je In (svijetle tačke na Slici 5 — desno) prisutan u obliku monoatomskih slojeva, na međusobno različitom rastojanju, odnosno koji razdvajaju vurcitne slabove različite debljine. Kristalografski i strukturni detalji dati su na Slici 6. Granica inverznih domena formira se tako što se jedan slab ZnO pomjera u odnosu na susjedni slab ZnO, za ⅓ duž pravca [100] vurcitne kristalne rešetke, pa se umjesto redosljeda slaganja baznih ravni AaBbAaBbA..., dobija sekvenca sa greškom u redosljedu: AaBbAcCaA, što za rezultat ima formiranje granice inverznih domena (IDB je skraćenica engleskih riječi: Inversion Domain Boundary).

U svim ispitivanim uzorcima samo su prisutne monoatomske ravni indijuma. U suštini, kao što je već rečeno, ove monoatomske ravni, formiraju se na (002) ravnima vurcitne rešetke, poprečno na pravac rasta nanožica, a sastoje se od oktaedarski koordinisanih InO_{2-} molekula (ili opštije MO_{2-}, gdje M može da bude Al, In, Ga, Fe...), za koje su proračuni na bazi prvog principa pokazali da predstavljaju najstabilniju konfiguraciju indijuma u ZnO vurcitnoj kristalnoj superrešetki [20].



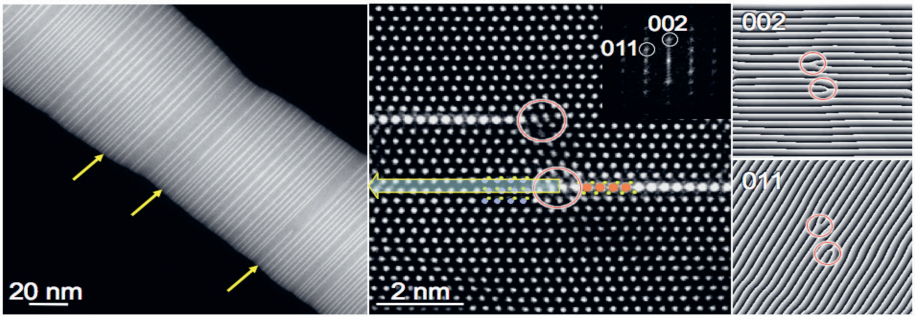
Slika 6. Eksperimentalna Z-kontrast slika atomske rezolucije na kojoj se vide kristalografski i strukturni detalji IGZO nanožica; IDB označava granicu inverznih domena

Energodosperziono spektroskopsko (EDS) mapiranje K linijama pokazalo je da galijum nije prisutan u monoatomskim slojevima, već samo indijum. Galijum je, u principu, prisutan u slabovima između monoslojeva na bazi indijuma i u cik-cak defektima, kao što se to može vidjeti na Slici 7.



Slika 7. HAADF slika i EDS mape elemenata prisutnih u nanožicama 10/10 IGZO. Interesantno je pomenuti, da je galijum predominantno raspoređen između monoatomskih slojeva na bazi indijuma, ili u cik-cak defektima, što se jasno vidi na EDS mapi galijuma.

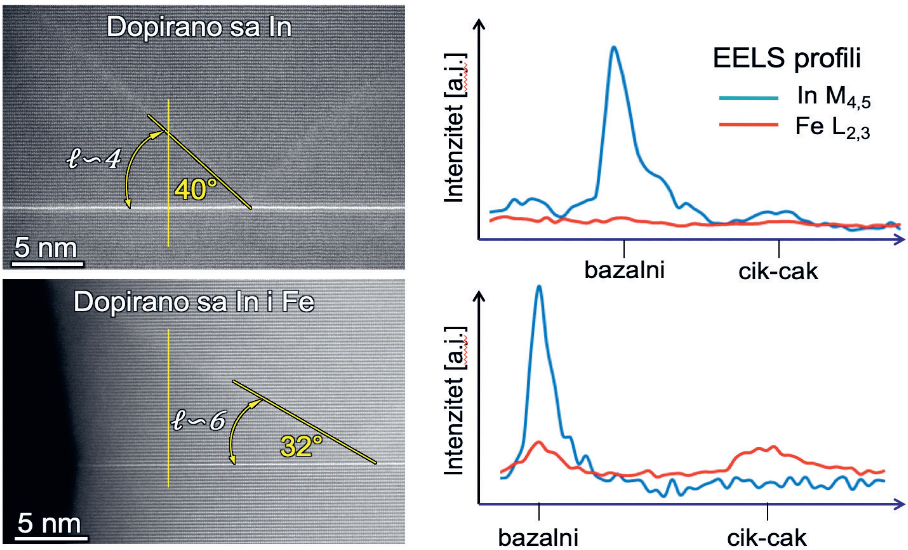
Formiranje monoatomskih uključaka na bazi indijuma u vurcitnoj kristalnoj rešetki nije bilo moguće objasniti samo mehanizmom zapreminske difuzije indijuma kroz rešetku ZnO. Razlog tome je bio što maksimalno rastojanje između bazalnih vurcitnih ravni iznosi 266 pikometara, dok je prečnik atoma indijuma oko 346 pikometara. Analiza velikog broja Z-kontrast mirofotografija omogućila nam je da ustanovimo da neke monoatomske ravni koje sadrže indijum, ne prostiru se po cijelom poprečnom presjeku nanožica (označene žutim strelicama). Ispostavilo se da taj fenomen sadrži ključ za razumijevanje nukleacije i širenja monoatomskih ravni indijuma po presjeku nanožica, kao što je to prikazano na Slici 8.



Slika 8. Na Z-kontrast mikrofotografiji male rezolucije (lijevo) žutim strelicama označeni su monoatomski uključci na bazi indijuma, koji nijesu prošireni po cijelom poprečnom presjeku nanožice. U sredini je Z-kontrast mikrofotografija atomske rezolucije, koja pokazuje raspored atoma indijuma i cinka na frontu rasta monoatomskih uključaka; Moare slike na desnoj strani, napravljene geometrijskom faznom analizom Z-kontrast slike atomske rezolucije, koristeći refleksije 002 i 011; krugovi označavaju područja u kojima se nalaze jezgra ivičnih dislokacija [19]

Parcijalno proširene monoatomske ravni indijuma (označene žutim strelicama) pokazale su da se na difuzionom frontu formira jezgro ivične dislokacije, za koje je mjerenjem ustanovljeno da ima prečnik oko 500 pikometara. Ovako veliki prostor jezgra ivične dislokacije dovoljan je za nesmetanu difuziju indijumovih atoma, njihovu ugradnju u vurcitnu rešetku, i formiranje monoatomskih ravanskih defekata paralelnih (002) ravnima, poprečno na osu nanožica. Moare slike (Slika 8 (desno), napravljene sa 002 i 011 refleksijama vurcita) takođe jasno pokazuju prisustvo ekstraravni, kao osnovnih pokazatelja prisustva ivičnih dislokacija u strukturi nanožica. Rast monoatomskih ravni na bazi indijuma i širenje kroz poprečni presjek nanožica sa vurcitnom rešetkom ZnO, bilans mase i električna neutralnost sistema zahtijevaju da dva monoatomska sloja cinka i jedan sloj kiseonika difunduju na površinu, što za posljedicu ima modulaciju prečnika nanožica.

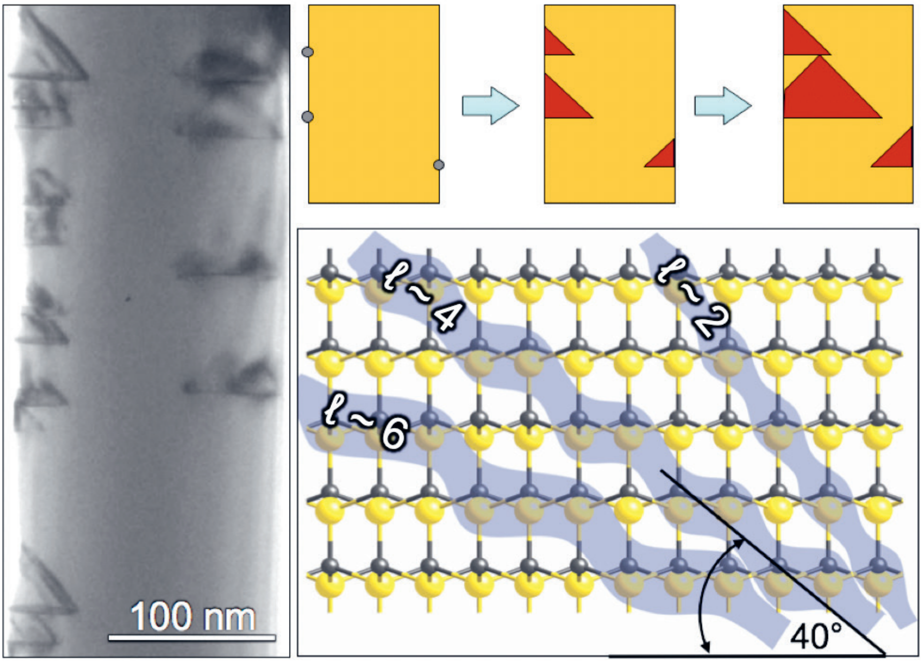
Ugao pod kojim se formiraju cik-cak defekti zavisi od vrste i kombinacije metala kojim se dopiraju ZnO nanožice. Što je prečnik dopirajućeg atoma veći to je veći ugao pod kojim se formira cik-cak defekat. Eksperimentalno je ustanovljeno (Slika 9) da u slučaju dopiranja samo sa indijumom cik-cak greške se formiraju pod uglom od približno 40° u odnosu na bazalnu ravan ZnO vurcitne kristalne rešetke. Ukoliko se nanožice dopiraju sa željezom i indijumom, tada se cik-cak greške formiraju pod uglom od približno 32° u odnosu na bazalnu ravan ZnO. Proračuni na bazi teorije funkcionalnih gustina (Slika 10) potvrdili su da ugao pod kojim se formiraju cik-cak defekti direktno zavisi od kombinacije elemenata kojima se vrši dopiranje ZnO nanožica. Proračun je pokazao da je u slučaju dopiranja sa indijumom, ugao cik-cak grešaka iznosi oko 39°, dok pri dopiranju sa indijumom i željezom, ugao je oko 32°, što je u prilično dobroj saglasnosti sa eksperimentalno izmjerenim vrijednostima [21].



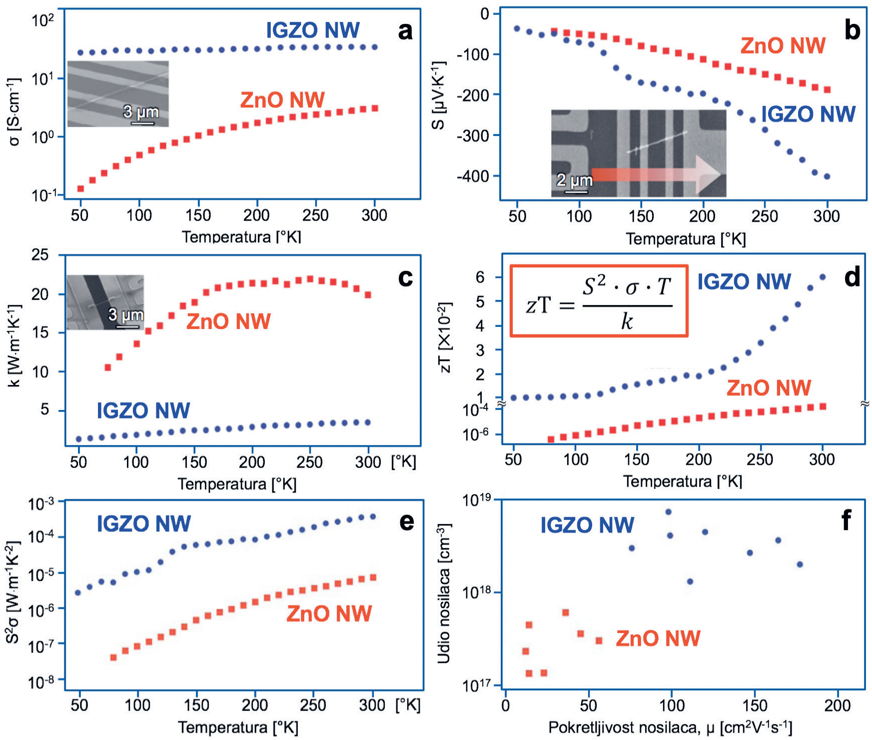
Slika 9. Mikrohemijska spektroskopska analiza na bazi izgubljene energije upadnih elektrona (EELS), raspodjele dopirajućih elementa u ZnO nanožicama dopiranim indijumom (lijevo) ili indijumom i željezom (desno); žute vertikalne linije označavaju rastojanje duž kojih je pravljena EELS analiza [21]

Eksperimentalni rezultati prikazani na Slici 9 pokazuju da ugao pod kojim se stvaraju cik-cak defekti zavisi od veličine dopirajućih atoma. Pored ravanskih grešaka, monoatomskih uključaka na bazi indijuma, u ovim strukturama je zapaženo formiranje cik-cak grešaka. Vrijednosti $\ell \sim 4$ i $\ell \sim 6$ na Slici 9 predstavljaju treći Millerov indeks u oznakama < $uv\ell w$ >, koje definišu nagib u odnosu na bazalnu ravan i pravac prostiranja cik-cak defekata.

Eksperimentalni rezultati i proračuni na bazi teorije funkcionalnih gustina (Slike 10) potvrđuju da formiranje ravanskih MO_{2-} defekata, koji predstavljaju granicu inverznih domena (IDB), istovremeno mora da bude praćeno formiranjem prateće granice inverznog domena, koji čini cik-cak sloj, koji ispravlja prvu IDB na bazalnoj ravni ZnO. Ispitivanjem formiranja ravanskih i cik-cak grešaka u ZnO nanožicama, u ranoj fazi žarenja uzoraka na 900°C, može da se zaključi da se ove greške formiraju istovremeno. Mehanizam nastanka i njihovo širenje kroz poprečni presjek nanožica sastoji se od ugradnje atoma legirajućih elemenata u površinu nanožice (shematski prikazano na Slici 10, desno, gore), formiranja jezgara ivičnih dislokacija i pojačane difuzije duž dislokacionih linija do ispunjavanja cjelokupnog poprečnog presjeka monoatomskim slojevima na bazi indijuma. Ustanovljeno je da se u cik-cak defekte ugrađuju oba legirajuća elementa, indijum i željezo.



Slika 10. Transmisiona elektronska mikrofotografija svijetlog polja IGZO nanožice na početku procesa žarenja (lijevo); shematski prikaz ugradnje dopirajućih atoma u površinu nanožica i stvaranja istovremeno ravanskih i cik-cak defekata (desno, gore); rezultati proračuna na bazi teorije funkcionalnih gustina (desno, donja slika) [21]



Slika 11. Mjerenje električnih osobina pojedinačnih nanožica: (a) električna provodljivost, σ, u funkciji temperature za ZnO (crvene tačke) i IGZO nanožica (plave tačke); (b) Zebekov koeficijent, *S*, za iste nanožice ispitivane pod (a); umetak na slici prikazuje SEM mikrofotografiju individualne nanožice u MEMS uređaju mjerenjem u 4 tačke; (c) toplotna provodljivost, *k*; (d) parametar valjanosti, *zT*; (e) faktor snage, $\sigma \cdot S^2$, u funkciji temperature za ZnO i IGZO nanožice; (f) odnos pokretljivosti nosilaca, μ, i njihove koncentracije [19]

Električna mjerenja (Slika 11), kao što je već ranije rečeno, vršena su na pojedinačnim ZnO i IGZO nanožicama, korišćenjem odgovarajućih MEMS ili uređaja konstruisanih u našoj laboratoriji. Sa Slike 11a se zapaža povećanje električne provodljivosti u IGZO, u odnosu na ZnO nanožice, koja je razumljiva jer se radi o blago dopiranom poluprovodničkom materijalu. Potpuno je jasno da prisustvo InO_{2-}, ili generalno MO_{2-} uključaka modifikuje oboje, termičku i električnu provodljivost politajpoidnih IGZO nanožica.

Nije neobično da ZnO demonstrira relativno veliku intrinzičnu električnu provodljivost, zahvaljujući stvaranju atomskih praznina kiseonika, koje imaju ulogu donora, pa mali dodaci dopirajućih elemenata, poput In, Ga, ili Al, uspješno povećavaju električnu provodljivost [22].

Kada se In i Ga ugrađuju u mjestima sa trigonalnom bipiramidalnom koordinacijom time obezbjeđuju elektroneutralnost sistema. Tako ugrađeni atomi In i Ga smanjuju broj nosilaca naelektrisanja. Pored toga, MO_{2-} uključci mogu da budu barijera za difuziju kiseonika, što može da vodi smanjenju provodljivosti zbog otežanog stvaranja atomskih praznina kiseonika [23].

Parametar valjanosti, *zT*, izračunat na osnovu mjerenja termoelektričnih parametara, Zebekovog koeficijenta, električne i toplotne provodljivosti, pokazuje povećanje do 2,5 reda veličine u IGZO u odnosu na nedopirane ZnO nanožice (Slika 15d).

Važno je pomenuti da procesiranjem, kojim se stvara politipoidna struktura ZnO nanožica, koje se zasniva na difuziji indijuma, galijuma i cinka kroz vurcitnu kristalnu rešetku, možemo da kontrolišemo gustinu raspodjele monoatomskih ravni na bazi indijuma, odnosno gustinu raspodjele MO_{2-} uključaka.

ZAKLJUČCI

Na bazi eksperimentalnih rezultata sinteze i karakterizacije oksidnih termoelektričnih nanožica i proračuna na bazi prvog principa, možemo da zaključimo da je razvijena nova metoda sinteze nanostruktura tipa $In_xM_{2-x}O_3(ZnO)_n$, (M = In, Ga, Fe...). Obezbijeđena je istovremena kontrola strukture i sastava, a time i mogućnost smanjenja termičke provodljivosti, povećanje električne provodljivosti i Zebekovog koeficijenta, odnosno faktora snage, što je dovelo do drastičnog povećanja parametra valjanosti, *zT*.

Ispitivanja su pokazala da je ovakve efekte moguće postići, koristeći alatke kvantne i talasne mehanike, koje su imale za cilj:

— Povećanje gustine stanja u blizini Fermijevog energetskog nivoa, čime se povećava vrijednost Zebekovog koeficijenta, a time i vrijednost faktora snage.

— Kontrolu kretanja fonona duž ZnO nanožica projektovanjem i kontrolom strukture defekata, čime se smanjuje toplotna provodnost rešetke, što vodi daljem povećanju parametra valjanosti, *zT*. Ukupno povećanje *zT* u $In_xM_{2-x}O_3(ZnO)_n$ nanožicama iznosilo je za preko 2,5 redova veličine u odnosu na čiste ZnO nanožice.

— Zahvaljujući prisustvu InO_{2-} uključaka, istovremeno je moguće primijeniti obje alatke kvatne i talasne mehanike, stvaranje inverznih domena, dimenzija koje omogućavaju pojavu kvantne stiješnjenosti, i kontrolu rasipanja fonona unošenjem odgovarajuće vrste i gustine monoatomskih defekata na bazi indijuma, što može da doprinese povećavanju parametra valjanosti, zT.

Razumijevanje fenomena i bolja kontrola strukture na atomskom nivou može da predstavlja ključni faktor u razvoju nove generacije termoelektričnih materijala, čime bi se doprinijelo da ovi zeleni izvori energije postanu konkurentniji toplotnim mašinama u kojima se vrši konverzija hemijske energije sagorijevanjem fosilnih goriva.

ZAHVALNICA

Autor ovog priloga želi da zahvali Srpskoj akademiji nauka i umetnosti za višegodišnje kontinuirano finansiranje istaživačkih aktivnosti u okviru projekta Fonda za nauku SANU broj F-141. Posebno zahvaljujem kolegi profesoru P. Jangu sa Univerziteta Kalifornije u Berkliju i njegovim studentima — doktorandima M. Muru, M. Fardi, Š. Endrjusu i A. Goldštajn, za dugogodišnju saradnju. Najveći dio rezultata ispitivanja atomske strukture termoelektričnih nanožica na bazi ZnO, prikazanih u ovom radu, realizovan je u Nacionalnom centru za elektronsku mikroskopiju Molekularne livnice, Lorens Berkli nacionalne laboratorije, u okviru projekta broj DE-AC02–05CH11231, koji je finansiralo Ministarstvo za energiju američke vlade, ali je dio rezultata dobijen u laboratorijama profesora E. Spikera na Erlangen univerzitetu i profesora G. Kotlajtnera na Tehničkom univerzitetu u Gracu, na čemu im najsrdačnije zahvaljujem.

LITERATURA

[1] D. M. Ronje, ed, *Thermoelectrics Handbook: Macro to Nano*, CRS Press; Boca Raton, FL, 2006.

[2] L. E. Bell, „Cooling, heating, generating power, and recovering waste heat with thermoelectric systems", *Science*, 321 (2008), 1457–1461.

[3] C. B. Vining, „An inconvenient truth about thermoelectrics", *Nature Materials*, 8 (2009), 83–85.

[4] G. Jeffrey Snyder and Eric S. Toberer, „Complex thermoelectric materials", *Nature Materials*, 7 (2008), 105–114.

[5] H. Julian Goldsmid, *Introduction to Thermoelectricity*, Second Edition; Springer Series in Materials Science, Springer, 2016.

[6] U. Ozgur, Ya. I. Alivov, C. Liu, A. Teke, M. A. Reshchikov, S. Dogan, V. Avrutin, S-J. Cho, and H. Morkoc, „A comprehensive review of ZnO materials and devices". *Journal of Applied Physics*, 98 (2005), 041301.

[7] J. R. Szczech, JM. Higgins, S. Jin, „Enhancement of the thermoelectric properties in nanoscale and nanostructured materials", J. Mater. Chem, 21 (2011), 4037–4055.
[8] L. D. Hicks, M. S. Dresselhouse, „Thermoelectric figure of merit of a one-dimensional conductor" Phys. Rev. B, 47 (1993), 16631(R).
[9] Y-M. Lin, M. S. Dresselhaus, „Thermoelectric properties of superlattice nanowires", *Phys. Rev. B*, 68 (2003), 075304.
[10] J. H. Lee, G. A. Galli, J. C. Grossman, „Nanoporous Si as an efficient thermoelectric material", *Nano Letters,* 8 (2008), 3750–3754.
[11] D. Donadio, G. Galli, „Atomistic simulations of heat transport in silicon nanowires", *Phys. Rev. Lett,* 102 (2009), 195901.
[12] T. C. Harman, P. J. Taylor, M. P. Walsh, B. E. LaForge, „Quantum dot superlattice thermoelectric materials and devices", *Science,* 297 (2002), 2229–2232.
[13] B. Poudel, et al,. „High-thermoelectric performance of nanostructured bismuth antimony telluride bulk alloys", *Science,* 320 (2008), 634–638.
[14] J. P. Heremans, et al., „Enhancement of thermoelectric efficiency in PbTe by distortion of the electronic density of states", *Science,* 321 (2008), 554–557.
[15] Boukai, A. I. et al., „Silicon nanowires as efficient thermoelectric materials", *Nature* 451, (2008), 168–171.
[16] Z. X. Huang, Z. A. Tang, J. Yu, S. Bai, „Thermal conductivity of nanoscale polycrystalline ZnO thin films", Physica B, 406 (2011), 818–823.
[17] M. Xia, Z. Cheng, J. Han, M. Zheng, C-H. Sow, J. T. L. Thong, S. Zhang, B. Li, „Gallium ion implantation greatly reduces thermal conductivity and enhances electronic one of ZnO nanowires", *AIP Advances*, 4 (2014), 057128.
[18] L. Shi, et al., „Measuring thermal and thermoelectric properties of onedimensional nanostructures using a microfabricated device", *J. Heat Transfer*, 125 (2003), 881–888.
[19] S. C. Andrews, M. A. Fardy, M. C. Moore, S. Aloni, M. Zhang, V. Radmilovic, and P. Yang, „Atomic-level control of the thermoelectric properties in polytypoid nanowires", *Chemical Science*, 2 (2011), 706–714.
[20] J. L. F. Da Silva, Y. F. Yan and S. H. Wei, „Rules of structure formation for the homologous $InMO_3(ZnO)$ n compounds", Phys. Rev. Lett, 100 (2008), 255501.
[21] A. P. Goldstein, S. C. Andrews, R. F. Berger, V. R. Radmilovic, J. B. Neaton, and P. Yang, „Zigzag Inversion Domain Boundaries in Indium Zinc Oxide-Based Nanowires: Structure and Formation", *ACS Nano*, 7 (2013), 10747–10751.
[22] G.-D. Yuan, et al, „Tunable n-type conductivity and transport properties of Ga-doped ZnO nanowire arrays", *Adv. Mater,* 20 (2008), 168–173.
[23] K. Nomura, et al., „Thin-film transistor fabricated in single-crystalline transparent oxide semiconductor", *Science*, 300 (2003), 1269–1272.

Velimir R. RADMILOVIĆ

ATOMISTIC PHENOMENA IN ZNO-IN_2O_3 NANOWIRES

Summary

To reach a proper balance between energy production and environmental protection, it is necessary to have a full transition to renewable energy sources and eliminating the use of fossil fuels which negatively affect the ecosystem of our planet. In order to solve this very difficult task, research has been taking variety of different avenues. One of such avenues is the field of thermoelectric nanomaterials, which demonstrate the conversion of the temperature gradient into electric current, with potential applications in sensors, consumer electronics, medical devices, etc. Thermoelectric $M_2O_3(ZnO)_n$ nanowires, where M could be In, Ga, Fe, synthesized using solid-state diffusion, enabled us to control their defect structure at atomic level and to create an aperiodic superlattice structure, in which complex oxide interfaces are formed between ZnO wurtzite crystal structure and monoatomic indium (In) containing defect layers. Two kinds of defects: planar, parallel to basal wurtzite planes and zigzag, parallel to pyramidal planes, facilitate decoupling of electrical and thermal properties. Both of these factors facilitate achieving a high figure of merit, zT, used to rank thermoelectric materials and devices. Atomic resolution scanning transmission electron microscopy of these nanowires confirmed the presence of In containing planar defects perpendicular to the [0001] direction. These defects separate wurtzite $GaZn_nO_{(n+1)}{}^+$ slabs of various thicknesses at nanoscale enabling quantum confinement effect to take place and act as obstacles for phonon propagation. Incomplete In monoatomic layers are associated with edge dislocations, providing fast diffusion paths for large indium atoms. It is apparent that these two quantum and wave mechanics tools, quantum confinement and phonon scattering, can be used for better control of thermoelectric properties, which could be the key in developing next-generation nanostructured materials.

Pavle R. ANĐUS[*]

INFORMATION FROM *IN VITRO* BIOMEDICAL DIAGNOSTICS — A ROAD TO PERSONALIZED MEDICINE

Abstract: In a number of neurodegenerative diseases (NDs) the production of the anti-neuronal immunoglobulin G (IgG) is a significant feature of the inflammatory process. It was shown that human IgGs may induce diverse physiological effects on neurons and glial cells of animal origin. In an ongoing project we propose to use IgGs for *in vitro* diagnostics of NDs. Based on already known cellular signaling responses recorded by fluorescent markers robust multipurpose processing of a single patient's IgG sample effect on seeded normal cells can give a complex physiological information. This is achieved with microfluidics and automated microscopy towards experimental/clinical personalized diagnostics. Such a medical device is based on 1) the development of procedures based on a lab-on-a-chip microfluidic system with intracellular light sensors; 2) defining the standardized *in vitro* personalized diagnostic protocols; 3) design of a small-scale pilot platform based on automated/miniaturized fluorescence microscopy. Most of these principles have already been partly tested though a EC-H2020 project and a national Innovation Fund project. The interdisciplinarity of our research comprises of the following approaches: Cellular neurophysiology, Biophysics of intracellular fluorescent indicators, Video microscopy of intracellular molecular signaling, Microfluidics and biochip design, Advanced custom-made optics for automated microscopy and Machine learning for signal analysis. The designed personalized diagnostics technology will be applicable for a variety of NDs for a sustained healthcare system.

INTRODUCTION

Although a vast literature exists on successful preclinical and even on some clinical studies of amyotrophic lateral sclerosis (ALS) therapy, this is still a fatal disease without a reliable cure. Often patients being conscious

[*] Pavle R. Anđus, Center for Laser Microscopy, Faculty of Biology, University of Belgrade, Studentski trg 3, 11000 Belgrade, Serbia

throughout the illness are demanding euthanasia. It is a consensus among researchers and clinicians alike that such a poor prognosis is caused because there is a lack of information or its organization for a robust diagnostic tool that can primarily predict but also follow the therapeutic procedure. Since ALS is in 90% of cases of sporadic nature the main problem is to recruit patients for screening early enough. This can be achieved first by screening patients for comorbidities (e. g. FTD or muscle cachexia) or recruiting them based on genetic or proteomic markers (e. g. TDP-43 and FUS/TLS genes) in order to subject them to a robust point-of care diagnostics platform for prediction of disease and/or its stage and later for the follow-up of therapy. Such a multifactorial information — rich diagnostic tool is expected to offer a personalized approach to the disease and better patient stratification. The diagnostic process of ALS immunoglobulin G (IgG) application on stained cell cultures for an automated recording system is of a highly innovative potential. Although experimentally confirmed and used in describing ALS pathogenesis, such phenomena were never used previously in practical terms. Our approach towards standardizing the procedure for a market pilot follows on an innovative path that could be linked to the published [1, 2] and patented technology [3, 4] of automated microscopy for multidimensional cell profiling for personalized diagnostics. This approach led us to the pilot platform for high-throughput multidimensional cell profile analysis upon ALS IgG challenge. In such a setup the multivariate information-rich single cell analysis is a starting point for identifying relationships among ALS IgG effects at a systems level and a step toward phenotypic/physiological profiling at the single-cell level.

BACKGROUND (BASIC) RESEARCH

In a number of neurodegenerative diseases (NDs) the production of the anti-neuronal immunoglobulin G (IgG) is a significant feature of the immunoinflammatory process. It was shown by us and others that, compared to control, IgG from NDs patients may induce diverse physiological effects not only on neurons but also on glial cells of animal origin. Thus, for example, studies on amyotrophic lateral sclerosis (ALS) have shown that human IgG increased intracellular Ca^{2+} in motor neurons, enhanced glutamate release from synapses residing on lower motor neurons and enhanced the release of acetylcholine from the axon terminal at the neuromuscular junction [5–7]. These crucial events in our own early experiments were shown not only on neurons (facilitation of synaptic activity and Ca^{2+} transient change) but also on astroglia (Ca^{2+} transients, vesicle trafficking) [8–10].

The current state of the art findings obtained with ALS IgG can be focused on the following discoveries:

a) Ca^{2+} imaging. Ca^{2+} imaging with fluorescent dyes provided information on intracellular calcium mobilization in response to ALS IgGs on several cultured cell types (neurons, glia, and cell lines [11]);

b) Reactive Oxygen Species (ROS) imaging. Pilot experiments [12] on microglial cell lines (BV-2) transfected with ROS-sensitive fluorescent constructs have shown that ALS IgGs induce acute ROS signaling.

c) Synaptic activity. The observed effect on hippocampal neurons in culture was a rise in frequency of post-synaptic currents not seen with IgGs from healthy or disease controls [8,9].

d) Vesicle trafficking. We have previously shown that ALS IgGs increase the mobility of acidic vesicles (mostly endosomes and lysosomes) in primary cortical astrocytes [10].

Moreover, in addition to ALS, in a number of other NDs such as, Lambert-Eaton — myasthenic syndrome, Guillain-Barré syndrome, Rasmussen's encephalitis, Systemic Lupus Erythematosus, Multiple sclerosis and Neuromyelitis Optica, production of anti-neuronal IgGs is a significant feature of the immune-inflammatory process.

We want to use this knowledge to design a microfluidic device in order to obtain a lab-on-chip for innovative and disruptive diagnostics for better segregation of patients of NDs. Further on we will present the main directions of this ongoing project (already funded by EU-H2020 program and the Innovation Fund of Republic of Serbia and the EIT Jumpstarter program).

APPLIED RESEARCH

The objectives addressed are: 1) Development of procedures based on the lab-on-a- chip microfluidic systems with light sensor probes within living cells reacting to IgG or total sera from patients as a diagnostic and prognostic technology related to diverse NDs. 2) Defining mark-up characteristics of the standardized in vitro approach for personalized diagnostic protocols. 3) Design of a small-scale pilot platform based on automated/miniaturized fluorescence microscopy.

The first step will be thorough standardization of already investigated and published data on biophysical processes monitoring of Ca^{2+} — transients and ROS generation by fluorescent dye imaging (on ALS IgGs [8–11] on Lupus erythematosus CSF and brain autoantibodies [13]). As an outlook of these studies IgGs or sera from different ND patients will be assessed on animal cell models (cultures of neurons, astrocytes, cell lines) throughout these biophysical processes. Further development should go towards the same tests on human cells derived from inducible pluripotent stem cells (iPSC). Using this combination of sensitive readouts, the goal is to correlate quantitative parameters of these physiological biomarkers (unlike existing molecular or biochemical) with disease specific parameters for the disruptive personalized *in vitro* diagnostics using machine learning protocols. This approach will thus introduce a novel and unforeseen diagnostic approach with physiological biomarkers as opposed to the existing „static“ molecular or biochemical markers. To accomplish this in a robust clinical/lab setup an innovative automated integrated microchip microscopy system is being designed as the first prototype.

The interdisciplinary approach will add an integrated added value to a compound solution that combines advanced biochip technology, microfluidics, cellular biophysics, custom-made micro-optics and software analytics into a genuinely novel area of theragnostic research and technology. The multivariate single-cell robust physiological analysis offered by the proposed in vitro diagnostic technology will provide the adequate approach for the personalized care of patients with a multifocal disease (such as NDs generally are).

NOVEL DESIGN OF THE OPTICS-ON-CHIP

The integration of optics within the microfluidic device towards optics-on-a-chip serves as a genuinely new technical idea. Namely, the following solutions have never been put in place in one device: 1) coded-aperture imaging and scanning holography — techniques requiring minimum amount of optical components, while providing 3D imaging and software-based autofocusing; 2) stabilized LED or diode laser for fluorescence-excitation illumination; 3) mobile phone computing power used for hologram reconstruction. As a contingency measure alternative designs are also envisaged: a stand-alone, miniaturized, microscope optical system or attachment to a standard microscope with an add-on mount.

MACHINE LEARNING PROTOCOLS

A specific software and novel data acquisition and processing capabilities have been developed. The software will acquire fluorescence image series and will discriminate the response pattern of cells treated with IgG or sera. Supervised machine learning will be employed to classify the traces. The extracted trace features will be fed into unsupervised machine learning-based software to identify the diseases itself as well as its progression phenotypes.

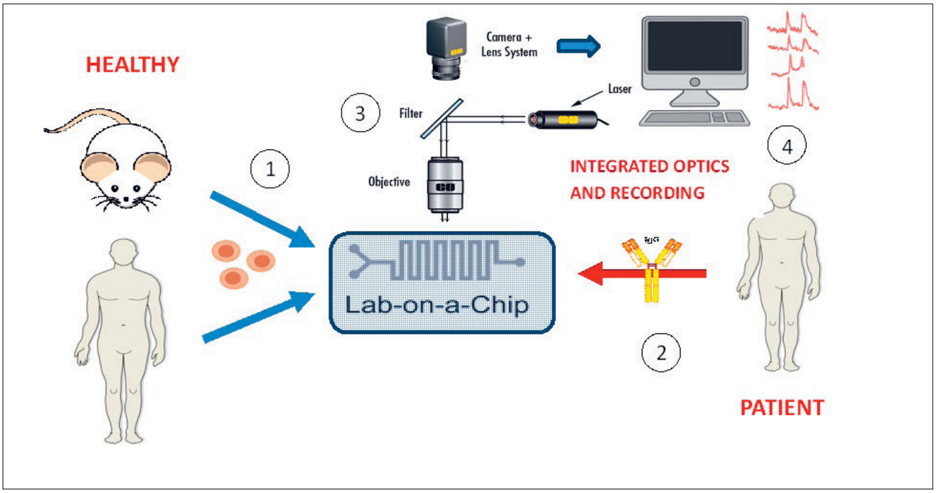


Figure 1. Proposed medical device workflow. 1) Seeding cells (either cultured animal cells or human iPSCSs — derived cells or cell lines) in the microfluidic lab-on-a-chip; 2) treatment with patient IgGs through the microfluidic flow-control system; 3) fluorescence recording system (laser — objective — lens system — camera). In the final system development the laser light source will be substituted with a LED system); 4) signal analysis using machine learning protocols for personalized medicine.

CONCLUSION

Although the effects of ALS IgGs are well studied and documented in the Belgrade laboratory and elsewhere the process of IgG application for *in vitro* diagnostics of ALS is completely novel and offers a practical multidimensional functional analysis for personalized medicine. Thus, based on a standard clinical blood sample and upon routine serum separation and IgG purification one can obtain under an automated microscope actually a pattern of activities in the form of fluorescence signaling in time and space thus establishing a personalized signature of the disease for the individual patient (Fig. 1). The combining of several recording modes raises the reliability of the obtained diagnostic pattern. On the other hand, this procedure can give early diagnosis of ALS since it has been shown that inflammation markers appear early in the disease model [14]. The complex multifactorial nature of ALS underlines the need for a personalized treatment and patient stratification. It is strongly believed that the multivariate single-cell analysis offered by the developed *in vitro* diagnostic technology may present exactly the adequate approach for the personalized patient care of such a multifocal disease. At the same time the pattern analysis of the multivariate single-cell response allows for a robust point-of-care diagnostics necessary for improved and efficient clinical decisions. This approach will contribute to the sustainability of the health care of ALS patients by drafting and planning a large-scale prototype in an operational environment. In addition, the designed personalized diagnostics technology of *in vitro* testing of IgGs from patient sera can be proposed for other motoneuron and neuroinflammatory neurodegenerative processes, thus allowing for sustainability of the health care system in the particular area of neuroinflammation as the common mechanism of neurodegenerative diseases.

Finally, the ongoing project will rise an opportunity for partnership with relevant SMEs that could contribute to the upgrade and strengthening of the designed technology in its many aspects from standardized cell culturing to the hardware and software design for automated microscopy.

BIBLIOGRAPHY

[1] Perlman ZE, Slack MD, Feng Y, Mitchison TJ, Wu LF, Altschuler SJ. Multidimensional drug profiling by automated microscopy. Science. 2004; 306: 1194–1198.
[2] Measuring Biological Responses with Automated microscopy. 2006; Methods in Enzymology Vol. 414, J. Inglese Ed. Academic Press.
[3] De La Torre-Bueno J, McBride J. US Patent US7,272,252, 2007
[4] Gough AH, Gough KA, Giuliano D, Lansing T. US Patent US 8,597,899, 2013.
[5] Appel SH, Engelhardt JI, García J, Stefani E. Immunoglobulins from animal models of motor neuron disease and from human amyotrophic lateral sclerosis patients passively transfer physiological abnormalities to the neuromuscular junction. Proc Natl Acad Sci U S A. 1991; 88: 647–651.
[6] Fratantoni SA, Weisz G, Pardal AM, Reisin RC, Uchitel OD. Amyotrophic lateral sclerosis IgG-treated neuromuscular junctions develop sensitivity to L-type calcium channel blocker. Muscle Nerve. 2000; 23: 543–550.
[7] Gonzalez LE, Kotler ML, Vattino LG, et al. Amyotrophic lateral sclerosis-immunoglobulins selectively interact with neuromuscular junctions expressing P/Q-type calcium channels. J Neurochem. 2011; 119: 826–838.
[8] Andjus PR, Khiroug L, Nistri A, Cherubini E. ALS IgGs suppress [$Ca2^+$]i rise through P/Q-type calcium channels in central neurones in culture. Neuroreport. 1996; 7: 1914–1916.
[9] Andjus PR, Stevic-Marinkovic Z, Cherubini E. Immunoglobulins from motoneurone disease patients enhance glutamate release from rat hippocampal neurones in culture. J Physiol. 1997; 504 (Pt 1)(Pt 1): 103–112.
[10] Stenovec M, Milošević M, Petrušić V, et al. Amyotrophic lateral sclerosis immunoglobulins G enhance the mobility of Lysotracker-labelled vesicles in cultured rat astrocytes. Acta Physiol (Oxf). 2011; 203: 457–471.

[11] Milošević M, Stenovec M, Kreft M, et al. Immunoglobulins G from patients with sporadic amyotrophic lateral sclerosis affects cytosolic Ca^{2+} homeostasis in cultured rat astrocytes. Cell Calcium. 2013; 54: 17–25.
[12] Milošević M, Milićević K, Božić I, et al. Immunoglobulins G from Sera of Amyotrophic Lateral Sclerosis Patients Induce Oxidative Stress and Upregulation of Antioxidative System in BV-2 Microglial Cell Line. Front Immunol. 2017; 8: 1619.
[13] Kapadia M, Bijelić D, Zhao H, et al. Effects of sustained i. c. v. infusion of lupus CSF and autoantibodies on behavioral phenotype and neuronal calcium signaling. Acta Neuropathol Commun. 2017; 5: 70.
[14] Beers DR, Henkel JS, Zhao W, Wang J, Appel SH. CD4+ T cells support glial neuroprotection, slow disease progression, and modify glial morphology in an animal model of inherited ALS. Proc Natl Acad Sci U S A. 2008; 105: 15558–15563.

Pavle R. ANĐUS

INFORMACIJA U *IN VITRO* BIOMEDICINSKOJ DIJAGNOSTICI — PUT KA PERSONALIZOVANOJ MEDICINI

Sažetak

Kod brojnih neurodegenerativnih bolesti (ND) proizvodnja antineuronskog imunoglobulina G (IgG) je značajna karakteristika inflamatornog procesa. Pokazalo se da humani IgG mogu izazvati različite fiziološke efekte na neurone i glijalne ćelije životinjskog porekla. U projektu koji je u toku predlažemo da se IgG koriste za *in vitro* dijagnostiku ND robusnom multifunkcionalnom obradom pojedinačnog uzorka (IgG pacijenta), na zasejanim normalnim ćelijama, koje mogu dati složene fiziološke odgovore zasnovane na već poznatim intraćelijskim signalima, snimljenim pomoću fluorescentnih markera u mikrofluidičkom čipu sa automatizovanom mikroskopijom, sve u pravcu kliničke personalizovane dijagnostike. Ovakav medicinski uređaj zasnovan je na: 1) razvoju procedura zasnovanih na mikrofluidičkom sistemu laboratorije na čipu sa intracelularnim svetlosnim senzorima; 2) definisanju standardizovanih *in vitro* personalizovanih dijagnostičkih protokola; 3) dizajnu male pilot platforme zasnovane na automatizovanoj/minijaturizovanoj fluorescentnoj mikroskopiji. Većina ovih principa je već delimično testirana kroz projekat EC-H2020 i projekat nacionalnog Fonda za inovacije. Interdisciplinarnost našeg istraživanja obuhvataju sledeće pristupe: ćelijska neurofiziologija, biofizika intracelularnih fluorescentnih indikatora, video-mikroskopija intracelularne molekularne signalizacije, mikrofluidika i dizajn biočipa, napredna optika po meri za automatizovanu mikroskopiju i mašinsko učenje za analizu signala. Predložena personalizovana tehnologija dijagnostike biće primenjiva na različite NB u održivom zdravstvenom sistemu.

INŽENJERSKE NAUKE I

ENGINEERING SCIENCES I

Marija ILIĆ*

DIGITISATION AND DISTRIBUTED AUTONOMY: THE PATH TO RELIABLE COST-EFFECTIVE DECARBONIZED ENERGY SERVICE

Abstract: This paper provides a short summary of the chang- ing energy resource and energy demand characteristics in diverse social-ecological energy systems (SES). Examples are given of worldwide challenges created by the societal objectives to divert climate change, and to, at the same time provide reliable energy services at reasonable cost. Given unique temporal and spatial challenges of these systems, we suggest that one must from the beginning consider an SES as a social ecological technical system (SETS), and re-think producing, consuming and delivering en- ergy as a new complex operations and planning problem. If this is not done, it will become impossible to align often conflicting objectives. We describe how, starting from the first principles, one can utilise a unified modelling of general energy systems for implementing transparent data-enabled distributed protocols for interactive power balancing and delivery at the right rate. We document how having such data-enabled protocols would significantly increase the utilisation of the existing electric energy systems infrastructure, and provide signals for its enhancing at value. Potential benefits measured in terms of social welfare, service quality and cost are illustrated.

Index Terms: *rolling blackouts, social ecological energy systems (SES), social ecological technical systems (SETS), protocols for distributed autonomous electricity service, digitisation, Ex- tended AC Optimal Power Flow (XOPF), plug-and-play arhci- tectures, Dynamic Monitoring and Decision Systems (DyMonDS) framework, cyber-physical systems (CPS)*

* Marija Ilić, Massachusetts Institute of Technology, Cambridge, MA

I. INTRODUCTION

Currently we are witnessing worldwide difficult, often volatile, energy provision problems. During holidays US customers were asked to conserve power and reduce their heat; not long ago Texas consumers had no power during cold winter snap; Puerto Rico population repeatedly loses power during extreme weather conditions; many developing countries have been accustomed for a very long time to use electricity only when available, namely service being available only when generation is available. Europe has most recently been deprived of expensive energy services due to geopolitical conflicts, as have other war torn countries. There have also been service interruptions caused by physical and cyber attacks on infrastructure.

To make matters worse, these problems are becoming more frequent although only very few conventional energy resources are actually getting offline for retiring. They have so far fallen under the category of „very low probability high impact" events. The problem is likely to worsen as attempts are being made to meet zero carbon energy services as the generation mix becomes more intermittent and the conventional power plants no longer supply power.

A. The need for rethinking energy operations and planning paradigm

In parallel with these highly visible service problems, there have been ongoing R&D e fforts toward reverting cli- mate change by deploying many distributed energy resources (DERs), such as households with their own roof top solar photo-voltaic (PVs) and back-up generators; price-responsive demand consumption; even utilising very small storage appliances, such as HVACs, electric water heaters, electric vehicles (EVs), for regulating frequency and voltage caused by power imbalances around the scheduled generation for predictable system demand [1]. Probably the best known and the earliest was the homeostatic control vision by the late Fred Schweppe with his colleagues at MIT more than three decades ago [2]. Much R&D effort has gone since into conceptualising sensing and automated control of DERs, so that they self-adjust as per needed basis.

These concepts have been shown using simulations mainly, and by carrying out very few small scale pilot projects. Utilities operating distribution systems have not considered systematically new ways of integrating the effects of very large number of DERs when assessing options to serve their customers. Reasons for this are numerous, including: 1) wanting to carry out top-down service by pre-programming the equipment based on historical power usages; 2) thinking of these effects as having insignificant effects on electricity service; and, 3) not seeing the need to fix something that is not broken. Unfortunately, the current challenges with energy services clearly indicate that something is broken, and that there exists an obvious need to rely on all possible resources as system conditions vary. Notably, some utilities are working with their customers on installing higher efficiency electic loads, such as heat pumps; these are most frequently not programmed to respond to system conditions and/or price of electricity [3]. This means that they are not equipped with automation for adjusting when most needed. Electric distribution grids, in particular, can and should consider deployment of data- enabled normally-open switches (NoS) and/or normally-closed switches (NcS) which are likely to become instrumental in re-routing power within the distribution grid to accommodate power delivery from other resources when local resources fail to do so due to intermittent power provision or due to equipment failures. Research has shown that it is possible to assess the impact of certain number of such switches on the improvements in electricity service, notably during low prob- ability high impact events [4], [5]. Aggregation of many small geographically dispersed DERs is often not implementable because distribution grid protection is still programmed mainly under the assumption that power flows uni-directionally from a substation to the end users connected to the local feeders.

On the bulk power systems (BPS) side, there has been on- going research toward enhancing just-in-time)JIT), just-in- place (JIP) data-enabled decision-making computer applications to assist system operators to make the most out of what is available [9], [10]. Instead of utilising generation for the worst-case single, or double outage (($N-1$) or ($N-2$) reliability standards [6]) and requiring much reserve in case these occur, methods were proposed for doing optimisation to compute quickly the key constraints to feasible power delivery during these contingencies and actions to overcome by most effective adjustment of the remaining controllable equipment ratter than creating „proxy" limits to physical limits of hardware, like transmission line flows [11], [12]. This can replace the need for conservative local reserve allocation. Instead, during extreme events reliable service can be provided at a regional level, rather than state by state. BPS operators and planners rarely account for effects of DERs at the point of connection (PoC) between them and distribution grids. Unfortunately, the same way as with distribution systems, implementing innovative solutions, in particular these low- hanging fruit decision-making software and statistical learning about the system demand and conditions, has been rather slow, almost non-existent. The (Independent)System Operators ((I)SOs) have continued to use DC power flow, and, in the electricity markets DC Optimal Power Flow (DC OPF), while fully aware of the fact that most binding constraints to delivering power over far electrical distances have been voltage/reactive power problems. The decisions on by how overcome voltage-related delivery problems between different supply-demand parts of the system are left to the (I)SOs; as such they cannot be reproduced, and have led to must run generation units, and out-of-market dispatch [8].

II. THE DANGER OF STATUS QUO AND POSSIBLE WAYS FORWARD

Utilities are still required to serve as the providers of the last resort. At the same time, they will have to, more often than in the past, deal with major weather-related service disruptions, because it is becoming harder and harder to accurately predict demand and plan for extreme events. Also, with the electrification, peak demand will increase, and probability of disruptions will increase significantly. This situation clearly points into the basic need to utilise all available resources as much as possible. In simple terms, it is insufficient to consider BPS level large-scale conventional generation as the only resources responsible for enabling reliable service. Given maturity of sensing and automation, utilities will do much better by transforming themselves into cyber-physical systems (CPS) in which data-enabled on-line decision-making is key to flexible utilisation, all else being equal. Reliable service and using clean resources do not have to be conflicting objectives. There exists a break-even point at which the cost of pollution balances with the ability to avoid service interruption. There exists also a way to assess different technologies, hardware and software, for their ability to enhance benefits in operations versus incremental cost of enhancing the infrastructure [13]. There exist even performance-based regulatory pricing, such as peak load pricing which establish value from these enhancements [14].

The time has come to think of electric energy service rather than kW hour as a volumetric quantity. Because of intermittency, and because of requiring both distribution systems and BPS to deliver power flow patterns for which they were not originally designed, the time has come to have operations and planning for best electricity service. Instead of building for assumed demand, it becomes essential to have interactive information exchange between grid users and the grid operators. This must be done at the level of granularity sufficient to extract the most out of what is available.

A. Today's approach to operations and planning for reliable and resilient electricity services

The bulk power systems (BPS) have different planning and operating practices than the distribution systems, as summarised next.

1) Basic functions performed by the control centres in BPS: Today BPS have established computer applications in their control centres for meeting the NERC/FERC reliability criteria [6]. To briefly review, and set the basis for understanding the evolution and new opportunities offered by the increasing penetration of DERs at the distribution system levels, we take a step back and recall that in today's industry electricity service is provided by performing five (5) functions. These are: 1) scheduling controllable conventional generation to supply predictable system demand; 2) supplying a bit more generation to compensate for Joule losses in the delivery grid;

3) scheduling so that the power delivery is feasible, namely that there is no grid congestion, namely the power delivery is within the allowable thermal line flow limits and equipment voltage constraints; 4) having sufficient regulation reserve to compensate hard-to-predict system demand deviations from predicted demand patterns; and, 5) having sufficient reserve to not interrupt customers at least for 30 minutes following any single, or double, BPS outage. The control centres of a BPS have computer applications for assisting system operators with performing these functions [15].

2) Basic practice in operating distribution systems: Notably, it has been shown that there is very little assurance that small distribution grid users will have the reliability standards required by their states, such as SAIDI and SAIFI, met based on the BPS control [16]. Despite this recognition, distribution systems are mainly based on pre-programming controllable equipment for seasonal demand, and under the assumption that the only power comes from the substations connected to the BPS and that these point of contact (PoC) are effectively ideal source which provide perfect AC frequency and voltage while sending power to the distirbution systems. Minor service interruptions in localised part of the grid are created mainly by trees touching equipment during bad weather. Utility, sooner or later, learns about the service interruption and sends engineers to repair the failed equipment. As a rule, there are no alternate paths from PoC to small end users, and when a substation itself gets disconnected, all customers lose power, the case of recent North Carolina service interruptions [17]. Substations are becoming favourite places for both physical and cyber attacks, and, they mainly lead to major service interruptions. The BPS cannot prevent this problem from happening since there is simply no viable alternate path. NoS and NcS are instrumental to supporting power delivery from other neighbouring substations [4], [5].

III. DIFFERENT WAYS OF ENABLING DECARBONIZATION

To move forward, it is necessary to establish systematic guidelines for integrating new technologies, in particular clean DERs within the legacy electric energy systems. These are being added to both the BPS and into lower-voltage level grids, such as distribution systems and even by building local microgrids. Since many of these new resources are often third-party-owned, grid planners and system operators need to establish transparent, technically- and economically- justifiable protocols for their integration. As described earlier in this paper, today's interconnection standards are system- specific and generally do not provide quantifiable ways of determining ranges of energy, power, reactive power and their rates of change so that feasible and robust electricity service is ensured. A particular challenge concerns very high-impact low-probability events, which are becoming more frequent as a result of extreme weather and hard-to-predict cyber- and physical-infrastructure attacks [18] One can not neglect possibility of wide-spread life threatening events which can be created by targeted cyber attacks and even electromagnetic pulse storms, documented through state exploration programs. While the industry is currently not charged with preparing for higher than (N – 2) events, it is important to be prepared for providing at least critical service during such extreme events. More generally, as system conditions vary, it is necessary to support gradual

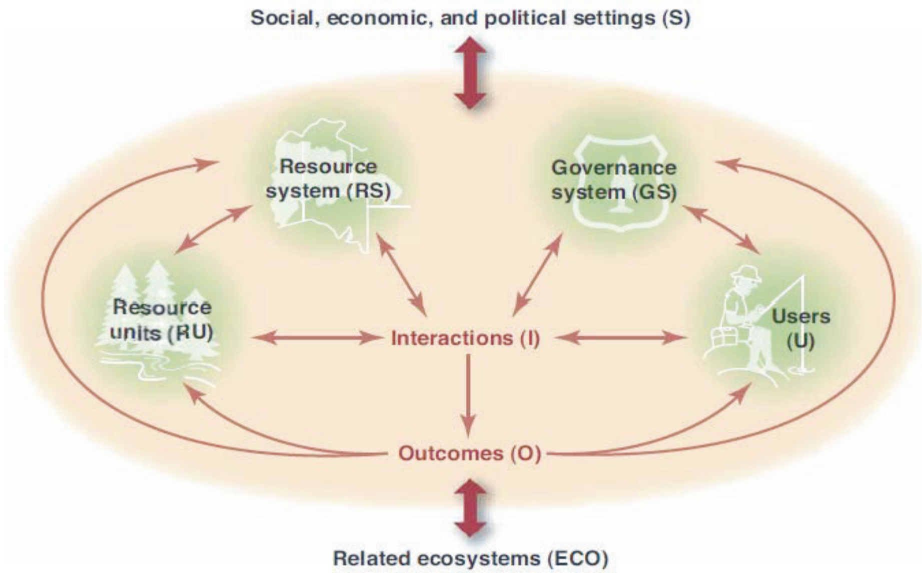


Fig. 1. Basic social-ecological system core variables and their interactions [19]

degradation of service rather than have wide- spread loss of service. Later in this paper we describe how this might become possible with the approach proposed here. Although one of the main ideas in this paper is to describe how distributed automation and digitisation can play a major role in making the most out of any given social ecological energy system, for this to be done in an effective way it is necessary to pose this objective in the broader context as with any other technologies.

IV. SOCIAL ECOLOGICAL ENERGY SYSTEMS (SES)

We start by posing the problem of enabling reliable and cost-effective electricity service as the problem of making energy systems sustainable in the same sense as any other SES. This is done by viewing these systems as complex systems comprising many diverse ecological groups of resources (RS); and socially-diverse groups of users (U), all governed by a particular governance system (GS). This view is much the same way as Elinor Ostrom used to introduce her general framework for assessing sustainability of an SES Figure 1 [19]. The RS, U, and governance system (GS) form core variables of an SES, and their attributes which determine sustainability represent second-level core variables. Notably, there exist key second-level variables contributing the most to the overall sustainability of an SES, Table 1 in [19].

In this section we introduce a generalized architecture of the evolving energy systems which, instead of RS, U and GS comprise physically interconnected subgroups of resources and users by means of man-made infrastructures, with their own sub-objectives and data-enabled decision-making and information exchange with other subsystems. We then describe how to use the general notion of second-order core variables introduced for any SES, and give examples of such variables for assessing performance of any given social-ecological- technical system (SETS). We then discuss potential role of governance system in SETS for enabling good performance, we give examples of three qualitatively different architectures and the interaction variables within these SETS.

A. Fundamental complexity of an end-to-end electric power interconnection SETS architecture

The electric power-, gas- and other energy carrier grids are man-made physical- and cyber-network infrastructures enabling the interactions between RS and U in a general electric energy system sketched in Figure 2 [9].

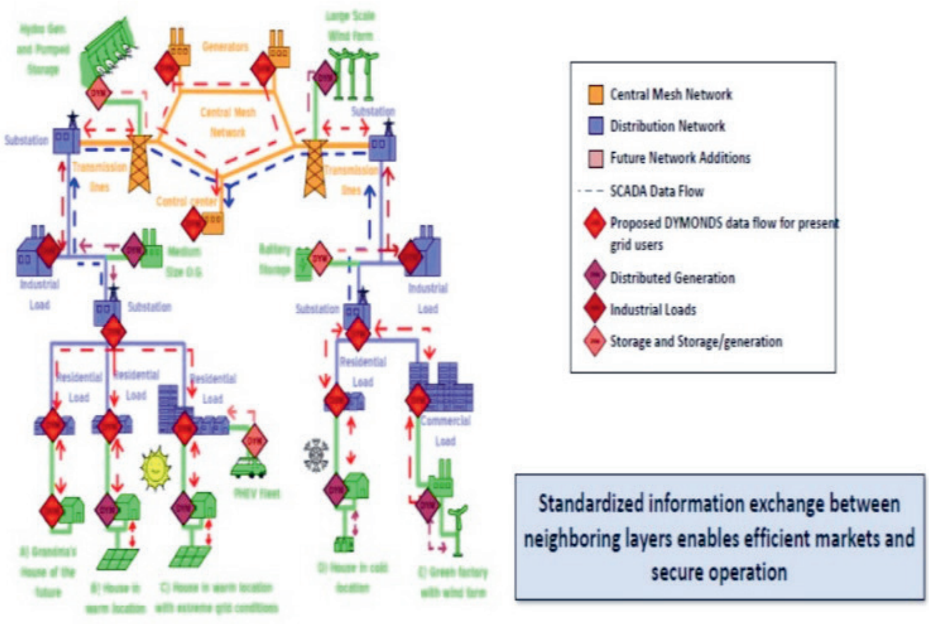


Fig. 2. Fundamental complexity of an end-to-end physical-/cyber electric energy system [9]

The man-made power grid makes the representation of an SETS more complex than the representation of an SES shown in Figure 1. Shown is the physical infrastructure which is generally a physical interconnection of the central extra-high-voltage (EHV) mesh network (orange solid lines), via step-down transformers to high voltage (HV) sub-transmission and medium voltage (MV) distribution radial grids (purple solid lines) and further down to low voltage (LV) distribution grids and microgrids [9]. Dotted red and blue lines represent the SCADA data flow, in today's bulk power systems and the missing in lower voltage networks, respectively. The red „dymonds" represent cyber (data sensing, processing in support of decision-making and control) embedded within the core variables. Today most of the BPS core variables (RS and U) have somewhat standardized cyber, including the Energy Management System (EMS) collecting SCADA data, using different computer applications (such as economic dispatch, power flow analysis, and alike) and sending commands for generation scheduling. Finally, it can be seen from Figure 2 that there are many new DERs with their own cyber, mainly embedded closer to the small end users. The spatial

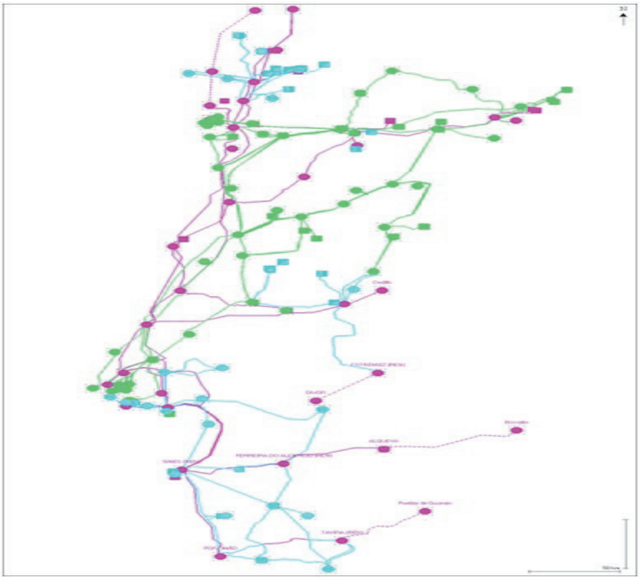


Fig. 3. Real-world distribution feeder in Portugal [9]

complexity of an SETS is two-fold: Horizontally, there are several interconnected BAs, as shown in Figures 3 and 5, each of which with the basic hierarchical structure shown in Figure 2. Vertically, in each BA there are many

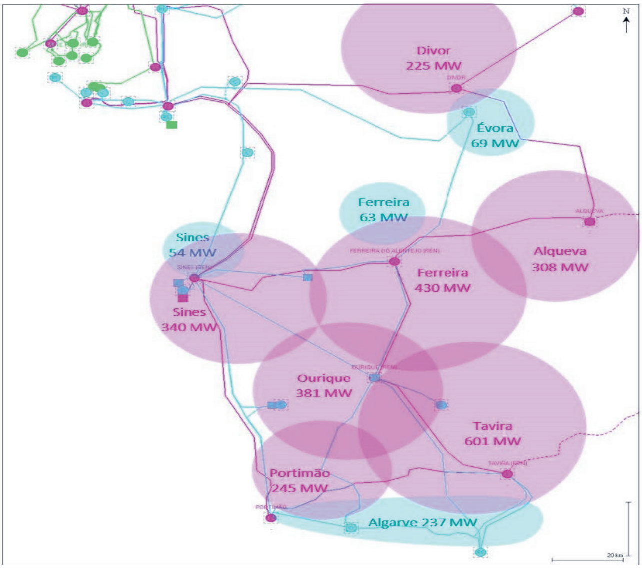


Fig. 4. Candidate solar-based iBAs seeking interconnection with the main BPS in Portugal

Fig. 5. The US BPS interconnection comprising several large BAs [6]

candidate iBAs. Shown in Figure 4 are candidate solar iBAs seeking connection to the main grid. In today's industry this very complex end-to-end SETS comprises large subsystems known as the Balancing Authorities (BAs) shown in Figure 5 for the US interconnected system. Integration of third-party sub-systems shown in Figure 4 leads to embedded nested architecture of smaller intelligent Balancing Authorities (iBAs) within the existing BAs. Shown in Figures 6 and 7 are a general nested architecture of the evolving SETS and a small example of transforming RS, U interconnected by electric power grid into interconnected iBAs, respectively [20], [21].

B. Key role of aggregating „flat“ SES to multi-layered inter- acting composition of iBAs

The hierarchical complexity in today's industry in systems comprising horizontally-composed BAs and vertically- organized transmission- and distribution-systems within BAs, has been managed by spatially and temporally decom- posing these entities and managing each level hierarchically under major assumptions, such as weak spatial coupling between horizontally-organized BAs, and temporally- decomposed vertically-organized entities. They cyber designed has greatly simplified under these assumptions [22].

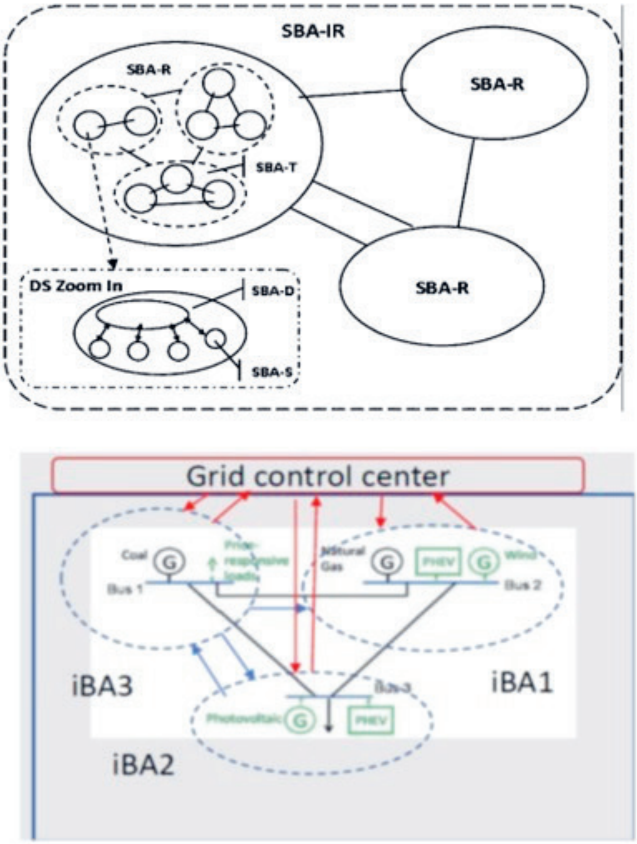


Fig. 6. Nested end-to-end architecture comprising lower levels iBAs and higher level larger BAs [9]

As the industry is evolving into a composition of interacting iBA s, it is no longer possible to ensure validity of these hierarchical arrangements. Instead, hierarchies are becoming nested, Figure 6 and are interacting more dynamically. Shown in Figure 7 is a sketch of a small BPS SETS comprising conventional RS and U (drawn in black). Embedded into this BPS are DERs of diverse types (green lines), including price responsive demand. it is then shown how this „flat" SETS with clearly defined RS and U, evolves into a nested system with aggregated iBAs.

In summary, wes tress that one of the most important second-level core variables are clearly identified boundaries of the groups of components with common performance objectives. This is teh main reason for having introduced a notion of technology-agnostic iBAs, and then defining their interaction variables with the rest of iBAs within an SES. Unique to the energy systems is that their alignments must be temporal, spatial and functional, otherwise the system may not be feasible and/or it may be overly sensitive, non- robust to even small deviations from assumed conditions and specifications.

Once this is understood, it becomes possible to assess different hardware and software technologies which can help the given SES become more sustainable. This then becomes the problem of designing an SETS for sustainability. This high-level conceptualisation helps tremendously in assessing potential of different technologies for making the most out of what is ecologically available and socially acceptable to the governance and the users.

V. UNIFIED APPROACH TO MODELLING THE INTERACTION DYNAMICS WITHIN AN ENERGY SETS

So far we have conceptualized an architecture of any SETS as a generalization of an SES into a multi-layered system comprising interconnected iBAs. In order to capture their unique temporal and spatial characterisation we next view them as complex end-to-end interactive multi-layered dynamical systems comprising diverse subsystems, named, iBAs. These iBAs, much the same way as the core variables in any SES, are characterized by their second-level core variables. We observe that the second-level core variables most desired to support sustainable SETS are the ones which align the technical, economic and environmental sub-objectives of iBAs with those of the other iBAs. Notably, selecting man- made technologies to support sustainable interactions requires modelling of the interactions of iBAs with other iBAs. To assess how well the interactions are aligned, we utilize a unified notion of interaction variables for any given iBA in terms of physically interpret-able energy dynamics. These are based on the first principles and are technology and system- agnostic and can support performance at value, discussed next.

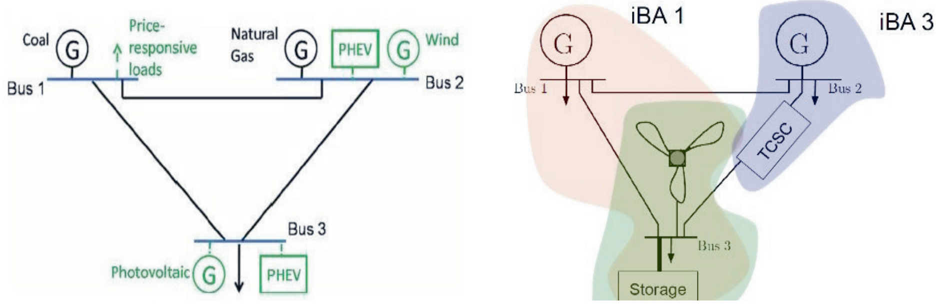


Fig. 7. The evoluting from „flat“ SES into a composition of interacting iBAs [21]

A. Technology- and system-agnostic interaction variable representing iBAs

The aggregate variables characterizing any iBA are stored energy and rate of change of stored energy. The dynamics of these aggregate variables have been shown to satisfy the following model

$$\frac{dE(t)}{dt} = -\frac{E}{\tau} + p(t) \quad (1)$$

$$\frac{dp(t)}{dt} = 4E_t(t) - \dot{q}(t) \quad (2)$$

where $E(t)$, and $p(t)$ are aggregate state variables, stored energy and rate of power coming into the iBA, respectively. The term $-\frac{E}{\tau}$ represents Joule losses of the iBA, the term $E_t(t)$ represents energy in tangent space, and it has an intuitive interpretation related to exergy, namely to the potential to do work [24]; the term $\dot{q}(t)$ represents rate of change of power which does not perform real work, namely it reflects the impossibility to do work without increasing entropy, known as anergy [24]. Power $p(t)$ can be sent from the neighbouring iBAs $P(t)$, or it can be injected internally by some local controller $P_u(t)$, namely

$$p(t) = P(t) + P_u(t) \qquad (3)$$

Similarly, rate of change of generalised reactive power [23] can be decomposed into the component stemming from the neighbouring iBAs $\dot{Q}(t)$ and/or from the internal controller $\dot{Q}_u(t)$, namely

$$\dot{q}(t) = \dot{Q}(t) + \dot{Q}_u(t) \qquad (4)$$

Our earlier work concerning interaction variables was recently summarized in detail [25]. It is described in this article how, looking in a hindsight, one can start from the first principles and define interaction variable as an aggregate variable whose properties are a direct expression of conservation of energy dynamics. In short, when disconnected from the rest of the system, an iBA conserves its own energy. When connected to the other iBAs, it must satisfy instantaneous power balancing and the rate of change of generalised reactive power It can be seen from the dynamics of aggregate variables $E(t)$ and $p(t)$ that they are function of themselves and the rate of change of iBAs interaction variable $\dot{z}(t)$ where

$$\dot{z}(t) = [P(t)\ \dot{Q}(t)] \quad (5)$$

While iBAs can be quite complex themselves, their interaction with the rest of the system can be captured by writing the dynamics of the aggregate variables in a modular way for each iBA, and then stating that in order for the interconnected system to be feasible and stable the rate of change of their interaction variables must balance. For the simplest case of one resource (RS), interconnected via transmission line (iBA 1) to one user (U) (iBA 2) the interconnected system model takes on the form

$$\frac{dE_1(t)}{dt} = -\frac{E_1}{\tau_1} + p_1(t) \quad (6)$$

$$\frac{dp_1(t)}{dt} = 4E_{t,1}(t) - \dot{q}_1(t) \quad (7)$$

$$\frac{dE_2(t)}{dt} = -\frac{E_2}{\tau_2} + p_2(t) \quad (8)$$

$$\frac{dp_2(t)}{dt} = 4E_{t,2}(t) - \dot{q}_2(t) \quad (9)$$

When there are no internal controls inside iBAs, $p_1(t) = P_1(t)$, $p_2(t) = P_2(t)$, $\dot{q}_1(t) = \dot{Q}_1(t)$ and $\dot{q}_2(t) = \dot{Q}_2(t)$. For the interconnected system to be feasible rate of change of interaction variables must be equal, namely

$$P_1(t) + P_2(t) = 0 \quad (10)$$

and

$$\dot{Q}_1(t) + \dot{Q}_2(t) = 0 \quad (11)$$

In our work up to date we have shown that it is sufficient to characterise outputs of any component comprising the system. Coupled model given in Equations (6)–(11) is the new unified energy dynamics model.

VI. PRINCIPLES OF DATA-ENABLED SENSING AND CONTROL FOR SETS

The piecemeal process of decarbonization is already under way, without having a holistic understanding of its objectives and feasibility of achieving it. The debates have become highly ideological and biased, ranging from extreme proposals that there will be one solution solving it all; for example, proponents of demand side efficiency and response are exchanging endless emails focusing on this particular aspect of solution [26]; there are also many expressing concerns that the climate change objectives are simply not achievable [27].

To overcome this endless non-constructive debate, in this paper we propose that it is possible to utilize the unified modelling of the evolving electric energy systems to assess their sustainability. Of particular interest is to understand potential of innovation, both hardware and software, in enhancing their performance. While the qualitative characterization of performance is very similar to the one put forward by Elinor Ostrom, the unique challenges in these systems come from the need for quantifying the desired second-level core variables, key to good performance of any SES. This can be done by assessing spatial, temporal and functional alignments of the interactions between R, U and GS, shown in Figures 1 and 2 represented as interactive mathematical models of type given in Eqns. (6)– (11). These models can be used to interactively in feed-forward manner check the power flow feasibility when iBAs are interconnected [28]. This recognition becomes the key to embedding most effective man-made technologies and transform any given SES into as sustainable as possible SETS. Enabling an SES with many man-made infrastructures and their data-faciliated functionalities, sketched in Figure 2 [9]. On the contrary, if these interactions are not aligned, much will be wasted. The main objective of distributed automation embedded within the R, U, GS is to enable implementation of these alignments to as large extent as possible.

The process of assessing potential of technological innovations in electric energy systems can be supported by proposing the following three basic principles [29].

— Characterize all (group of) component(s) (subsystem) in terms of transparent, unified outputs for their technical, economic, and environmental potential performance when connecting to the system.

— Generalize today's Supervisory Control and Data Acquisition (SCADA) of today's BPS to support protocols (interactive multi-layered information exchange) over multiple time horizons, in a feed-forward and feedback manner, over a stratum of time horizons for which performance of the system is assessed.

— Establish data-enabled assessment tools based on conditions which are sufficient to ensure feasible and robust preference of the interconnected systems. Shown in Figure 2 for any time interval T in terms of a triplet interaction outputs energy ET, power PT, and rate of change of reactive power $\dot{Q}_T$ [28].

VII. GENERAL DYNAMIC MONITORING AND DECISION SYSTEMS (DYMONDS) FRAMEWORK FOR ASSESSING PERFORMANCE OF ANY SETS

As described above, any given electric energy system can be thought of as a data-enabled complex interactive dynamical system comprising many iBAs. Conceptualizing the architecture this way is extremely helpful because it becomes possible to understand performance sub-objectives of different iBAs, and model their interactions. It becomes possible to think of the problem of feasible and sustainable SETS as the problem of technological enhancements most effective in aligning interacting resources, groups of resources, groups of users, within a given governance system. Governance rules, rights and regulations (3Rs) themselves can be assessed and further evolved to give incentives for supporting the second- order variables most important for sustainability.

A. Assessing different architectures for their support of decarbonization

Having taken this view it becomes quite straightforward to assess next the same SES for its performance depending on the technological solutions. Just to illustrate, we consider next

1) an SES without any technological solutions utilized; 2) SES with today's centralised operations and planning; 3) entirely decentralised; and 4) enabled by an interactive DyMonDS which we have conceived some time ago [9], [30]. It is quite important to recognize that there is no universal technology which makes any SES system most sustainable. This, in turn, implies that different man-made power grids for enabling physical interactions between R and S within a given GS lead to different performance.

— **Social ecological energy system** without any man-made infrastructures is clearly not feasible nor robust with respect to even small disturbances. Only energy from local resources to the users can be provided.

— **SES with today's centralised operations and planning** Shown in Figure 2 is a typical modern-day electric energy system with large conventional resources, interconnected by the EHV/HV electric power grid and connected further through substations to the large number of users in local distribution systems. In the past users have consumed energy without receiving much feedback from the utilities concerning the need to adjust their consumption. These systems have provided electricity service by unidirectional sending power from resources to the users. A quick look at their second-order variables reveals that there has not been significant temporal nor spatial alignment between the resources and users. The alignment has been achieved mainly by building a physically- strong electric power grid and not relying on just-in-time (JIT), just in place (JIP) just-in context (JIC) dynamic alignments of resources and users. Many other second- level variables key to making the SETS sustainable are generally not a part of planning and operating practices in today's industry. The operating and planning standards are fundamentally based on the worst-case deterministic approach which requires setting aside large reserves [6]. The reserves are typically not used in real-time except when certain worst case event, such as large resource or transmission line contingency, takes place. Much has been written by this author to document missed opportunities from not relying on more dynamic allocation of such reserve as the operating conditions change [20], [33]–[35].

As the conventional power plants are being retired, the more environmentally-friendly intermittent resources, such as large utility-scale wind- and solar-power plants are being deployed, the need for more flexible utilization of what is available by means of data-enabled software for aligning interaction variables within the changing SETS becomes self-evident [7], [32]. Notably, even the next wave of smaller-scale nuclear reactors and other promising nuclear technologies will require more flexible system integration to avoid tripping these units through unaligned interactions with the rest of the system. The problem of data-enabled integration of growing scale of safe and reliable nuclear power deployment in support of decarbonization is an important R& D question which must be given major attention.

This paper is written in part by wanting to point out this low-hanging opportunities which can be made a reality in relatively short term manner. Pursuing top-down highly centralised deterministic approach will make it hard to implement gradual degradation of service during extreme rare-events such as hurricanes [32]. Planning will require much large-scale storage to manage temporal uncertain- ties and huge investments in power grid infrastructure to deliver power from often distant intermittent resources, including EHV DC backbone grid estimated to cost more than $200 billion dollars in the US and similarly in Western Europe.

— **Competitive decentralised operations and planning** is considered as an alternative architecture in which many distributed, often small-scale energy resources (DERs), including demand response [26], are placed locally close to the users and are intended to serve them by forming local microgrids. This approach of defecting the utility grid and each household, or neighbourhood, becoming an iBA and serving itself using solar PVs, back-up generation and storage (thermal, EVs), is appealing to many. However, this architecture may end up experiencing significantto periodically post the so-called hosting capacity for such resources [36].

NERC BA standards for AGC in place today are currently being extended to non-utility-owned iBAs. Such stan- dards effectively require each iBAs to cancel the changes seen through the interactions with the neighbouring iBAs. This architecture would become the one of many decoupled iBAs, a highly inefficient and unsustainable solution [31]. This architecture itself goes against the very funda- mentals of supporting cooperative alignment of second- order variables needed for efficient decarbonization. Becasue of this, the emerging decentralzied architectures require much re-thinking about governance and design of 3Rs in which all iBAs have clearly defined boundaries and are incentivized to go beyond strictly competitive integration. A drastic example is the problem of expecting today's utilities to still remain providers of the last resort, when local resources fail to provide expected service. Utilities can simply not do this and stay financially feasible as the scale of DER deployments increases to meet the environmental objectives, another major R&D question which must be studied.

— **DyMonDS-based architectures** are a natural extension of today's utility systems needed to plan and operate in a sustainable manner large amounts of clean DERs within low voltage distribution grids and re configurable microgrids. This must be done without endangering the backbone bulk power system production, consumption and delivery of power from large utility-scale, often electrically far from the substation level points of con- tact (PoC) with lower voltage level substations. As the utility BPS level power flow patterns change due to changing generation and demand mix, there are already major concerns that the reserves required in the past are significantly reduced. An obvious short-term solution to this problem is to begin to enable lower-level DERs and users, and the grid itself with the distributed interactive data-enabled minimally coordinated autonomy [29]. Instead of setting rigid interconnection standard constraints on hosting capacity, it is important to have a DyMonDS-based system in place in which self- adaptation and protocols using minimal information ex- change in terms of interaction variables support their cooperative alignments. Due to space limitations we do not discuss how such protocols led themselves to highly distributed autonomous multi-layered control. The higher layers utilize unified energy dynamics for deriving control logic needed to align the rate of change of interaction variables $\dot{z}(t)$ defined in Eqn. (5). The energy dynamics of the interconnected systems shown in Equations(6)–(11) lends itself to defining the control problem operating problems in real time without embedded data-of aligning interaction variables $\dot{z}(t)$ among iBAs as enabled support to self-adjust and align local resources and users. At present, there are interconnection standards which limit the presence of such local DERs in the distribution systems. Utilities, on the other hand, are required a provable linear control problem, it becomes possible to have provable performance in an otherwise highly complex multi-layered system, and build confidence in the performance of novel re configurable microgrids, for example [37]–[39]. The higher layer control design of aligning energy dynamics is technology agnostic. The Implementation of the actual internal control of $P_u(t)$ and $Q^{\cdot}\ u(t)$ (recall Equations (3) and (4)) is technology-specific, and can be kept proprietary to the manufacturers as long as they can specify the ranges of rate of change of their interaction variables prior to interconnecting and, moreover, participate in self-adjusting to the rest of the system.

VIII. A PATH FORWARD TO SUSTAINABLE DECARBONIZATION: DISTRIBUTED AUTONOMY WITHIN THE DYMONDS FRAMEWORK

DyMoNDS architecture helps make the case for embed- ding systematically data-enabled management of the changing electric energy systems. Having a flexible JIT, JIP, JIC service is paramount to having desired second-order variables in an SETS. Characterising the iBAs, as these evolve and become nested within the existing utilities, requires major data-enabled internalization by diverse iBAs themselves and minimal infor- mation exchange in terms of their interaction variables. This begins to form the kernel of distributed interactive autonomy, in which having key information contributes in significant ways to decarbonization. There is a rapidly growing awareness of the need for end-to-end information exchange. Without systematic foundations for protocols this quickly leads to an open ended complexity in which information is not utilized for the right functionalities required to have a sustainable SETS. In DyMoNDS-based cyber-physical ecosystem data begins to be utilized systematically, to list just a few examples:

— Use of weather information for predicting clean power generation; equipment status; system demand

— Adaptive load management for dynamic balance of sup- ply and demand with least expensive and cleanest re- sources

— Minimizing the need for stand-by reserves while still ensuring reliable high QoS electricity service (implementable preventive and corrective reserves)

— Minimizing the need for long-term capacity

— Enabling many DERs to participate in grid congestion management

— Data-enabled management of controllable T&D equipment to support most sustainable social ecological energy systems

— Resilient service during extreme events

The vision put forward in this paper represents qualitatively different ways of data-enabled electric energy services. To manage more complex design and operations objectives, one must understand the trade-off between communication/control complexity; market, technical outcomes; and

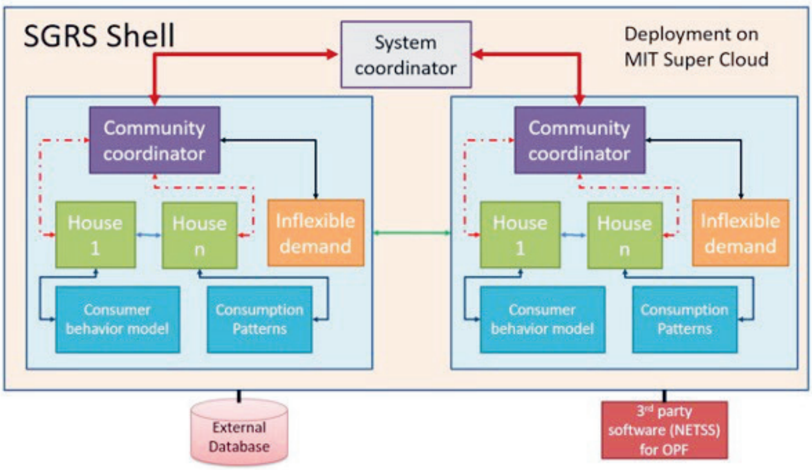


Fig. 8. Embedded IoT/ML/AI DyMonDS-based computer platform [42]

environmental effects. These are truly exciting complex systems problems whose solutions offer hidden value of high tech without having to be a domain expert in specific generation, distribution, consumption technology. Instead there is much need and opportunity to use knowledge from dynamical systems, control, numerical methods, analytics, machine learning and, ultimately, AI to enable feasible, efficient and clean services to the extent possible.

Shown in Figure 8. is a sketch of our Smart Grid in a Room Simulator (SGRS) which we have worked on developing throughout the years of building our knowledge in support of DyMonDS framework. [40]. It has become clear over the years of this research that the grid analytics becomes crucial for effective use of digitalization. Pursuing a systematic path from gathering data to IT-enabled protocols in support of targeted system performance can have a potentially large impact on decarbonization process. Shorter term it can significantly help with resilient/robust electricity service at the affordable price. Notably, DyMonDS framework lends it self quite naturally to cyber secure implementation of data-enabled re-integration of DERs into legacy electric energy systems [41]. There are some examples of pilot proof-of-concept use of digitalization. How- ever, scaling up in transparent ways requires sound principles for setting minimally required information exchange proposed in this paper. DyMonDS framework supports solving a long- standing practical problem of growing households in a plug- and-play manner [43] without causing reliability problems, and doing this in a cooperative efficient manner. This knowledge offers a major opportunity which may not be ever materialized unless a focused effort to setting up protocols based on the three simple principles put forward in this paper is pursued. Perhaps the most urgent is to focused efforts on transforming a SGRS into a digital twin which might work [42]. Having such facility will make many perhaps difficult-to-absorb concepts put forward in this paper tangible illustrations using a carefully designed digital twin.

REFERENCES

[1] Rohlfing, Eric. „The ARPA-E Mission." In 2015 ERC. 2015.

[2] Schweppe, F. C., Tabors, R. D., Kirtley, J. L., Outhred, H. R., Pickel, F. H., & Cox, A. J. (1980). Homeostatic utility control. IEEE Transactions on Power Apparatus and Systems, (3), 1151–1163.

[3] Love, Jenny, Andrew ZP Smith, Stephen Watson, Eleni Oikonomou, Alex Summerfield, Colin Gleeson, Phillip Biddulph et al. „The addition of heat pump

electricity load profiles to GB electricity demand: Evidence from a heat pump field trial.“ Applied Energy 204 (2017): 332–342.

[4] Junlakarn, Siripha, and Marija Ilic. „Toward implementation of the reconfiguration for providing differentiated reliability options in distri- bution systems.“ In 2014 IEEE PES General Meeting— Conference & Exposition. IEEE, 2014.

[5] Junlakarn, Siripha, and Marija Ilic. „Provision of differentiated reliabil- ity services under a market-based investment decision making.“ IEEE Transactions on Smart Grid 11, no. 5 (2020): 3970–3981.

[6] Watson, W. F. (2017). NERC mandatory reliability standards: A 10- year assessment. The Electricity Journal, 30(2), 9–14.

[7] Cvijic, S., Ilic, M., Allen, E., & Lang, J. (2018, October). Using extended ac optimal power flow for effective decision making. In 2018 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe).

[8] Mukerji, R., Merrill, H. M., Erickson, B. W., Parker, J. H., & Fried- man, R. E. (1991). Power plant maintenance scheduling: optimizing economics and reliability. IEEE Transactions on Power Systems, 6(2), 476–483.

[9] Ilic, M. D. (2010). Dynamic monitoring and decision systems for enabling sustainable energy services. Proceedings of the IEEE, 99(1), 58–79.

[10] Ilic, Marija. „3Rs for Power and Demand.“ Public Utilities Fortnightly Magazine 61 (2009).

[11] Ilic, M., & Jaddivada, R. (2022). Modeling and Control of Multi- Energy Dynamical Systems: Hidden Paths to Decarbonization. arXiv preprint arXiv: 2207.08370.

[12] Ilic, Marija D., and Pedro MS Carvalho. „From hierarchical control to flexible interactive electricity services: A path to decarbonisation.“ Electric Power Systems Research 212 (2022): 108554.

[13] Ilic, M., & Gough, M. (2022). Interactive Planning and Operations using Peak Load Pricing in Distribution Systems. arXiv preprint arXiv: 2212.02145.

[14] Crew, M. A., Fernando, C. S., & Kleindorfer, P. R. (1995). The theory of peak-load pricing: A survey. Journal of regulatory economics, 8(3), 215–248.

[15] Ilic, M., Galiana, F., and Fink, L. (Eds.). (2013). Power systems restructuring: engineering and economics. Springer Science & Business Media.

[16] Arce, J. R., Ilic, M. D., & Garce ́s, F. F. (2001, July). Managing short-term reliability related risks. In 2001 Power Engineering Society Summer Meeting. Conference Proceedings (Cat. No. 01CH37262) (Vol. 1, pp. 516–522). IEEE.

[17] https://nsjonline.com/article/2023/01/n-c-utilities-commission-hears-from-duke-energy-on-christmas-blackouts/

[18] Weiss, Joseph. Protecting industrial control systems from electronic threats. Momentum Press, 2010.

[19] Ostrom, E. (2009). A general framework for analyzing sustainability of social-ecological systems. Science, 325(5939), 419–422.

[20] Ilic, M., and Liu, S. (2012). Hierarchical power systems control: its value in a changing industry. Springer Science & Business Media.

[21] Baros, S., and Ilic, M. (2014, July). intelligent Balancing Authorities (iBAs) for transient stabilization of large power systems. In 2014 IEEE PES General Meeting— Conference & Exposition (pp. 1–5). IEEE.

[22] Ilic, M. D., & Zaborszky, J. (2000). Dynamics and control of large electric power systems (p. xviii). New York: Wiley.
[23] Wyatt, J. L., and M. Ilic. „Time-domain reactive power concepts for nonlinear, nonsinusoidal or nonperiodic networks." In IEEE interna- tional symposium on circuits and systems, pp. 387–390. IEEE, 1990.
[24] Ilic, Marija D., and Rupamathi Jaddivada. „Multi-layered interactive energy space modeling for near-optimal electrification of terrestrial, shipboard and aircraft systems." Annual Reviews in Control 45 (2018): 52–75.
[25] Marija Ili/'c, Interaction variables-based Modelling and Control of Energy Dynamics, in Women in Power, Springer Nature Switzerland AG, 2023 (to appear)
[26] electricity-brain-trust@googlegroups. com
[27] https: //www. wsj.com/articles/the-christmas-electric-grid-emergency—11672091317
[28] Ilic, Marija D., and Rupamathi Jaddivada. „Fundamental modeling and conditions for realizable and efficient energy systems." In 2018 IEEE conference on decision and control (CDC), pp. 5694–5701. IEEE, 2018.
[29] Marija Ilic and Donald Lessard, A distributed coordinated architec- ture of electrical energy systems for sustainability, EESG@MITWP- dec302020.
[30] Ilic, Marija D. „Toward a unified modeling and control for sustainable and resilient electric energy systems." Foundations and Trends® in Electric Energy Systems 1.1–2 (2016): 1–141. Ilic, M. D., & Liu, Q. (2012). Toward sensing, communications and control architectures for frequency regulation in systems with highly variable resources. In Control and optimization methods for electric smart grids (pp. 3–33). Springer, New York, NY.
[31] M. Ilic, R. S. Ulerio, E. Corbett, E. Austin, M. Shatz, and E. Limpaecher, „A framework for evaluating electric power grid improve- ments in Puerto Rico," 2020.
[32] Ilic, M., Xie, L., & Liu, Q. (Eds.). (2013). Engineering IT-Enabled sustainable electricity services: The tale of two low-cost green Azores islands (Vol. 30). Springer Science & Business Media.
[33] Ilic, M., Cvijic, S., Lang, J. H., Tong, J., & Obadina, D. (2015, July). Operating beyond today's PV curves: Challenges and potential benefits. In 2015 IEEE Power & Energy Society General Meeting (pp. 1–5). IEEE.
[34] Ilic, Marija, Sanja Cvijic, Jeffrey H. Lang, and Jiangzhong Tong. „Op- timal voltage management for enhancing electricity market efficiency." In 2015 IEEE Power & Energy Society General Meeting, pp. 1–5. IEEE, 2015.
[35] Ilic, M. D., Jaddivada, R., & Korpas, M. (2020). Interactive protocols for distributed energy resource management systems (DERMS). IET Generation, Transmission & Distribution, 14(11), 2065–2081.
[36] Jaddivada, Rupamathi, and Marija D. Ilic. „A feasible and stable distributed interactive control design in energy state space." In 2021 60th IEEE Conference on Decision and Control (CDC), pp. 4950–4957. IEEE, 2021.
[37] Miao, Xia, and Marija D. Ilic. „High Quality of Service in Future Electrical Energy Systems: A New Time-Domain Approach."IEEE Transactions on Sustainable Energy 12, no. 2 (2020): 1196–1205.
[38] Ilic, M. D., & Jaddivada, R. (2021). Making flying microgrids work in future aircrafts and aerospace vehicles. Annual Reviews in Control, 52, 428–445.

[39] Ilic, Marija, M. Wagner, F. Franchetti, K. Bachovchin, R. Jaddivada, S. Ray, Jonathan Donadee et al. „Smart grid in a room simulator-sgrs." National Institute of Standards and Technology, Tech. Rep (2016).
[40] Marija Ilic, Cyber secure Dynamic Monitoring and Decision Systems, MIT patent filing, 2022.
[41] Ilic, M., Jaddivada, R., & Gebremedhin, A. (2023). Unified modeling for emulating electric energy systems: Toward digital twin that might work. In Research Anthology on BIM and Digital Twins in Smart Cities (pp. 107–135). IGI Global.
[42] M. D. Ilic, Miao, X. and Jaddivada, R., „Plug-and-Play Reconfigurable Electric Power Microgrids," U. S. Patent 10,656,609, issued May 19, 2020.

Ayhan IRFANOGLU*

FACING UNCERTAINTY: THE PRACTICE OF EARTHQUAKE ENGINEERING

Abstract: The realm of earthquake engineering practice is rife with uncertainties. The constraints imposed by regulations on this creative profession are often for good reason as they, sometimes, help tame the forces from finance with realities known from Nature. However, there are assumptions built into the current practice of earthquake engineering, whether stemming from or feeding into some of the regulations, which go against the maxim „know what you know and what you do not know" which defines engineering. While many examples may be given from earthquake engineering practice around the World, the dominating attitude in two contrasting cases could be used to illustrate how uncertainty is taken into consideration. One is the case of identifying low to mid-rise buildings with seismic vulnerabilities in large urban areas. This is a case where vast inventories of buildings, and as such, at least an order of magnitude more people, are impacted by the decisions of the earthquake engineers involved. The other case is the practice of designing high-rise buildings in earthquake country where the subject matter is often a single, monumental structure. Both cases are defined by how uncertainties are perceived, quantified, and taken into consideration by various parties involved (owners, local jurisdiction/government, and engineers) and by the actions taken, or not taken, in the face of these uncertainties. Typical outcomes in these cases may be surprising to the untrained in the current practice of earthquake engineering.

Key Words: *earthquake engineering, uncertainty, risk, decision making*

INTRODUCTION

Earthquake engineering as a well-defined profession is about a hundred years old. Earliest modern engineering approaches to seismic design of structures were developed in Italy and Japan, in early 20th century [1]. Spurred by the 1924 Kanto (Japan) and 1933 Long Beach (USA) earthquakes, with the

* Ayhan Irfanoglu, Professor and Associate Head, Lyles School of Civil Engineering, Purdue University, West Lafayette, Indiana 47907, U. S. A.

first strong earthquake ground motion records obtained during the latter event, there was keen interest in expressing fundamental concepts governing earthquake demands and response of buildings to ground motions, and that with stunning insight [2, 3, 4]. These original works were, unfortunately, forgotten or ignored for several decades. Later, the ideas and insight presented in them are proven to be true by independent researchers [5, 6, 7, 8].

Meanwhile, following the first World conference on earthquake engineering in 1956 and accelerated by growth in the science of seismology as well as engineering research, laboratory data and field observations, earthquake engineers around the World have started using similar models to characterize earthquake demands, expressions for building seismic performance, and guidelines/provisions for design and construction. The resulting lingua franca has allowed rapid exchange of ideas. However, at the same time it has increased the pressure to adopt similar approaches to design and field implementation. Certainly, it is not true that all seismic design provisions around the World are identical. But apart from the acceptable seismic performance levels considered in different countries with seismic design codes, one could see that not only the language but more importantly the quantification of earthquake hazards (tied to uncertainties in earthquake occurrences as well as representation of earthquake demands) and the tools used in earthquake engineering practice (modeling structures, analysis, simulation, design principles, construction guidelines and specifications) are practically the same around the World. It is important to ponder whether such nearly uniform thinking and implementation are in the best interest of all affected by the resulting outcomes.

TWO ILLUSTRATIVE CASES REGARDING FACING UNCERTAINTY

The most common task one comes across in earthquake engineering practice is, by far, designing a low to mid-rise building, say, up to ten stories tall following a local design code. Perhaps a distant second most common task is carrying out detailed inspection of an existing building, again, low to mid-rise, for possible seismic vulnerabilities. While there are many other tasks earthquake engineers may do occasionally, two of the much less common tasks but with very high impact on the society in earthquake country appear diametrically opposite in the spectrum of earthquake engineering practice. One is inspecting large inventory of buildings, i. e., tens to tens of thousands of buildings, for possible seismic vulnerabilities and, if necessary, taking action to retrofit or demolish them. The other one is designing high-rise buildings.

Case I — Inspecting Large Inventory of Buildings

After every strong earthquake in an urban area with a large population, there is heightened interested in inspecting buildings, in the same region and sometimes elsewhere, to identify the most vulnerable buildings rapidly for further, detailed study, and, if need be, for retrofit or demolition. Most earthquake engineers approach this task with the training and mentality they have acquired from designing new buildings. Meaning, they use a „high-pass filter“ and work to identify those buildings that are not vulnerable. This often results in vast majority of the buildings in the inventory to be categorized as vulnerable [9]. Certainly, the element of fear, namely the „fear of failure“ is what drives such thinking [10], and it originates from the engineers acting conservatively in the face of uncertainty. Ironically, and unfortunately, such thinking and action by the engineers often guarantee inaction by the powers-to-be, i. e., political leaders and financiers.

In dealing with large inventory of buildings with varied characteristics and facing numerous uncertainties, one should take a simplifying and practical approach [11, 12] and, effectively, execute a „low-pass filter“ to identify the most vulnerable buildings. Such an approach is more likely to result in action as the political will and financial means may be possible to build or find.

Case II — Designing High-Rise Buildings in Earthquake Country

The design of high-rise buildings is different from that of low- or mid-rise buildings in that, often, the resulting design of such large, monumental structures is reviewed, and a recommendation is made to the authority with jurisdiction, by a panel of peer reviewers.

There are several orders of magnitude more low- to mid-rise buildings than high-rise buildings. Our experience with low- and mid-rise buildings, with regards to their design basis and how they perform during strong earthquake ground motions, is more robust and firm-footed simply because we have large collections of evidence gather over the decades and in different parts of the World. Even on the engineering seismology front, we have greater experience: low- and mid-rise buildings are influenced primarily by seismic waves with relatively shorter period. The amplitudes of these shorter period waves are known to saturate with earthquake magnitude. In other words, they are capped. Meaning, even in the case of rare events considered in design of low- and mid-rise buildings, such as those earthquakes represented with 500 year or even 2,500 year return periods, the earthquake demands could be estimated and accounted for with reasonable confidence [13, 14].

What we know about the shorter period waves and how low- and mid-rise buildings respond to them is unlikely to change by much with new data.

Statements like those made above for low- and mid-rise buildings cannot be made for high-rise buildings. Our experience with of high-rise buildings is limited. Data about long-period wave characteristics of near-source ground motions from large earthquakes are also limited. Simply, we do not have enough empirical data to calibrate our understanding, put a cap on demands, or even assign a true level of confidence on what we know about these demands let alone what we do not know about them. There is justified skepticism about what is claimed to be known about these long-period demands during intense, near source ground motions. For example, we do not know what type of probability distribution might be appropriate to use to model these near-source long-period seismic waves, or simply how big they may get [13, 14]. As a result, we are not quite sure how high-rise buildings affected by them might perform. Yet, some of the dominant concepts currently used in design of low- and mid-rise buildings are applied practically as-is in designing high-rise buildings [14]. Despite such limited understanding, in the face of uncertainty and against such odds with dire consequences if realized, high-rise buildings are being built with reportedly high confidence.

CONCLUSION

Uncertainty is unavoidable in earthquake engineering practice. When making decisions in the presence of uncertainty, earthquake engineers need to weigh the possible consequences of their actions, and inactions, against the odds of being wrong and tested in the future versus missing an opportunity to help reduce future losses. Whatever they do, they must know what they know and what they do not know.

BIBLIOGRAPHY

[1] R. K. Reitherman: *Earthquakes and Engineers: An International History*. American Society of Civil Engineers, 2012.

[2] G. E. Howe GE: „Requirements for buildings to resist earthquakes" American Institute of Steel Construction, 1936.

[3] H. M. Westergaard: „ Measuring earthquake intensity in pounds per square foot." *Engineering News Record,* 20 April 1933, p. 504.

[4] H. M. Westergaard: „Earthquake-shock transmission in tall buildings." *Engineering News Record,* vol. 111, no. 22, 1933, pp. 654–656.

[5] M. A. Sozen: „Review of earthquake of R/C buildings with a view to drift control.“ In *Proc. Of the 7th World Conference on Earthquake Engineering, State of the Art Panel Reports,* 8–13 September 1980, Istanbul, Turkey, pp. 383–418.
[6] J. F. Hall, T. H. Heaton, M. W. Halling, D. J. Wald: „Near-source ground motion and its effects on flexible buildings“ *Earthquake Spectra,* vol. 11, no. 4. 1995. pp. 569–605.
[7] M. A. Sozen: „On first reading Westergaard's „Measuring Earthquake Intensity...“ Engineering News Record, 20 April, 1933.“ In *Proceedings of the Otani Symposium*, 2003, pp. 317–320.
[8] M. A. Sozen: „The velocity of displacement“ In *Seismic Assessment and Rehabilitation of Existing Buildings,* Springer, Dordrecht, 2003, pp. 11–28.
[9] T. Paret, A. Irfanoglu: „From complexity to simplicity — How Mete Sozen arrived at Priority Index“ in *11th National (U. S.) Conference on Earthquake Engineering,* Los Angeles, June 25–29, 2018.
[10] M. A. Sozen: „Fear of Failure“ In *World Bank Meeting on Assessment and Repair,* Ankara, Turkey, 5–6 May 2000
[11] A. Hassan, M. A. Sozen: „Seismic vulnerability assessment of low-rise buildings in regions with infrequent earthquakes“ *ACI Structural Journal,* vol. 94, no. 1.1997, pp. 31–39.
[12] A. Irfanoglu, S. Pujol: „Earthquake-vulnerability evaluation of buildings“ in *17th World Conference on Earthquake Engineering,* Sendai, Japan, September 27—October 2, 2021
[13] M. Yamada, A. H. Olsen, T. H. Heaton: „Statistical features of short-period and long-period near-source ground motions.“ *Bulletin of the Seismological Society of America,* vol. 99, no. 6, 2009, pp. 3264–3274.
[14] T. H. Heaton: „Will performance-based earthquake engineering break the power law?“ *Seismological Research Letters,* vol. 78, no. 2, 2007, pp. 183–185.

Željko TORBICA[*]

PURDUE UNIVERSITY'S UNIQUE MODEL OF CONSTRUCTION MANAGEMENT EDUCATION — LESSONS LEARNED

Abstract: Construction Management program at Purdue University has introduced a unique educational model for its undergraduate degree. The curriculum transformation process has been an extremely complex endeavor with a set of unique challenges for which there was no reference in the existing literature. The article briefly describes Purdue's educational approach and provides a summary of some of the major challenges that have been encountered. The main goal is to encourage the discussion and exchange of ideas aimed at advancing the quality of undergraduate education regardless of the field of study. Although it is recognized that the complexity of the Purdue model may be too complicated to be adopted in its entirety (as a „whole"), it is believed that there are many aspects of the model that other institutions can study and potentially benefit from.

Key words: *construction management, undergraduate education, integrated curriculum*

1. INTRODUCTION

Purdue University's reputation as a premier institution in engineering, science, and technology, is well established both nationally and internationally. In a complimentary fashion, the School of Construction Management Technology (SCMT) has been building the reputation as a premier institution for the advancement and dissemination of knowledge in the field of construction management. The School is one of six academic units in the Purdue Polytechnic Institute (PPI) which is one of the 10 academic colleges of Purdue University. The School is offering a full spectrum of degrees including baccalaureate (BS), masters (MS) and doctorate (Ph. D.) degrees. The subject of this paper is the ongoing transformation of the undergraduate

[*] Željko Torbica, School of Construction Management Technology, Purdue University, USA

construction management program. An overview of Purdue's unique educational approach is provided, with a summary of some of the challenges encountered, results accomplished and lessons learned. The goal is to encourage the discussion and exchange of ideas aimed at advancing the quality of undergraduate education regardless of the field of study.

A Bachelor of Science (BS) degree at Purdue University was officially introduced in 1967. An important milestone in the program's evolution was obtaining the accreditation from the American Council for Construction Education (ACCE) in 1979. The program was one of the first two programs in the country that were accredited by the ACCE. The School has demonstrated its systematic and sustained efforts of providing a quality education through an application-based classroom, laboratory, and immersive learning experiences aimed at enabling students to become problem solvers and critical thinkers. Purdue construction management graduates are the most sought-after entry-level construction management professionals in the country as they have consistently demonstrated their readiness for work in any sector of the construction industry, for any size of company, nationally and internationally. Purdue CM career fairs are the largest construction management fairs in the country typically attended by about 200 companies.

2. PURDUE CM UNDERGRADUATE CURRICULUM TRANSFORMATION

Purdue Polytechnic Institute (PPI) has gained a national reputation for its teaching philosophy that was reformed around *10 Elements of Transformation* in response to a changing economy and a changing student base. The 10 Elements include: (1) Theory-Based Applied Learning; (2) Team Project-Based Learning; (3) Modernized Teaching Methods; (4) Integrated Learning-In-Context Curriculum; (5) Integrated Humanities Studies; (6) Competency Credentialing; (7) Senior Capstone Projects; (8) Internships; (9) Global/Cultural Immersions; and (10) Faculty-to-Student Mentorship. The curriculum transformation undertaken by the School of Construction Management Technology (SCMT) is in direct response and alignment with the 10 Elements of Transformation, but the extent of the transformation has been significantly broader and more complex. In fact, the construction management curriculum transformation has been one of the most complex and unique initiatives in a long history of Purdue University; we are not aware of any other academic program that has attempted to implement the integration of this magnitude encompassing the entire four-year curriculum. Since there is no roadmap to follow, the SCMT faculty have had to demonstrate a great deal of creativity and ingenuity in implementing this transformation.

The impetus for the transformation was the fall 2015 School's retreat meeting in which the faculty were asked „*If you could start over from scratch, what would you do*?" The faculty took the challenge and set a goal to transform curriculum into an innovative learning environment that creates a „*seamless transition from college to industry*". After 2 years of preparation the new curriculum was launched in the Fall 2017 semester and subsequently the first cohort of students following the new curriculum graduated in December 2020, one semester ahead of schedule.

The „old" curriculum was organized in a traditional way with the Construction Management Body of Knowledge (CM BOK) being „compartmentalized" in a number of individual courses, typically, 3-credit-hour courses, each covering a single subject area, and taught in particular year/semester in isolation of other courses. The new curriculum, on the other hand, attempts to „de-compartmentalize" the CM knowledge, by combining material from several single-discipline courses into new „integrated" multi-discipline courses („horizontal" integration). The material from any „old" single-discipline course in now being spread over a number of courses and is covered in multiple semesters („vertical" integration). Figure 1 graphically depicts that process.

As an example, the subject area of „Estimating", which, in the „old" curriculum was covered in a single 3-CH Estimating course in the third (Junior) year, is now taught in (at least) 5 courses throughout the entire four-year curriculum; in CM 100 (1st year), in CM 200 (2nd year), in CM 300 (3rd year), and in CM 400 and CM 450 (4th year) (see Figure 2).

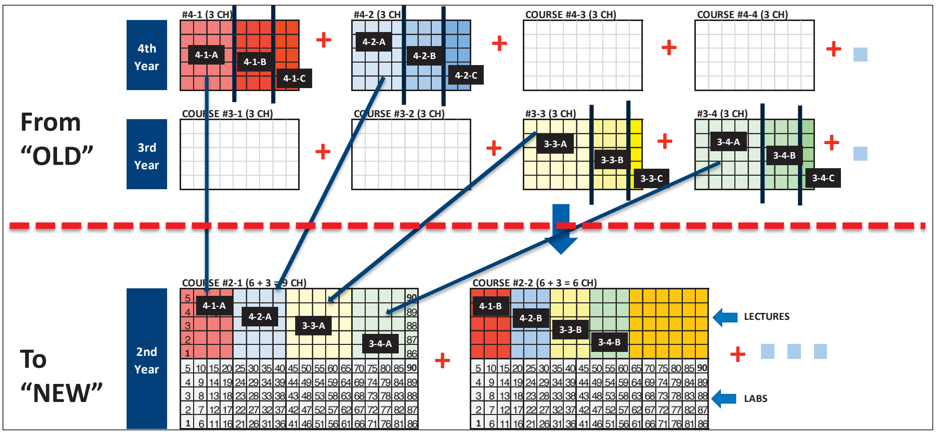


Figure 1: CM knowledge „de-compartmentalization"

ESTIMATING	CM 100	CM 200	CM 300	CM 400	CM 450

Figure 2: Courses that include „Estimating" subject area

One of the results of this transformation is that the single-discipline 3-CH courses are „replaced" by bigger, 6-CH and 9-CH, multi-disciplinary courses, such as CM 200: Intermediate Pre-Construction Management. As it can be seen from Figure 3, this 9-CH course is covering material from (at least) 15 different subject areas, and is taught by a team of 10 + instructors.

CM 200	ESTIMATING	PLAN READING	ACCOUNTING	CAREER PREPARADNESS	COMMUNICATIONS
	SCHEDULING	SUSTAINABILITY	HISTORY	PROJECT MANAGEMENT	DESIGN MANAGEMENT
	STRENGTH OF MATERIAL/ SOILS/ STRUCTURES	SAFETY	MEP	COMPANY MANAGEMENT	OTHER

Figure 3: Subject areas covered in CM 200

3. „NEW" LEARNING ENVIRONMENT

The contents of the old and new curriculum are practically identical; we believe the program content realistically represents the contemporary Construction Management Body of Knowledge (CM BOK). What has changed is the „manner" in which the content is taught and learned. Heavy emphasis is placed on Active Learning (anything course-related that students in a class session are called on to do other than simply watching and listening to a lecture and taking notes[1]), **P**roject-**B**ased **L**earning (PBL) (a teaching method in which students gain knowledge and skills by working for an extended period of time to investigate and respond to an authentic, engaging, and complex question, problem, or challenge); utilization of documents and project files from actual projects („case studies"); and utilization of co- and team-teaching.

We claim that the new, vertically and horizontally integrated curriculum, which is studied in an authentic, project-based, team-taught environment, provides students more opportunities to synthesize material earlier in their college career rather than waiting for a capstone course in their senior year. We also believe that the students retain more of their knowledge and education by scaffolding construction content throughout the four-year

[1] Felder, R., and Brent, R. Teaching and Learning STEM — A Practical Guide. Wiley, 2016.

program rather than condense all information into „stand-alone" 3-credit-hour courses.

4. CHALLENGES

Some of the most complex challenges of the transformation process are listed below.

Need for More Resources (including Faculty Lines)

Working on the transformation has taken a significant toll on faculty and staff. The first three years after the launch of the new curriculum were particularly demanding as the School was teaching-out the „old" curriculum, while developing and introducing new courses. In addition, introduction of co-teaching and team-teaching imposed some additional challenges, one of which was the determination of „actual" faculty teaching efforts („loads"). A new methodology for calculating actual efforts was developed in fall 2019 as a result of a comprehensive analysis that was conducted over the period of 4 months. One of the findings was that some faculty members

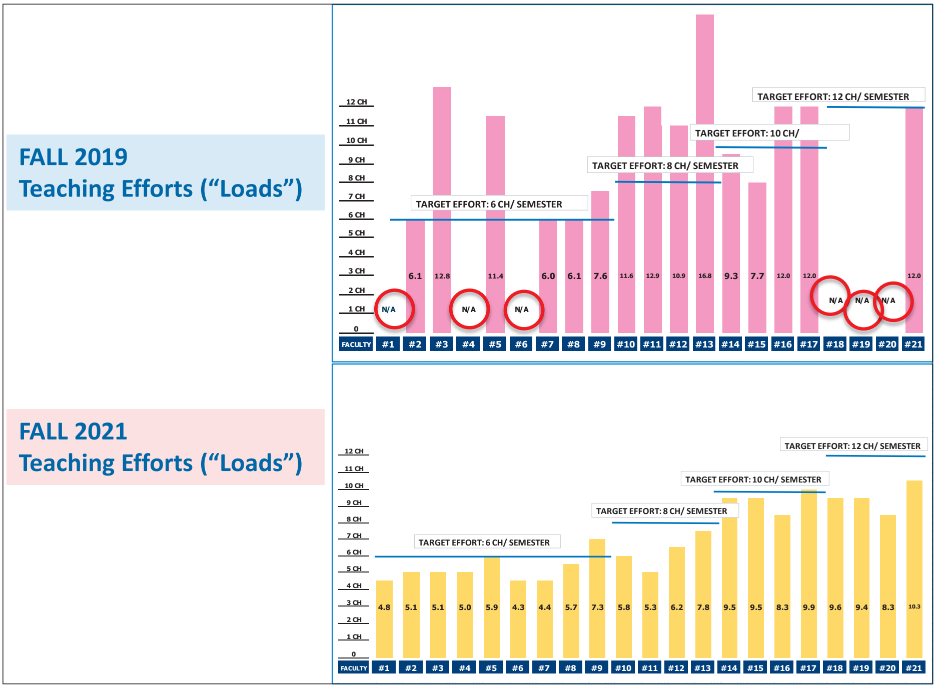


Figure 4: Comparison of teaching efforts („loads"): Fall 2019 and Fall 2021

had had consistently high teaching loads efforts („loads“). We were also able to make a strong case for additional faculty lines. More faculty members and the utilization of the new methodology have helped us improve the fairness of the process of sharing teaching responsibilities among faculty members (see Figure 4).

Utilization of *Co*-teaching and *Team*-teaching

One of the distinguishing outcomes of the transformation is that each course is now being taught by a team of faculty members who are responsible for the instruction of materials in their areas of expertise. In addition to courses being team-taught, some individual lecture and lab sessions are also jointly taught („co-taught“) by multiple faculty members. Figure 4 depicts the difference between the two teaching arrangements, co-teaching and team-teaching. Two or more instructors conducting a joint session to the same group of students is considered co-teaching. If a group of students is split into smaller groups (teams or sub-groups), and each sub-group is taught

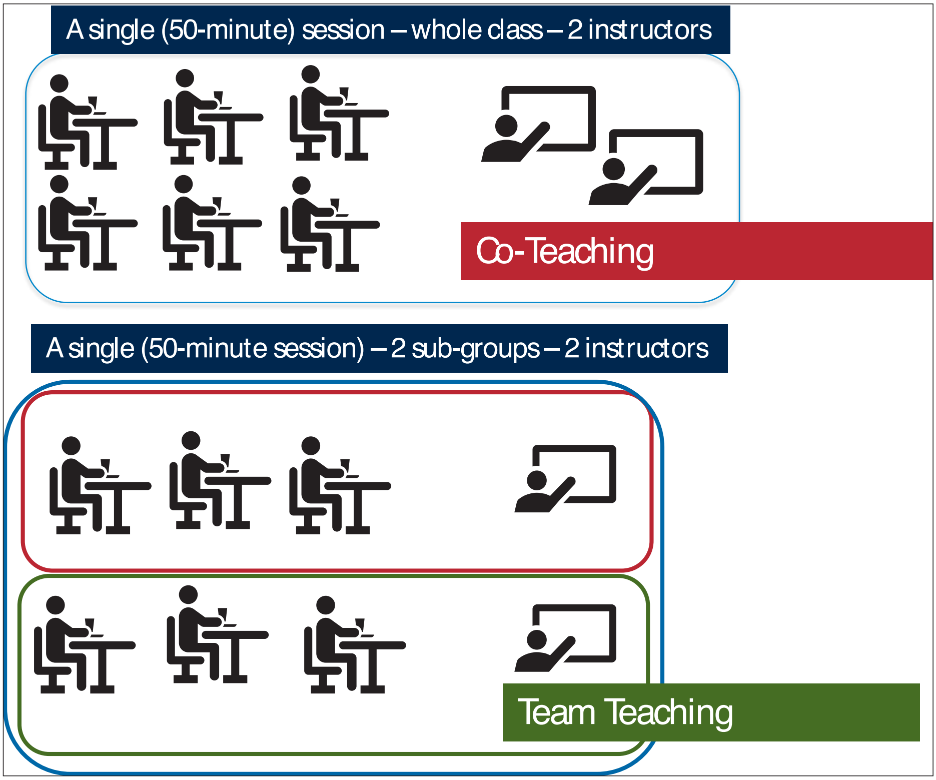


Figure 5: Co-teaching vs. Team-teaching

by a different single instructor, this arrangement is considered as team-teaching. We differentiate „co-teaching" from „team-teaching", in recognition that the „co-teaching" is typically a more complex teaching arrangement.

Identifying which sessions/ topics should be co-taught, what is the best way to conduct the joint sessions, which instructors should team up for a particular joint session/topic — these are some of the dilemmas that we are facing on a daily basis.

Management of Individual Courses

New courses have become much bigger and more complex to manage. A composition of a typical 9-CH course is shown in Figure 6.

There are 90 50-minute „lecture" sessions, and 90 50-minute „lab" sessions. As it was mentioned before, teams of instructors, in some cases with more than 10 of them, are involved in teaching a course, making the course coordination exceptionally complex. As a result, there is a need for frequent coordination meetings between team members, and also for each course it was necessary to assign one or two „Course Managers" who are responsible for the overall course coordination.

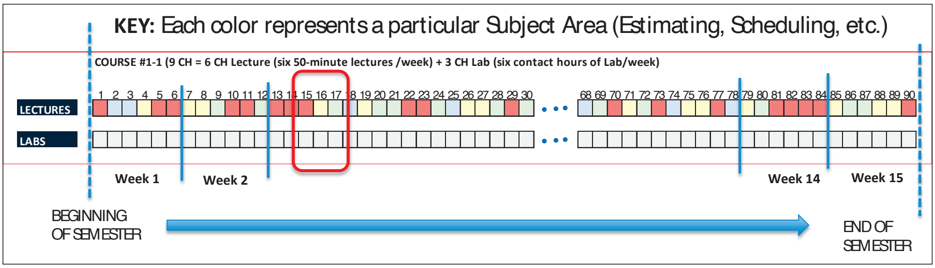


Figure 6: Typical structure of a 9-CH course

Content and Topic Integration of Individual Courses

The content of each course was originally determined based on an analysis conducted by an assigned faculty team. It was anticipated that this would be just a starting point and that more work will be needed to find the „optimal" course content and to determine the „best" way to integrate the various subject areas within a given course. Using an imaginary course shown in Figure 6 as an example, we should be able, for example, to provide a convincing rationale for why and how the specific topics for lectures 15, 16, and 17 (representing three different subject areas) were chosen? Also, are these sessions synchronized in a meaningful way from the students' learning point of view? Fine-tuning of individual courses and the search for the „optimal" integration of topics has been an „unending" work.

More Expensive System

This new educational model is more expensive than the previous model mainly due to a significant lab component (which requires splitting classes in multiple sections), co-teaching and team-teaching, complex course coordination, and the need for course managers. For example, as it can be seen from Figure 7, delivering CM 300 course, which is a 9-CH („Published" CH) course, actually requires more than 30 CH („Actual" CH) — an increase of 300%.

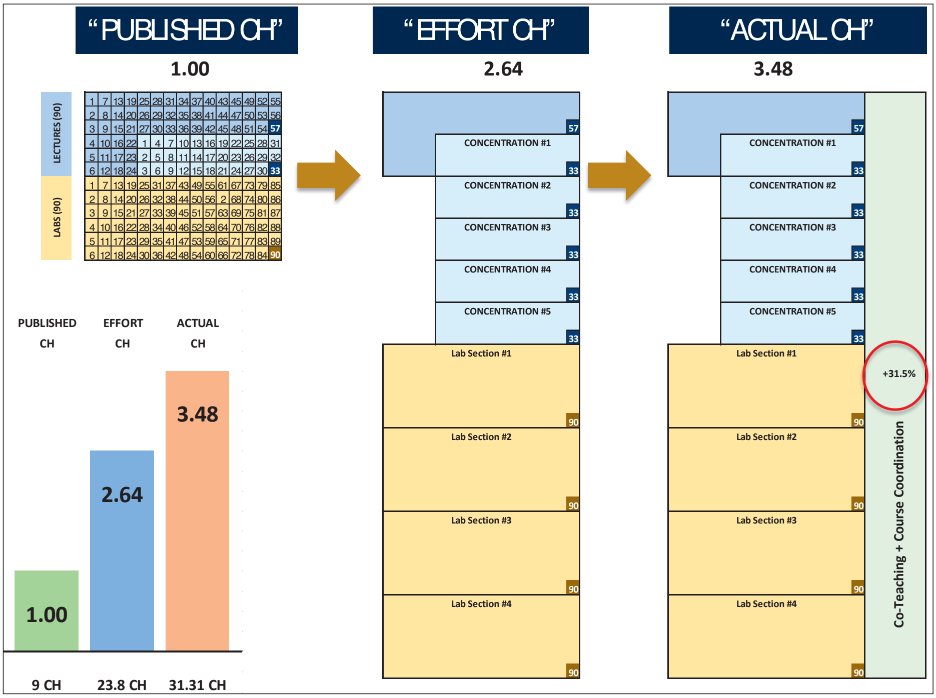


Figure 7: Actual cost of delivering CM 300

5. NEXT STEPS

As it has been suggested in the preceding discussion, the „fine-tuning" of this unique educational model will take years of additional work. As an example, just in the Fall 2021 semester there were 12 meetings in which we worked on a number of topics related to the curriculum transformation.

We invested 695 person-hours in these discussions with additional 300 + hours of preparation time. One of the outcomes was a 300-page document of original scholarship material.

Some of the most pressing items that will be the focus of our attention in the next 1–3 years are: improving and increasing the use of co-teaching and team-teaching; refining the integration of various subject areas within a single course and across the entire curriculum; finding the optimal lecture-lab ratio; consolidation of the number of instructors involved in teaching a single course; eliminating 9-CH courses and limiting all courses to 6-CH; further improvement of course „Project" assignments; more efficient utilization of actual projects' documentation from the Project library; and ensuring that technology integration in curricula and study plans are up-to-date, responsive to market needs, relevant, and effective in preparing students for future careers.

6. CONCLUDING REMARKS

At the moment, we have insufficient data to provide a reliable answer to the questions like: Does the new model lead to improved learning? or, Are the „new" CM graduates better/ more competent than their predecessors following the „old" model? It will take, perhaps, 5 more years before we will be able to fully assess the success of this new educational model. Based on our experience with several cohorts that have already graduated, we feel optimistic about the quality of our graduates and the prospect that they will quickly develop in top-notch professionals. We have, however, limited „hard" data to support our optimism, mainly the data coming from the direct measurement of 20 ACCE Student Learning Objectives (SLO) which we conduct each semester. First round of *Employee* and *Alumni* surveys will be conducted in 2023 (these surveys are typically conducted 3 years after graduation) and with that data and more input from the industry we will be able to assess the direction and success of our transformation. We do have some anecdotal evidence that we are on the good path. For example, an industry executive who manages two of ours Construction Work Experience courses shared his impressions about our students in the email that was addressed to the SCMT faculty: „*You ...should be extremely proud of the product of students you are putting out into the Construction Industry. The future of construction is in great hands with students from this department and it is easy to see why over 200 companies come to the career fair in search of Purdue CM students.*" Another encouraging indicator is that the first two cohorts of students who followed the new curriculum produced the winners of the prestigious national essay competition sponsored by the Associated

General Contractors (AGC) two years in a row (2021 and 2022), which is an unprecedented accomplishment.

Even if we „confirm" that the proposed model is indeed superior and that it brings clear advantages compared to a more „traditional" approach, the question that remains to be answered is, how suitable it is for adoption by other institutions? We do recognize that the complexity of the model may make it very difficult for other institutions to adopt. However, even if the suggested model proves to be too complicated for adoption as a „whole", we believe that there are some aspects of our approach worth studying that other institutions can benefit from.

Željko TORBICA

JEDINSTVENI MODEL OBRAZOVANJA MENADŽERA U GRAĐEVINARSTVU SA PURDUE UNIVERZITETA — STEČENA ISKUSTVA

Sažetak

Program Menadžment u građevinarstvu na Univerzitetu Purdue je razvio unikatan model obrazovanja za studente na dodiplomskim („redovnim") studijama. Proces transformacije načina studiranja se pokazao ekstremno komplikovan sa nizom unikatnih izazova koji nisu obrađeni u postojećoj naučnoj literaturi. Članak ukratko opisuje pomenuti model obrazovanja sa listom glavnih izazova. Cilj diskusije je da se razmene iskustva i ideje koje bi mogle da unaprede dodiplomske studije nezavisno od discipline studiranja. Usvajanje ovog modela, kao celine, od strane drugih institucija može da se okarakteriše kao komplikovano, međutim, izučavanje i implementacija nekih aspekata modela mogu da budu od potencijalne koristi za zainteresovane institucije.

Ključne reči: upravljanje u građevinarstvu, obrazovanje, integrisani nastavni program

Siniša ĐUROVIĆ*

THE UK POWER AND ENERGY CHALLENGES TOWARDS NET ZERO — A BRIEF TECHNOLOGICAL OUTLOOK

Abstract: The report aims to present an overview of the key challenges and possible solutions for the growing needs of the UK power and energy sectors and their ongoing decarbonisation. The material explores how the UK electricity demand and electrification of services challenges may be satisfied and the 2050 Net Zero target achieved, including some of the considered energy system and power device changes associated with achieving Net Zero.

1. INTRODUCTION

The United Kingdom's (UK's) energy system is today one of the fastest growing globally in terms of the diversity and the composition of the utilised energy sources. This inevitably raises a number of considerable technical challenges on successful and stable integration of new sources into an increasingly decarbonised power grid. The advent of large scale renewable power generation is particularly demanding in this respect, as the intermittent availability of renewables is disruptive to conventional techniques of operating the electric grid and could lead to power shortages if not addressed adequately.

The UK's legal commitment to delivering Net Zero by 2050 imposes further demands to provide what needs to be an intrinsically flexible and resilient energy system of the future. Its structure will fundamentally need to allow reliable real-time manipulation and matching of new and emerging demand sources to renewable generation, and will be particularly challenged

* Siniša Đurović, Department of Electrical and Electronic Engineering, University of Manchester, UK

by the balance of, on one side, the changing peak demands and uncertainty in their prediction, and on the other a peak system capacity increasingly compromised by decarbonised generation. To deliver this capability a number of technological breakthroughs and solutions are required that have hitherto not been conceived or proven on this scale. Furthermore, the energy system must enable the economy to decarbonise including negative emissions provision to offset those sectors that are difficult and more time demanding to fully or partially decarbonise. As carbon neutral and renewable resources are expected to play a key role in this decarbonisation it will be vital for the feasibility of the future Net Zero network to identify new reliable sources of ancillary services, which are essential to effective functioning of the power network and are today largely delivered from sources such as gas and/or nuclear generation.

This study looks to provide a brief overarching review of the current state of the UK energy system and the general requirements for its successful decarbonisation going forward. Different possible scenarios put forward by the UK's electricity system operator to achieve the national Net Zero targets for the electricity system are then overviewed. Finally, some of the key underlying electrification technology challenges to do with electrification of transportation are reviewed to provide an insight into how some of the wider energy system decarbonisation targets translate into those to do with device design and utilisation.

2. PATHWAYS TO A FULLY DECARBONISED POWER SYSTEM

The UK energy power system has traditionally been dominated by conventional carbon rich energy sources, much like those of most developed countries. However in the past few decades an initially slow but a now rapidly accelerating transition towards integration of large scale renewables has been taking place. The current state of the system in 2022 is of an approximate total installed capacity of 105 GW out of which almost 50 GW is renewables, made up mostly of wind power generation but also of a rising percentage of solar and some marine generation [1]. The current peak demand is approximately 50–55GW, and has moderately varied in the past decade driven by a nationwide push for a switch to higher energy efficient appliances and lighting and some reduction in industrial activity [2]; this is widely expect to considerably increase going forward.

The variability of renewables and the challenges this imposes in exploiting the renewables installed capacity was illustrated in 2021 during which, despite an overall increase in installed renewables capacity, the generation from renewables was reported to decrease by close to 9% caused by the reported less favourable conditions for wind, solar and hydro generation [3]. This was most prominently observed for wind generation, whose contribution reduced by 14% compared to 2020, due to unusually low average winds during 2021. Despite this reduction however, the proportion of electricity generation from renewables in the UK in 2021 was second only to that originating from fossil fuels, fully illustrating the importance and the sizeable contribution renewables make to the modern UK electricity network. The 2022 status of the UK's installed electricity generation capacity composition and its corresponding generation output contributions are illustrated in Fig. 1, clearly highlighting the large renewables input in the power generation mix.

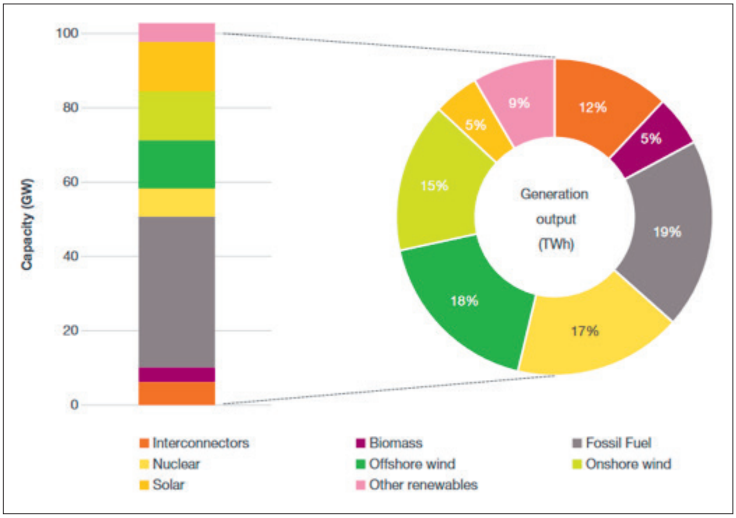


Fig. 1 UK electricity generation capacity (GW) and output (TWh, approx. 333.2 TWh in total in 2021), 2021 [4]

2.1 *The aim: a fully decarbonised network*

The UK has committed its energy future to elimination of emissions and a carbon free economy with the ultimate goal of achieving Net Zero by 2050, and on the road to this a decarbonised power system as early as 2035. The national statistics on emissions by source, shown in Fig. 2, present a growing reduction trend in the past two decades, that is picking up pace in recent years with the increased proliferation of renewables. Perhaps unsurprisingly the biggest net emissions contributor in 2020 was transport

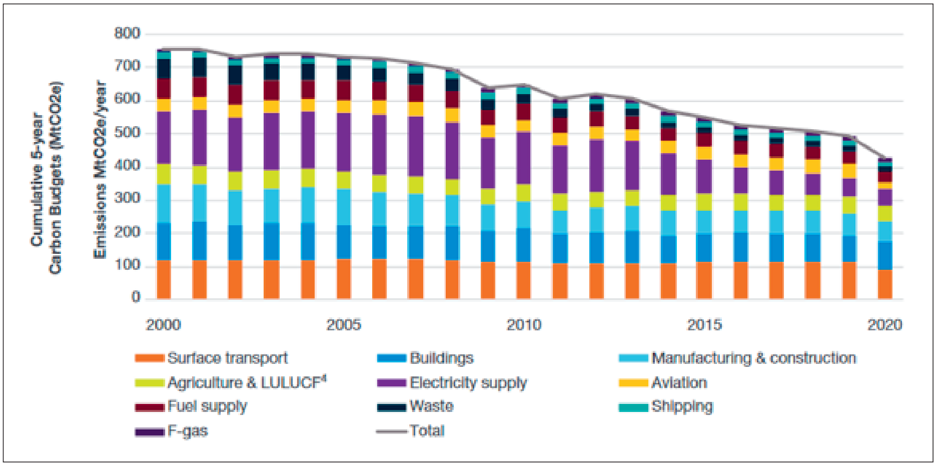


Fig. 2 UK historic emissions by sector [5]

which contributed to almost a quarter of all emissions. The transportation sector is consequently set to undergo a large scale decarbonisation through full and/or partial electrification of ground and eventually of marine and air transport. Similar solutions are being sought for decarbonising the buildings and manufacturing sectors whose carbon footprint is also significant.

The UK desires to significantly accelerate the transition to a carbon free energy system, including the delivery of a reliable, secure and low (ideally zero) carbon power system. To achieve this a number of technological advances will need to be developed and adopted [4] as identified by the national electricity network operator. Some of the key strategic aspects of the required innovation in the electricity and energy system are listed below:

— *Advanced demand side strategy* will be needed to incentivise the consumers to adopt a more flexible energy consumption and thus alleviate the pressures in balancing what is expected to be a challenging to predict and intermittent renewables contribution and a very flexible demand.

— *Large scale energy storage* will be key to success of the future low carbon electricity network and its ability to balance the supply and demand with variable and intermittent generation input. This will be vital to increasing system resilience and also the avoidance of potentially having to waste large amounts of renewable generation output due to it coinciding with low demand/oversupply periods. This will include the provision of large scale storage solutions, either electrical assuming the technology can be delivered in time, or by exploring alternatives such as hydrogen uptake.

— *Improved energy efficiency* can provide rapid benefits on the energy cost and security fronts through demand reduction. This can be achieved through usage of more efficient devices on the network (e. g. lighting, heating, home appliances etc.) but also on a more fundamental level through, e. g. better building insulation to reduce heating related energy consumption.

— *Decarbonisation of heat* is highly pertinent to the UK system which is highly reliant on natural gas for heating, and needs to be undertaken urgently with consideration of regional differences in the UK, and in particular the differences in proximity to energy infrastructure between these.

— *Digitalisation and innovation* will be needed to enable the customer demand and consumption profiles to be better aligned with the changing energy cost and availability profiles. Consumer data availability will be key to enabling developments in smart technology on this front that can facilitate automated or quasi automated adjustment of consumption with cost of energy. This would be beneficial to both the customers and the grid operator/provider and would enable the consumers to take advantage of low cost energy in the period of high renewable output which does not necessarily coincide with conventional peak demand timing. Ensuring both the consumer data security and cyber security would be crucial and integral to this process.

— *Flexibilty*, to securely, effectively and stably operate the future low carbon energy system with high renewables penetration will require a much higher level of flexible capacity than is the case today. This will be needed to offset the loss of traditional ancillary service provided by e. g. natural gas, which is essential to grid stability, but also to allow the optimal contribution from flexible assets in the grid to ensure supply security.

2.2 Possible pathways to Net Zero energy future

The UK national electricity system operator has outlined a number of different possible pathways toward the Net Zero energy future and the UK 2050 target, naming these the future energy scenarios [6]. Four different energy scenarios are identified which follow different trajectories and are all deemed to be potentially credible predictions of future developments in the UK energy sector. The explored scenarios aim to consider the future energy needs, how these may be met and which solutions they may require to realise. The underlying aim of the analysis is to provide an advisory framework for the UK energy industry stakeholders that can be used to guide and underpin the energy network development, support investment decisions and guide policy development.

Four different scenarios for 2050 were defined, exploring different possibilities for how the power system could evolve between now and 2050 through changes in infrastructure innovation, technology and behaviour.

The purpose is to look at different possibilities in decarbonising the energy network (e. g. societal change or energy system technological transformation) and attempt to predict what implications and impact this may have and what future routes it may create. The scenarios are defined as:

1. *Falling short / Steady Progression:* which assumes the slowest possible decarbonisation progression rate scenario that does not meet the 2050 Net Zero target.

2. *Consumer Transformation:* looks at ways of reaching Net Zero by 2050 through changing the consumer behavioural patterns, i. e. the way energy is used and consumed.

3. *System Transformation:* looks at reaching the 2050 target through changing the way we generate and supply energy.

4. *Leading the Way:* looks at the fastest credible way of reaching Net Zero through a combination of consumer engagement and the assuming availability of world leading technology.

Each of the defined scenarios is presented with a corresponding predicted annual energy flow diagram as shown in Fig. 3. The diagrams show the linkage between energy source (diagram left hand side, in TWh) and energy consumption (diagram right hand side, in TWh) per primary energy type. They are conceived to enable the understanding of the annual supply and demand and its different composition possibilities based on the growth scenario adopted.

The Falling Short scenario depicted in Fig. 3a assumes the minimum amount of difference with respect to the current state of the power network and fails to reach decarbonisation targets. This scenario assumes that domestic heating decarbonisation has been very limited and by 2050 still relies heavily on fossil fuel usage (i. e. largely natural gas). It also assumes limited progress with energy efficiency and usage of hydrogen, but does assume

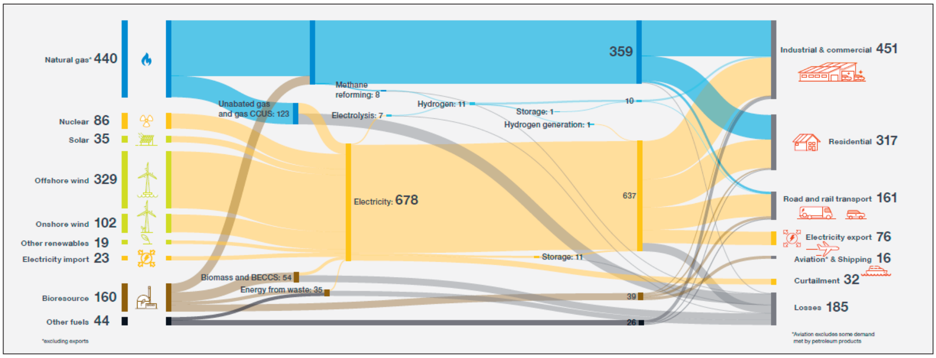


Fig. 3. a Falling Short scenario energy flows, 1237 TWh total

progress has been made with fully electrifying small passenger vehicles, while large transportation vehicles remain fossil fuel powered.

The Consumer Transformation scenario in Fig. 3. b assumes the transport, industry and heating are largely electrified and high levels of renewables are available in the system with some integration of hydrogen. High adoption of energy efficiency is assumed and combined with successful large scale electrification predicted to lead to the lowest end user energy demand out of the four scenarios. The System Transformation scenario, Fig. 3 c, maintains a high natural gas intake but however assumes the highest proportion of hydrogen integration and use for powering the industry, heating homes and operating high girth transportation vehicles. This is assumed to be followed by moderate progress with electrification.

Finally, the Leading the Way scenario assumes that industrial and heating needs are largely met by 2050 through a combination of usage of hydrogen and electricity. A high grid flexibility is attained through hydrogen manipulation and direct air carbon capture and storage systems used for

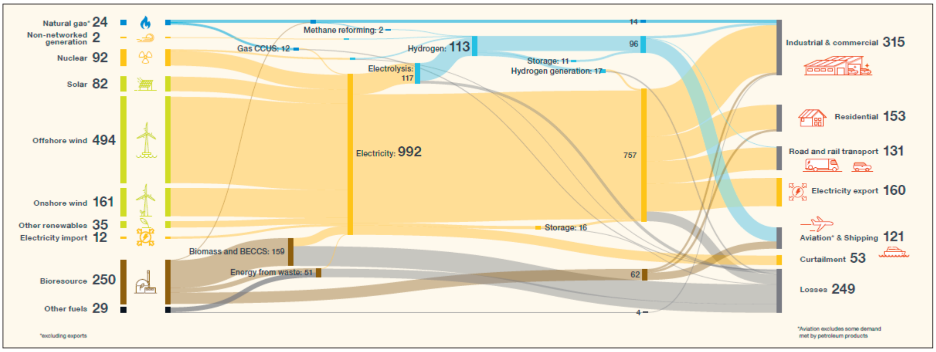


Fig. 3. b Consumer Transformation scenario energy flows, 1182 TWh total

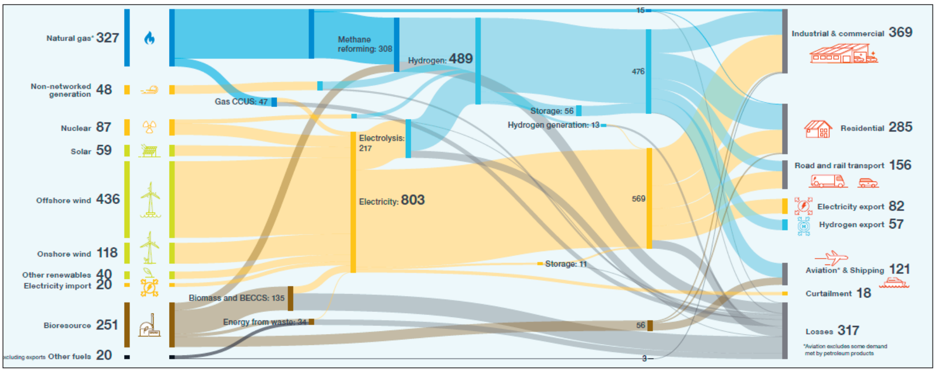


Fig. 3. c System Transformation scenario energy flows, 1406 TWh total

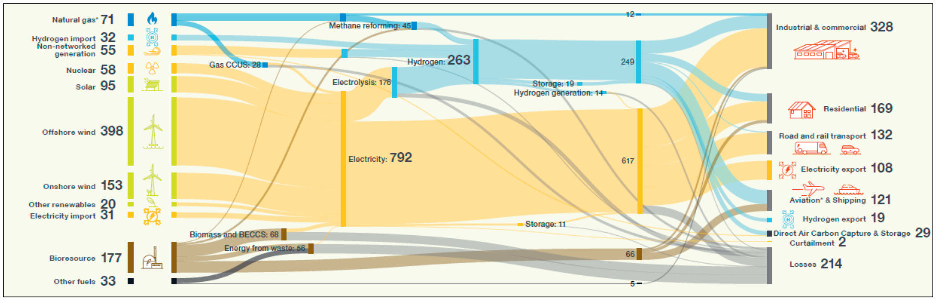


Fig. 3. d Leading the Way scenario energy flows, 1123 TWh total

Fig. 3 The UK National Grid's future energy scenarios energy flow [4, 6]

negative emissions, i. e. to offset the emissions from the remaining reduced amount of fossil fuel used in generation.

All four scenarios are deemed to be realistic and possible, and detail the implications of different decisions in the process of power network decarbonisation. Their findings indicate that there are different potential paths that could deliver Net Zero, but also that the viability of this delivery is vitally dependent on the adoption and success of large scale electrification of predominantly the transportation and heating sectors, and reduction of any leftover fossil fuel usage related emissions.

3. TRANSPORT ELECTRIFICATION: ONE OF KEY INGREDIENTS TO NET ZERO

Transport electrification remains one of essential barriers to wider Net Zero adoption in the UK and one that requires imminent and extended duration action to be addressed. The UK is strongly committed to changing this landscape and moving towards full road electrification at a reasonably rapid pace. As part of this push the UK has decided to end the sales of all diesel and petrol cars in the country by 2030 and to focus on use of electric vehicles (EVs) exclusively beyond [7].

There has already been steady progress on facilitating increased usage of EVs with the energy infrastructure, however this is still at relative infancy. Fig. 4 illustrates the growth in EV charging infrastructure in the UK over the past few years, which is seen to be rapidly accelerating with the early uptake of EVs. However while the pace of charging infrastructure rollout is increasing it is currently deemed too slow to cater for fast rising EV users needs, and requires further intensification [8].

The large scale transport electrification targets on the national level translate into targets for advances in the underpinning technology needed to

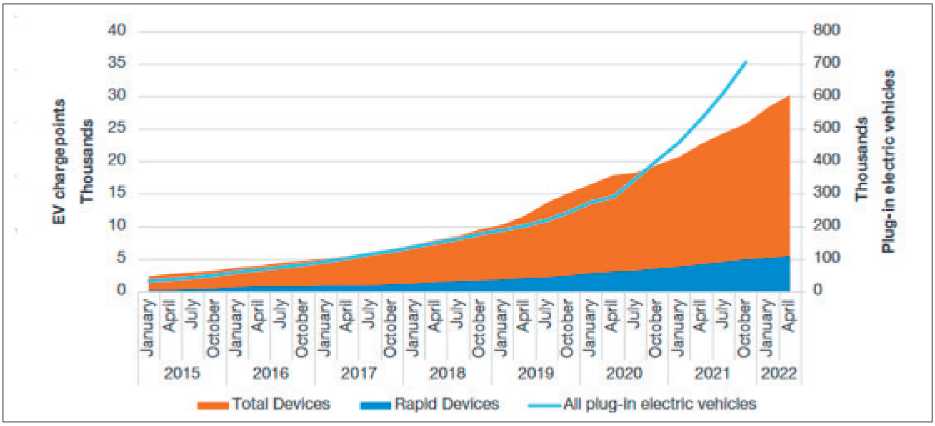


Fig. 4. EV public charge point installed in the UK [1]

deliver the technically and economically viable EVs in required numbers. The relevant automotive technology roadmaps look at mapping the future developments of specific enabling technologies and set out performance target that align with target performance desired from EVs [APC, US ev everywhere]. At the core of these is the electromechanical power conversion system of the vehicles making up their powertrain and containing the propulsion electric motor, the power electronics to drive it and the energy source to supply it from.

The electrical energy storage system remains the main challenge to large scale commercialisation of EVs. This is pertinent to both the battery power/energy density and its cost, as while these are continuously improving neither are still not quite at the levels required to ensure the desired mass adoption of EVs. The targets set out by the UK Automotive propulsion council stipulate that, to ensure long term EV viability, by 2035 the power and energy density of EV batteries need to rise from the current 3kW/kg and 280Wh/l, to 12kW/kg and 100Wh/l, respectively, while at the same time the cost is expected to reduce by more than 50%. The battery technology is largely expected to remain dominated by the traditional Lithium-ion solutions however alternative chemistries are researched. To attain the desired improvement in energy/power density and cost, improved cell chemistry, battery management and manufacturing processes will need to be developed.

Another key element of the EV powertrain is the power electronic converter, whose performance is essential for ensuring optimised power flows between the propulsion motor and the battery, and hence the delivery of desired traction. EV power electronics will continue to heavily utilise traditional silicone based semiconductor devices. However it is widely expected that the emerging class of wide band gap semiconductor devices will be increasingly finding implementation in EV traction drives as they can offer improved electrical and thermal performance. For the potential of these devices to be utilised fully new solutions will be needed around their manufacturing, integration and cost, but also around conceiving new circuit topologies that can best utilise their performance potential. In addition, the thermal management of power electronics is widely expected to remain one of key challenges in their improved utilisation in this area. The set targets for EV power electronics power density is to reach 50kW/kg by 2035 rising from the current 15kW/kg, while also achieving around 40% cost reduction at the same time.

The electric motor is at the heart of the EV system and its key propulsion enabler. The EV e-motors performance is also the subject to mandatory improvements, and in particular where power density and efficiency is concerned. Their manufacturing also remains a challenge with the mass market maturity in this sector being fairly low still, and needing advances for the mass market application costs to be brought down. The e-motor performance improvements are expected to be delivered through exploring the application of new materials and closer integration with the power electronics elements of the power train and to an extent through exploration of increased speed designs. The targets for e-motor power density are to increase to 9kW/kg by 2035 from the current average of around 3 kW/kg.

4. CONCLUSIONS

This paper has presented a brief overview summarising the key challenges in decarbonising the UK energy network. Viable means for network decarbonisation were reviewed in various network growth scenarios, as put forward by the UK national electricity network operator. The success of UK decarbonisation by 2050 is found to depend on a complex combination of a range of technologies: it fundamentally revolves around electrification of heating and industry as well as electrification of transport, and around large scale integration and exploitation of renewable power generation. If further requires the currently unavailable large scale energy storage technology, as well as a host of additional advances to maintain the Net Zero power network stable, flexible and resilient. Finally the delivery of set decarbonisation targets is underpinned by requirements for improvements in the underlying enabling technology. This is perhaps best illustrated in the set electrification of ground transportation targets, which effectively translate in requirementt for considerable technological advances in the underpinning powertrain technology, ranging from much better small scale energy storage, to high power density power electronics and traction e-motors. To deliver these will need not only significant investment but also a range of innovative engineering techniques that remain to be conceived and proven.

REFERENCES

[1] 'UK Energy in brief 2022 'The UK Gov. Department for Business, Energy & Industrial Strategy, Aug 2022.
[2] A Vaughan, 'UK electricity use falls–as rest of EU rises 'The Guardian, 2018.
[3] 'Digest of UK Energy Statistics 'The UK Gov. Department for Business, Energy & Industrial Strategy, Aug 2022.
[4] 'Future Energy Scenarios 2022 Report 'UK National Grid, 2022.
[5] 'Final UK greenhouse gas emissions national statistics: 1990 to 2020', The UK Gov. Department for Business, Energy & Industrial Strategy, June 2022.
[6] 'FES in Five 'UK National Grid, 2022.
[7] 'Government takes historic step towards net-zero with end of sale of new petrol and diesel cars by 2030 'The UK Gov. Department for Transport, Office for Low Emission Vehicles, Department for Business, Energy & Industrial Strategy, Nov 2020.
[8] 'Taking charge: the electric vehicle infrastructure strategy 'The UK Gov. Department for Transport, March 2022.

Bratislav STIPANIĆ[*]

ADVANCES IN BRIDGE ENGINEERING

Abstract: The paper treats overall aspects of advances in bridge engineering, affected by rapid development of natural sciences and technical innovations in design & structural analysis (static and dynamic due to wind & earthquake), building materials (high-strength & high-performed concrete, high-strength steel) and construction technology (building procedures, prefabrication, robustness). Consequently, advancements have not been only made in improved realizations of classical structural types (beam, frame, truss, arch & suspension bridges), but futhermore in application of inovative structural types as: integral bridges, cable-stayed and extradosed bridges. Nowadays the Building Information Modeling (BIM) is referred to as a process that connects engineers (involved in bridge design, construction, supervision & management) very efficiently in the various stages of construction.

The prestressed concrete bridges, steel bridges (with orthotropic deck) and steel-concrete composite bridges are dominantly applied nowadays. The number of bridges with super-long (over 500m) and ultra-long spans (over 1000m) has been considerably increased in this century. The tables of longest spans for arch, cable-stayed and suspension bridges are given in the paper.

As the particular new bridge achievements, built in ex-YU region, are noted in the paper: Roadway Cable-Stayed Bridge Ada in Belgrade (2011), Roadway-Railway Arch Bridge in Novi Sad (2018), Highway Beam-Frame Bridge Moračica at Podgorica (2022) and Roadway Multy-Span Cable- Stayed Bridge Pelješac (2022). The Cable-Stayed Bridge Solidarity in Plock (2007), with 375m span (the longest one where pylons are fixed to girder and the longest bridge span in Poland), is designed by the paper author, as the co-author. The paper author's scientific works are the contributions to analysis of cable-stayed and box beam bridges.

[*] Bratislav Stipanić, Faculty of Civil Engineering — University of Belgrade, Project Biro UTIBER Novi Sad

INTRODUCTION

Bridge engineering is a branch of civil engineering, which includes: planning, design, construction, operation and maintenance of bridges, in order to provide safe and operative transportation of vehicles for people & goods. Transportation infrastructures — roadways and railways are vital factor for economy and overall quality of human life. In the frame of transportation networks, bridges are crucial for connecting people and delivering goods.

Bridge is a structure built over some physical obstacle (natural as: river, lake, sea strait, valley, canyon; or artificial as: roadway, railway, waterway, building), for the purpose to provide the passage by crossing over the obstacle.

Bridges are mainly classified according to: function (roadway, railway, footway, cycleway), structural system (beam, truss, frame, arch, cable-stayed, suspension) and building material (concrete, steel, timber, masonry).

Ancient bridges were built as masonry (or wood) arch structures and beam wood structures only. The use of steel as replacement for wrought iron (firstly applied for the Coalbrookdale Bridge in 1779) was extended in last quarter of 19th century: the steel Arch Bridge on Missisipi River (159m span), the Brooklyn Bridge — suspension bridge (486m span) with cables of hard steel-draw wires and the Forth Rail Bridge (521m span) with steel truss cantilever construction. The first half of 20th century was noted by steel trusses for beam, arch and suspension bridges; while arch and beam structures dominated as reinforced concrete bridges.

The development of building material production, design modelling, prestressing technology, welding technology and overall building technique had a crucial influence on bridge constructions from middle 20th century.

Consequently, advancements have been not only made in the realization of previously applied structural types (beam, truss, arch & suspension bridges) but futhermore in application of inovative structural types as: integral bridges, cable-stayed bridges, and extradosed bridges.

The pre-stressed concrete bridges, steel bridges (with orthotropic deck) and steel-concrete composite bridges are nowdays dominant in application.

BRIDGE DESIGN PROCESS

The bridge design process traditionally contains the stages: concept design, preliminary design, detailed design and construction design.

The most challenging part of bridge design reflects to an optimal choice of concepts with regard to structural systems, span lengths and cross sections.

The successful concept design results not only from solid scientific-technical knowledge of bridge design engineer, but also from experience, awareness of visual form and creative fantasy.

The preliminary design elaborates further the best proposed concept; the feasibility of the selected concept should be ascertained and its cost estimates should be refined.

The details of the bridge structure are finalized in detailed design, in order to be sufficient for tendering and construction design.

Finally, the construction design provides step-by-step procedures for the building of the bridge according to realization methodology of the chosen contractor.

Nowadays the infrastructure tenders are often published in design-build form. It means that the employer (engaging expert consultancy) should prepare so-called design for tender and the construction contractor, chosen by tender, will make the construction design. The design for tender should be based on the previous concept design and preliminary design, that were made by a respectable design office chosen by employer upon a published tender, an invitation or design competition for most significant bridges.

Starting from the end of the previous century, the FIDIC contracts are the most commonly used standard form of international construction contracts for infrastructure projects.

The so-called FIDIC Red Book is construction contract form in infrastructure projects where the design is provided by the employer, following the traditional procurement route of design, bid and build. The contractor to be paid on measurement basis of the actual quantities of performed work, that were based on the design estimation.

The so-called FIDIC Yellow Book, is a standard contract where the design is carried out by the contractor. The Yellow Book is also known as Design-Build contract. The contractor to be paid on a lump sum basis.

The design engineer primary should take care of bridge structural reliability, that means safety, serviceability and durability; and furthermore: constructability, construction & maintenance costs, as well as architecture appearance & urbanistic integration (paying attention to the client's aspiration).

The requirements, that should be fulfilled in process of modern bridge design, can be classified as:

— technical: resistance (loads, actions, building material), serviceability (deflections, vibrations, traffic functionality, easy maintenace) & other (codes, norms, standards)

— architectural & urbanistic: overall layout, structural elements appearance, non-structural elements appearance & details, accessibility, functionality, integration in environment

— constructional: material availability, contactor availability, building technology

— economical: cost efficiency (construction cost, maintenance cost) and time efficiency (construction time, prefabrication time)

— service life (durability, maintenance, monitoring, inspections)

— environmental (sustainability, produced waste, effect on habitat & nature, efficiency in use of resources)

— legal: (building law).

The input data should be collected for design process, such as: traffic data, site data, topographic data, geological & geotechical data, hidraulic data, climatic data (wind, temperature, rainfall, snowfall, ice, floods), seismic data... The new scientific-technical development enabled higher quality of the input data for design.

In design procedure of a bridge it shall be considered a set of influential factors in order to achieve a best bridge solution, such as: structural system, materials, dimensions, foundations, aesthetics, local landscape, environment. The bridge design engineer is required to provide the most effective structural solution.

STRUCTURAL DESIGN — APPLICATION OF EUROCODES

By the end of previous century, it is established a set of new standards for structural design, so-called Eurocodes — related to different type of structures applied in civil engineering. The Eurocodes, published by CEN, are widely applied for structural design in Europe in 21st century. The Eurocodes provide common structural design rules for everyday use in the design of whole structures and component products.

The Eurocode set [1] comprises the following standards for structural design, consisting of a number of parts, that are periodically updated.

EN 1990 Eurocode 0: Basis of structural design

EN 1991 Eurocode 1: Actions on structures (Part 2 refers to traffic loads on bridges)

EN 1992 Eurocode 2: Design of concrete structures (Part 2 refers to concrete bridges)

EN 1993 Eurocode 3: Design of steel structures (Part 2 refers to steel bridges)

EN 1994 Eurocode 4: Design of composite steel and concrete structures (Part 2 refers to bridges)

EN 1995 Eurocode 5: Design of timber structures (Part 2 refers to bridges)

EN 1996 Eurocode 6: Design of masonry structures

EN 1997 Eurocode 7: Geotechnical design

EN 1998 Eurocode 8: Design of structures for earthquake resistance (part 2 refers to bridges)

EN 1999 Eurocode 9: Design of aluminum structures.

In Basis of structural design (EN 1990) [2] it is explained the complete new „philosophy" of structural design, enabled by the scientific-technical development of structural analysis.

As the requirements, firstly are defined the basic requirements, where the most significant items are the following ones.

„Structure shall be designed and executed in such a way that it will, during its intended life, with appropriate degrees of reliability and in an economical way: sustain all actions & influences likely to occur during execution & use, and remain fit for the use for which it is required."

„Structure shall be designed to have adequate: structural resistance, serviceability, and durability."

„Potential damage shall be avoided or limited by appropriate choice of one or more of the following: avoiding, eliminating or reducing the hazards to which the structure can be subjected;

selecting a structural form which has low sensitivity to the hazards considered; selecting a structural form and design that can survive adequately the accidental removal of an individual member or a limited part of the structure, or the occurrence of acceptable localized damage; avoiding as far as possible structural systems that can collapse without warning; tying the structural members together."

„The basic requirements should be met: by the choice of suitable materials, by appropriate design & detailing, and by specifying control procedures for design, production, execution, and use relevant to the particular project."

Afterwards, the general assumptions are listed as: „the choice of the structural system and the design of the structure is made by appropriately qualified and experienced personnel; execution is carried out by personnel having the appropriate skill and experience; the adequate supervision and quality control is provided during execution of the work, i. e. in design offices, factories, plants, and on site; the construction materials and products are used as specified in EN 1990 or in EN 1991 to EN 1999 or in the relevant execution standards, or reference material or product specifications; the structure will be adequately maintained; and the structure will be used in accordance with the design assumptions."

The requirements are related as well to reliability management, design working life (100 years for bridges), durability and quality management.

The reliability required for structures shall be achieved by design in accordance with Eurocodes and by appropriate execution and quality management measures.

„The levels of reliability relating to structural resistance and serviceability can be achieved by suitable combinations of:

a) preventative and protective measures (e. g. implementation of safety barriers, active and passive protective measures against fire, protection against risks of corrosion such as painting or cathode protection);

b) measures relating to design calculations (representative values of actions and the choice of partial factors;

c) measures relating to quality management;

d) measures aimed to reduce errors in design and execution of the structure and gross human errors;

e) other measures relating to the following other design matters: the basic requirements, the degree of robustness (structural integrity), durability, including the choice of the design working life, the extent and quality of preliminary investigations of soils and possible environmental influences, the accuracy of the mechanical models used and the detailing;

f) efficient execution, *e. g.* in accordance with execution standards referred to in Eurocodes;

g) adequate inspection and maintenance according to procedures specified in the project documentation.

The measures to prevent potential causes of failure and/or reduce their consequences may, in appropriate circumstances, be interchanged to a limited extent provided that the required reliability levels are maintained."

The standard EN 1990 explains the principle of limit state design, with distinction to ultimate limit state and serviceability limit state.

„States prior to structural collapse, which, for simplicity, are considered in place of the collapse itself, may be treated as ultimate limit states."

„The following ultimate limit states shall be verified where they are relevant:

— loss of equilibrium of the structure or any part of it, considered as a rigid body;

— failure by excessive deformation, transformation of the structure or any part of it into a mechanism, rupture, loss of stability of the structure or any part of it, including supports and foundation;

— failure caused by fatigue or other time-dependent effects." „The serviceability limit states that concern:

— the functioning of the structure or structural members under normal use;

— the comfort of people;

— the appearance of the construction works,

shall be classified as serviceability limit states."

„Design for limit states shall be based on the use of structural and load models for relevant limit states."

„It shall be verified that no limit state is exceeded when relevant design values for: actions, material properties, or product properties, and geometrical data are used in these models."

„The verifications shall be carried out for all relevant design situations and load cases".

„Design situations are classified as persistent, transient or accidental."

„The selected design situations shall be considered and critical load cases identified."

„For a particular verification load cases should be selected, identifying compatible load arrangements, sets of deformations and imperfections that should be considered simultaneously with fixed variable actions and permanent actions."

It is defined the classification of actions (permanent, variable & accidental) with its characteristic values, material and product properties and geometrical data.

It follows structural analysis by structural modelling with models of static & dynamic actions.

It is introduced verification by partial factor method, with determination of design values of actions & effects of actions, design values of material & products properties and design values of geometrical data.

It follows the determination of design resistance.

Afterwards it can be carried out the verification of the static equilibrium and the resistance according to limit state design.

Limit state of static equilibrium of the structure, shall be verified that

$$E_{d,\,dst} \leq E_{d,\,stb}$$

where: $E_{d,\,dst}$ is *design value of the effect of destabilizing actions*; $E_{d,\,stb}$ is *design value of the effect of stabilizing actions*.

Limit state of rupture or excessive deformation of a section, member or connection, shall be verified that:

$$E_d \leq R_d$$

where: E_d is *design value of the effect of actions* such as internal force, moment or a vector representing several internal forces or moments; R_d is *design value of the corresponding resistance*.

Serviceability limit state shall be verified that:

$$E_d \leq C_d$$

where: E_d is *design value of the effects of actions* specified in the serviceability criterion, determined on the basis of the relevant combination; C_d is *limiting design value* of the relevant serviceability criterion.

According to Eurocode 1, the following design loading should be considered: self-weight and imposed loads, wind, thermal actions, actions during execution, accidental actions (impact loads) and traffic loads. As well, it should be considered the other actions, such as: concrete creep & shrinkage, settlements & earth pressures and seismic actions if needed.

Part 2 of Eurocode 1 is related to traffic loads on bridges, containing: classification of actions (variable actions, actions for accidental design situations), design situations, road traffic actions & other actions specifically for road bridges, actions on footways, cycle tracks & footbridges, rail traffic actions & other actions specifically for railway bridges and 8 informative annexes.

Part 2 of Eurocode 2 treats the design and detailing rules of concrete bridges, containing: basis of design, materials, durability & cover to reinforcement, structural analysis, ultimate limit state, serviceability limit state, detailing of reinforcement and prestressing tendons, detailing of memebrs & particular rules, additional rules for precast concrete elements & structures, lightweihgt agggreate concrete structures, plain for lightweight reinforced structures, design for the executing stages and 17 informative annexes. In structural anlaysis, besides the linear elastic analysis with partial distribution, it is introduced plastic analysis, non-linear analysis and analysis of second order affects with axial load; as well it is appropriately treated: geometric imperfections, idealization of the structure and prestressed members & structures. In relation with ultimate limit state it is considered: bending with & without axial force, shear, torsion, partially loaded areas, fatigue and membrane elements. In relation with serviceability limit state it is treated: stresses, cracks control and deflection control.

Part 2 of Eurocode 3 treats the design of steel bridges, containing: basis of design, materials, connecting devices, durability, structural analysis, ultimate limit states, serviceablity limit states, fasteners / welds / connections / joints, fatigue assesment, design assisted by testing and 5 informative annexes (including recommendations for the structural detailing of steel bridge decks). In structural analysis it is introduced: structural modelling for analysis, global analysis, imperfections, methods of analysis considering material non-linearities and classification of cross sections. In relation with ultimate limit state it is considered: resistance of cross sections, buckling resistance of members, built-up compression memebers and buckling of plates. In relation with serviceability limit state it is treated: calculation models, limitations for stress, limitation of web breathing, limitations for clearance gauges, limits for visual impression, performance criteria for railway bridges, performance criteria for road bridges, performance citeria for pedestrian bridges, performance criteria for the effects of wind, accesabilty of joint details & surfaces and drainage.

Part 2 of Eurocode 4 treats the design and rules for composite bridges, containing: basis of design, materials, durability, structural analysis, ultimate limit states, serviceability limit states, precast concrete slabs in composite bridges, composite plates in bridges and an informative annex. In structural analysis it is introduced: structural modelling for analysis, structural stability, imperfections, calculation of action effects (methods of global analysis, linear elastic analysis, non-linear global analysis for bridges, combination of global & local action effects and classification of cross sections. In relation with ultimate limit state it is considered: resistance of cross sections of beams, filler beam decks, lateral torsional buckling of composite beams, transverse forces on webs, shear connection & connecters, composite coloumns & composite compression members, fatigue and tension members in composite beams. In relation with serviceability limit state it is treated: stresses, deformations in bridges (deflections & vibrations), cracking of concrete and filler beam decks.

Part 2 of Eurocode 8 treats design of bridge structures for earthequake resistance, containing: basic requirements & compliance criteria, seismic action, analysis (modelling, methods of analysis), strength verification (materials, resistance verification of concrete sections, resistance verification for steel & composite members, foundations), detailing (cocncrete piers, steel piers, foundations, structures of limited ductil behaviour, bearings & seisimic links, concrete abutments & retaining walls) and bridges with seismic isolation. As methods of analysis can be applied: linear dynamic analysis — response spectrum method, fundamental mode method, alternative linear methods, non-linear dynamic time-history analysis, static non-linear analysis (pushover analysis).

SOFTWARE PACKAGES IN BRIDGE PROJECTS

The fast development of new technologies and materials initiated the major changes in how bridges are designed. In the past, the tools to provide accurate models or detailed analysis of bridges were very limited. Nowadays, tools and software are available for bridge engineers, following the advancements in technology, structural analysis, construction methods, building materials, etc.

In today's bridge engineering ICT is present in design modelling, building material production, construction technology and bridge menagement.

The bridge structural design analysis has been largely advanced by the effective application of finite element method (FEM), enabled by a rapid development of computer hardware from the seventies of last century. The theoretical basis of finite element method was elaborated in scientific papers published by the middle of last century. FEM software is based on mesh discretization of a continuous domain into a set of discrete sub-domains. FEM has been generalized for the numerical modelling of physical systems in many engineering disciplines including bridge structural design. Modern FEM applications software offer variety of simulation options for modelling and analysis.

In modern bridge engineering, FEM algorithms were embedded in many powerful design tools, contributing to raising the standards of engineering and significantly improving the design process. FEM software enables performing simulations of bridge structures, including linear and nonlinear static analyses. The FEM software component allows analyzing of the dynamic response of a structure subjected to time-dependent loads and displacements. The dynamic analysis features support the investigation of response spectrum and time history analysis. As another structural analysis software feature, modal analysis can help determine the eigenvalues and mode shapes of a structure due to vibration. The possibility of analyzing time dependent and rheological materials of bridge structures is present as well.

Modern FEM software packages increasingly develop comprehensive modules incorporating 'wizards' for bridge modelling and construction stage analysis, such as: SAP2000, SOFISTIK, MIDAS Civil, STAAD Pro, CSI Bridge, LUSAS Bridge, ALLPLAN Bridge, BENTLEY (but not only limited to them) are widely used in bridge engineering. Today state-of-the-art engineering software sets the new standards for bridge design and analysis. It is enabled to establish more accurate and reliable bridge design models.

Project management software Primavera and MS-Project are being widely used for project planning and execution for reducing and to increase productivity and efficiency. Construction management software is designed to monitor and track project progress in terms of workforce management, scheduling, time, cost and quality management.

Building information modeling (BIM) is referred to as a process that connects engineers, involved in bridge design, construction, supervision & management, very efficiently in the various stages of construction. BIM engages all participants in a multi-disciplinary collaboration for better insight into the project they are working on. The BIM technology generates instant and accurate outcomes for all phases of the bridge project. Further development of building information modelling (BIM) software goes in the direction from 3D which implies 3-dimensional geographical structures towards 4D and 5D which integrate respectively time scheduling, cost estimation and budget analysis. Benefits of 4D BIM are outlining of duration, timeline and scheduling data while 5D BIM enable real-time cost visualisation with notification of changes in cost simplifying cost and budgetary analysis.

CONSTRUCTION MATERIALS

The modern bridges are mainly constructed as concrete, steel or combined concrete-steel composite bridges. The wooden bridges today are built for pedestrian/cycle bridges in countryside, recreation areas and temporary bridges, having small spans. The masonry (stone or brick) bridges are now rarely built; the rehabilitation or reconstruction of some old bridges can be present today.

The choice of bridge types, referred to construction material, depends on several factors, such as: bridge total length & span lengths, structural type, kind of traffic, bridge location, available material supplies; all connected with economical, executive, architectural and durability aspects.

Concrete

The concrete material (classification, requirements for concrete & methods of verification, specification of concrete, delivery of fresh concrete, conformity control, production control) is defined in standard EN 206–1.

The mechanical and deformation characteristics of concrete (strength, elastic deformation, creep & shrinkage), reinforcing steels, prestressing steels & devices are defined in part 1–1 of Eurocode 2 (EN 1992–1-1).

The relation between stress σc strain εc (compressive stress and shortening strain for short term uniaxial loading) is presented schematically in Figure 1a, as given in part 1–1 of Eurocode 2. The parabola-rectangle diagram for

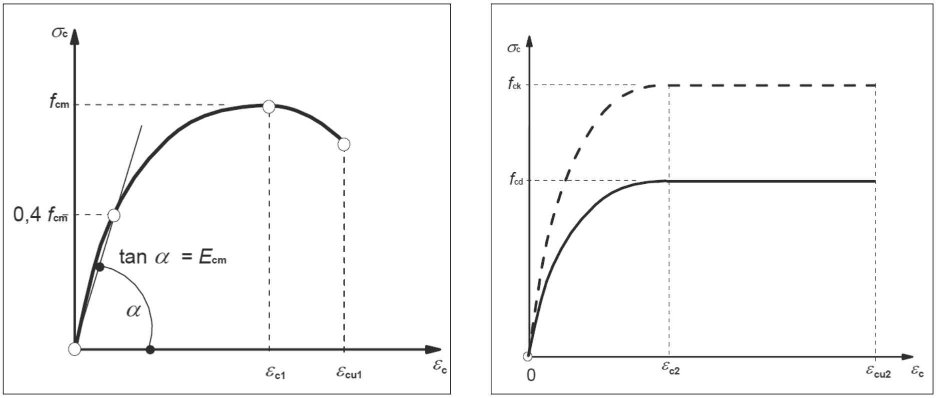


Figure 1a, b: Stress-strain diagrams for concrete under compression

concrete under compression, representing an adoption for design of cross-sections is shown in Figure 1b.

A special emphasis is given to durability in different environmental conditions, classified by description of environments (specified by different exposure classes) as: no risk of corrosion attack, corrosion induced by carbonation, corrosion induced by chlorides, corrosion induced by chlorides from see water, freeze/thaw attack and chemical attack. The values of minimum concrete cover (depending on exposure class & structural class), with regard to durability for reinforcement steel and prestressing steel, are now precisely specified. The corrosion protection of steel reinforcement depends on density, quality and thickness of concrete cover and cracking. The cover density and quality is achieved by controlling the maximum water/cement ratio and minimum cement content, and may be related to minimum class of concrete. The minimum thickness of concrete cover for bridge structural components is nowadays significantly increased — for reinforcement not less than: 40mm for superstructure & piers above soil, 50mm for piers in soil; and for cables not less than 50mm.

The bearing capacity of concrete is directly depending on its compressive strength C, defined in by strength classes of concrete. The minimum strength classes are: C 20/25 for bridge piers, C25/30 for RC (reinforced concrete) superstructures and C 30/37 for PC (prestressed concrete) superstructures. Nowadays the following concrete classes are mostly applied for bridge constructions: C 25/30, C 30/37, C 35/45, C 40/50, C45/55 and C50/60.

The enormous advances in concrete technology, starting from this century, relates to the possibility to obtain ready-mixed concrete with strengths as high as 100 MPa and even much more.

The concrete, according to its strength, nowadays can be classified as follows: normal-strength concrete (20–55 MPa), high-strength concrete (55–100 MPa), ultra high-strength concrete (100–150 MPa) and especial concrete (strength ≥150 MPa).

The application of high-strength concrete may be appropriate for the largely compressed bridge components with exceptional heights, such as pylons (cable-stayed bridges) and towers (suspension bridges), and nowadays as well there are some applications for some superstructures of pedestrian and small-medium span bridges. Manufacture of high strength concrete includes an optimal use of the basic ingredients that constitute normal-strength concrete. It is needed an optimal selection of high-quality portland cement, aggregates, as well as an appropriate combination of materials with selected proportions of cement, water, aggregates, and admixtures. The aggregate shall have the remarkable strength, the optimum size & surface characteristics and the bond with cement paste. The pozzolans, such as fly ash and silica fume, are the usual mineral admixtures in high-strength concrete. These materials introduce an additional strength to the concrete by reacting to portland cement hydration, that generates an additional C-S-H (calcium — silicate — hydrate) gel, which is responsible for the concrete strength. The applied chemical admixtures include a superplasticizer in combination with a water-reducing retarder. The superplasticizer gives the concrete an adequate workability at low water-cement ratios, resulting in greater strength of concrete. The water-reducing retarder slows the hydration of the cement and enables more time to place the concrete.

Starting from the end of previous century, the innovation has been made in application of UHPC (ultra high performance concrete) for certain number of bridges, and more for bearing components of bridges (prestressed beams, arch ribs, joints, deck pavements, etc), constructed in technically most developed countries (USA, Canada, Japan, China, South Korea, France, Germany, etc). High performance concrete is defined basing on performance criteria, such as: high durability, high strength and high workability. UHPC is composed of the following material components: cement, well-graded fine sand, quartz sand, silica fume & other mineral admixtures, steel fiber and superplasticizer. The absence of coarse aggregate can improve the homogeneity of UHPC. The high density of UHPC is improved by application of well-graded fine sand, quartz sand and silica fume, which can reduce the porosity of the UHPC. The steel fiber has a different tensile stress, which effectively reduces the occurrence of concrete cracks. In order to make less the amount of water and increase the strength, a significant amount of effective superplasticizer is added. UHPC has excellent mechanical properties and durability, which can upgrade the connection integrity of bridge component joints, reduce the deformation and crack problems of bridge pavement and increase the load-carrying capacity of bridges. UHPC is currently applied for small and medium-sized bridges or pedestrian bridges in technically most developed countries.

Reinforcing Steel and Prestressing Steel

The reinforcing steel material is defined in standard EN 10080. Reinforcement steel is produced in the form of bars (6–40 Ø) and de-coiled rods (6–16 Ø) of quality B500.

The relation between tensile stress σ and strain ε for typical reinforcing steel (hot rolled steel) is presented schematically in Figure 2a, as given in part 1–1 of Eurocode 2.

The reinforcement shall have adequate ductility, as defined by the ratio of tensile strength to the yield stress $(f\mathrm{t}/f\mathrm{y})k$ and the elongation at maximum force εuk.

The prestressing steel material is defined in standard EN 10138. Wires, strands and bars are applied as prestressing tendons in concrete structures. The cold drawn wires products are used as: Y1860C (4,0 or 5,0mm Ø) and Y1770C (5,0 or 6,0mm Ø). The 7-wire strands are produced as: class A — Y1860S7 (7–16mm Ø) & Y1770S7 (15,2, 16 or 18mm Ø) and class B — Y2160S7 (6,85mm Ø), Y2060S7 (7mm Ø) & Y1860S7 (9mm Ø). The 7-wire strands compacted are produced as: class A — Y1860S7G (12,7–15,2mm Ø), Y1820S7G (15,2mm Ø) & Y1770S7G (18mm Ø).

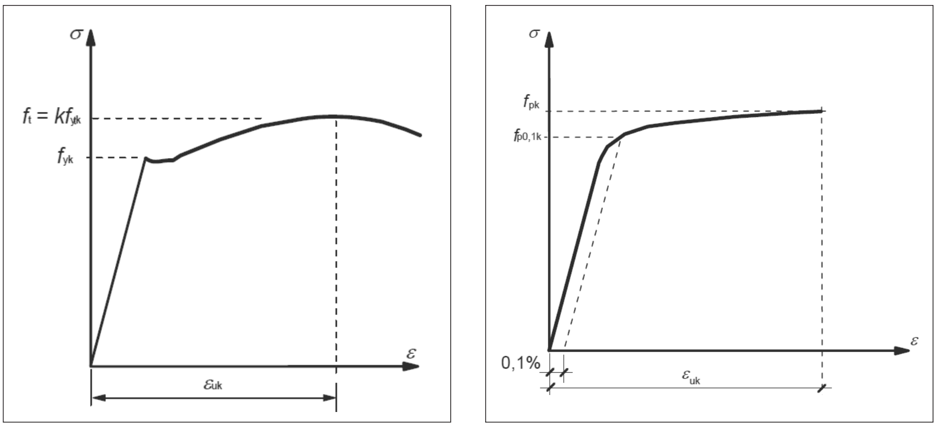


Figure 2a, b: Stress-strain diagrams: a) typical reinforcing steel, b) typical prestressing steel

The relation between tensile stress σ and strain ε for typical prestressing steel is presented schematically in Figure 2b, as given part 1–1 of Eurocode 2, The prestressing tendons shall have adequate ductility, and adequate fatigue strength, as specified in standard EN 10138.

Structural Steel

The hot-rolled products of structural steels are specified in standard EN 10025, where the following products relevant to bridge steelwork are: non-alloy structural steels, fine grain structural steels, weathering steels and quenched & tempered steels. The non-alloy structural steels are the conventional structural steels, nowadays usually applied for bridge structures in 4 yield strength classes 275/355/420/ 460 MPa (steel grades: S275, S355, S420 & S460). Generally, S275 steel grade is often applied for railway bridges, in the case when stiffness rather than strength governs the design, or where fatigue is the critical design factor. S355 steel grade is very often applied for highway bridges, because one gets an optimum balance between stiffness and strength. S460 steel grade application gives the advantages where it is needed to minimize the plate thicknesses or overall self-weight, in the case when the fatigue, stiffness and instability of slender members are not the governing design factors. The thicknesses of plates from 10mm till 40mm are mainly used for steel bridge structures; the thicknesses above 40mm (40mm — 80mm) may be applied for some particular structural components.

The requirements for structural steel for bridgeworks, relating to: material properties, ductility requirements, fracture toughness, through thickness properties and tolerances are contented in part 2 of Eurocode 3 (EN 1993–2). The selection of materials for fracture toughness and the selection of materials for through-thickness properties shall be appropriately made (EN 1993–1-10). The execution shall be carried out according to the technical requirements given in the part 2 of standard EN1090.

The structural steels have to be protected for corrosion by coating system. The majority of steel bridge structures are protected against corrosion by application of paint coatings. The efficient modern coating systems have been improved, applicable for different exposed conditions. Nowadays the advances in coating technology enable the application of protective systems which are expected to last well thirty years (or even more) without maintenance need. Modern duplex coating system has been used, where the paints are applied over thermally sprayed metal coatings. Hot-dip galvanizing is as well a durable coating, but because of the nature of the application process its use is limited to relatively small bridges. The use of unpainted weathering steel, high strength law alloy steel forming adherent protective rust — patina to inhibit further condition, is increasingly popular for modern bridges. The bridges fabricated from unpainted weathering steel can achieve a 100-year design life, with almost no maintenance. Nowadays it has been increased the application of stainless steel in bridge construction, especially in footway bridges and as well for structural components of roadway bridges that are susceptible to corrosion. The advantages of using stainless steel in bridges are: durability in aggressive (acidic & alkaline) environment conditions, high strength to weight ration and overall the aesthetically appealing. The disadvantage of stainless steel application is its relatively high cost, which often limits its use in main bridge supporting parts.

Materials for Protection and Repair of Concrete

The products and systems for the protection and repair of concrete structures are contented in standard EN 1504, covering: general principles for the use of products and systems, surface protection systems for concrete, reinforcement corrosion protection, structural & non-structural repairs, structural bonding, concrete injection, anchoring of reinforcing steel bar, site application of products & systems, and quality control of the works. The main chemical types and constituents of protection and repair products and systems are: additions, additives for hydraulic binders, additives for reactive polymer (plasticizers, flexibilizers, accelerators, retarders, materials regulating the rheology, pigments and fillers), admixtures, coatings, hydraulic binders, hydraulic mortars & hydraulic concretes, hydrophobic impregnation, impregnation, polymer hydraulic cement mortars or polymer concretes, polymer mortars and polymer concretes, reactive polymer binder (epoxies, unsaturated polyesters, acrylics and one or two-component polyurethanes).

Innovative Materials

New building materials have been appeared nowadays.

One of these innovative building materials is self-healing concrete. It is well known that because of the various loads acting on bridges, concrete is prone to cracking. New concrete mixtures with inclusion of limestone-producing bacteria are being developed to fill the cracks as they form. This new technology can prevent the costly repairs of the cracks in concrete.

In order to minimize the damages of bridge structural elements due to earthquakes, which require costly repairs, the new materials are now being applied, such as so-called superelastic reinforcement made from shape memory alloy (SMA). It replaces the classical steel reinforcement in reinforced concrete. The steel reinforcement in concrete subjected to stresses beyond its yield point is plastically permanently deformed, while superelastic SMA return to original shape even after high stressing.

The new materials have been initiated in bridge constructions, to be lighter and relatively stronger.

The steel reinforcements can be replaced with the fiber-reinforced polymer (FRP) bars. The FRP reinforcements besides their high strength to weight ratio, have other advantages such as: non-corrosiveness durability, low thermal conductivity, non-electrical conductivity, and non-magnetic property. The FRP reinforcements can be made from fibers of glass, carbon, basalt, etc. Carbon fiber reinforced polymer (CFRP) have been successfully applied in the USA for construction of prestressed concrete bridges. The FRPs products are nowadays applicable for bridges in different forms, such as: rods, cables/strands, fabric, laminates. The FRP materials are today widely applied for retrofitting and rehabilitation of the bridges.

STRUCTURAL EQUIPMENT

Bearings

In the second half of last century the innovative bearings were invented. Classical bridge bearing concepts as: rollers, rockers, pin steel plate and concrete hinges are today historical ones, i. e. rarely applied. The requirements for design, manufacture and installation of bridge bearings have been completely changed. Besides of traditional steel material used in modern bridge bearings, nowadays are present the materials as: austenite steel, synthetic rubber, polytetrafluoroethylene (PTFE), silicone grease, polyurethane and composite materials.

The family of standards EN 1337, named Structural Bearings, published in first years of current century, is nowadays essential for modern bridge bearings, covering: general design rules, sliding elements, elastomer bearings, roller bearings, pot bearings, rocker bearings, spherical & cylindrical PTFE bearings, guide bearings & restraint bearings, protection, inspection & maintenance and transport, storage & installation.

The elastomeric, pot, spherical/cylindrical PTFE, guide and restraint bearings are considered as the modern bridge bearings. They contain sliding element composing of polish austenite steel plate and PTFE dimpled sheet supplied with silicone grease for easy sliding and preventing an excessive wear of PTFE. These modern bearings have, besides the functional one, the other advantages as well, such as: lower height, lower weight, easier transport, easier installation & replacement, lower cost, etc. easier installation & replacement, lower costs etc.

Expansion Joints

Ten years ago were published the Guideline for European Technical Approval of Expansion Joints for Road Bridges (ETAG 032), specifying all types of modern expansion joints as: buried, flexible plug, nosing, mat, cantilever, supported and modular expansion joints. Buried expansion joints are used for displacements maximum 25–30 mm and flexible plug expansion joints are used for displacement maximum 40 mm; the execution is directly on site. Nosing expansion points are used for gap displacements maximum 80 mm; mat expansion points for gap displacements maximum 330 mm and cantilever (finger) expansion points for gap displacements from 80 to 800 mm. Modular expansion joints have the possibility of large movements (over 800 mm). Today, the most applied are nosing, modular and cantilever expansion joints — the latter can be as the supported ones. The modern expansion joints are waterproofing. The noise reducing elements can be installed in modular expansion joints when it is environmental requirement.

Anti-Seismic Devices

In the second half of the previous century began the protection of bridge structures subjected to the risk of earthquakes. Firstly, it was applied only the passive protection of earthquakes based on the plastification of bridge elements chosen in advance. However, after intensive earthquake it often follows major repairs of the damaged protective elements, although generally the bridge structure resists design earthquakes and the protection of human lives are enabled.

Nowadays in seismic zones it has been introduced so-called positive protection; it means that the new bridge structures are equipped with anti-seismic devices to absorb or limit the effects of earthquakes. The positive protection of earthquakes in principle is based on three operational modes: isolation, connection and dissipation. The anti-seismic devices are defined as the elements which contribute to modifying the seismic response of a structure by isolating it, by dissipating energy or by creating permanent or temporary restraints via rigid connections. The standard EN 15129 „Anti-seismic devices“ specifies functional requirements & general design rules of the devices, material characteristics, manufacturing & testing requirements, as well as assessment and verification of constancy of performance, installation and maintenance requirements. This standard covers the types of devices and combinations thereof. The choice of anti-seismic devices depends on various parameters: the seismic level of the site, the type and characteristics of the bridge structure and the maximum response permitted.

REALIZATIONS OF MODERN BRIDGES

The modern bridges can be nowadays realized as: short span bridges (5–50m span length), medium span bridges (50–150m span length), long span bridges (150–500m span length), super-long span bridges (500–1000m span length) and ultra-long span bridges (above 1000m span).

The most used structural systems of modern bridges nowadays are realized as: beam bridges (simple or continuous beam), truss bridges (simple or continuous beam), frame bridges, arch bridges, extradosed bridges, cable-stayed bridges and suspension bridges.

Referring to building material, the actual realization of bridges can be classified as: concrete bridges (reinforced concrete or prestressed concrete bridges), steel bridges (plated girders, box girders or truss girders) and composite (concrete/steel) bridges. Today the wooden bridges may be suitable only for short span footway/cycle bridges in recreation areas.

Beam Bridges

The reinforced concrete slab bridges are used only for very short span bridges (up to 10m) and culverts. Nowadays reinforced concrete beam bridges (without pre-stressing or post tensioning) are only applied for minor bridges of short spans. The beam bridges can be constructed of steel, concrete or composite (steel girder with concrete deck). The concrete bridge components may be reinforced, pre-stressed or post-tensioned. Steel beam bridges can be different types of plate girders or box girders. The pre-stressed or post tensioned concrete I beam elements can be economically applied for main girders up to maximum 40m span and the concrete box girder and composite girder, with constant depth, are suitable from 40m till 80m spans. The concrete beam bridges, with variable depth, can be economically applied up to about 200m main span and steel box beam bridges, with variable depths till about 250m. Nowadays the continuous beam bridges having main span more than 200m (250m) are mostly constructed as cable-stayed bridges.

The future of beam bridges will be oriented to the application of light, strong, and long-lasting materials as: reformulated concrete with high performance characteristics, fiber reinforced composite materials, electro-chemical corrosion protection systems etc. The modern concrete beam bridges use pre-stressed concrete girders that combine the high tensile strength of steel and the superior compression properties of concrete. The modern steel bridges use the material of high quality steel grades and they have the innovated better corrosion protection. The box girders are mostly applied for concrete, composite and steel beam bridges with longer spans, highway bridges and roadway & railway bridges in curvature.

As special type of beam bridges, the integral bridges have been introduced, in which the conventional superstructure and substructure are integral with each other and the bearings are excluded. It is enabled by more sophisticated design analysis with the interaction of superstructure, substructure and soil where the bridge is founded.

Steel trusses, as simple and continuous beams nowadays are mainly applied for railway bridges or road-railway bridges, the applications for modern road bridges are very rare.

Frame Bridges

Frame bridges have the superstructure and substructure rigidly connected, i. e. the piers and girder are one solid structure. The cross sections of beams are usually box-shaped or I-shaped. The frame bridges have been applied from third decade of 20th century, as reinforced concrete and steel single span bridges, later in prestressed concrete. Nowadays in the application are as well the batter post and V-shaped frame bridges. Single span frame bridges may be effectively used for inner city highways or express ways. The batter post frame bridges are well suited for river and valley crossings because their supports run from the deck to the abutments at an angle; consequently, the abutments are either larger or additional foundations are placed next to the abutments. V-shaped frame bridge is efficiently used when a longer span is not feasible. Each V-shaped pier provides two supports to deck, requiring only one foundation. Gazelle Bridge in Belgrade (1970), with 250m span, is today ranked on 3rd place in the category of batter post bridges.

Arch Bridges

The ancient arch bridges were built as masonry (stone or brick) or wood arch structures. The use of steel as replacement for wrought iron (firstly used for the Coalbrookdale Bridge UK in 1779) was extended in last quarter of 19th century, when (1874) it was built the Steel Arch Bridge in Saint Louis USA over Missisipi River, with 159m span — record one for that time. The first concrete arch bridges were built by the end of 19th century.

Nowadays the materials used for arch bridges are: concrete, steel and combination of both as named CFST (concrete filled steel tube).

Three main types of arch bridges are nowadays governing: deck arch bridges, through arch bridges, and tied-arch (bowstring arch) bridges.

Deck arch bridges are those ones where the deck is completely over the arch rib, supported by the columns rising from the arch.

Through arch bridges are characterized by the arch which base is below and top above the deck, and deck passes through the arch. The deck is partly (or whole) below the arch suspended by cable ties, and if it is partly above the arch than these deck sections are supported by the columns.

Tied-arch bridges have deck connected to two opposite ends of the arch serving as tie. The deck is suspended from the arch rib by cable ties (vertical or inclined). The tie-arch bridge, suitable for 100–250m spans, is nowadays frequently applied for the roadway (or railway) bridges across the rivers. The layouts of modern tie-arch bridges, with different forms of arch rib(s), have an advantageous architecture impression.

The first ten arch bridges today, with the longest spans, are listed in Table 1. It can be noticed that the building material is either steel or CFST. Among 30 arch bridges with longest spans, 25 bridges were built in China, and under construction are now 25 arch bridges with spans above 300m.

The Krk Arch Bridge (1980), with 416m span — ex record (till 1997), is today ranked on 3rd place for arch concrete bridges, and 23rd place in overall ranking for arch bridges with the longest spans.

Table 1: Arch Bridges — Longest Spans (data: Wikipedia)

Bridge name	Location	Material	Span length	Year opened
Pingnan Third Bridge	Guanxi (China)	CFST	575 m	2020
Chaotianmen Bridge	Chongqing (China)	steel	552 m	2009
Lupu Bridge	Shanghai (China)	steel	550 m	2003
Bosideng Bridge	Sichuan (China)	CFST	530 m	2012
New River George Bridge	Fayetteville, Virginia (USA)	steel	518 m	1977
Bayonne Bridge	Bayonne, New Jersey (USA)	steel	510 m	1931
Zigui Yangtze River Bridge	Chongqing (China)	CFST	508 m	2019
Hejiang Yangtze River Bridge	Sichuan (China)	CFST	507 m	2021
Sidney Harbor Bridge	Sidney (Australia)	steel	503 m	1932
Wushan Bridge	Chongqing (China)	CFST	460 m	2005

Cable-Stayed Bridges

The modern cable-stayed bridges have been constructed from the second half of 20th century; firstly, in Germany, later worldwide in Europe, Japan, USA, Asia, etc. In 21st century the numerous cable-stayed bridges have been realized in China. This structural type of bridge contains the main girder elastically supported at intervals along its length by the inclined cable stays anchored in pylon(s). The first cable-stayed bridges were constructed mainly as 3-span bridges, main span supported by limited number of inclined cable stays equal with the number of cable stays in back spans; all stays anchored in high pylons and girder. Today the cable-stayed bridges are constructed with multi-stayed cables as 2-spans, 3-spans and multi-spans structures, having one, two or several pylons. Nowadays, the wide variety of cable-stayed bridges has been built all over the world, with steel, concrete or composite deck and steel or concrete pylons. This structural form gives to bridge a prestige and effective architecture appearance, because of the layout combinations related to: configuration of cable stays (harp, modified harp, fan or star) in single or double „planes“ (with back-stays or without); type of steel or concrete pylon (single, double, portal, A-shaped, H-shaped, Y-inverted, M-shaped) — vertical, inclined or spaced; and type of steel, concrete or composite girder (box beam with or without struts, deck of 2 or multi beams) with 2, 3 or multiple spans.

First stays were done as locked coil cables which were subjected to corrosion deterioration with fatigue problems of cable anchorages and vibration problems. Later follows the application of prefabricated stay cables of parallel wire strands, packed in large drums with heavy transport and erection. Nowadays the parallel strand system is largely applied, where the stay cable is compound on site as a bundle of prefabricated 7-wires strands. The stay cable installation is simple, made strand by strand, where each strand is individually anchored by jaws in anchorage block, enabled to sustain remarkable fatigue effects. Each strand is individually protected against corrosion by two redundant corrosion barriers. The strand bundle is protected by a colored polyethylene outer duct. The stay cables can be equipped by special dumpers to prevent the vibrations.

Due to new developments in the modelling and analysis of dynamic behaviour and the use of sophisticated damping against oscillations, the cable-stayed bridge constructions have been expanded to spans in excess of 1000m, before 21st century the span range only applicable for suspension bridges. The advantages of the cable-stayed bridges compared to suspension bridges are the following ones: greater stiffness, easer construction — cable-supported cantilevering procedure and the balance of horizontal forces achieved without need of large ground anchorages.

The notable multi-span cable-stayed bridges have been built in 21^{st} century. Nowadays the cable-stayed bridges are suitable for the spans from 200m till 1200m. Also they are applicable for bridges with modest spans where deck depth should be very limited and for footway/cycle bridges.

Ten cable-stayed bridges, with the longest spans, are listed in Table 2. The record span is doubled in last twenty years and the number of constructed cable-stayed bridges has been largely increased.

Table 2: Cable-Stayed Bridges — Evaluation of Record Spans (data: Wikipedia)

Bridge name	Location	Deck / Pylon	Main span	Record period
Stromsund Bridge	Stromsund (Sweden)	steel / steel	182 m	1956–1957
Theodor Heuss Bridge	Dusseldorf (Germany)	steel / steel	260 m	1957–1959
Severin Bridge	Cologne (Germany)	steel / steel	302 m	1959–1969
Knie Bridge	Dusseldorf (Germany)	steel/concrete	319 m	1969–1971
Duisburg Bridge	Duisburg (Germany)	steel / steel	350 m	1971–1974
Saint-Nazaire Bridge	Saint-Nazaire (France)	steel / steel	404 m	1974–1983
C. F. Casado Bridge	Leon (Spain)	concrete	440 m	1983–1986
Alex Frazer Bridge	Vancouver (Canada)	steel/concrete	465 m	1986–1991
Skarnsund Bridge	Inderoy (Norway)	concrete	530 m	1991–1993
Yangpu Bridge	Shanghai (China)	steel/concrete	602 m	1993–1995
Normandy Bridge	Le Havre (France)	steel/concrete	856 m	1995–1999
Tatara Bridge	Imabari (Japan)	steel/concrete	890 m	1999–2008
Sutong Yangtze Bridge	Suzhou-Nantong (China)	steel/concrete	1.088 m	2008–2012
Russki Bridge	Vladivostok (Russia)	steel/concrete	1.092 m	2012–2020
Hutong Yangtze Bridge	Suzhou-Nantong (China)	steel/concrete	1.104 m	from 2020
Changtai Yangtze River Bridge	Changzhou-Tiazhou (China)	steel/concrete	1.176 m	from 2024 in construction

Extradosed Bridges

The modern extradosed bridges have been constructed from the last decade of 20^{th} century in Europe and Japan and afterwards largely worldwide in 21^{st} century. They are suitable for concrete bridges with medium spans from 100 to 250m. Visually they are similar to cable-stayed bridges with short concrete pylons and with cable stays of shallow angles to concrete deck. Cable stays elastically support the deck and they are as well the external prestressing tendons. Therefore, the extradosed bridges may be considered as a hybrid structure originated from cable-stayed bridge (with short pylons) and cantilever-girder bridge (with shallow deck).

Suspension Bridges

The suspension bridges have been applied for very long spans, nowadays for spans over 1000m. In first half of 19th century the chain suspension bridges were built. From the middle of 19th century wire-cable suspension bridges in USA started the construction of wire-cable suspension bridges. The main structural components of suspension bridge are: long supporting cables, towers, suspender cables, bridge deck (stiffening girder) and anchorage blocks (for supporting cables).

The first suspension bridge having span over 1000m (1067m) is George Washington Bridge (1931), later Golden Gate Bridge with 1280m span (1937) and Verazzano Narrows Bridge with 1298m span (1964); all roadway bridges with steel truss deck as stiffening girder. Later from 8th decade of 20th century, after aerodynamic tests in wind tunnels, the roadway suspension bridges with airfoil box sections of stiffening girder have been designed and constructed, firstly in UK — including Humber Bridge with 1410m span (1981). The suspension bridges with truss deck have been constructed for roadway and railway traffic in two levels, such as: Great Belt Bridge with 1624m span (1998) and Akashi Kaikyo Bridge with 1991m span (1998).

Table 3: Suspension Bridges — Evaluation of Record Spans (data: Wikipedia)

Bridge Name	Location	Traffic / Girder	Span Length	Record Span
George Washington Bridge	New York — New Jersey (USA)	roadway steel truss	1.067 m	1931–1937
Golden Gate Bridge	San Francisco (USA)	roadway steel truss	1.280 m	1937–1964
Verrazano Narrows Bridge	New York City (USA)	roadway steel truss	1.298 m	1964–1981
Humber Bridge	Yorkshire (UK)	roadway steel box beam	1.410 m	1981–1998
Akashi Kaikyo Bridge	Japan	roadway & railway steel truss	1.991 m	1998–2022
1915 Canakkale Bridge	Turkey	roadway steel box beam	2.023 m	from 2022
Zhagjiagang-Jingjian-Rugao South Yangtze River Bridge	Zangjigang Jiangsu (China)	roadway steel box beam	2.300 m	from 2028 in construction

The evolution of record spans (over 1000m) is presented in Table 3. The present record span of 2023m belongs to recently constructed the 1915 Canakkale Bridge (2022) [3], with innovative bridge deck consisting of two stiffened closed steel box girders spaced 9m apart and connected by cross-girders every 24m. The 9m air gap between the two box girders enables the aerodynamic stability of the bridge deck in strong winds. Messina Bridge Project design [4] has been developed as roadway-railway suspension bridge, with huge main span of 3300m and 382m height pylons. It is designed triple box deck (with gaps between box girders), connected by cross girders.

CONSTRUCTION OF CONCRETE BRIDGES

Nowadays in Europe the actual concrete bridge structures are constructed according to the standard EN 13670 „Execution of concrete structures“, that gives common requirements for execution of concrete structures related to: execution management, falsework & framework, reinforcement, prestressing, concreting, execution with precast concrete elements and geometrical tolerances.

Today, concrete bridge construction techniques are present as cast-in-situ or precast construction. The *cast-in-situ construction* is a procedure whereby the segments are progressively cast on site in their final positions within the structure. The *precast construction* is a procedure whereby the segments are prefabricated at casting plant, either on site or transported from the factory, and erected as a completed unit in their final positions. The construction methodology may be classified as: *conventional* (on falsework / scaffold), *balanced cantilever*, *incremental launch*ing, *movable scaffold system* (MSS) / *advanced shoring* and *span-by-span* techniques.

Balanced cantilever construction methodology is usually performed as cast-in-situ and can be precast also. The concrete structure is constructed outward from a fixed point, without temporary support, as a cantilever structure, using staged cast-in-situ construction. The balanced cantilever construction method is realized when two opposing free cantilever structures are attached as a single structure and erected in the same step. It is often appropriate and cost-effective for the construction of long span concrete bridges where because of height, topography or geotechnical conditions the application of conventional formwork is uneconomical. Nowadays the economical span lengths for cast-in-situ cantilever construction are considered from 70m and extends to beyond 250m. The precast balanced cantilever construction is suitable for 45m — 135m spans.

Incremental launching is a suitable construction method of continuous post-tensioned multi-span bridges. It includes casting 15–30m long sections of the bridge superstructure in a stationary formwork behind an abutment and pushing a completed section forward with jacks along the bridge axis. The sections are cast contiguously and then stressed together. The superstructure is launched over temporary sliding bearings on the piers. The steel launching nose is connected to the front of the bridge deck, in order to reduce bending of the superstructure during construction. The incremental launching method is suitable for restricted and limited sites over deep valleys, rivers, highways, railways, as well as in poor soil conditions or environmentally protected areas.

Movable scaffold system (MSS) is developed for multi-span bridges over difficult terrain or water where scaffolding is not suitable. It comprises an application of launching girder that moves forward on the bridge piers, span-by-span to allow placing of the cast-in-situ concrete. This methodology can be adapted for a wide range of spans and types of superstructure.

Nowadays, the *conventional falsework / scaffold* is suitable for construction of bridge structures with single spans up to 80m This construction methodology today is not convenient for longer bridges with multiple spans, because the scaffolding needs to be moved between the different sections of the bridge during construction. Therefore, it is developed the *advanced shoring* technique. The construction method uses a movable supporting beam, gantry, for the falsework that can reach over the length of one or two spans. The special roller bearings and launching jacks enable that the gantry can move forward along the bridge as the construction proceeds. The travelling gantry system can be economically applicable for spans from 30m to 60m.

The *precast span-by-span* bridge construction methodology can enable a significant speed of construction. It is often applied in conjunction with an erection truss under the bridge segments or an overhead erection gantry to guide the precast elements into position. The precast segmental span-by-span method is suitable for construction of long viaducts with spans of similar length. This method is appropriate for spans from 25m to 45m.

Heavy lifting is a special hydraulic cable lifting technique developed for exceptionally heavy loads. This technique, as timely and economic solution, is particularly applicable for the projects based on modular construction methods where large, heavy, pre-fabricated elements are present. The heavy lifting technique can be applied to: lifting and lowering of heavy precast beams and entire structural elements by means of strands and hydraulic jacks; lifting of bridge structures for the erection or repair of bearings; sliding of bridge structures from the assembly area to final position.

CONSTRUCTION OF STEEL BRIDGES

Nowadays in Europe the steel bridge structures are constructed according to standard EN 1090 „Execution of steel structures and aluminium structures“, that regulates the fabrication and assembly as well. It comprises the requirements for conformity assessment for structural components and technical requirements for the execution. The technical requirements for execution of steel structures are specified for: constituent products (structural steel products, steel castings, welding consumables, mechanical fasteners, studs & sheer connectors, grouting materials, expansion joints for bridges, high strength cables & rods & terminators, structural bearings), preparation & assembly (identification, handling & storage, cutting, shaping, holing, cut outs, full contact bearing surfaces, assembly, assembly check), welding, mechanical fastening, erection, surface treatment, geometrical tolerances, inspection, testing and corrosion protection.

Construction of steel bridges consists of fabrication of steel structure in the workshop (from the delivered steel plates and profiles) and the assembly with the erection on site. In the steelwork workshop is performed the production of steel segments suitable for transport by roadway, railway or waterway. The steel structure can be produced in large blocks, or in segments convenient to be packed and transported in containers. On site it can be prior carried out the preassembling and (or directly) assembling in erection units. The trial assembling of erection units is carried out in the workshop. Nowadays the site assembly joints and segment splices are performed by welding or by pre-stressed high strength bolts. Besides the segments that are produced as welded in the factory, today often mostly (or completely) the entire steel bridge structure is assembled by welding at site from the delivered segments.

The modern workshops are today equipped for computer added manufacturing (CAM) and the workshop drawings are produced by computer added design (CAD). The modern construction sites today can be equipped for erection by all kinds of cranes of high capacities: mobile cranes, derick cranes, portal cranes, tower cranes & floating cranes, as well as huge hydraulic jack systems.

The erection of steel bridge is the most sensitive operation of bridge construction. During different stages of construction, the bridge structure passes through different states of stresses and deformations. The appropriate choice of construction stages can be crucible for dimensioning and final alignment of bridge structure.

As it is well known, the usual types of erection of steel bridges are: installation on the temporary piers, free cantilever erection, incremental launching, span/segment lifting and adequate combination of the cited types.

The choice of steel bridge erection procedure depends on: structural system of bridge (beam bridge, arch bridge, frame bridge, batter-post bridge, cable-stayed bridge, suspension bridge), bridge type (plated, box or truss bridge), bridge deck (steel or composite), bridge size; terrain configuration and type of obstacle; transport possibilities from the structure workshop to the construction site; available equipment for erection and contractor's experience.

NOTABLE BRIDGES RECENTLY BUILT IN EX-YU REGION

Ada Bridge in Belgrade

The Ada Roadway (and light railway) Bridge across the Sava River [5] on Belgrade inner city semi-ring road, was completed in 2011. Bridge length: 996m. The main structure is designed as cable-stayed bridge with originally shaped single concrete pylon, steel deck in 376m main span and concrete deck in 200m back span (as well in 388m side spans & 50m end span). The 376m span is ranked as one of the longest span in the category of cable-stayed bridges with single pylon.

Arch Bridge in Novi Sad

The Railway-Roadway Arch Bridge across the Danube in Novi Sad [6], was completed in 2018 on the foundations of former arch concrete bridge. Bridge length: 474m. The main bridge structure consists of two steel tied network arches (l/h=219/42m & l/h=177/34m) with composite deck. The 219m span is the longest one in the category of tied network arch bridges with two rail tracks.

Moračica Bridge in Podgorica

The Moračica Roadway Bridge [7] across the Morača River canyon, was completed in 2022, near Podgorica on the highway section Podgorica-Mateševo. The bridge is designed as a ballanced cantilever concrete bridge structure of beam-frame type, with spans: 95 + 170 + 3x190 + 125m. The maximum height of the bridge above the terrain exceeds 175m i. e. 205m above the riverbed.

Bridge Pelješac on Adriatic Coast

The 2404m long Pelješac Roadway Bridge [8], across Mali Ston Bay of Adriatic sea, completed in 2022, is designed as a multi-span cable-stayed bridge with a semi-integral hybrid structure (steel deck & concrete pylons) with five 285 m long main spans. The Pelješac Bridge, located in high seismic area, belongs to the most attractive European bridges at sea cost.

Contribution of Paper Author to Bridge Engineering

The paper author has been actively engaged in bridge engineering in three fields: professional (design engineer & supervision engineer), scientific (cable-stayed bridges, wide-flanged box girder structures, thin-walled structures) and educational (lectures on steel structures & bridges).

The Solidarity Roadway Bridge across the Vistula River [9] in Plock (Poland), open for traffic in 2007, is designed by the paper author, as the co-author (with N. Hajdin) and the main design & supervision engineer. The entire bridge, with 1200m total length, consists of 615m long main bridge part and 585m long access bridge part. The main bridge structure is designed as steel cable-stayed bridge with two pylons, having 375m main span and side spans 2 x 60 m each (Photo 1).

The main span of 375 m belongs to the longest ones in the category of cable-stayed bridges with the cables in a single plane; it is the bridge with the longest span built up to now in Poland.

The concept design proposal of Roadway Bridge across the Bokokotorski Bay (location: Rt Sveta Nedjelja — Rt Opatovo), submitted by the co-authors (B. Stipanić, N. Hajdin & M. Maletin) to the international competition of Montenegro government for the choice of location & concept solution of Bokokotorski Bay Crossing, was awarded by third prize (1999).

Photo 1: Cable-stayed bridge over Vistula River in Plock (Poland)
— open for traffic in 2007

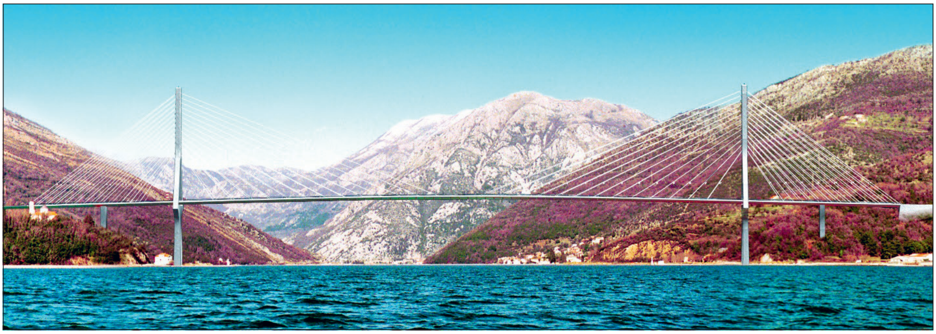


Photo 2: Visualisation of Cable-stayed bridge across Bokokotorski Bay — concept design solution

The design proposal of 960m long entire bridge, consists of 690m long main bridge part and 270m long access bridge part. The main bridge structure is designed as cable-stayed bridge with steel deck and two concrete pylons, having 450m main span and both-sided 2x60m spans (Photo 2).

CONCLUSION

Thanks to the rapid development of natural sciences and technical innovations, the advances have been achieved and permanently up-dated in all fronts of bridge enginering: planning, design (software packages), structural analysis — static & dynamic (due to actions of traffic, wind, earthquake), building materials (high-strength, high-performed & self-healing concrete, high-strength & weathering steel, SMA superelastic reinforcements, FRP bars/strips/cables/strands, new products for protection & repair of concrete structures), construction technology (building procedures, prefabrication, robustness), monitoring, maintenance and retrofitting of bridges.

Consequently, the advancements have not been only made in improved realizations of classical structural types (beam, frame, truss, arch & suspension bridges), but futhermore in application of inovative structural types as: integral bridges, cable-stayed bridges (with 2, 3 & multiple spans), extradosed bridges and hybrid bridges. The application of cable-stayed bridges has been largely increased for roadway bridges with spans from 200m to 1200m and for footway bridges; all designed with wide range of pylon shapes. This year it was completed the suspension bridge with main span longer than 2000m, and now in construction is a suspension bridge with 2300m main span. Two years ago, it was built 165km long bridge for high-speed railways.

Today the digital technology is impacting and modifying the construction industry. The artificial intelligence, virtual reality, drones, robots, monitor devices, GPS surveillance, etc. are now being used in construction and fabrication sites.

In today's bridge engineering ICT is present in design modelling, building material production, construction technology and bridge menagement.

Recently, the Building Information Modeling (BIM) has been introduced as a process which very efficiently connects engineers (involved in bridge design, construction, supervision & management) in the various stages of construction.

BIBLIOGRAPHY

[1] EN 1990 „Eurocode" — EN 1999 „Eurocode 9", CEN, Brussels.

[2] EN 1990: 2002 + A1 „Eurocode: Basis of structural design", *CEN. Brussels.* December 2005. p. 1–118.

[3] J. Pihl: „The design of the 1915 Canakkale Bridge", *Proceedings of 16th ASES Congress, September 2022. p. 71–75.*

[4] F. Brancaleoni, G. Diana, E. Facioli, G. Fiamenghi et al: „The Messina Strait Bridge: A challenge and a dream", *Stretto di Messina S. p. A, Taylor & Francis Group.* London. 2010. p. 1–324.

[5] B. Stipanić, N. Hajdin, Z. Rubinjoni and M. Kozlović: „Project of new Belgrade Bridge across Sava River in the scope of the Inner City Semi-Ring Road". *Proceedings of 7th international conference on bridges across the Danube — Bridges in Danube basin: Theory & practice in bridge engineering.* Sofia 2010. p. 333–342.

[6] A. Bojović, A. M. Munoz, Z. Marković and N. Novaković: „Network arches over the Danube — Railway-Road Bridge in Novi Sad", *Baingenieur.* No 93. March 2018. p. 110–115.

[7] Ž. Ličina, L. Huatu, V. Ćipranić and Y. H. Gang: „Main design and construction of Moračica Bridge on highway E-763, section Smokovac-Mateševo". *Proceedings of 16th ASES Congress,* September 2022. p. 9–14.

[8] M. Pipenbaher: „Pelješac Bridge, Croatia — Design and construction", e-mosty 1/2022. p. 1–24.

[9] B. Stipanić, N. Hajdin, J. Krawczyk: „Project realization of Roadway Bridge across Vistula River in Plock". *Zbornik 30 zborovanja gradbenih konstruktorjev Slovenije.* October 2008. p. 35–50.

Bratislav STIPANIĆ

GRADITELJSTVO MOSTOVA DANAS

Sažetak

U ovom radu se razmatraju sveukupni aspekti napretka u graditeljstvu mostova, proistekli iz ubrzanog razvoja prirodnih nauka, kao i tehničkih inovacija u pogledu: projektovanja i računske analize statičke i dinamičke (dejstvo vetra i zemljotresa), građevinskih materijala (beton visokih čvrstoća i visokovredni beton, visokovredni čelik), kao i tehnologije izvođenja radova (metodologija izgradnje, prefabrikacija, robusnost). Sledstveno, značajan napredak se ne ogleda samo u poboljšanim realizacijama klasičnih konstrukcijskih tipova mostova (gredni, ramovski, rešetkasti, lučni i viseći mostovi), već osim toga i u primeni inovativnih konstrukcijskih tipova mostova kao što su: integralni mostovi, mostovi sa kosim kablovima i ekstradosni mostovi. Danas se primenjuje BIM informaciono modelovanje izgradnje, po kome su tokom raznih faza građenja sveobuhvatno vrlo efikasno povezani inženjeri uključeni u projektovanje, izvođenje, nadzor i menadžment.

Prednapregnutobetonski mostovi, čelični mostovi (sa ortotropnom pločom kolovoza) i spregnuti (čelik-beton) mostovi su danas dominantni u primeni. U ovom veku broj izvedenih mostova sa vrlo velikim rasponima (preko 500 m) i ultradugim rasponima (preko 1000 m) je u znatnom porastu. U radu su date tabele mostova najvećih raspona za lučne, mostove sa kosim kablovima i viseće mostove.

Kao izuzetna nova dostignuća u mostogradnji, ostvarena u eksjugoslovenskom regionu, izdvojeni su: putni most sa kosim kablovima na Adi u Beogradu (2011), putno-železnički lučni most u Novom Sadu (2018), autoputni gredno-ramovski most Moračica kod Podgorice (2022) i putni višerasponski most sa kosim kablovima Pelješac (2022). Putni most sa kosim kablovima Solidarnost u Plocku (2007), sa rasponom od 375 m (najveći za most sa pilonima uklještenim u gredu i most sa najvećim rasponom u Poljskoj) je koautorski isprojektovan od autora ovog rada. Autor ovog rada je u naučnom smislu dao doprinose analizi mostova sa kosim kablovima i grednih sandučastih mostova.

Vladan ĐOKIĆ*

ADVANCES IN URBANISM

Abstract. This paper aims to provide an experience-based overview of current advances in urbanism, with a specific focus on the higher education sphere impacted by numerous challenges. On the one side, international projects provide the framework for fostering a connection between research and education, and on the other hand, demonstrate the importance of acquiring sources of funding for the realization of research activities that are very relevant and interesting in the context of „small countries“). In this sense, the following four projects would be presented: 1. HERSUS — an example of a strategic partnership, focused on intellectual results and activities, 2. ECOBUILT — an example of a project that correlates research in the early phase of the conception of new study programs, 3. RMB — an example of a joint master's program project and its influence on the reformulation of the national accreditation framework, and 4. Circle U. — project focused on building a European University by co-creating knowledge and solutions across disciplines. In addition, in the thematic sense, all four projects illustrate current topics in the field of advances in urbanism: from the relationship between sustainability and heritage (HERSUS), to ecological and environmentally sensitive approaches (ECOBUILT), reuse of existing building stock and reuse, rehabilitation, reconstruction, (RMB), and public health, well being (CIRCLE U.).

1. INTRODUCTION

We are experiencing an era of accelerated growth in the higher education sphere impacted by lots of challenges derived from different layers of contemporary urban life. Having in mind that urbanism, in its scientific and exploratory nature, has spatial values at the core of study and that all urban challenges have repercussions on space, urban studies and related disciplines and scientific fields are faced with the occupation of formulating new paradigms and precedents for action. Looking at the current strategic

* Vladan Đokić, Faculty of Architecture, University of Belgrade, Rector of the University of Belgrade

and action documents, a number of priorities are found that are positioned and have an affirmative influence in the practice of urbanism: (1) The 2030 Agenda for Sustainable Development highlights the goal of 'making cities and human settlements inclusive, safe, resilient and sustainable as one of 17 sustainable development goals (SDGs), (2) New European Bauhaus which highlights a culture- and design-led approaches as an important strategy for contemporary urban transformation continuously shaped by cultural conditions and dynamic patterns, and (3) European Green Deal Initiative which aims to improve the well-being and health of citizens and future generations through renovation and energy-efficiency of buildings.

This paper advocates a bottom-up approach that starts from the belief that a sustainable and fundamental practice of urbanism can be built exclusively through enhancing higher education and innovating study programs that rely on current priorities and challenges. The paper proposes a '3I' approach which implies: (1) investigation — direction towards the conception of curricular and extracurricular activities that are based on scientific research and critical dialogue with practice, (2) innovation — direction towards a unique pedagogical agenda that implies the development of new study programs that focus on certain narrowly specialized areas or skills which can contribute to the improvement of the practical scope, and (3) integration — direction towards the establishment of an integrated university that enables a high level of interactivity and exchange of knowledge based on mobilities and initiatives that stimulate creative and critical thinking both for students and teaching staff.

This paper is by its nature experience-based while according to the type and structure it is a review. It will present the experiences of the University of Belgrade, through the initiation and implementation of Erasmus + projects co-funded by the European Commission at the Faculty of Architecture. The main goal of the review is to point out the importance of higher education in the field of architecture and urbanism, as a leading pillar for the problem-based and innovation-oriented practice of urbanism. The paper will present three program actions that are implemented within the current Erasmus + cycle (2021–2027) through the demonstration of four projects that are currently being implemented at the University of Belgrade, for which the Faculty of Architecture is a coordinator or partner.

2. STRATEGY FOR ENHANCING HIGHER EDUCATION OF ARCHITECTURE AND URBANISM

The current development of the University of Belgrade is faced with numerous challenges. Higher education plays a unique role in creating a sustainable framework for modern society. The demand for highly qualified, socially engaged experts is increasing and changing in accordance with the numerous global challenges facing society as a whole. According to the European agenda for higher education, it is predicted that in the period until 2025, half of the existing professions and jobs require high qualifications, and the existence of numerous omissions in the education process and skills acquired within higher education is recognized. Under the influence of numerous challenges that include social transformation, climate change, globalization, urbanization, environmental threats, and increasing pressure on public services, the health system, infrastructure, and housing, as well as the growth and affirmative positioning of information technologies and the digital transition, various jobs, and professional domains are becoming increasingly flexible and complex. Such circumstances require future experts in various fields to be entrepreneurially oriented, manage complex information and systems, think creatively and problem-solving, develop an awareness of sustainable planning and development, the health of the population and the environment, use resources wisely and communicate effectively, expressing their opinions based on knowledge. Moreover, the National Sustainable Development Strategy of the Republic of Serbia recognizes applied knowledge, education, and science as dominant factors in the development of modern economy and society, which is why the continuous improvement of the education system is one of the leading directions of change for the sustainable development of society.

Well-designed programs and curricula, focused on the needs of students, are crucial for the effective development of professional skills. A wider range of elective courses, including innovative master's and specialist programs, and opportunities for continuous professional development, help higher education better respond to the needs of society. In this framework, it is necessary to harmonize study programs in accordance with European academic standards and regulations, with a special focus on the compliance of these requirements with the local context and newly identified priority areas for the labour market in a national and international environment. Numerous nationally and globally established acts confirm that effective education systems are the foundation of open and democratic societies and sustainable growth and employment, and accordingly identify education and skills as a priority for international cooperation. Accordingly, the necessity of encouraging new study programs, as well as joint study programs with reference to international universities and institutions, is also recognized. In this sense, this paper represents the belief that the gradual realization of

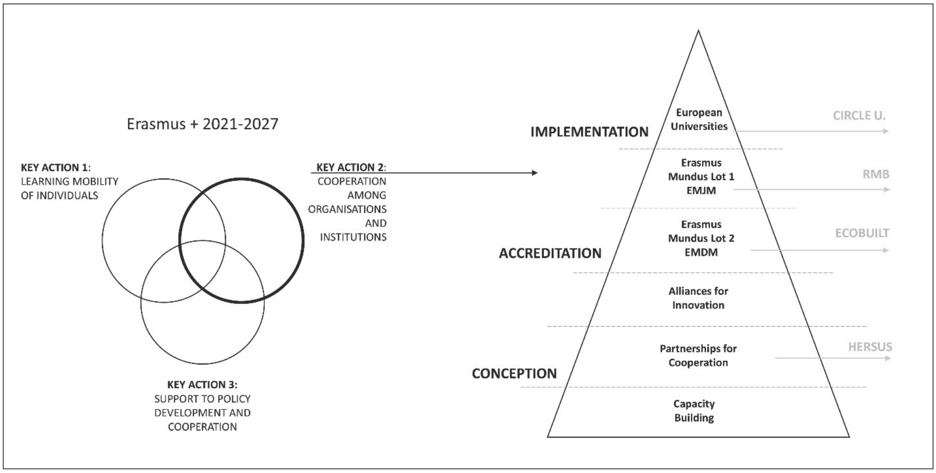


Figure 1. The structure of the Erasmus + program for the period 2021–2027 with the positions of UB-FA projects.

this long-term and the strategic goal would be in feedback with the permanent need to modernize study programs and methodologically operational ways of working at faculties.

In order to achieve a broader perspective of the aforementioned strategic goal, the University of Belgrade and the Faculty of Architecture have adopted the Erasmus + program as a specific system for capacity building and innovation within which a series of actions positioned in a hierarchical manner are carried out — from conceptualization (actions of capacity buildings and strategic partnerships) towards accreditation process (Erasmus Mundus Design Measures) and implementation (Erasmus Mundus Joint Master and European Universities). Figure 1 shows the structure of the Erasmus + program for the period 2021–2027 with the indicated positions of the 4 projects that are currently being implemented at the University of Belgrade: (1) Strategic Partnerships in Higher Education — HERSUS project, (2) Erasmus Mundus Lot 2 — Design Measures — ECOBUILT project, (3) Erasmus Mundus Lot 1 — Joint Masters — RMB project, and (4) European Universities — CIRCLE U. project. In the continuation of the paper, all four projects will be presented through (1) general background and aim, and (2) formulation of project contribution for the long-term realization of a set strategic goal.

3. PROJECTS REVIEW

3.1. HERSUS project

Enhancing of Heritage Awareness and Sustainability of Built Environment in Architectural and Urban Design Higher Education (HERSUS) is Strategic Partnership in Higher Education project implemented in the period from 2020–2023 and coordinated by UB-FA. Together with the UB-FA, HERSUS engages 4 European architectural schools: Iuav University of Venice, Aristotle University of Thessaloniki, University of Cyprus, and Universidad de Sevilla.

The main objective of the project is to create and pilot new innovative courses/groups of courses/extracurricular activities within existing study programs of participating schools, which can contribute to bridging the gap between sustainability and heritage. Moreover, HERSUS aims to promote greater awareness about the types and methods of innovative and sustainable-directed teaching and learning through (1) publishing Intellectual Outputs (Review of best practices, Survey on state-of-the-art, Statements for educating sustainability and heritage, Digital sharing platform, Book of innovative courses, and Handbook on Research and Design for the Sustainability of Heritage) and (2) implementing Learning/teaching/training activities (Seminar for Teaching through design for Sustainability, three Student workshops on Sustainable Reconstruction in Urban Areas, Adaptive Reuse and Resilience, and Training for Teachers on design studio practice). The project is designed to result in the conceptual framework for further formal development through the accreditation of new innovative study

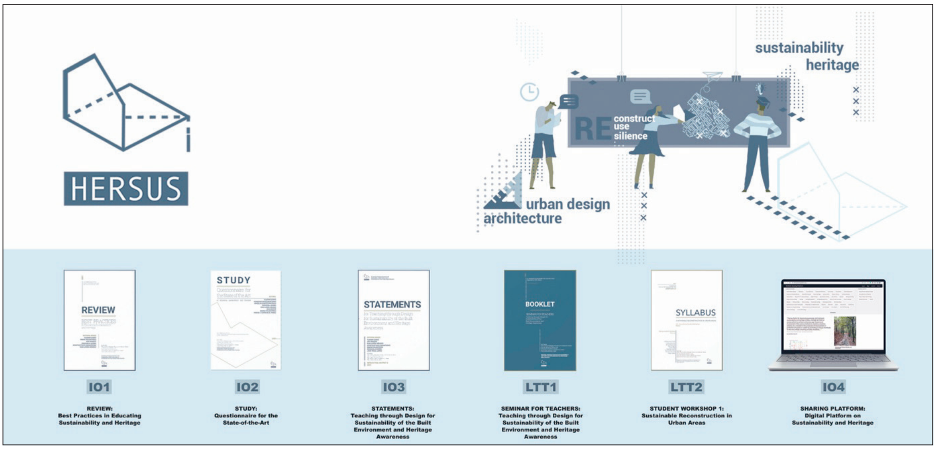


Figure 2. HERSUS Project illustration of intellectual outputs

programmes or through piloting knowledge alliances that will enhance the arena of education and arena of practice. HERSUS contribution is twofold and it is reflected in the enhanced existing study programs at the master level, and the achieved stable and sustainable education framework complementary to the globally established goals in the field of architectural and urban studies education.

3.2. ECOBUILT project

Eco-friendly Built Environment (ECOBUILT) is Erasmus Mundus Lot 2 — Design Measures project implemented in the period from 2022–2023. In this project, UB-FA is engaged as a partner institution together with the Riga Technical University (coordinating university) and the University of Genova who have agreed on a process-based, problem-based and multilevel approach based on a theoretical and practical focus on the specific areas of eco-oriented architecture and its implementation in architectural and urban design. The main objective of the project is to develop an integrated Master program in Eco-friendly Built Environment aimed at training highly qualified professionals who will be able to design and restore buildings and landscapes in urban and rural settings with a minimal impact on human health and environment.

The core strategy of ECOBUILT is to focus on specific skills related to specialized tracks: green architecture, urban design, landscape design, and renovation design. Moreover, ECOBUILT aims to raise awareness of eco-paradigm in architecture urbanism through (1) organizing scientific conferences and open houses (events directed to the wider public), and (2)

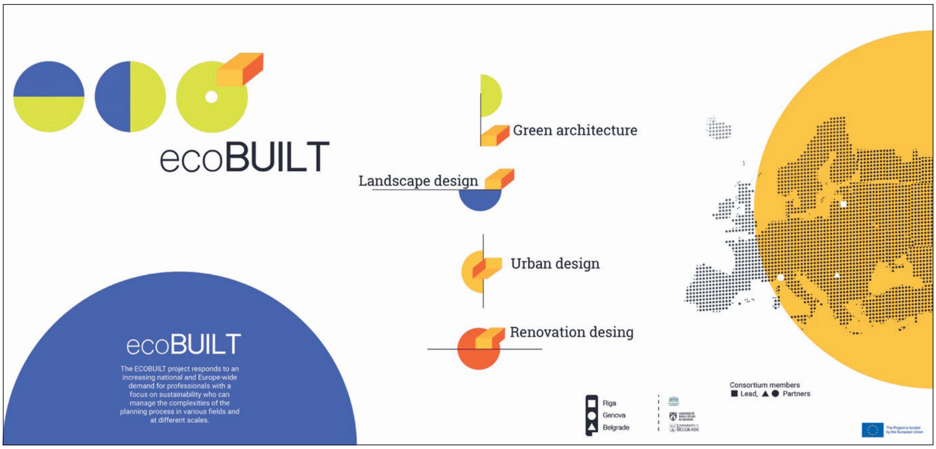


Figure 3. ecoBUILT Project illustration of main topics

developing common procedures for the accreditation process, QA, and promotion of future study programme. The project is designed to mainstream the Master's programme into universities portfolio/accreditation with the further intention to upgrade it within Erasmus + programme hierarchy. ECOBUILT's contribution can be primarily recognized in gaining experience when it comes to accreditation procedures and mechanisms according to EU standards, which at the same time enables the perception of all complementarities and differences with other institutions on the international scene of education.

3.3. RMB project

Reuse of Modernist Buildings (RMB) is forced to be implanted within Erasmus Mundus Lot 1 — Joint Master action in the period from 2023–2028. In this project, UB-FA is engaged as a partner institution together with the TH-OWL (coordinating university), the University of Antwerp, the University of Coimbra, and The Technical Institute of the University of Lisbon. The main objective of the project is to provide deep and qualified academic expertise and knowledge of procedures of a conceptual / design, technical/constructive, and methodical kind in the areas of building conversion and reuse of modernist buildings. The need for the RMB study programme is based on the individual experiences of individual partners, market surveys, and current EU initiatives that show a general concern for heritage at loss, the growing awareness of reuse as effective in fighting climate change, and development in the building industry.

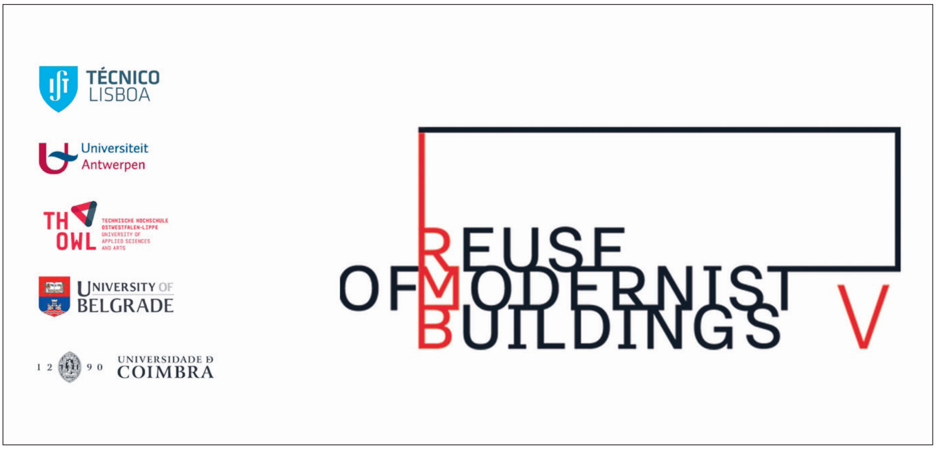


Figure 4. RMB Project illustration of partner institutions

The core strategy of RMB is to (1) deal with built heritage issues and reuse them confidently in a methodologically clear and respectful way and (2) to approach conservation, transformation, and reuse from a broad perspective, including spatial patterns, cultural heritage, climate and construction principles, social and technical assessment, and monitoring of built spaces. RMB's contribution is reflected in the broadening of institutional international activities and networking which is expected to result in an opportunity for students to apply to a new joint international Master programme in high demand, in enhanced collaboration between participating countries with the impact on the regional environment — developed regional hubs, in enhanced teaching skills of key staff members, as well as in enhanced university-business partnerships through combining knowledge into a multidisciplinary knowledge pack.

3.4. CIRCLE U. project

Circle U. European University (CIRCLE U.) is an inclusive, research-intensive and interdisciplinary European university formally implemented in the period from 2020–2023. In this project UB is engaged as a partner institution together with the Aarhus University, Humboldt-Universität zu Berlin, King's College London, Université Paris Cité, University of Louvain, University of Oslo, Università di Pisa and Universität Wien.

The main objective of the project is grounded in a dedication to ethical principles and a sustainable society, links together the missions of all participating universities in education, research, innovation, and service to society, and paves the way for the universities of the future. Moreover, CIRCLE

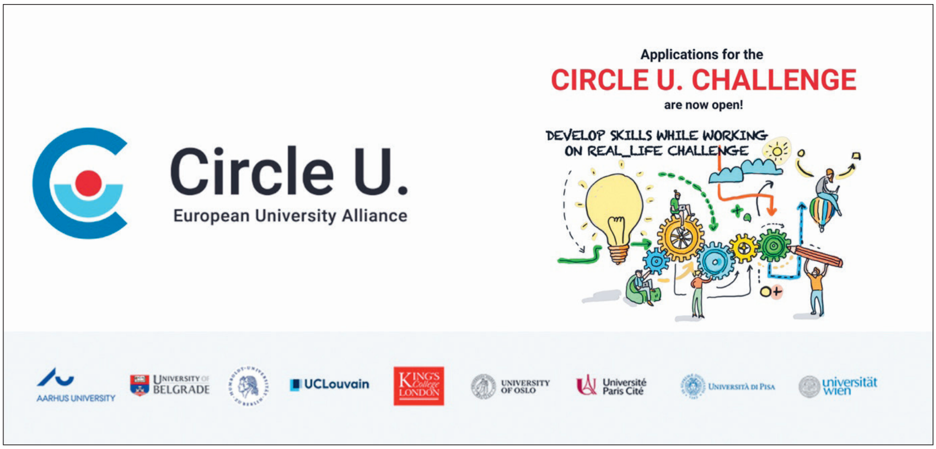


Figure 5. Circle U. Project illustration of partner institutions and one of the activity

U. aims to promote European values, knowledge, and identity through (1) Knowledge Hubs as physical and virtual spaces for students and staff to co-create knowledge and solutions across disciplines with external stakeholders in three main fields: Climate, Democracy, and Global Health, and (2) Academic Chair Programme as a network for the implementation of Open School of Public Governance, in the Student Led Sustainable Innovation and, in the Think, and Do Tank. CIRCLE U. contribution is recognized in the capacity of the transformational changes in teaching methods and programmes, and their cross-cutting with critical thinking and problem-based research which in long-term impact contribute to the development of a specific profile of students and teachers in the European Education Area recognized as 'active co-creator.

4. DISCUSSION AND CONCLUDING REMARKS

Sustainability and responsibility toward people and environment, followed by culture and design-led approaches and green strategies present main priorities in the field of urbanism. Following this line of reasoning, all four projects illustrate current topics in the field of advances in urbanism: from the relationship between sustainability and heritage (HERSUS), to ecological and environmentally sensitive approaches (ECOBUILT), reuse of existing building stock and reuse, rehabilitation, reconstruction, (RMB), and public health, well-being and co-creation and interdisciplinarity (CIRCLE U.).

Conceptualization, development, accreditation, and implementation of the presented projects can significantly contribute to encouraging the adaptation of the legal framework that would enable more effective cooperation on the joint implementation of interdisciplinary study programs that can further respond to increasingly intense changes in the labour market and contribute to the formation of new highly qualified professionals. The establishment of the Framework for monitoring and evaluating quality in accordance with the Strategy for the Development of Education and Training in the Republic of Serbia until 2030 is a prerequisite for continuous and systematic monitoring and evaluation of the quality of education, as well as the creation of purposeful educational policies based on accurate and up-to-date data. Continuous monitoring of quality would contribute to a greater degree of transparency and integrity at the University of Belgrade. On the other hand, improving and evaluating quality would contribute to the development and efficiency of the quality management system through a multiple quality control system that will include all relevant aspects for the management and work of a modern University — science, teaching, finance, and international cooperation.

Branislav MITROVIĆ[*], Jelena MITROVIĆ[**]

OD SUVOMEĐE DO SAVREMENOG OBLIKA: MORFOLOŠKA ULOGA ZIDA U CRNOGORSKIM PRIMORSKO-PLANINSKIM PEJZAŽIMA

Sažetak: Ova studija ispituje morfološke uslove za savremenu arhitekturu u prirodno-istorijskom kontekstu crnogorskog primorsko-planinskog predela. Istraživanje obuhvata teme koje se kreću od analize značenja izvorne strukture suvomeđe i primarne terasaste izgradnje radi postizanja obradivog tla, kao i od autentičnih mesta okupljanja u nepristupačnim krajevima, do savremenih arhitektonskih rešenja, koja u sebi sadrže metamorfoze oblika zida kao odgovor na originalnost pejzaža u ekspanziji.

Projektantska pozicija predstavljena je na osnovu tri studije slučaja: 1) svetlost i senka — likovna vrednost kontrasta: *Royal Gardens Paradise*, Budva; 2) transformacija strukture u istorijskom kontekstu: *Hotel Kamenovo*; 3) otvoreni oblik nadogradnje: *Helidrom/gumno/bistijerna*, Markovina kod Cetinja. Pored odgovornosti prema programskim zahtevima novih sadržaja, prikazana uloga projektanta usmerena je ka očuvanju nasleđenih prostornih vrednosti. Studija teži da pokaže kako artikulacija ovih vrednosti koja novim oblicima daje dublji smisao i pokazuje njihov istinski potencijal, može postati individualni zadatak i izazov za arhitektonsku investiciju u prostoru koja predstavlja mnogo više od doslednog odgovora na utilitarne aspekte.

U studijama slučaja istaknuta je originalnost građevina koje interpretacijom primarnih oblika dobijaju širu vrednost kulturoloških, prostorno-urbanističkih označitelja. Objekti su tumačeni kao relacioni sadržaji okruženja, koji putem razmene postižu sopstveni identitet. Tematski kontekst analize koji u širem smislu predstavlja fenomenologija zidova, sa jedne strane je određen istraživanjem metaforičkih značenja primarnih formi, a sa druge tehničkim iskustvima procesa izgradnje.

Ključne reči: *suvomeđa, zid, prostor, svetlo — senka, istorijski kontekst, oblik, nadogradnja*

---

[*] Branislav Mitrović, profesor emeritus, Srpska akademija nauka i umetnosti; Univerzitet u Beogradu, Arhitektonski fakultet

[**] Jelena Mitrović, slobodni istraživač

UVOD: SIMBOLIKA ZIDOVA

U nekoliko studija koje se bave savremenim modelima intervencije u zaštićenom predelu [1][2], definisani su oblici koji ne narušavaju tradicionalnu građu, zasnivajući se na kaskadnoj kompoziciji, koja morfologiju vrtača[1] odražava strukturno kroz novoizgrađenu teritoriju, i funkcionalno sa ozelenjenim terasama (proces je od ranije poznat u južnoameričkim civilizacijama, čijom su primenom Inke izgradile carstvo). Cilj daljih teorijskih razmatranja ostaje posvećen ovim tehnikama i unapređenju kritičkog pristupa u savremenoj arhitektonskoj praksi. Savremena izgradnja na Crnogorskom primorju u svojoj ukupnosti ima odlike nekontrolisanog procesa, a posledice se ogledaju u nestajanju autentične forme, kao i u narušavanju prostornih i istorijskih odnosa u kojima su se ogledali utemeljenost i kontinuitet.

Podzidi[2] su igrali fundamentalnu graditeljsku ulogu u strukturi tradicionalnih naselja, u onom smislu u kome je kuća sa okućnicom predstavljala osnovni element prostorne organizacije na ravnom tlu [2]. Mesta su se preko vrtača uklapala u predeo, što je bio vid horizontalnog širenja kultivisanog zamljišta na vertikalnoj i goloj podlozi. Ambijentalno prožimanje prirodne i građene forme istorijskog primorsko-planinskog pejzaža u zaleđu obale proizlazilo je iz osnovnih iskustava narodnog graditeljstva, koje je najpre služilo da obezbedi obradivo tlo za uzgajanje poljoprivrednih kultura i, prema navodu Aleksandra Frojdenrajha (Aleksandar Freudenreich), „nije sadržalo ništa što nije konstruktivno ili funkcionalno motivisano". [3] U prošlom veku moderni pokret iznova je definisao uporišta savremene arhitektonske funkcije, pozivajući se na povratak strogoj jednostavnosti neimarskog graditeljstva. Međutim, koliko god da je naša strukovno-graditeljska motivacija u skladu sa tradicionalno-logičkom jednostavnošću zidova i iz njih proistekle „znalačke, pravilne i raskošne igre oblika okupljenih na svetlosti", [4] u realnosti preovladavaju profili koji su odraz sasvim drugačijih proračuna, a čiji je nedostatak tipološke artikulacije najneposredniji izraz uslova na terenu. Obuhvatajući studije slučajeva, bazirane na praktičnim realizacijama izvedenim tokom protekle decenije, ovo istraživanje

[1] *Vrtača* je originalni naziv za površinski oblik kraškog reljefa. U zavisnosti od kraja u kome se sreću, odnosno dijalekta lokalnog stanovništva, sinonimi za vrtaču su dolac, duliba, vrtop ili vrteč, dok je termin „dolina" ušao u svetsku geomorfološku literaturu. U ovom radu, reč *vrtača* korišćena je u izmenjenom značenju kakvo je u najčešćoj upotrebi na Crnogorskom primorju, prevashodno u paštrovskim krajevima, gde označava terasaste veštačke zemljane zaravni ograđene međama, pripremljene tako da omoguće sađenje i uzgajanje maslina i ostalih biljnih kultura.

[2] *Podzid* je zid koji služi kao ograda ili podupirač.

preispituje merila proistekla iz modernističkog pojma funkcionalnosti radi stvaranja nove perspektive u odnosu na aktuelni gubitak prostorno-programskih normi. Pretpostavka je da je gubitak plastičnosti takođe faktor razgradnje primorskog ambijentalnog kontinuiteta. Ovaj problem se ogleda u zgusnutoj masi dimenzionalno i prostorno preeksponiranih objekata, podignutih bez ikakvog geometrijskog odnosa prema izgrađenoj pozadini.

Plastičnost je svojstvo arhitekture najdirektnije vezano za profil zida i njegove metamorfoze, čija se simbolička vrednost ne može u potpunosti objasniti ispunjavanjem prostorne funkcije. Problematizacija urbanističke celine kao *vizuelnog prostora* zahteva da se iznova razmotre narativi *funkcije* i *značenja*. To su međusobno uslovljene kategorije, koje se u strukovnom jeziku pogrešno poistovećuju, imajući u vidu da se realizuju na sasvim različitim nivoima. *Funkcija* po definiciji predstavlja razlog zbog kojeg neka građevina nastaje, dok *značenje* koje joj se pripisuje ne zavisi samo od aktivnosti koja se u njoj obavlja, već i od recepcije koja se formira na osnovu svakodnevnog prisustva arhitektonskog objekta u vizuelnom prostoru grada. Estetska vrednost ne nestaje time što struktura izgubi inicijalnu namenu, kao što su istorijske zidine gradova, koje više nemaju odbrambenu funkciju, ali se njihovo značenje u strukturi primorskog pejzaža, kroz različite kulturne ili prosto samo istorijske procese, uvek iznova aktuelizuje. Iskustva praktične realizacije potvrđuju da je obrnuti odnos takođe moguć, te da su napori u pravcu postizanja plastičke vrednosti objekta doveli do novih perspektiva u pogledu upotrebljivosti prostora. Doživljaj zidova u primorskom kontekstu može biti prevashodno koloristički, sveden na valerske ritmove, gde se kvalitet materijala u potpunosti apsorbuje direktnim svetlom koje prostorne konfiguracije pretvara u čiste likovne kategorije. Na tom elementarnom nivou fenomenologija urbane organizacije, čak i kad se suprotstavlja stogoj tradicionalnoj tipologiji, zadržava svoje dodirne tačke sa fenomenologijom običaja. U vremenu kvantitativnog nadmetanja bez arhitektonskih, urbanističkih i tehnoloških standarda (na kojima se u prošlom veku još uvek bazirao projektantsko-graditeljski proces) oslobađaju se tačke gledišta za inovativnu vizuelnu artikulaciju. U tom smislu, morfološki kapacitet objekta nije diferencirana, već relaciona vrednost, koja prevazilaženjem odgovarajućih dimenzija prelazi na širi prostorni nivo, noseći u sebi poruku o relokaciji, kao što su gradske utvrde, odvajanjem sofisticirane arhitektonske konstrukcije od prvobitnog značenja i označavajuće forme suvomeđe[3] ipak sačuvale kulturološki smisao označitelja.

---

[3] *Suvomeđa* (ili *suhomeđa*) je kulturna baština nastala iz nekadašnje uslovljenosti i prilagođavanja čovjeka podneblju. Karakteristična u kraškim predelima, ova veština zidanja kamenom bez vezivnog materijala se i dalje neguje, pri čemu

Profil suvomeđe, zaslužan za temeljnu arhitektoničnost *kulture vrtača*, nosilac je kompleksne simbolike prepletene u inicijalnim formama potpore i granice. U neimarskom graditeljstvu pitanje forme razrešavalo se prenošenjem opšteg u pojedinačno, dok se razvoj objekta u, bilo sistematskom ili stihijskom, modernom projektovanju i planiranju kreće generički, od pojedinačnog ka opštem. Označavajući krajnju granicu kultivisanog područja, uključujući svaki pojedinačni deo i celinu, tradicionalni podzidi, iz kojih su se razvili morfološki obrasci kuće i grada, nisu dozvoljavali apsolutnu identifikaciju sopstvenog mesta ili parčeta teritorije. Istorijski modeli lokalizacije unutar višeg geometrijskog totaliteta označavali su, ne toliko mnoštvo jedinica, koliko jedinstvenost. Princip suvomeđe prenosio se na svaki razgraničeni komad zemlje, opisujući profilnu liniju na kojoj je, prethodeći tehnički i oblikovno usavršenoj arhitektonskoj formi, počivao univerzum identiteta. Oblik krajnje instance prostora organski je bio povezan sa njegovim unutrašnjim sadržajem. Siluete su govorile oku o suštini stvari: kontura gumna[4] svojom sažetošću stabilizuje aktivnost za koju je predviđena, okupljajući članove zajednice ravnopravno pozicionirane u savršenom krugu. Zidovi su se nalazili u funkciji zajedničkog života iz koje proizlazi arhitektura naselja kao što je Sveti Stefan, koja je posedovala snagu da kultivisanu prostornu formu zaštiti čak i kad je eksponirana usred neprijateljskog okruženja. Od tog istorijskog doba do danas, urbanistička celina postupno je gubila primarno značenje morfološkog omotača, a takva i izglede za ponovno naseljavanje. U današnjem vremenu, u uslovima širenja bez forme, samo se povećanim naporima imaginacije spoljašnja kontura objekta može uobličiti kao likovno motivisani element panorame.

Iz perspektive konstrukcije zidova, suvomeđa jedina poseduje svojstvo nedovršene građevine, čija figurativna propustljivost omogućava aluzivno preklapanje strukture i prirode, upućujući na tajanstvenu simboličku razmenu između tla i prostora. Poroznost koju omogućava čvrstoća kamena ogleda se u pukotinama koje propuštaju vodu i iz kojih proviruje rastinje, kao da poručuje da struktura nije definitivna i posvećena trajanju, već živi u vremenu i prirodi kao sve druge stvari koje postoje na svetu. Zidovi u primorskom pejzažu, zahvatajući prostor svojim masivnim prisustvom, nisu jednoobrazni sistemi definisani pozicijom i funkcijom, već tajanstvene, neprozirne, u sebe

---

predstavlja idealan način za uređenje okućnica i dvorišta, nezavisno od namene objekta. Pored parternog uređenja podzidom, suvomeđom se mogu zidati zidovi kuća, volati itd, [5].

[4] *Gumno* (ili *guvno*) je mesto na kome se vrše žito, a u nastanjenim planinskim vrletima mesto na kome se ljudi sastaju, druže, dogovaraju i igraju. To je mali, kružni horizontalni plato, ograđen najčešće kamenom, koji je odredila nepristupačna i negostoljubiva konfiguracija tla, [6].

zatvorene konstrukcije u kojima svaki otvor igra ulogu istovremeno sadržaja i znaka. Produbljivanje percepcije „kroz zidove“ neodvojivo je od psihičke dimenzije dubine koju Merlo-Ponti (Merleau-Ponty) naziva „iskonskom“ [7], a u skladu sa kojom zatamnjeni prodori asociraju na tajni život unutar siluete, omogućavajući zaštićene tokove vazdušne cirkulacije profilisane i apsorbovane u solidnosti kamenih elemenata. Prema Džonu Loku (John Locke, 1690) neposredni doživljaj zidova ne može se odvojiti od njihovog vizuelnog učinka upravo zbog čvrstoće koja se otkriva posredstvom fizičkog pritiska, predstavljajući jedino svojstvo koje je suštinski neprobojno i koje se ne može zamisliti ni pronaći nigde drugo osim u materiji. Fizička čvrstoća ne zavisi od tvrdoće kao osobine površi, niti od konstrukcije na koju se pozivaju moderna tumačenja Vitruvijevog mesta, već obuhvata materijalnu punoću i psihološku dubinu, omogućavajući senzaciju sigurnosti „čvrsto povezanu i od suštinskog značaja za telo“ [8]. Čvrstoća zidova i tla je u primorskoj tradiciji građenja predstavljala osnovnu kvalitativnu vrednost strukture koja se posredstvom dodira i pritiska prenosi na telo, pružajući osećaj sopstvene jačine i postojanosti forme. Ona je uobličila vidljivu i skrivenu arhitekturu primorskog predela kao prostorni uslov za razvoj kulture, čijem je svetu omogućila stabilnost i trajnost nasuprot nesigurnosti prirodnog okruženja. Hegelova (Hegel) definicija arhitekture, na koju se nadovezuje Lakan (Lacan), identifikujući prazan prostor ili simboličko nemesto u unutrašnjosti piramide, denotira idealistički iskorak modernog doba ka sinteznoj geometrijskoj funkciji omotača i figurativnom karakteru objekta. Time se arhitektonsko znanje udaljilo od istorijskog pojma oblika — elementa koji je bio gradivan u odnosu na prostor, a ne čisto simbolički, odgovarajući vitruvijanskom razumevanju „umetnosti građenja“. Arhitektonski problem koji donosi aktuelno doba vezan je za gubitak analognog odnosa arhitekture i geometrije, uslova građevine koje je modernizam najpre iz osnove preformulisao, nepovratno ih razdvajajući od tradicionalnih značenja. Gubitak strukturalne vrednosti, zasnovane na modularnom sistemu, vodio je do konačnog rastapanja istorijske urbane konfiguracije, ka pojmu fasade i grada kao sistema informacija nespojivog sa dubinom tradicionalnih kompozicija i njihovom unutrašnjom plastičkom hijerarhijom.

1. SVETLOST I SENKA — LIKOVNA VREDNOST KONTRASTA: *ROYAL GARDENS PARADISE*, BUDVA[5]

Tradicionalna arhitektura Crnogorskog primorja istovremeno je ogoljena u svom izrazu i organska u svom odnosu prema kontekstu; svi elementi

[5] Autori: Branislav Mitrović, dia, Jelena Kuzmanović, dia; saradnici: Siniša Tatalović, dia, Danijela Bakić, M. arch. Projektovano: MITarh, 2014, [10].

koji je sačinjavaju su konstruktivne i funkcionalne prirode — jasnoća definiše arhitekturu. Kamen daje ozbiljan, a donekle i monumentalni izraz čak i najmanjim objektima arhitektonskog i graditeljskog nasleđa. Trajnost gradivnog materijala kao što je kamen, za razliku od drugih materijala koji se koriste u narodnom graditeljstvu (drvo, ilovača ili opeka) doprinela je u znatno većoj meri održavanju tradicionalnih arhitektonskih motiva, jer su uzori i iskustva u kamenu sačuvani od davnina i danas gotovo u neprekidnoj živoj vezi i upotrebi. Pitanje savremene izgradnje na Crnogorskom primorju u funkciji modaliteta lečenja prostora koji je u najvećoj meri devastiran, jeste sledeće: kako se arhitektonskim oblikovnim sredstvima, koja su uvek u vezi sa strukturom i identitetom nove arhitektonske forme, uslovi optimizacije savremenih stambenih, uslužnih i komercijalnih sadržaja mogu uskladiti sa karakterom i morfološkim aspektima predela u kome su sadržani i u kome obrazuju prepoznatljiva mesta [2]?

Metamorfoza kamenog zida ogleda se u modernističkoj arhitekturi nadovezanoj na njegovu izvornu racionalnost, samo što je modernizam klasičnu korelaciju između značenja tog elementa i njegove pozicije u prostoru zamenio geometrijskim standardom koji oblik povezuje sa funkcijom. Le Korbizijeova definicija arhitekture tek na mediteranskom suncu dobija kontekstualni smisao, ali se istovremeno njena plastička kategoričnost razbija tonalnim kontrastima između zida i otvora koji dopuštaju da dva različita prostorna intenziteta poprime istu formalnu vrednost i da svako traženje plastičkog prelaza od svetlog ka tamnom zameni jukstapozicija. [9] U savremenim okolnostima projektovanja i izgradnje prostorno-logička veza opstaje samo kao mogućnost i uporište da se dva arhitektonska oblika, koja nastaju iz iste ravni, postave jedan do drugog kao dva tona iste boje. Jukstapozicija oblika (punog i praznog) od početka je sadržala odnos svetlog i tamnog, ali se opažajni efekat oslobađa prostornog značenja i više ne posmatra u kvantitativnoj, već u kvalitativnoj tonalnoj vrednosti. Arhitektonski koncept rezidencijalnog kompleksa *Royal Gardens* u Budvi zasniva se na prebacivanju od plastičkog odnosa svetlo — senka na čistu valersku suprotstavljenost, vizuelnim negiranjem prostora kao dubine i njegovim nadoknađivanjem vrednošću plohe kao apsolutne frontalnosti. Razgradnja volumena u službi je panoramskog efekta i pojavljivanja likovnog ritma u haotičnoj slici građevinske ekspanzije, posledici masovnog nedostatka urbanističke strukture u kojoj je profil starog grada relativizovan. Eksponirana *frontalnost* objekta nije, kao kod Le Korbizijea, potvrda klasične perspektive, već iznenadno preokretanje prostornih elemenata na površini radi odlučnog hromatskog i ritmičkog potvrđivanja. Podizanje primarnog oblika na visoke stubove iznad prve etaže projektantski je čin, koji omogućava da se sa

Slika 1. / Slika 2. Rezidencijalni kompleks *Royal Gardens*, iz projektne dokumentacije autora

gornjih spratova uživa u lepom pogledu, ali gledano iz pozicije na tlu, kao postupak da se najveća težina podigne što je moguće više, omogućavajući dramski intenzitet hromatski suprotstavljenih elemenata (Slika 1).

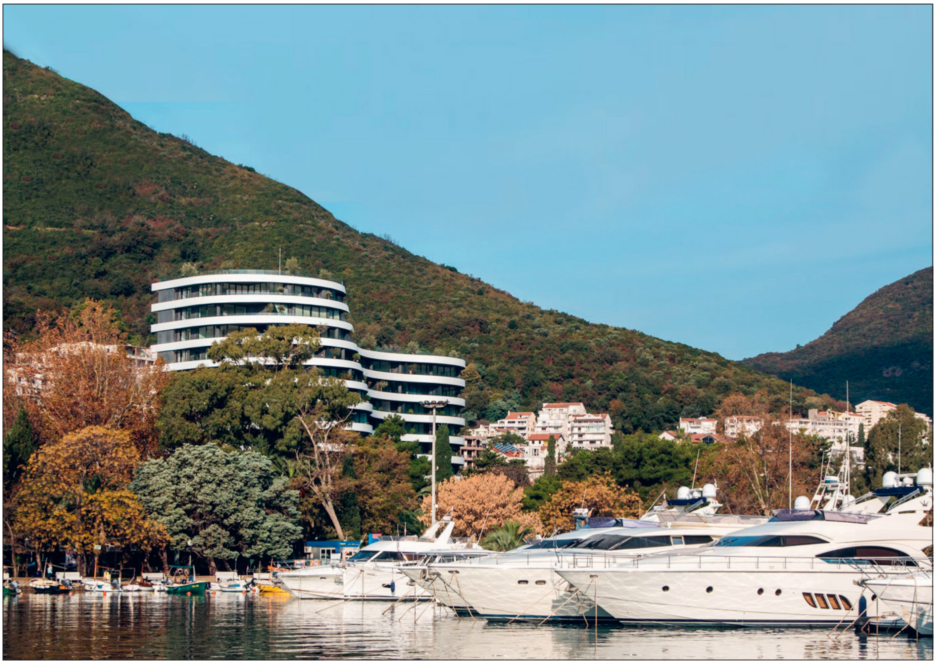
Rezidencijalni kompleks *Royal Gardens* (Slika 2) nalazi se na urbanistički istaknutoj pristupnoj lokaciji prema najužem gradskom jezgru i Starom gradu Budva, a u neposrednoj blizini marine i gradske luke *Pizana*. Između objekta i šetališta uz obalu nalazi se gradski park, tako da se objekat frontalnom pozicijom nadovezuje na niz građevina koji se od ulaza u Stari grad i *Hotela Avale* prostire u prvoj liniji fronta prema moru, sa direktnim pogledom na luku i marinu. Zgrada formira kompaktnu vizuelnu celinu između Trga Republike, Ulice Mila Milunovića i Slovenske obale, koju čini vertikalni poligonalni postament, prateći regulaciju i građevinsku liniju na parceli, i gornji, etažni deo (od 1. do 11. sprata), čija je frontalna ravan oblikovana kao zalučena, kontinualna geometrijska površ, artikulisana naglašenim tonalnim kontrastom horizontalnih ploha. Elementom slobodnih „visećih" platformi, ovaj deo asocijativno upućuje na brodsku arhitekturu. Horizontalno zakrivljene linije fasade, formirane nadnesenim profilima terasa, meandriraju na delu objekta orijentisanom ka moru, evocirajući stepenaste palube (Slika 2), a kulminiraju kružnim oblikom, stvarajući akcenat ugaonog elementa ka pristupnoj saobraćajnici prema centru grada. Autori su želeli da u urbanističkoj gustini centra Budve, gde uglavnom niču objekti od stakla i betona, postignu drugačiji prostorni koncept koji svojom oblikovnom kompozicijom omogućava različit programski sadržaj, a strukturalnom organizacijom nivoa unosi nov likovni kvalitet i dubinu u zatečeni kontekst. Pored zaobljenosti brodske forme i beline mediteranskog zdanja, duboke terase sa zelenim baštama, treba da naglase i apostrofiraju senzibilitet primorskog ambijenta, animirajući element pejzaža kroz savremenu interpretaciju horizontalnog obrasca vrtače. Pogledom sa mora korpus objekta je

Slika 3. / Slika 4. Rezidencijalni kompleks *Royal Gardens*,
iz projektne dokumentacije autora

volumetrijski dekomponovan, a njegov gabarit ubedljivo minimiziran isturenom, frontalno osvetljenom plohom, izbrazdanom tamnim linijama senki ispunjenih zelenilom. Dekompozicija objekta na osnovu samog svetla ispostavlja se kao mogućnost pejzažnog kontinuiteta u kojem je na fonu brda Spas iznova uspostavljena sličnost između hridi, zidova i brodova — prepoznatljivih silueta koje pripadaju obali (Slika 5).

U prostorno-programskoj strukturi, koju grade jake terase ogledaju se sve karakteristike primorskog ambijenta. Oštrina i kontrast, belina nasuprot

Slika 5. Rezidencijalni kompleks *Royal Gardens*, pogled iz kompleksa na park,
marinu i gradsku luku Pizana. Autor fotografije: Relja Ivanić



Slika 6. Pogled na kompleks *Royal Gardens* iz gradske luke,
iz projektne dokumentacije autora

(* *Objekat ima tri podzemne etaže, sa garažama i tehničkim prostorijama. Rampom je obezbeđen ulaz u garažu iz Ulice Mila Milunovića. U prizemlju i na prvom spratu projektovani su komercijalno-ugostiteljski sadržaji. Između ulice Trga Republike i parka sa šetalištem ispred objekta postoji natkrivena veza — prodor kroz sredinu objekta koji formira gradsku pasarelu u parteru. Na svim gornjim nivoima projektovani su stanovi veličine 50–200 m^2, a posebni prostorni i ambijentalni kvalitet imaju rezidencijalne jedinice u kaskadama između 8. i 11. sprata.*) [2]

dubokim senkama, zelenilo i život na terasi, kao i širok pogled prema pučini, čine strukturalnost i identitet mediteranskog predela sadržan u slici objekta koji u sebi produžava prostor i daje urbanoj lokaciji karakter mesta. Kvalitet stambenog prostora je time dodatno unapređen: svaka jedinica je orijentisana prema moru i ima duboku nadstrešnicu za odbranu od jakog letnjeg sunca, kao i element autohtonog primorskog zelenila, mimoze, oleandra, lavande ili limuna. Sa naglaskom na ozelenjenoj terasi, kao na palubi broda, prožimanje spoljašnjosti i unutrašnjosti je aktivirano u oba pravca (Slika 6). Arhitektonski konept je postavio naglasak na ozelenjenim horizontalnim platoima kao primarnoj ideji izgleda objekta. Suština ovakve prostorne koncepcije je da u preterano izgrađenom i zgusnitom okruženju centra Budve, izgradnjom novog objekta, ne poveća dodatno građevinsko i vizuelno *zagušenje*, već da, uprkos nezanemarljivom gabaritu, stvarajući kaskade terasa i ozelenjavajući sve horizontalne površine, započne vraćati gradu deo oduzetih zelenih zastora iz partera, a njegovim fasadama naglašeni javni karakter, koji omogućava privatnost unutrašnje strukture. Zeleni zastori, koji su poređani po visini, stvaraju utisak visećih vrtova, što bi mogao da bude osnovni moto objekta. Autori su smatrali da prihvatanje ovakvog pionirskog poduhvata na ovim prostorima, a što je definitivno savremeni trend u Evropi, pogotovo u Francuskoj, gde „mrtvi“ materijali građevina moraju da „vrate“ svu oduzetu mogućnost biljnog potencijala zauzete parcele, može da kroz buduću, sličnu praksu, nadoknadi gradu deo potrošenog prirodnog prostora.

2. TRANSFORMACIJA STRUKTURE U ISTORIJSKOM KONTEKSTU: *HOTEL KAMENOVO*[6]

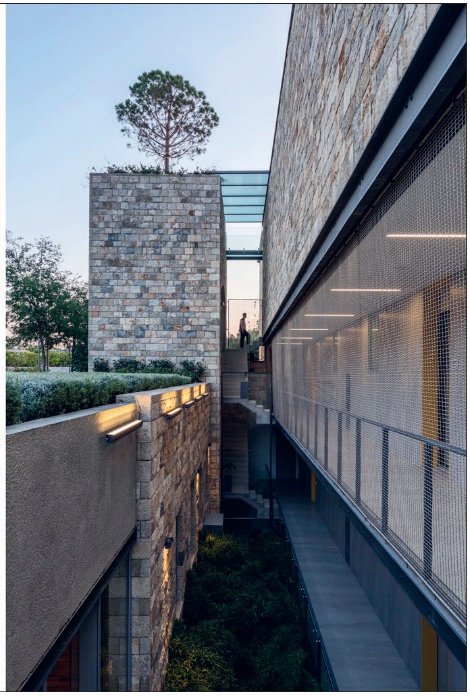
Jedan od najvećih arhitektonskih izazova u primorju, kao i u drugim autentičnim i autohtonim sredinama, prikladna je primena tradicionalnog građevinskog materijala u savremenom, automatizovanom građevinskom procesu. Udaljavanjem od izvornog konteksta, prenošenje neimarske veštine i znanja ubrzano nestaje, a sa tim i mogućnost da se u modernom detalju obuhvati misao i lepota suvomeđe kao usvojene tradicionalne neimarske koncepcije, a koja proizlazi iz svojstvene tehnike obrade i prvobitnog graditeljskog sloga. Uklapanje u tradicionalnu strukturu njenim podražavanjem više je problem improvizacije nego doslovne reprodukcije, zahtevajući da se tip veze posredstvom niza projektantskih odluka konkretizuje i prilagodi ne samo programskim činjenicama, već i sasvim drugačijim formalnim mogućnostima savremene građevine. Izgradnja turističkih sadržaja u zaštićenom predelu vrtača u obalskom zaleđu pokreće ista pitanja kao nadogradnja i rekonstrukcija tradicionalnih urbanih celina. U slučaju hotela na strmini iznad plaže Kamenovo autorski pristup ogleda se u kamuflaži, tako što je korpus zgrade, umesto volumetrijske razgradnje, istanjen i ugrađen u nagib terena (Slika 7). Geometrija koja objekat svodi na dve paralelne ravni omogućava da se projektantski problem kamuflaže reši njihovom (de) materijalizacijom. Pristupna *gornja* fasada tretirana je kao kameni podzid, pozivajući se na fenomenologiju suvomeđe kao primarnog konstruktivnog elementa na tlu (Slika 12), dok je *donji* front ka obali projektovan u vidu transparentnog *ekrana* sa dvoslojnim, staklenim i aluminijumskim, zastorima, utapajući se u okruženje posredstvom različitih nivoa refleksija (Slika

[6] Branislav Mitrović, dia, Siniša Tatalović, dia, Ognjen Krašina, M. arch (u fazi izrade tehničke dokumentacije glavnog projekta); saradnici: Vanja Miletić (autor projekta enterijera). Projektovano: MITarh, 2015, [11].

Slika 7. *Hotel Kamenovo*, iz projektne dokumentacije autora

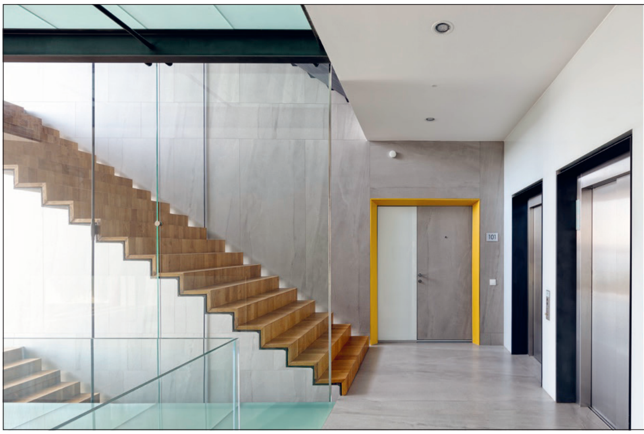
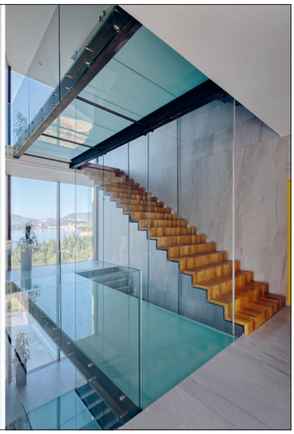
13). Detalj kamenog zida, orijentisanog prema prilazu na nivou magistrale, predstavlja projektantski čin na kome se zasniva celokupni kontekstualni lik arhitekture. U njegovoj hromatskoj kompoziciji sažet je vizuelni ritam stroge ravni, čiju teksturu svetlost dovodi do usijanja. Vertikalna struktura zida, pozicionirana između dubokog useka u steni i enterijera koji celom širinom „gleda" na more, predstavlja u izvesnoj meri insceniranu konstrukciju, koja umesto masivnog podupirača linijski artikuliše kretanje i organizaciju unutrašnjeg prostora, zadržavajući označiteljski odnos prema pejzažu. Tonalne razlike između nijansi iste boje postignute na osnovu komplementarnog odnosa sa okolinom (Slika 8), obrazuju intenzivni kolorit u kamenu, koji se spektralno razlaže pred očima prolaznika, doprinoseći utisku da konstrukcija postoji samo na svetlu i za svetlo. Uronjena u nemirne kontraste okruženja, ona tako stvara nešto uvek drugačije i sugestivno.

Hotel Kamenovo (Slika 7) je realizovani autorski projekat na karakterističnoj, isturenoj lokaciji koju čini nepristupačna zaravan obalskog rta na desnom kraju velike uvale poznate kao plaža Kamenovo. Samo mesto obeleženo je kulturno-istorijskim simbolima, blizinom ambijentalne celine paštrovskog sela Radovići i manastira Duljevo koji potiče iz XV veka. Naglašena topografija litice — krša, kao i značajni arhitektonski spomenici u susednoj okolini, imali su presudni značaj za formiranje tipologije objekta. Arhitektonski koncept se zasniva na opredeljenju da se umesto arhetipa kuće oponaša struktura suvomeđe svojstvena prostorima maslinjaka (Slika 8), čime je istaknuta geometrija vrtače karakteristična za lokalni prostor i čitavo primorsko okruženje sa visokim zaleđem. Prenošenjem logike tradicionalnog načina zidanja na strmom i stenovitom tlu pronađena je forma mimikrije koja

Slika 8. / Slika 9. *Hotel Kamenovo* — detalj zida, iz projektne dokumentacije autora

hotelski objekat svodi na osnovne površinske elemente i tako ne narušava tektonski profil brda dodatom volumetrijom. Prostorno-programska struktura objekta, uključujući liftove i stepeništa, pristupni nivo, kuhinju i restoran na donjim etažama, rezidencijalne jedinice na gornjim, kao i rekreativne

Slika 10. / Slika 11. *Hotel Kamenovo* — detalj enterijera,
iz projektne dokumentacije autora

Slika 12. *Hotel Kamenovo* — pogled sa magistrale, iz projektne dokumentacije autora

(* *Objekat ima dve podzemne etaže, sa kuhinjom, restoranom i tehničkim prostorijama. Auto-liftom je obezbeđen ulaz u garažu iz novoprojektovane Ulice SK2. U prizemlju i na prvom spratu projektovani su recepcija i javni hotelski sadržaji. Između ulaza na nivou magistrale i najnižeg platoa omogućena je pešačka pasarela do plaže, mimo komunikacija u funkciji samog objekta. Na svim gornjim nivoima projektovani su apartmani približne veličine od 140* m^2*, a posebni prostorni i ambijentalni kvalitet imaju apartmanske jedinice na drugom nivou Trakta 1 sa izlazom na denivelisane terase krova, bazen i krovnu baštu.*) [2]

Slika 13. / Slika 14. *Hotel Kamenovo* — detalj fasade ka obali, iz projektne dokumentacije autora

Slika 15. *Hotel Kamenovo* — fasada ka obali, iz projektne dokumentacije autora

sadržaje na terasama i krovu, pozicionirana je između dva karakteristična, paralelno postavljena vertikalna plana od kojih je prvi, kao glavna osa komunikacije kroz objekat, projektovan u formi tradicionalnog kamenog podzida sa kulama, uključujući ravan *ukopa* sa prorezom između tehničke fasade i tla kojim je unutrašnji volumen dodatno *istanjen* po dubini (Slika 9). Sa suprotne strane, gde se objekat celom površinom fasade okreće ka moru, otkrivajući svoju stvarnu veličinu, panoramska ravan vizure tretirana je kao istovremeno reflektujući i programski ekran, definisan pokretnim aluminijumskim žaluzinama ispred kontinualnog stakla (Slika 14).

Arhitektonska forma *Hotela Kamenovo* nastala je metodom *mimikrije* ili utapanja u prirodu. Mimikrija je, prema projektantima ovog rešenja, jedan mogući odgovor na pitanje: kako se odnositi prema zahtevima specifične lokacije u odnosu na krupne prostorne zahteve velikih investicija? U podražavanju se ogleda autorski stav koji kompenzuje zahtevan projektni program i funkcionalnu šemu raslojavanjem korpusa građevine i vizuelnim uklapanjem u postojeću strukturu predela, vodeći do konkretizovanja projektantskih tehnika i relacija, na osnovu kojih se neguje morfološki kontinuitet autentičnog pejzaža. Pozicioniranje objekta ima višestruki smisao za lokacijske uslove karakteristične za Crnogorsko primorje, polazeći od činjenice da se glavne strane hotela dijametralno razlikuju, kao i da je, pored pogleda iz samog kompleksa, značajan i pogled na kompleks sa mora i plaže gde se čitava površina fasade otvara u širini litice i u punoj veličini izlaže perspektivi posmatrača (Slika 15). Volumen objekta je vizuelno negiran dvostrukom artikulacijom *zida*, čime su definisana dva lica kuće kao odrazi odgovarajućih lica pejzaža, a istovremeno se zahtevani program logično savladava, ne opterećujući formu ni gabarit.

3. OTVORENI OBLIK NADOGRADNJE: „HELIDROM/GUMNO/BISTIJERNA", MARKOVINA KOD CETINJA[7]

Problem tradicije i poteškoće njenog *autentičnog* oponašanja zahtevaju šire razumevanje savremenih prostornih, programskih i tehnoloških pojava u projektovanju i realizaciji, imajući u vidu ambivalentnost društvenog konteksta bez jasne pozicije u odnosu na istorijske događaje, kao i gubitak standarda koji područje vrednosti čini neodređenim i neuhvatljivim. Jednostavnost programa svedenog na infrastrukturnu nadogradnju u okviru privatnog rezidencijalnog imanja u Markovini kod Cetinja pruža slobodu za pejzažnu intervenciju definisanjem odgovarajućeg arhitektonskog jezika, koji se više, nego na oponašanju tradicionalnih oblika i tehnika, poziva na njihovu interpretaciju za potrebe sadašnjosti. Arhitektonski koncept ogleda se u narativima helidroma, gumna i bistijerne, stvarajući tradicionalno-moderne dijalektičke momente u napetosti njihovog istovremenog odnosa. Infrastrukturni karakter i saobraćajna funkcija helidroma, čiji je gabarit određen dimenzijama i statikom, definiše osnovu za projektovanje gumna i bistijerne uzdignute na simbolični nivo platforme kao pozornice ili spomenika (Slika 16). Kružni plato koji je služio za okupljanje, sastanke i igru, i u tradiciji bio definisan nepristupačnom konfiguracijom tla, u višeslojnom odnosu prema pejzažu koji omogućava transponovanje značenja dobija teatrični karakter prostora predstave čija je pozornica krajolik. Stepenaste terase u čijim je kaskadama usečena rampa koja se spušta do helidroma, sledeći zamišljene vodotokove i povezujući strukturu vrtače sa temom visećih vrtova, omogućava sagledavanje čitave kompozicije kao nadogradnje predela sa pozicije vidikovca, gde se kružna kontura gumna još jednom pojavljuje sa tradicionalnom kamenom klupom oivičenom kamenom. Nedovršenost tog monolita — klupe koja kao da je odsečena nožem — pokazuje da je struktura od samog početka zamišljena kao eksponat tradicije, zajedno sa podzidima, koji sa jedne strane, otkrivaju geološku slojevitost, a sa druge, najezdu puzavica (Slika 19). Plato helidroma je kružna betonska konstrukcija

[7] Autori: Branislav Mitrović, dia, Marija Miković, m. arch, Siniša Tatalović, dia. Pojekat sobraćaja: Petar Milosavljević, ing. saobraćaja, Slobodan Gvozdenović, ing. saobraćaja. Projektovano: MITarh, 2014. Nagrada Ranko Radović 2016, [12].

Slika 16. Pogled na *Helidrom*, iz projektne dokumentacije autora

sa travnatim krugom u središtu i profilisanim omotačem koji je u neimarskom maniru iskrojen od krupnih, nepravilnih komada kamena. Ideja je bila da se geometrija nedovršenog predela, kao označitelj izvornog crnogorskog pejzaža, nadovezujući se na ostatke suvomeđa na stenovitoj padini u koje je usađen volumen helidroma (Slika 20), produži preko podzida sa raščlanjenim osnovama vrtača i njihovom postepenom ekstenzijom u arhitekturu parka (Slika 18).

Veliko izolovano imanje i teren sa kog se pruža pogled na šume okolnih brda je veoma poetičan prostor, a takvi krajolici su kao stvoreni za primarne, pa i za tradicionalne forme. Helidrom je od prvog dana projektovan kao crnogorsko mesto okupljanja, iz čijeg je okvira izveden objekat koji svojom svedenošću i lepotom kamenog kruga dodaje na poetičnosti i mistici tradicionalnih crnogorskih naselja (Slika 21). Prvobitna zamisao je bila da se u unutrašnjem bazenu sakuplja i drži tehnološka voda, ispunjavajući ekološku funkciju cisterne (bistijerne), projektujući metafizičku priču o vodotoku pod rampom, koji pada duž planinskog venca i spušta se niz kaskade do akumulacionog *jezera* zamišljenog u središtu konstrukcije. Poseban akcenat je stavljen na obradu materijala, tehniku slaganja kamena, livenja betona i ispitivanje njihovog odnosa u detaljima čime je realizacija potvrdila autorski koncept.


Slika 17. / Slika 18. Pristup *Helidromu*. Slika 19. Pogled na gumno i helidrom, iz projektne dokumentacije autora

Slika 20. / Slika 21. Helidrom/gumno/bistijerna — pogled odozdo, iz projektne dokumentacije autora

(* *Helidrom iznutra predstavlja snažnu betonsku konstrukciju jer je izgrađen na terenu koji je u padu, tako da su dve trećine objekta vidljive. Osmišljen je unutrašnji sistem odvodnjavanja jer je središnji deo planiran kao zeleni travnati krug. Postoje jasni uslovi koji se tiču signalizacije, osvetljenja i navigacije, što su kao deo projekta definisali i rešili saobraćajni inženjeri.*)

ZAKLJUČNA RAZMATRANJA O FUNKCIJI ZNAČENJA TRADICIONALNOG GRADITELJSKOG ELEMENTA U SAVREMENOM ARHITEKTONSKOM KONTEKSTU

Tri studije slučaja realizovanih projekata biroa MITarh u ovom radu predstavljaju specifična scenarija savremene gradnje u kontekstu Crnogorskog primorja: 1) građenje u urbanom ambijentu primorskog grada u ekspanziji; 2) građenje u zaštićenom prirodnom zaleđu obale; 3) nadogradnja infrastrukture na tradicionalnoj međi u brdskom, nepristupačnom području u sklopu privatnog imanja. Svaki od slučajeva nosi izazove promišljanja i projektantskog postupanja u odnosu na pojedinosti neposrednog konteksta, iz kojih se mogu izdvojiti sledeće zaključne opservacije.

1. Minimiziranjem volumetrije građevine — eksponiranjem frontalne površine i uvlačenjem gabarita, odnosno, optičkom razgradnjom konture objekta — oslobođene su linije terasastih platoa i omogućen sistem *vertikalnih vrtova* koji javnom tlu grada vraća ono što mu je gradnjom uzeto.

2. Gabarit hotela u zaštićenoj zoni vrtače između magistrale i obale dekomponovan je usecanjem terena, isticanjem frontalnih ravni i savladavanjem programa po dubini. Prevođenjem volumetrije u dvostruki odraz pejzaža, sa jedne strane mimikrijom, a sa druge refleksijom, sačuvano je pojavno jedinstvo predela i postignut skladan odnos arhitekture i terena.

3. Infrastrukturna nadogradnja na tradicionalnoj međi formira simbolične i likovne veze sa okruženjem, a tehničko rešenje objekta-platforme

preduslov je za arhitekturu pejzaža na osnovu oblika, *gumna*, koji pored kulturološkog smisla označitelja (helidrom) ostaje znak koji pripada endemskom krajoliku.

U ovim projektima, i sličnim iste vrste najveći je izazov biti autentičan. Tim MITarh se u svim odlukama trudio da do kraja bude veran ideji originala. U savremenoj arhitekturi koja nam donosi epohu bez univezalnih pravila, do živih uvida dolazi se prevashodno kroz pokušaje. Arhitektonski objekti predstavljeni u ovom izlaganju projektovani su i tumačeni kao relacioni sadržaji okruženja, što se ogleda u njihovim karakterističnim vizuelnim identitetima, polazeći od optičkih dimenzija prizora umesto od projektantskih i planerskih činjenica, a vodeći do fenomenologije oblika zidova inspirisanih sa jedne strane, metaforičkim značenjem, a sa druge, tehničkim iskustvima same izgradnje.

BIBLIOGRAFIJA

[1] Mitrović, B, P. Stamenović, „Tipološka geneza nasleđenih prostornih obrazaca u savremenim okolnostima građenja u Crnoj Gori". N. Martinović, (ur.), N. Vuković, (et al) *Paradigme diskursa kulture u Crnoj Gori*. Crnogorska akademija nauka i umjetnosti, Podgorica, 2019, p. 397–421. (/ISBN 978–86–7215–456–6, COBISS. CG-ID 38820112/)

[2] Mitrović, B., Mitrović, J. „Objekti arhitekture u nastajanju u savremenim okolnostima građenja na Crnogorskom primorju". V. Radonjanin, R. Folić (ed.) *Међународна конференција. Савремена грађевинска пракса, 2017, Андревље. / Conference Contemporary Civil Engineering Practice 2017, Andrevlje, 25. и 2. мај 2017.* FTN, Grafički centar GRID, Novi Sad, 2017, p. 7–24. (/ISBN 978–86–7892–917–5; COBISS. SR-ID 314297095/)

[3] Freudenreich, A. *Narod gradi na ogoljenom krasu. Zapažanja — snimci i crteži arhitekata*. Savezni institut za zaštitu spomenika kulture, Zagreb-Beograd, 1962.

[4] Le Korbizije, *Ka pravoj arhitekturi*. Građevinska knjiga, Beograd, 1977.

[5] (*https://www.gradnja.me/clanak/339/Suvomedja*)

[6] (*https://sr.wikipedia.org/sr-el/Гумно*)

[7] Merleau-Ponty, M. *Fenomenologija percepcije.* IP „Veselin Masleša", Sarajevo, 1978.

[8] Locke, J. *An Essay Concerning Human Understanding.* Peguin Classics, London, 1998.

[9] Argan, G. C. *Arhitektura i kultura*. Logos, Split, 1989.

[10] (*https://www.mitarh.rs/index.php?p=project&project_id=274*)

[11] (*https://www.mitarh.rs/index.php?p=project&project_id=269*)

[12] (*https://www.mitarh.rs/index.php?p=project&project_id=250*)

Branislav MITROVIĆ, Jelena MITROVIĆ

FROM DRY-WALL TO MODERN FORM: THE MORPHOLOGICAL ROLE OF THE WALL IN THE MONTENEGRO COASTAL AND MOUNTAIN LANDSCAPES

Abstract

This study examines the morphological conditions for contemporary architecture in the natural-historical context of the Montenegrin coastal-mountain landscape. The research includes topics that range from the analysis of the meaning of the original dry-wall structure and primary terraced construction in order to achieve arable land, as well as from authentic gathering places in inaccessible areas, to contemporary architectural solutions that contain metamorphoses of the wall as a response to the originality of the expanding landscape.

The design position is presented on the basis of three case studies: 1) Light and shadow — the artistic value of the contrast: *Royal Gardens Paradise*, Budva, 2) Transformation of the structure in the historical context: *Hotel Kamenovo* and 3) Open form of upgrade: *Helidrom/Gumno/Bistijerna*, Markovina near Cetinje. In addition to the responsibility according to the programmatic requirements of the new content, the presented role of the designer is aimed at preserving the inherited spatial values. The study aims to show how the articulation of these values, which gives new forms a deeper meaning and shows their true potential, can become an individual task and a challenge for an architectural investment in space that represents much more than a consistent response to utilitarian aspects.

In the case studies, the originality of buildings is highlighted, which through the interpretation of primary forms acquire a wider value as cultural, spatial and urban signifiers. Objects are interpreted as relational contents of the environment that achieve their own identity through exchange. The thematic context of the analysis, which in a broader sense represents the phenomenology of walls, is determined on the one hand by researching the metaphorical meanings of primary forms, and on the other by the technical experiences of the construction process.

Key words: dry-wall, wall, space, light-shadow, historical context, form, superstructure

INŽENJERSKE NAUKE II

ENGINEERING SCIENCES II

Zhilbert TAFA[*]

AIR POLLUTION MONITORING AND MANAGEMENT USING WIRELESS TECHNOLOGIES

Abstract: Air pollution (AP) is one of the main causes of lung cancer and stroke. It has also been correlated with cardiac, ophthalmologic, psychiatric, and many other diseases. In order to minimize the negative health impacts, AP should be properly monitored and managed. Conventional systems are expensive and sparsely deployed, hence cannot provide the required spatiotemporal resolution. This paper reviews the emerging (wireless) technologies for real-time AP monitoring. The implementation of machine learning (ML) in AP monitoring and management is also considered.

1. INTRODUCTION

The development of human societies and the accompanying growth in the consumption of natural resources, especially over the past couple of centuries, has given rise to a multitude of human-induced environmental problems [1]. Air pollution is becoming one of the greatest threats to the human health. Traditionally, it has been monitored by using sparsely deployed expensive instruments. However, the AP has a much higher spatiotemporal resolution. Precisely, due to the metrological (temperature, humidity, wind, etc.) and terrain conditions, the pollutant concentration may quickly vary over a small portion of an area and may not be necessarily accumulated in the close vicinity to the AP source. Some experimental justifications can be found [2] and [3]. Consequently, data acquired from just one station or a few devices may be insufficient to describe the pollution distribution over a given region. The development of high-sensitivity multi-parameter, densely

[*] Zhilbert Tafa, Indiana University, Bloomington, USA and UBT College, Priština, Kosovo

deployed, and inexpensive real-time monitoring systems appear to be of crucial importance to the appropriate pollution prevention and/or management.

The advances in wireless communications and low-cost embedded systems have opened opportunities of ubiquitous computing [4]. Small smart sensors can sense, locally process, and transfer environmental data, virtually from anywhere. Being densely deployed, network can provide a more precise AP distribution over a given region. Such a georeferenced and time-stamped data can further be processed with ML techniques for AP source localization and AP distribution prediction.

In order to eliminate, decrease, or prevent the AP; the monitoring and management information systems work in the inverse feedback manner. They provide real-time sensing and data transmission, remote data visualization and analytics, AP forecast, early warning, etc. Main aspects of such a system may be summarized in: (a) data acquisition, (b) data dissemination, (c) energy management, (d) data utilization.

This article presents a short review on data acquisition and dissemination for AP monitoring. After the introduction, the target parameters and the overall architecture of an AP monitoring system are presented. The modules of a sensor node are described along with the main issues related to the design and implementation. The available wireless technologies and possible topologies are presented in the similar manner. Energy-related issues, as being an important aspect of the Wireless Sensor Networks (WSNs), are also shortly presented. In order to exemplify the aforementioned concepts, two prototype systems developed in the laboratory are presented as well. Some notes on using ML for different tasks in AP monitoring and management are given before the conclusion. Finally, last section concludes the article.

2. WSNS FOR AIR QUALITY MONITORING

Air quality analysis involves the examination of the biological, physical, and chemical properties of the air. Although public concern has been mostly focused on urban areas, due to cooking, spraying pesticides etc., indoor areas may be also highly contaminated. Traditional air quality analysis involves periodical sensing and data reporting from the accurate and reliable instruments. These instruments are expensive (of the order of few thousand dollars) and of large physical dimensions (Fig. 1). As such, only one or a few of them are typically used to cover an urban area (e. g., a city).

WSN nodes can be placed anywhere. They are low-cost (of order of few hundred dollars), power autonomous, and of small physical dimensions. As such, they can provide sensing systems of high sampling and spatial

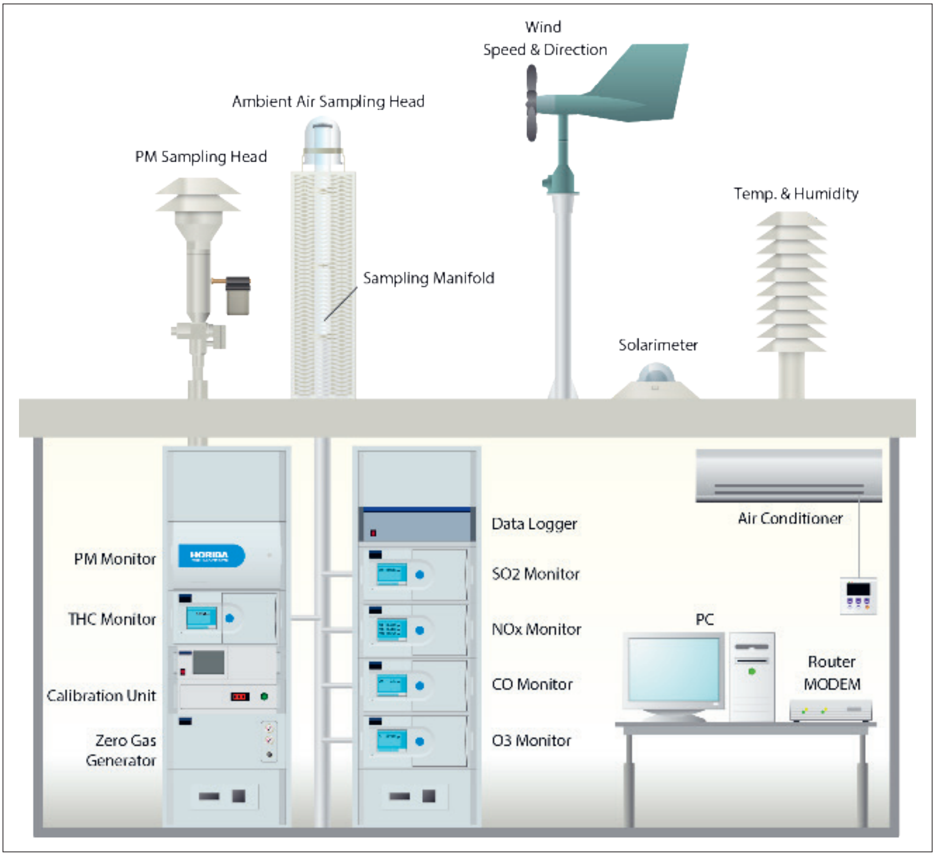


Figure 1: A conventional AP monitoring station [5]. The stations are expensive and of relatively large physical dimensions (typically occupy an area of few m^2).

resolution. Equipped with wireless transceivers, they can report readings in real time. However, the implementation of the WSNs has limitations, such as energy issues, wireless link instability, the environmental and electromechanical influences, etc.

2.1 Air Quality Indicators

Different countries and organizations have set different standards for AP evaluation. Most of the outdoor systems are focused on measuring: carbon monoxide (CO), nitrogen dioxide (NO_2), ground level ozone (O_3), ammonium (NH_4), particulate matter (PM), sulfur dioxide (SO_2), lead (Pb), temperature, and humidity. Indoors, carbon dioxide (CO_2), Particulate Matter (PM), Volatile Organic Compound (VOC), temperature, and humidity, are mostly measured. In order to provide a single understandable information, so-called Air Quality Index (AQI) is extracted from the measured parameters. AQI is usually measured as worst index of separately calculated indices for each pollutant. US EPA categorizes QA in six categories, in scale from 0 to 500. In EU countries, Air Quality Framework Directive specifies AQI in range of 0–100. Some countries (e. g., Canada and UK) use 10-point scale to quantify the overall AQI etc.

2.2 The Elements of a WSN-based Systems for AP Monitoring and Management

The emerging systems for AP monitoring and management rely on WSN infrastructure. A WSN is a data acquisition and dissemination platform that enables data reporting from smart sensors to a web server and/or to data center. The information system should provide the means of data visualization in real time. Also, it should contain modules for data analytics, prediction and early warning. The workflow scheme of such a system is depicted in Fig. 2.

In order to minimize the negative influence of the AP to human health, after identifying the pollution sources, or possible future pollution hotspots, the authorities can act towards the elimination or minimization of the contaminant's concentrations or AP sources.

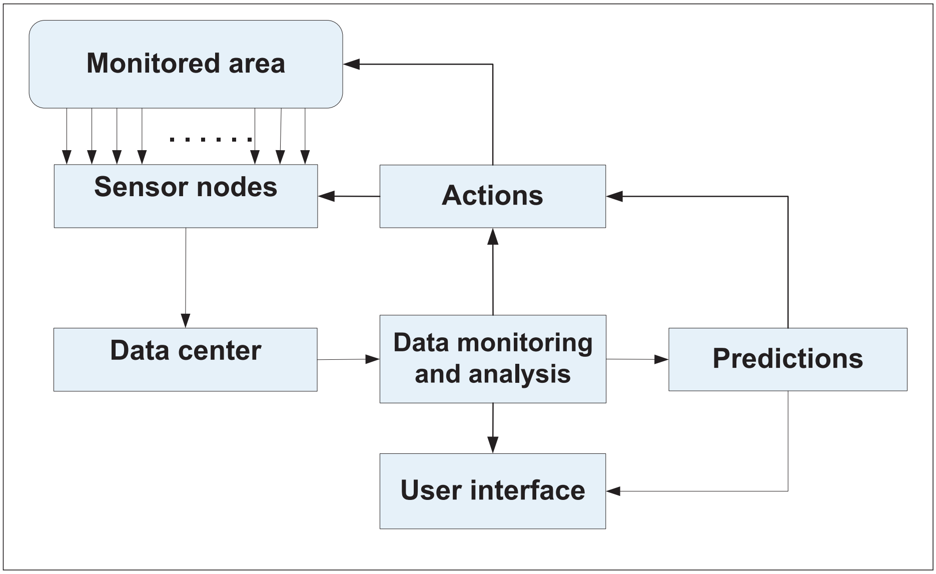


Figure 2: An AP monitoring system contains data acquisition module, data dissemination infrastructure, data center for data visualization and utilization, user interface, and (sometimes) actuators for AP and node's management.

2.3 Data Acquisition Architecture and Issues

Main components of a sensor node are: (a) sensors, (b) signal conditioning module, (c) computing module, (d) communication module, (e) energy management module. The core elements are depicted in Fig. 3.

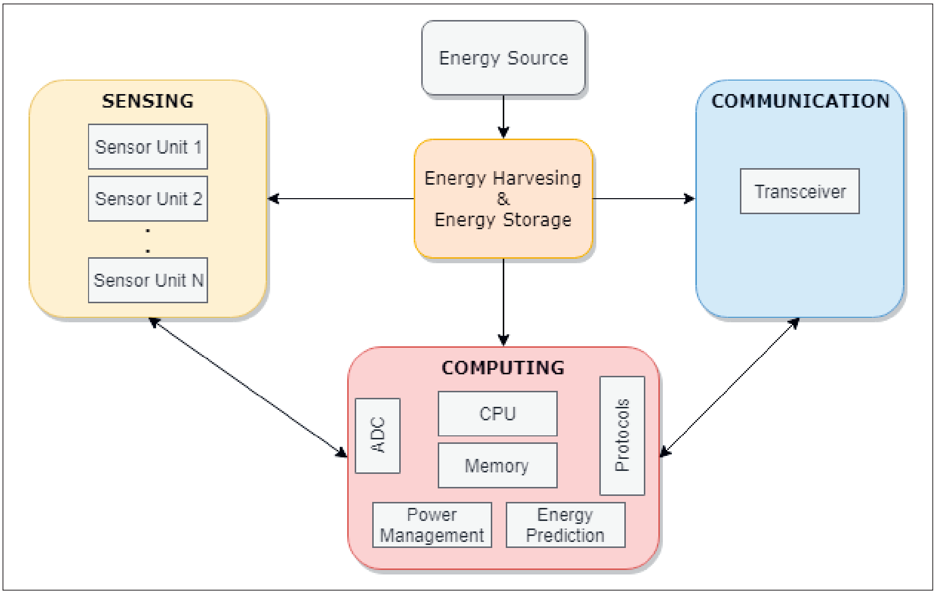


Figure 3: Signals are conditioned, sampled and digitized. Then, they are forwarded to the communication module for wireless transmission. Energy management module is focused on the energy optimization.

The existing AP low-cost monitoring nodes are either based on off-the-shelf components or are delivered as fully integrated plug-and-play solutions. The last ones are easier to integrate and have robust electro-mechanical design, but lack in flexibility (Fig. 4).

Continuous AP reporting based on low-cost sensors has several constrains. Most of them are related to sensors' performances. Precisely, sensors come with somewhat unknown and unpredictable settings. Prior to their integration into AP measurement system, they need to be evaluated in terms of accuracy, selectivity, sensitivity, and precision. The evaluation should be followed with the calibration process. Because of the aging effects, they should also be periodically rechecked on the aforementioned metrics and recalibrated.

The sensor performances vary from sensors to sensors. Also, some of them might show good performances regarding specific parameters, but they

Figure 4: Libelium Air Quality station [6] — an example of integrated low-cost sensor for AP monitoring. This module also enables for the online calibration by using artificial intelligence.

may be not as accurate to measure other parameters. For instance, for NO_2, data acquired from Alphasense [7] show high R^2 correlation with data obtained from standard instruments. On the other hand, the correlation is not satisfying regarding the O_3 parameter.

Low-cost sensors for AP monitoring are highly sensitive to the operating conditions and impose aging drift in time. Their readings are influenced by wind, temperature, humidity, etc. They are also influenced by external or internal noise, such as mobility and low sensitivity, respectively.

The electromechanical gas sensors consume a considerable amount of energy. As compared to the humidity or temperature sensors, they are much greater energy consumer. This questions the nodes' autonomy in some applications. The energy-saving schemes often include lowering the duty-cycle. However, turning off the gas sensor for a longer time is not suitable, because they need some time to heat-up before they become operable again.

Regardless of the aforementioned limitations, a study performed in [8] shows that, if the reason of using low-cost sensors is not to measure the absolute concentration of AP values, but to indicate the quality of the atmospheric environment through different health impact levels (such as AQI), then low-cost sensor devices may successfully fit this purpose.

The radio-characteristics of a wireless communication module has direct impact to the network coverage and topology, but also directly influence the node's lifetime. For instance, as compared to ZigBee technology, 3G/4G modules can provide much broader coverage, but they consume much more energy.

In contrast to gas sensors and wireless transceivers, modern microcontrollers have very low power consumption. Some of the open-source electronic prototyping platforms integrate high performance low-power microcontrollers (ATmega, PIC, MSP430 etc) to provide design's flexibility and modularity.

Energy management module manages power consumption and power supply of a node. Software-triggered routines enable dynamic duty-cycling as well as dynamic sampling and transmission rate adjustment. Also, when communication is achieved via multi-hop transmissions, energy-aware MAC (Media Access Control) and routing algorithms are often required to minimize power consumption. Finally, energy management modules may encompass some energy-harvesting techniques, such photovoltaic, wind, vibration, etc.

2.4 Wireless Transmission

With power-autonomous nodes, the design of a data dissemination topology and the selection of a most suitable wireless technology is of a crucial importance. Precisely, without considering other (heavy) consumers such as electrochemical gas sensors, a wireless transceiver typically consumes around 70% of the node's energy. The choice of wireless technology and network topology impact the optimal balance between the network coverage, energy efficiency, and network performance. For instance, an „energy-efficient" wireless routing protocol aims to route via the most energy-efficient path (instead via the shortest path). These protocols might impose long end-to-end delays or may fail to find a route [9].

In most of the emerging AP monitoring applications, smart sensors can be power-supplied from the existing power distribution systems, i. e., AP readings can be sent from the positions where some kind of stabile energy source is available. For instance, nodes can be installed on traffic lights, the roof of the buildings, vehicles, etc.

If energy consumption is not a system's limitation, AP systems may use some of the long-range technologies (3G, 4G, GPRS, etc.) to transfer data to the data center. If nodes are energy-constrained, there are two near-optimal sub-scenarios. If a local gateway can be supplied from the power distribution system, then data are locally transferred (to the gateway) via some low-power short-range technology (such as Zigbee, Bluetooth, 6LoWPAN, etc.). The gateway may then use some WAN technology (e. g., 4G, or even cable modem) to transfer data to the data center. If the power distribution system is far from the points of installation, some of the point-to-point Low Power WAN technologies (such as LoRa, NB-IoT, or Sigfox) may be used.

Wireless infrastructure for AP data dissemination can be based on: Static Sensor Networks (SSNs) [10,11], Community Sensor Networks (CSNs) [12,13], Vehicle Sensor Networks (VSNs) [14,15], or combination of the aforementioned [16]. A comprehensive review and comparison of the wireless architectures and topologies for AP monitoring is presented in [17].

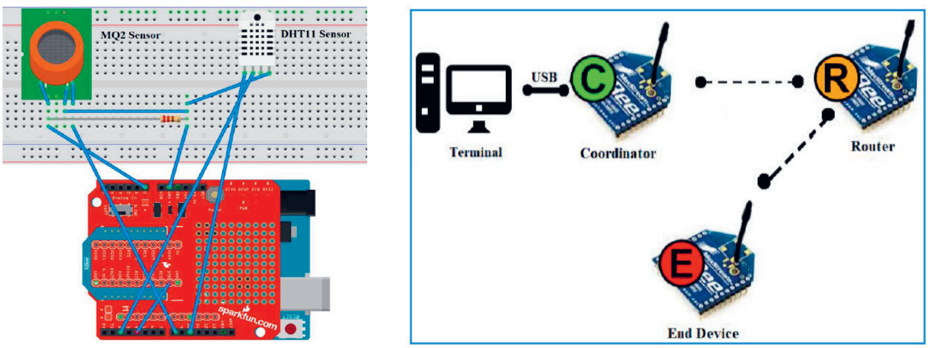
3. EXAMPLES — WSNS FOR AP MONITORING

In the proceeding subsections, two experimental studies on the implementation of WSNs for AP monitoring are shortly presented. The first one was designed for indoor implementation, while the second one has been implemented in an urban area.

3.1 A Multi-Hop Indoor AP Visualization System

According to the Environmental Protection Agency (EPA), indoors, air may be 2–5 times more polluted than outdoors. An indoor off-the-shelf system for AP monitoring, developed in UBT laboratory, is presented in [18]. The data acquisition module is based on Arduino Uno platform, which utilizes Atmel ATmega328P microcontroller and interfaces a MQ2 gas sensor, DHT temperature/humidity sensor, and a ZigBee transceiver (Fig. 5a). In order to cover a larger indoor area, the data dissemination is performed in multi-hop manner (Fig. 5b).

Although the proposed system provides a satisfying accuracy for the measurement of LPG, CO, and smoke, it shows some of the general aforementioned weaknesses related to the implementation of low-cost gas sensors. For instance, current drain in active mode is ≈140mA. With a specific duty cycle, system can achieve power autonomy of a few weeks. The system should be thoroughly tested and improved in accuracy.



(a) (b)

Figure 5: A system for indoor air quality monitoring.
(a) Sensor node, (b) Multi-hop wireless network.

3.2 A VSN-based AP Monitoring System

A number of VSN-based systems for AQM have been proposed in literature. They are either implemented as standalone VSNs or combined with SSNs. Different vehicles have been used to carry nodes, and different technologies have been used to sample, process, and transmit data. Some examples are given in [19, 20].

An experimental study has been conducted in the area of Prizren, Kosovo [21]. The MQ gas, temperature and humidity sensors are interfaced to the Arduino Uno board and are attached on roof of the Taxi vehicles (Fig. 6). Data are sent to the server via GPRS technology and are stamped in space and time with GPS. The system combines SSNs and VSNs to collect data for real-time visualization, historical view, and further analysis. The nodes are powered-up from the vehicle's power supply system. Hence, the system is not power constrained. Micro SD card module is installed on the board to store samples when the GPRS connection is unavailable.

A server receives data and stores them in a database for further web-based visualization and data analysis. Google API and JQuerry library are used for map and value entries visualization (Fig. 7a). The interface also provides the preview of the historical data (Fig. 7b)

The system shows satisfying performances in terms of overall functionality and continuity, with a small number of transmission errors. It provides

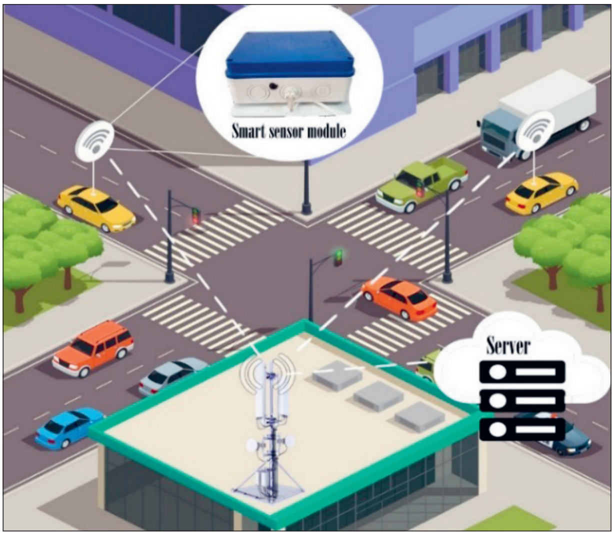


Figure 6: Sensor are installed at the roof of the taxi vehicles and are power supplied from the vehicles' battery. The readings are combined with those obtained from a conventional AP monitoring station in the region of Prizren, Kosovo.

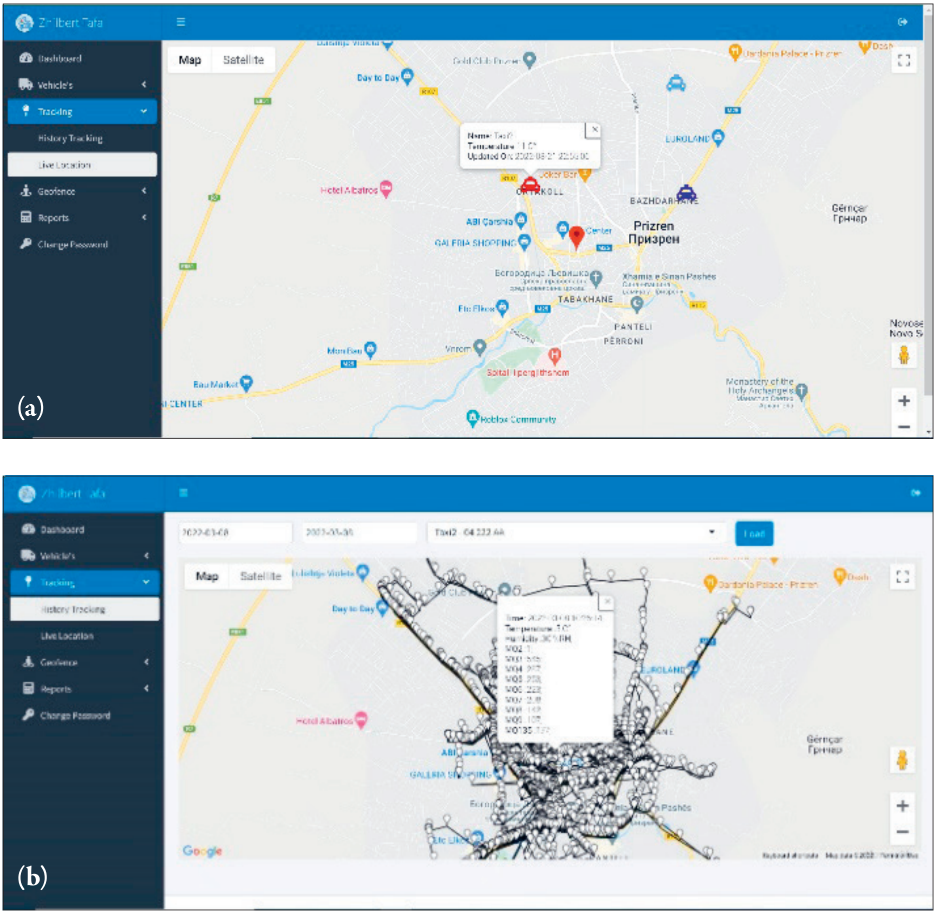


Figure 7: (a) User-friendly interface for real-time AP visualization, (b) Historical view

a satisfying spatiotemporal resolution for the region of Prizren, of an area of cca. 640 km^2. Future work will cover the implementation of ML for accuracy improvement and AP forecasting.

4. ML ALGORITHMS IN EMERGING AP MONITORING SYSTEMS

Calibration consists of setting a mathematical model to describe the relationship between low-cost sensor data and reference measurements [22]. Recently, calibration has been mostly performed by using statistical and ML approaches, such as multi-linear regression, Support Vector Machine (SVM), Artificial Neural Networks (ANN) etc. [23,24]. The ML techniques take into account the influences of the temperature, humidity, air pressure, solar radiation etc. on the deviations of readings.

Machine learning techniques have been also widely investigated for anomaly detection [25], AP estimation [26], and AP prediction [27, 28]. The follow-up of the study presented in [21] will include the aforementioned extensions. Future work will also combine sparse data acquired from fixed and mobile nodes for ML-based estimation of the AP for every point on the map.

5. CONCLUSION

WSNs have a potential to fill the gap between AP data acquisition requirements and the respective economically feasible technological solutions. Although some issues (as presented in this article) have slowed down the wider implementation of these technologies, recent extensive research show that WSNs can successfully improve spatiotemporal sensing resolution at a relatively low cost of implementation. Moreover, ongoing research show that implementation of the ML techniques can greatly contribute to the improvement and optimization of the WSN-based AP monitoring and management information systems.

REFERENCES

[1] V. Evagelopolous et. al, „Cloud-Based Decision Support System for Air Quality Management," *Climate,* 10(3), 39, 2022.

[2] G. Gualtieri et al., „An integrated low-cost traffic and air pollution monitoring platform to assess vehicles' air quality impact in urban areas, „*Transportation Res. Proc.*, vol. 27, pp. 609–606, 2017.

[3] A. M. Popoola et al., „Use of networks of low-cost air quality sensors to quantify air quality in urban settings," *Atmospheric Environment*, vol. 194, pp. 58–70, 2018.

[4] Z. Tafa, „Ubiquitous Sensor Networks," in *App. and Multidisciplinary Aspects of Wireless Sensor Networks*, L. Gavrilovska, et al., Eds., ed: *Springer London*, 2011, pp. 267–268.

[5] Online: https://www.horiba.com/pol/scientific/products/detail/action/show/Product/aqms-1560/ (accessed on Sept, 2022)

[6] Online: https://www.libelium.com/iot-products/air-quality-station/ (accessed on Sept, 2022)

[7] Online: https://www.alphasense.com/ (accessed on Sept, 2022)

[8] G. C. Spyropoulos, P. T. Nastos, and K. P. Moustris, „Performance of Low-Cost Sensors for Air Pollution Measurements in Urban Environments. Accuracy Evaluation Applying the Air Quality Index (AQI)," *Atmosphere*, 12(10), 1246, 2021.

[9] Z. Tafa, „WSNs in environmental monitoring: Data acquisition and dissemination aspects.“ *Advances in Computers*, vol. 126, 2022.
[10] D. Yaswanth & S. Umar, „A study on pollution monitoring system in wireless sensor networks,“ *International Journal of Computer Science & Engineering Technology*, vol. 3, pp. 324–328, 2013.
[11] P. J. A. John, „Wireless Air Quality and Emission Monitoring,“ *Master Thesis*, Uppsala University, 2016.
[12] Available online: https://airqualityegg.com/home (accessed on Sept, 2022)
[13] M. Nyarku et al. „Mobile phones as monitors of personal exposure to air pollution: Is this the future?“ *Plos ONE*, vol. 13, no. 2, pp. 1–18, 2018.
[14] A. Anjomshoaa, B. Duarte, D. Rennings, T. Matarazzo, P. de Souza, and C. Ratti, „City Scanner: building and scheduling a mobile sensing platform for smart city services, „*IEEE Internet of things journal*, pp. 2327–4662, 2018.
[15] S. Kaivonen & E. C.-H. Ngai, „Real-time air pollution monitoring with sensors on city bus,“ *Digital Communications and Networks,* vol. 6, iss. 1, pp. 23–30, 2020.
[16] S. Kaivonen and E. C. H. Ngai,“ Real-Time Air Pollution Monitoring with Sensors on City Bus,“ *Digital Communications and Networks*, vol. 6, pp. 23–30, 2020.
[17] W. Y. Yi, K..M. Lo, T. Mak, K. S. Leung, Y. Leung, and M. L. Meng, „A Survey of wireless sensor network based air pollution monitoring systems,“ *Sensors*, vol. 15, no. 12, 2015.
[18] Z. Tafa &F. Cakolli, „Design of a Multi-Hop Wireless Network to Continuous Indoor Air Quality Monitoring.“ *Journal of Communications* 14.5 (2019).
[19] A. Anjomshoaa, B. Duarte, D. Rennings, T. Matarazzo, P. de Souza, C. Ratti, „City Scanner: Building and Scheduling a Mobile Sensing Platform for Smart City Services,“ *IEEE IoT Journal*, pp. 2327–4662, 2018.
[20] O. Saukh, D. Hasenfratz, A. Noori, T. Ulrich, and L. Thiele, „Demo Abstract: Route selection of mobile sensors for air quality monitoring,“ *in: 9th European Conference on Wireless Sensor Networks*, pp. 10–11, 2012.
[21] B. Zherka & Z. Tafa, „A Vehicle Sensor Network for Real-Time Air Pollution Analysis,“ *Accepted, to appear in Journal of Advanced Information Technology*, 2022.
[22] F. Karagulian, M. Gerboles, M. Barbiere, A. Kotsev, F. Lagler, and A. Borowiak, „Review of sensors for air quality monitoring,“ EUR 29826 EN, Publications Office of the European Union, Luxembourg, 2019.
[23] C. -C. Chen *et al.*, „Calibration of Low-Cost Particle Sensors by Using Machine-Learning Method,“ *2018 IEEE Asia Pacific Conf. on Circuits and Systems*, 2018, pp. 111–114.
[24] D. Topalovic et al, „In Search of an Optimal Calibration Method of Low-Cost Gas Sensors for Ambient Air Pollutants: Comparison of Linear, Multilinear and Artificial Neural Network Approaches,“ *Atmospheric Environment,* vol. 213, pp. 640–658, 2019.
[25] D. -D. Dajic & N. -R. Grigoric, „Reliable Low-Cost Air Quality Monitoring Using Off-The-Shelf Sensors and Statistical Calibration,“ Elektronika ir Elektrotechnika, vol. 26, no. 2, pp. 32–41, 2020.
[26] K. Hu, A. Rahman, H. Bhrugubanda and V. Sivaraman, „HazeEst: Machine Learning Based Metropolitan Air Pollution Estimation from Fixed and Mobile Sensors,“ in *IEEE Sensors Journal*, vol. 17, no. 11, pp. 3517–3525, 2017.

[27] I. U. Samee, M. T. Jilani and H. G. A. Wahab, „An Application of IoT and Machine Learning to Air Pollution Monitoring in Smart Cities," *2019 4th International Conference on Emerging Trends in Engineering, Sciences and Technology (ICEEST)*, 2019, pp. 1–6.
[28] K. Kumar & B. -P. Pande, „Air Pollution Prediction Eith Machine Learning: A Case Study of Indian Cities," *International Journal of Environmental Science and Technology,* 2022.

Zhilbert TAFA

ZAGAĐENJE VAZDUHA: NADGLEDANJE I UPRAVLJANJE PRIMJENOM BEŽIČNIH TEHNOLOGIJA

Sažetak

Zagađenje vazduha je jedan od glavnih uzročnika raka pluća i moždanog udara. Ono se povezuje i sa drugim bolestima kao što su kardiološke, oftalmološke, psihijatrijske itd. Da bi se smanjio negativan uticaj na zdravlje, zagađenje se mora pratiti. Konvencionalni sistemi za praćenje kvaliteta vazduha su skupi zbog čega se instaliraju u relativno malom broju. Kao takvi, oni ne pružaju željenu vremensko-prostornu rezoluciju mjerenja. U ovom radu se predstavljaju bežične tehnologije za praćenje zagađenja vazduha u realnom vremenu. Osim infrastrukture, u radu su predstavljene i neke mogućnosti korišćenja mašinskog učenja u nadgledanju i upravljanju zagađenjem vazduha.

Ilir ÇAPUNI*

A RELIABLE TURING MACHINE — A DIGEST

Abstract. Reliability of computation concerns itself with the computation using a computing machine that is subjected to some noise. In this paper we survey the key elements of the construction of a reliable Turing machine which, with some moderate overhead, can perform arbitrarily large computations even when its operation is subjected to noise.

Since the construction is surprisingly complex, the outline presented here is suitable as an introduction for those who want to go to the depths of the construction and a substitute for those who want to get a low resolution picture of the construction with enough details to understand the core principles of it.

Key words and phrases. Turing machine, reliability, faults, errors.

1. Introduction

This is an exposition paper for a complex hierarchical construction of the Turing machine that, with some modest overhead can simulate any other Turing machine G, even when it is subjected to some noise which causes errors of the head that occur independently of each other with small probability. This construction was first published by the author in [1] and then in [2].

Historically the first such construction of a one-dimensional array of cellular automata that can perform arbitrarily large computations even though at each step each automaton makes an error independently of each other with some small probability is given in [3]. Clearly, in the case of the Turing machine faults do occur only where the head and information far from it does not decay spontaneously. However, even if only with small probability, occasionally a group of faults can put the head into the middle of a large segment of the tape rewritten before in an arbitrarily "malicious" way.

These constructions produce an infinite hierarchy of systems in which each layer simulates the next layer, which in turn has the same code as the previous layer.

An intuitive recipe of such constructions is as follows: We first construct a Turing machine M_1 that can withstand isolated bursts of errors of size β that

* Ilir Çapuni, chair, Department of Computer and Applied Sciences, Barleti University, Tirana, Albania

are followed by an error-free time period of at least V beta steps. Clearly, for this construction, we need to add some redundancy in space and computation, and organize the simulation in a way that even after the burst, the state of the simulation can be easily restored. For the general probabilistic noise, where errors occur independently of each other with some probability ϵ, this program needs the core modification which consists on "forcing" that the machine, instead of simulating the given machine G it actually simulates itself (by writing its code on the tape and simulating it). This will create an array of Turing machines M_1, M_2, $\dots M_k, \dots$ in which M_1 simulates machine M_2 and can withstand bursts of size β_1 separated by at least V_1 error-free steps (aka level 1 noise). Machine M_2 simulates machine M_3 and can withstand larger bursts of size β_2 and are separated by at least V_2 error-free steps (aka level 2 noise), and so on, machine M_k simulates machine M_{k+1} and can withstand large burst of size β_k separated by V_k error-free steps (aka level k noise). Each of these machines, writes its own code on the tape, and the program of M_1 is hardwired and cannot be corrupted by noise.

This is rather a simplistic description of the skeleton of the construction, and unlike in one-dimensional array of cellular automata where each cell is active, the simulation would work "right away". In the case of Turing machine model of computation, there are many problems that need to be addressed.

Recall, a Turing machine consists of an infinite array of constant sized cells that we call the tape and the head which is positioned over such a cell. One computational steps of a Turing machine consists of the following actions:

(1) Read the content a of the current cell and the content s of the registers of the head.
(2) Using the transition function (aka the program), infer the next state s', the new content a' of the current cell, and the direction $d \in \{-1, +1\}$ where the head should move.
(3) Write a' on the current cell, update the state to s', and move the head in the direction d.

The head of the machine M_1 is a part of the hardware. However, the heads of all other machines M_2, $M_3, \dots$ are actually "virtual" and not in hardware. They are defined in the information stored on the tape of the machine it is simulated by and henceforth, a large burst of errors can wipe out the information that defines the head of the simulated machine. Clearly, loosing a head can never occur at the first level (because the head is the part of the hardware). But this actually calls for modifications of the program of the machine M_1 to deal with this possibility and even for generalizing the definition of the Turing machine to allow for the head to vanish and appear. We will spell out other difficulties later in the sequel.

We will outline all the details of the construction and will also show the standard way of "switching" from bursts to probabilistic noise, and give a general structure of the proofs.

2. Intuitive explanations of the terminology

A standard definition of the Turing machine i

$$M = \langle \Gamma, \Sigma, \delta, q_0, F \rangle$$

where

(1) Γ is the finite, non-empty set of states;
(2) Σ is a finite, non-empty set of tape alphabet symbols;
(3) q_0 is the initial state;
(4) F is the set of final states;
(5) $\delta_M : (Q \setminus F) \times \Sigma \to Q \times \Sigma \times \{L, R\}$ is the transition function.

The set F of final states has the property that whenever M enters in a state in F, it can only continue from there to another state in F without changing the tape.

A *configuration* is a tuple

$$(q, h, x),$$

where $q \in \Gamma$ is the state, $h \in \mathbb{Z}$ is the position of the head, and $x \in \Sigma^{\mathbb{Z}}$ is the array of all the cells, that is, it is the tape. The content of the cell at position p is $x[p]$.

The work of the machine can be described as a sequence of configurations $C_0, C_1, \ldots$, where C_t is the configuration of the machine at time t. We are interested in a particular sequence of configurations that have the following properties:

(1) $q(0) = q_0$
(2) Nothing changes on the tape except possible where the head is: $x(t+1)[n] = x(t)[n]$ for all $n \neq h(t)$.
(3) On each step, the head does not jump for more than two positions: $h(t+1) - h(t) \in \{-1, 0, 1\}$.

We will refer to this sequence as a *history of* machine M.

If the transition from one configuration to the next one in a history is not obtained by applying the transition function, we say that *a fault* occurred at that time.

2.1. **Codes.** Let Σ_1 be the alphabet of M_1 and Σ_2 the alphabet of M_2. A *block code* is given by a positive integer Q – called the block size and a pair of two functions

$$\phi_* : \Sigma_2 \to \Sigma_1^Q, \phi^* : \Sigma_1^Q \to \Sigma_2$$

with the property

$$\phi^*(\phi_*(x)) = x$$

.

Function ϕ_* is the *encoding function* and we use it to encode one letter of M_2 to a Q letter word of M_1.

Function ϕ^* is the *decoding function* and we use it to decode a Q letter word of M_1 to the corresponding letter of M_2.

As an example, let us consider the *repetition code.* Suppose that $Q \geq 3\beta$ is divisible by 3, $\Sigma_2 = \Sigma_1^{Q/3}$, $\psi_*(x) = xxx$. If $y = y(1)\ldots y(Q)$, then $x = \psi^*(y)$ is defined by $x(i) = \mathbf{majority}(y(i), y(i+Q/3), y(i+2Q/3))$. For all $\beta \leq Q/3$ Clearly, this code can correct effects of a single burst of length β. If we repeat it 5 times, then we can correct 2 such bursts.

3. Statement of the theorem

Let us consider any machine G which at time t writes a value $y \neq *$ into the cell at position 0.

For any given Turing machine G, there exists some constants $\alpha_1, \alpha_2 > 0$ such that for any input size n there is a block code of size $O((\log n)^{\alpha_1})$ a fault bound ϵ and a Turing machine M_1 such that the following holds. Suppose that the machine M_1 starts working from the initial configuration (obtained by using the mentioned block encoding) with the head in the position 0. Suppose that M_1 runs and faults occur independently of the previous one with probability ϵ.

Then at any time greater than $t(\log t)^{\alpha_2 \log\log\log t}$, the tape symbol a of machine M_1 at position 0 will have the designated output field equal to y with probability at least $1 - O(\epsilon)$.

Remark 3.1. As we can see, the machine uses very moderate space: the block code is of size $O((\log n)^{\alpha_1})$.

The slowdown also is very moderate: if G halts in t steps, then M_1 will stop in $t \cdot g(t)$, where

$$g(t) = (\log t)^{\alpha_2 \log\log\log t},$$

which is a very slowly growing function.

Later, when defining the noise, we will see that this tower of machines M_1, M_2, ... will not be high.

4. Machine M_1

Let us consider two machines: M_2 and M_1. We want to define M_1 that simulates M_2 even when it is subjected to noise that cause burst of no more than β_1 consecutive faults to occur that are separated with at least V_1 error-free steps.

We will first give the bare bones construction of M_1 and then, in the subsections below, we will introduce difficulties and problems and upgrade our initial construction accordingly.

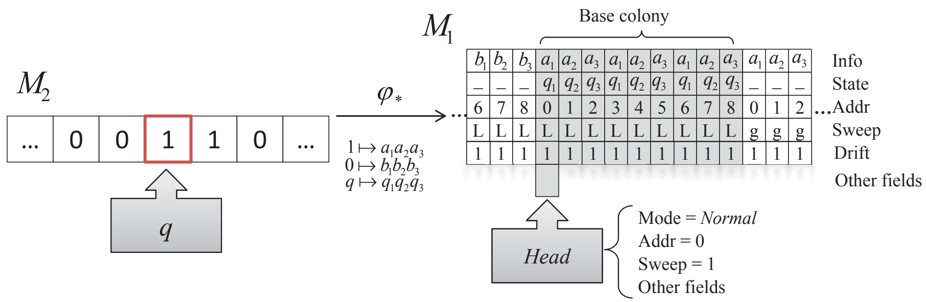


FIGURE 1. Encoding one cell of M_2 onto a colony of M_1. Each cell of M_1 has many fields

As depicted in Figure 1, we encode the content of a cell of machine M_2 using a repetition code onto Q consecutive cells of M_1. This "group" of consecutive cells that correspond to one cell of M_2 is called a *colony*. Each cell of a colony has a unique address ranging from 0 to $Q - 1$. The address is stored in the `Addr` field. The encoded information from M_2 is stored in the `Info` field.

The colony corresponding to the active cell is called the *base colony*.

To simulate one step of the machine M_2, machine M_1 first needs to sweep the base colony and compute the majority of the `Info` to infer the content of the cell of M_2 and its state. Then, it consults the transition function of M_2 and writes the results in a $\texttt{Hold}_1$ track. It does this three times, storing the results in $\texttt{Hold}_2$ and $\texttt{Hold}_3$ tracks. Then, it updates the `Info` field and the `Drift` field (where the direction of M_2 is stored) of each cell by computing the majority of the `Hold` fields of the cell. It repeats this two times.

To track the progress of the simulations, the M_1 contains the `Sweep` counter in its head. Similarly, to distinguish exactly the position of the head of M_1 it also needs to store the `Addr` of the cell where the head of M_1 is positioned. While doing these steps of the simulation, a single fault that changes these two registers will disturb the simulation. In order to restore the state of the simulation, we will write the `Sweep` and the `Addr` at each cell of the colony. Now, if a fault occurs, a specific healing procedure can reinstate the state of M_1 and move on with the simulation of machine M_2.

The last stage of the simulation of one step of M_2 is to transfer the head of M_2 in the colony determined by the `Drift` field: -1 for the left and 1 for the right colony. Recall that the state of M_2 is stored in the `Info` track of the base colony. That information needs to be "planted" in the state portion of `Info` track of the neighboring colony determined by `Drift`. This needs to be done again in a way that guarantees reliability. For this, the head will sweep `Info` track of the base colony and compute majority of the `State` and plant this information in the $\texttt{Hold}_1$ track of the designated neighboring colony. It repeats this another two times storing the result in $\texttt{Hold}_2$ and $\texttt{Hold}_3$ tracks. Then, state information at each cell will be computed by computing the majority of the `Hold` fields of each cell. This will be repeated two times. This completes the *transferring* stage and one work period of

$$U = O(Q)$$

steps.

At every single step of M_1 before the machine executes the above program, it first checks if the `Sweep` value on the current cell is one less than the value of the `Sweep` register in the head of the machine and that the `Addr` field on the head and on the cell match. If they match, after the corresponding part of the above program is executed, the M_1 updates the `Sweep` value on the current cell and moves the head. If they do not match, `Alarm` is called. This will initiate the healing procedure which we will explain later.

The time diagram of the head movement in time during one work period assuming no faults occur is depicted in Figure 2.

4.1. **What a single burst can do?** A single fault can change the content of the head (aka the state) of M_1. It can also change the content of the active cell. Therefore, after a burst, the state and the cells that were visited by the head during the burst have arbitrary content.

Intuitively, the entire simulation is devised in a way that it can be restored by doing local repairs of the parts of the colony and retrieving of the normal state of the machine before the burst occurred or before an island is encountered.

To see whether consistency – that is the basic tape pattern supporting simulation – is broken somewhere, a very local precaution will be taken in each step: each step will check whether the current cell-pair is allowed in a healthy configuration, which we will elaborate later.

If not then a *healing* procedure will be called; we will also say that *alarm* will be called. On the other hand, the *rebuilding* procedure will be called on some indications that healing fails.

4.2. **Zigging.** Assume that the head is in the middle of the designated right neighbor colony moving to the right during the transfer stage. A single fault occurs and it only increments the `Sweep` number in the head of M_1. The head now turns left because it thinks that it is performing the next sweep. From the cell where the fault occurred until the right end of the designated colony, we have wrong `Sweep` number on the tape. `Alarm` will be called only when the head returns to this area, possibly never.

Our goal is to put enough checks to prevent the machine to cause disproportionate damage from a short burst of faults or a single fault. We want our machine to return to the simulation as soon as possible.

To prevent this extensive damage to occur, we will introduce *zigging*: every Z steps forward are accompanied by Z steps backward and forward, where Z is a constant defined in the paper. It checks that `Addr` and `Sweep` registers of the head correspond to those on the tape. If not, `Alarm` is called.

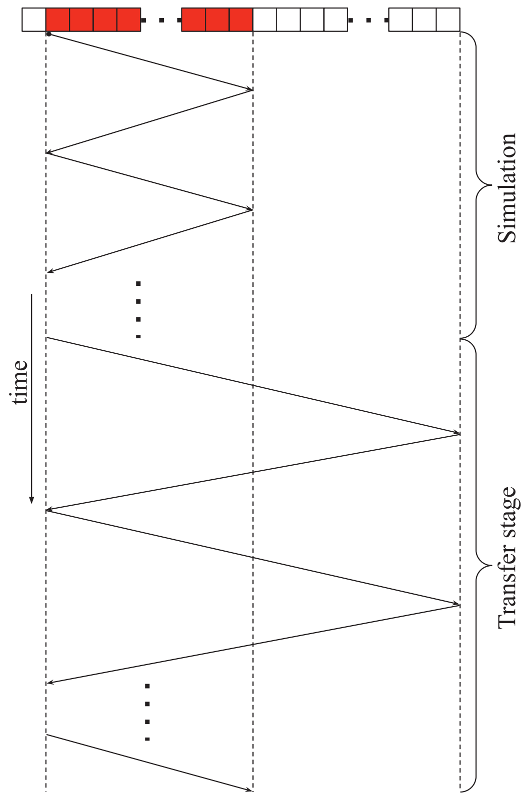


Figure 2. Illustration of the head movement in one work period

4.3. **Healing.** The healing procedures `Heal`, `RebuildHeal` and the rebuilding procedure `Rebuild` look as if we assumed no noise or disorder.

Healing performs only local repairs of the structure: for a given (locally) admissible configuration, it will attempt to compute a satisfying (locally) healthy configuration. If it fails—having encountered an inadmissible configuration—then the rebuilding procedure is called, which is designed to repair a larger interval.

Every healing operation starts with a survey zig around its starting point and marks the surveyed area appropriately. If the survey finds some possible healing to do then it performs one step of it, and returns. Otherwise the "attempt" *fails* and in this case, it will build a rebuilding an interval which is larger and is defined with the help of a new field special field.

Using majority, it heals the fields defining the state of the simulation (say `Addr`, `Sweep`, `Drift`, etc.), and not of the simulated machine.

4.4. **Problems caused by big bursts.** In the next two subsections we will consider some challenges that we need to deal with and are caused by big bursts.

4.4.1. *Conceptual issues.* Consider a big burst which have dislocated two neighboring colonies that correspond to two cells of M_2. The colonies are removed from each other and the space in between is filled with $< Q$ empty cells. Actually this means that two cells have some gaps in between. Clearly, this cannot happen according to our definition of the Turing machine and therefore we need to generalize the definition to allow for this possibility.

Consider a large burst of faults that spans multiple colonies.

By changing the content of the `Info` tracks, it can create more than one heads of M_2 in each colony, or erase the head completely. This is also something which we will need to address on a conceptual level.

A big burst can also erase completely the colony structure. From the point of view of the machine M_2, this means that a cell of M_2 can be completely removed. Again, the plain definition of the Turing machine does not allow this.

Finally, when two cells of M_2 are misaligned and have a gap in between, the need to remove a cell is accompanied with the need to create one.

We will take into account all these by adopting the notion of a *generalized Turing machine.*

4.4.2. *Entrapment.* Suppose that a large burst (spanning multiple colonies) occurs creating some intervals $I_1, I_2, \ldots, I_n$ consisting of several colonies of machine M_1, and each interval containing a simulated head whereas the neighboring intervals have no relations to each other with respect to simulation. A burst can send the head from one interval to the next one, and from the next one to the previous one forever causing head entrapment.

4.4.3. *What heals the effects of big bursts?* We reiterate that a big burst essentially may ruin intervals of cells spanning multiple colonies. Our healing procedure is very local and essentially spans a tiny fraction of a colony. The goal is to repair the simulation structure and carry on the simulation. Why is this so? Also, what heals these big intervals?

The answers to these two questions are related: A big island (of size β_2 – spanning multiple colonies of M_1) is essentially a small island of M_2. But M_2 is a simulated machine and it exists only on the tape as a part of the information saved in the colonies. For this reason we need to establish the simulation of M_2 by M_1 as soon as possible. If it is not possible, this will be achieved by the `Rebuild` procedure. Colonies corresponding to (possibly misaligned) cells of M_2 may be created. What part of M_2 is it simulating now over this new cell? Depending on the content of the colony, most probably M_2 is executing some healing procedure.

4.5. **Self-simulation and universality.** A tricky issue is "forced self-simulation". Each machine M_k can be implemented on a universal machine using as inputs the pair (p, k) where p is the common program and k is the level. Eventually, p will just be hard-wired into the definition of M_1, and therefore faults cannot corrupt it. While creating p for machine M_1, we want to make it simulate a machine M_2 that has the same program p. The method to achieve this has been applied already in some of the cellular automata and tiling papers cited, and is related to the proof of Kleene's fixed-point theorem (also called the recursion theorem).

Forced self-simulation can give rise to an infinite sequence of simulations, achieving the needed robustness. Let us point out that fixing the program of self-simulation does not prevent universality. A special track (which in the paper is called `Payload`) is set aside for simulating the given arbitrary machine G. If this simulation of G does not finish in a certain number of steps, a built-in mechanism will *lift* its tape content to the `Payload` field of the simulated cell-pair, allowing it to be continued in a colony-pair of the next level (with the corresponding higher reliability).

5. Noise

The main theorem of the paper talks about faults that occur independently of each other with probability ϵ whereas our treatise and construction are talks about big and small bursts. How can we "switch" from combinatorial noise (bursts) to probabilistic noise?

The set of faults in the noise model of the theorem is a set of points in time. It turns out more convenient to use an equivalent model: an ϵ-bounded *space-time* set of points. Let us make this statement more formal.

We want to deal with *bursts* (rectangles of space-time containing `Noise`) that are bigger in size and sparser in frequency. To derive such combinatorial constraints from the our probabilistic model we stratify `Noise` as follows.

We will have two series of parameters: $B_1 < B_2 < \cdots$ and $S_1 < S_2 < \cdots$, where B_k is the size of cells of M_k as represented on the tape of M_1, and S_k is a (somewhat increased) bound on the time needed to simulate one step of M_k.

For some constants $\beta, \gamma > 1$, a *burst* of noise of type (a, b) is a space-time set that is coverable by a rectangle of size $a \times b$. For an integer $k > 0$ it is of *level* k when it is of type $\beta(B_k, S_k)$. It is *isolated* if it is essentially alone in a rectangle of size $\gamma(B_{k+1} \times S_{k+1})$ First we remove such isolated bursts of level 1, then of level 2 from the remaining set, and so on. It is shown in the paper that with not too fast increasing sequences B_k, S_k, with probability 1, this infinite sequence of operations completely erases `Noise`: thus each fault belongs to a burst of "level" k for some k.

As stated before, machine M_k will concentrate only on correcting isolated bursts of level k and on restoring the framework allowing M_{k+1} to do its job. It can ignore the lower-level bursts and will need to work correctly only in the absence of higher-level bursts.

The rest of treatise on this matter is simple and is well treated in the paper and even some earlier publications on these topics. Arguments and statements given there allow a doubly exponentially increasing sequence U_k, resulting in relatively few simulation levels as a function of the computation time, which is important for establishing the time estimate of the time overhead given in the theorem: if the given machine G halts in t steps, then we can read its result from the designated cell (by construction) after

$$t(\log t)^{\alpha_2 \log\log\log t}$$

steps.

6. A roadmap of the proof

Recall that a big burst may cause cells of (simulated) machine M_2 to have space in between, or cells be dislocated, or a head to disappear or many of them to be created. For this reason and for the sake of proof, we need to define the generalized Turing machine. Then, we need to define the history of such a machine.

6.1. **Healthy configuration.** The definition of a healthy configuration is given by describing the conditions that a configuration must have. These conditions are given in axioms (H1) through (H5). To begin with, as we have explained before, we cannot count that the tape consists of a single interval of contiguous cells. No cell in these intervals should be marked by `Rebuild` marks and if removed from each other, they can contain a bridge of cells in between. In a healthy interval, the `Drift` marks on the cell always point to the front, that is a cell from where simulation can be carried out. Also, it is natural to require that all the `Drift` marks in the colonies in a healthy interval containing the head should point to the base colony. Other conditions establish the so called D-zone, an interval of $Z/2 \pm 1$ cells at the front or ahead of it which contains the head or is adjacent to it and contain zigging marks.

6.2. **Simulation. Healing.** A trajectory is a history in which transitions are done according to the transition function.

To validate the construction it is necessary to prove that the decoding map which is defined and used in the construction takes a trajectory to a trajectory.

Healthy configuration is an ultimate goal. We relax this definition with the notion of "almost healthy" or *admissible configuration*.

Informally, an admissible configuration may differ from a healthy one in a small number of intervals we will call "islands". Even a healthy configuration may contain some intervals called "stains": places in which the Info track differs from a codeword. These pose no obstacle to the simulation, and if they are small and few then will be eliminated by it, via the error-correcting code.

We introduce also the notion of *annotation* which interprets parts of a history, "covering up" small segments that are not quite healthy and leaves other parts uninterpreted.

A central and novel part of the argument is the *annotation game*. There are two players: the annotator and the range extender and they are not adversaries, but rather collaborators. A *range of the annotation* is a subset of $\mathbb{Z} \times [0, t)$. Essentially, the range extender is challenging the annotator to extend its reach. It is actually proven in the paper that the annotator can always respond to the challenge.

References

1. Capuni, I. (2013). A fault-tolerant Turing machine. Boston University.
2. Capuni, I., & Gacs, P. (2021). A reliable Turing machine. arXiv preprint arXiv:2112.02152.
3. Gacs, P. (1983, December). Reliable computation with cellular automata. In Proceedings of the fifteenth annual ACM symposium on Theory of computing (pp. 32-41).

Karolj SKALA,[*] Zorislav ŠOJAT[**]

A GLOBAL DISTRIBUTED SERVICE SYSTEM FOR NEW INDUSTRIAL REVOLUTION AND SOCIAL DEVELOPMENT

Abstract: The rapid development of the IoT and mobile devices without and with internet connection represent the most common area of application. The trend is that the most significant information processing around us is at the lowest possible level, directly connected to the physical environment and mostly directly controlling our immediate human environment. The most significant information processing around us takes place at the lowest possible computer level, directly connected to the physical environment and mostly directly controlling our immediate human environment. We find these „invisible" information processing devices in home entertainment systems, in traffic control systems, in industry and in industrial products. These devices, which are neither in the cloud nor at the mobile edge, but at the physical edge of computing, form the basis of the Dew Computing paradigm [1]. The benefits of seamlessly integrating devices into the Cloud — Fog — Dew computing hierarchy are enormous, for individuals, the public and industrial sectors, the scientific community and the commercial sector, by enhancing the physical and communicative, as well as the intellectual, immediate human environment. It is therefore imperative to explore the possibilities of Dew Computing through research, innovation and development, to solve the fundamental problems of integrating the Dew-Fog-Cloud hierarchy, with particular attention to the need for information processing (and not just data) and communication, and to demonstrate the feasibility and high effectiveness of the developed architecture in various areas of human endeavour through real-world implementations.

Dew computing differs from classical cloud and edge computing in that it brings devices closer to the end user and enables autonomous processing independent of the Internet, but is still able to interact with other devices to exchange information over the Internet. The difference is also expressed in terms of scalability, as edge and cloud providers

[*] Karolj Skala, Institute Ruđer Bošković, Zagreb, HR, Centre for Informatics and Computing; Member of Hungarian Academy of Science

[**] Zorislav Šojat, Institute Ruđer Bošković, Zagreb, HR, Centre for Informatics and Computing; Member of Croatian Academy of Technical Sciences

can provision (almost endless) resources, whereas scalability in dew computing has to be realised at the device level rather than at the server level. Such devices can be ubiquitous and pervasive devices in our environment, embedded processors, cyber-physical systems or other IoT devices.

The rapid development of IoT and mobile devices without and with internet connectivity is the most common application area. The trend is for the most important information processing around us to be at the lowest possible level, directly connected to the physical environment and mostly directly controlling our immediate human environment. These devices, which are not at the cloud or fog edge, nor at the mobile edge, are the foundation of the dew computing paradigm.

In the distributed cloud dew service architecture, a cloud server and many dew servers work together as a distributed application to provide controls or services.

Multiple clouds in symbiosis result in a cloud federation that connects the cloud service environments. Federated cloud services provide global Rainbow services based on the Dew server connecting directly to a user [4].

In this paper, three global grids are introduced: the energy grid, the information grid (www) and the service grid (Rainbow). These three grids will define the technical civilisation and infrastructure for artificial intelligence on Earth and shape the new industrial revolution in the future.

Key words: *Distributed Computer Services, Rainbow service, Dew-Computing, Fog-Computing, Cloud-Computing*

INTRODUCTION

Modern computing paradigms foster a vast community of participants from almost the entire spectrum of human endeavour. For computing and data processing there are individual computers, clusters, grids and finally the clouds. For pure data communication, there is the Internet and for human comprehensible information communication, there is, for example, the World Wide Web [5]. The rapid development of mobile handheld devices with high computing capacities and internet connections made it possible to „downgrade" certain parts of the clouds to so-called „thin clients". This led to the development of the Fog computing paradigm and the concepts of the „Internet of Things (IoT)" and the „Internet of Everything (IoE)".

However, most of the information processing around us takes place at the lowest possible computing level, directly connected to the physical environment and mostly directly controlling our immediate human environment. We find these „embedded" information processing devices everywhere (cars, air conditioners, vending machines, traffic controls...) and comprehensive products throughout industry. These devices, which are neither at the edge of the cloud nor at the mobile edge, but at the physical edge of computing, are the basis of the Dew Computing paradigm.

Horizontal scalability is an approach to capacity expansion based on connecting multiple hardware or software units so that they operate as a single logical distributed service unit. Horizontal scalability contrasts with vertical scalability, where capacity is increased by adding more resources and functional capabilities to the service environment. The vertical Dew-Fog-Cloud-Fog hierarchy represents the upward scaling of services, while horizontal scalability is often referred to as downward scaling, where functional capability is expanded.

The Dew computing paradigm focuses primarily on devices that control the physical environment, i. e. lighting, traffic control, heating, cooling, power distribution, etc., where human control of the environment must take precedence over, or at least be coordinated with, possible higher-level requirements without disrupting the immediate human environment. For this reason, Dew Computing has two basic concepts that do not exist in the rest of the hierarchy: Fog and Cloud: self-sufficiency and cooperation [2][4].

The Fog computing paradigm aims to enable the use of a cooperative multiplicity of end-user clients or user-facing edge devices (e. g., including the mobile edge) to greatly enhance and extend the available communication, computing and data resources, with particular attention to the spatial proximity of end-users and their specific goals [1].

The cloud computing paradigm aims to provide ubiquitous and convenient on-demand network access to a shared pool of configurable computing and data resources.

THE DEW COMPUTING

At the level of what is often referred to as the „Internet of Things (IoT)", there is a whole range of extremely heterogeneous devices, some of which are highly complex and whose full processing power often exceeds the capabilities of workstations and servers from the beginning of this millennium, to very simple devices connected via a network. However, the network movement inherent in the term IoT does not include the simplest specialised devices with extremely low processing power (such as processors with 32 bytes of memory and 1K of programme, e. g. in simple sensors, simple environmental control devices, etc.). It is important to realise that most of the information processing around us takes place at the lowest possible computer level, which is directly connected to the physical environment and mostly directly controls our immediate human environment. We find these „invisible" information-processing devices in the engine of our car, in the refrigerator, in the gas boiler, in air conditioners, vending machines, musical instruments, radio receivers, home entertainment systems, traffic control systems, theatres, lights and everywhere in industry. These devices, which are neither in the cloud nor at the mobile edge, but rather at the physical edge of computing, are the foundation of the tau computing paradigm. The main challenge is to enable the seamless integration of these devices with higher-end devices in the Fog and in the Cloud, leading to new development perspectives and new usage scenarios.

The fundamental Dew computing element is the Dew Dropslet, which consists of a self-organising, cooperative communication layer, an ontological „interpreter“ and individual physical Dew devices that contain all the necessary sensors, effectors and required algorithms to autonomously perform their task, as well as a physical communication layer. In other words, the individual physical microcontrollers may or may not produce information (i. e. ontological context messages), but must at least have the ability to receive „suggestions“ and transmit data. Of course, specific application solutions need to be developed to physically and informationally connect Dew devices and Dew drops in the Dew computing ecosystem [4].

DEW, FOG, CLOUD HIERARCHY SERVICE LAYERS

As far as the technological aspect is concerned, in computer science we have recently been able to observe a paradigmatic hierarchical system consisting essentially of three levels [4][5].

Let us start with the lowest level, the level of devices and appliances that are directly responsible for our physical environment and control everything from soil moisture to public lighting, from the temperature in the house to traffic lights — in recent years we have recognised the paradigm of Dew Computing [1]. In general, Dew Computing is responsible for the layer of devices that resides below the Internet boundary and is directly responsible for certain physical aspects of our environment. In Dew, devices can coordinate or be coordinated with each other, but they cannot be „ordered“ outside of the parameters set by the human user or a natural process. Therefore, Dew Droplets, the fundamental components of Dew, must follow two important principles that do not exist in the rest of the hierarchy — self-sufficiency and cooperation. At this level, information processing is crucial.

When the dew evaporates to slowly become a cloud, or the cloud touches the earth to spread the freshness of the dew over nature, the fog is in between. Fog computing is the hierarchical layer between the heterogeneity of dew devices, the complexity of cloud processing and the arbitrariness of human users.

The key areas that Fog computing needs to address are ergonomics, human-computer interaction, service delivery support, multi-service edge computing, adaptive application support, communication mobility and redirection, etc. Given Fog's central role as the primary link to humans, a consistent and extensible ontology would greatly enhance human-computer interaction capabilities by easily integrating with specific human language forms and enabling seamless two-way communication. Although devices at the Tau level must be directly accessible for manual control, it is easy to imagine that most of the monitoring, control and coordination of these devices could fall into the realm of the nebula.

The third level in the emerging computing hierarchy is the level of Clouds. Clouds are far away, somewhere, nobody knows where because it is irrelevant. Cloud computing is responsible for networking, monitoring, modelling and control (suggestive and/or directive), mass storage, extended use... There is High Performance Computing (HPC), High Throughput Computing (HTC), the evolving field of High Productivity Computing (HProC), the realm of artificial intelligence.

Against this backdrop, clouds are the most important glue of the global system. However, they are still primarily data-centric. This is easily seen in the scientific environment, where the use of results obtained by others can be extremely complicated due to inconsistent global data formats and the often complete lack of ontologically systematised metadata. This is also painfully evident in the enormous amount of special, specific and „proprietary" data formats, each of which is defined programmatically or even syntactically, but not ontologically and semantically.

THE RAINBOW

The other two overarching aspects of the activities that computer science must incorporate through its philosophy, human cooperation and technological development, are the humanistic and naturalistic aspects of its endeavours. It is true that computer science is still seen as a technical science. However, it is clear from the preceding discussion that developing a philosophy of computer science and basing future „naturo-humano-technological" developments on a full multidisciplinarity across the spectrum of sciences is critical to shaping the envisioned global information processing environment as a sustainable, human- and nature-friendly, self-organising global ecosystem. A first look at the computer architecture of such a future rainbow computing environment is presented in more detail.

The use of colours, as in healing, as symbols in philosophy, is ancient. Let us therefore take some symbolic meanings of colours to illuminate the all too complex field of computing from a distance. Note that each of these symbolic colours has a basic meaning and must of course cover a wide range

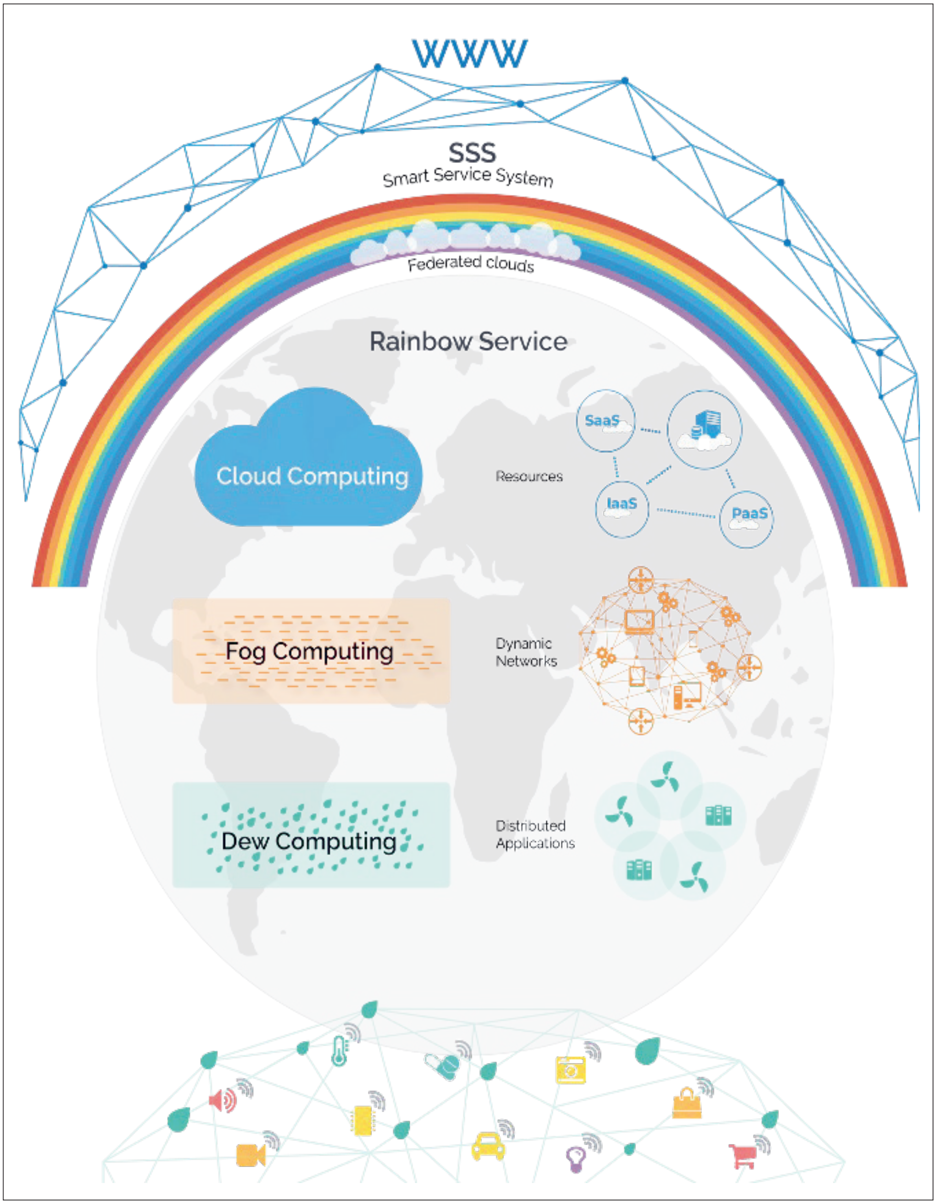


Fig 1. Vertical Distributed Rainbow Smart Service System

of elements. Like sunlight itself, all the symbolic rainbow colours are closely interwoven. This means that any single study or development of one of these colours must also take into account all the others.

THE MEANING OF RAINBOW COLORS REGARDING OF ICT

— **Red** — *Hardware: Architecture, memory, operations.*
— **Orange** — *Creativity: Stimulation, ideas, education.*
— **Green** — *Nature: Environment, health, well-being, backup systems, global ecosystem.*
— **Blue** — *Communication: Information, knowledge, human-computer interaction, languages.*
— **Indigo** — *Cooperation: Ethics, information use, redundancy, knowledge gathering and preservation.*
— **Violet** — *Interference: Security, limits of expansion, human-computer interference, interference processing.*,
— **Ultraviolet** — *Visions: Wisdom, prudence, conscience, responsibility, holism.*

THE GLOBAL SERVICE PROCESSING ENVIRONMENT

In the possibility of developing integrated home management/entertainment/maintenance systems, self-organising traffic-control systems, intelligent driver suggestion systems, coordinated building/car/traffic pollution control systems, real-time hospital systems with all patient and equipment status and control collaborating with the medical staff, fully consistent synesthetic artistic performances including artists and independent individuals („active public") from wide apart, power distribution peek filtering, self-reorganisation and mutual cooperation systems based on informed behavior of individual power consumption elements, etc., the Dew-Computing paradigm shows the way towards the Distributed Information Services Environment, and finally towards the present civilisation's aim of establishment of a Global Service Processing Environment, *Rainbow service* [5].

The *Rainbow service*, seamlessly connecting all computing hierarchy layers, will have to be established through horizontal self-organising coordination of individual Dew-droplets (i. e. the dew devices with communication/self-organisation possibilities) and vertical Fog/Cloud integration based on the newest developments of the Fog computing paradigm. The collaboration of Dew-droplets shall primarily be established on the level of their information exchange and self-organising coordination by (intra-) networking them, using proper ontologies/protocols and security/privacy and QoS based channels [3]. By having a common ontological basis on all hierarchical levels, i. e. throughout the „rainbow" of computing, and by developing common interaction protocols and information/instruction languages, such self-coordinating and collaborating systems of individual Dew devices will then, as micro services, be seamlessly integrated higher up into the computing hierarchy through Internet gateways into the Fog and Cloud, enabling the coordination of the physical edge control through common strategies and multi-level intelligent behavior, including adaptive-learning, predicting and other aspects of cognitive computing, with multi-modal user interfaces to facilitate appropriate Human/Computing interaction [3].

The merits of seamlessly integrating those „dew" devices into the Dew-Fog-Cloud computing hierarchy in form of *Rainbow service* are enormous, for individuals, the public and industrial sectors, the scientific community and the commercial sector, by bettering the physical and communication, as well as the intellectual, immediate human environment.

THE GLOBAL ENERGY, INFORMATION, AND FUTURE SERVICE GRID SYSTEM

The discovery of the three-phase energy system at the end of the 19th Century, the inventions of Nikola Tesla created the first Analogue Grid or Energy network on the globe. This grid provides the distribution of energy as desired and to every point of the globe. It started the second Industrial revolution, the beginnings of intensive development of technical civilization. In 1980, began developing the Worl Wide Web (www), which became the information global distribution digital network system on earth. Today through the www we are retrieving information globally. The global distribution network www system played a major role in the proven third industrial revolution. At the turning point of the millennium 2000, computer grids were announced as distributed, integrated computing network. This creates a distributed service system whose system architecture is addressed in this article, Fig1. Launching Dew computing services on the elemental service level with possible symbiosis and integration through Fog services (Dynamic Network Reconfigurable Services) to the Cloud computing service. Such virtual distribution Cloud services in form of federations can create Global Service Systems which in this paper called Global Rainbow Service. This global three-grid (energy, information, service) distributed system is the technological basis for the fifth Industrial revolution in which these three global Grids will be supported for the functioning of Artificial intelligence and advanced global cooperative systems (distributed robotics), in order to achieve the results and effectiveness of globalization and to establish highly utilized technical intelligence on earth. This

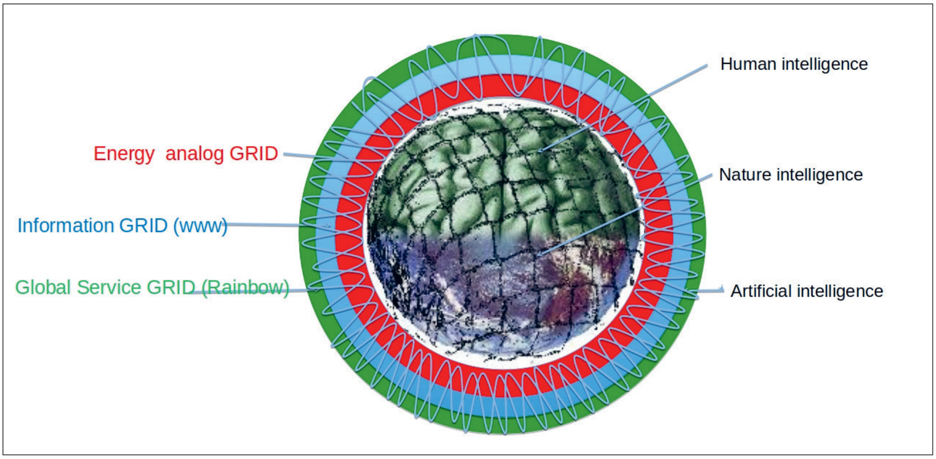


Fig 2. Artificial intelligence as an upgrade of Nature and Human Intelligence

Artificial intelligence is an upgrade of Nature and Human intelligence that creates new social environments, living echosystem and advanced perspectives on earth, Fig 2.

CONCLUSION

According to Dew, the Fog and Cloud computing extend the computing environment in a geographically distributed and hierarchical organization all over covering service technology.

The paper presents three global grids: the energy grid, the information grid (www) and the service grid (Rainbow). These three grids will define the technical civilization and infrastructure for Artificial intelligence on earth and mark the new Industrial revolution in the future.

The Dew-Computing Paradigm may well be the final missing ingredient to the computing development, transforming the all-pervading clusters, grids, clouds and fogs of computers into a human-helping Global Service Information and Processing Environment called *Rainbow Global Service* [5].

REFERENCE

[1] Skala, K., Davidović, D., Afgan, E., Sović, I., Šojat, Z.: *Scalable distributed computing hierarchy: cloud, fog and dew computing*. Open Journal of Cloud Computing (OJCC), 2 (1). pp. 16–24. ISSN 2199–1987, 2016.

[2] Y. Wang, „*Cloud-dew architecture,*“ International Journal of Cloud Computing, vol. 4, no. 3, pp. 199–210, 2015.

[3] Zorislav Šojat, Karolj Skala, „The dawn of Dew: Dew Computing for advanced living environment“, MIPRO 2017, pp. 375–380, DOI: 10.23919/MIPRO. 2017.7973447.
[4] Karolj Skala, Zorislav Šojat, „*The Rainbow Global Service Ecosystem*“, Proceedings of the 3rd International Workshop on Dew Computing (DewCom 2018), IEEE DewCom STC, Toronto, Canada, 2018, pp. 25–30.
[5] Zorislav Šojat, Karolj Skala, „*Views on the Role and Importance of Dew Computing in the Service and Control Technology*“, MIPRO 2016, pp. 175–179, DOI: 10.1109/MIPRO. 2016.7522131.
[6] Pooja Kukreja, Deepti Sharma: *A Detail Review on Cloud, Fog and Dew Computing*, International Journal of Science, Engineering and Technology Research (IJSETR), Volume 5, Issue 5, 1412, 2016.
[7] Afgan, P. Bangalore, K. Skala: *Application Information Services for Distributed Computing Environments,* Journal of Future Generation Compute Systems, Volume 27, Issue, 2, pp. 173–181, 2011.
[8] Pooja Kukreja, Deepti Sharma: „*A Detail Review on Cloud, Fog and Dew Computing*“, International Journal of Science, Engineering and Technology Research (IJSETR), Volume 5, Issue 5, 2016.

Veljko MILUTINOVIĆ,* Jakob SALOM**

SINERGIZOVANJE ČETIRI RAZLIČITE RAČUNARSKE PARADIGME ZA MAŠINSKO UČENJE I ANALITIKU VELIKIH PODATAKA

Sažetak: Ovaj članak predstavlja i analizira četiri računarske paradigme, koje su prisutne u današnjem svetu IT programiranja — Control Flow, Data Flow, Diffusion Flow i Energy Flow. Porede se njihova glavna svojstva, ukazuje na svrhu svakog od njih i opisuje koje su njihove prednosti i mane. U trećem delu ovog članka autori spekulišu o mogućoj arhitekturi superkompjutera na čipu, a u četvrtom predlažu optimalnu raspodelu resursa za određeni skup aplikacija u građevinarstvu.

Ključne reči: *Data Flow, Control Flow, Diffusion Flow, Energy Flow, Maxeler DFE, WSN, BioMolecular computing, QuantumMechanical computing, kompjuterska paradigma*

1. UVOD

Računarska scena danas uključuje četiri različite računarske paradigme i srodne modele programiranja. Neke od paradigmi i modela su u usponu, a druge su na stabilnim osnovama. Ove četiri paradigme su Control Flow (MultiCore — sa više jezgara kao kod Intela i ManyCore — sa puno jezgara kao kod NVidie), Data Flow (fiksni ASIC[1] — bazirani kao kod Google TPU[2] -a i fleksibilni FPA[3] — bazirani kao u početku sa Maxeler DFE[4], a u poslednje vreme i sa mnogim drugim), Diffusion Flow (npr. sa IoT-om

---

* Veljko Milutinović, Indiana University, Bloomington, USA, inostrani član Crnogorske akademije nauka i umjetnosti

** Jakob Salom, IPSI Ltd, Belgrade

[1] ASIC — Application-Specific Integrated Circuit

[2] TPU — Tensor Processing Units

[3] FPA — Programabilni hardver

[4] DFE — Data Flow Engine (mašina)

— internetom stvari i VSN-ima — bežičnim senzorskim mrežama) i Energy Flow — protok energije (kao kod biomolekularnog i kvantnomehaničkog računarstva). Za više detalja pogledajte reference [1...9]

Svaka od ovih paradigmi ima različite karakteristike u smislu (a) brzine, (b) snage, (c) veličine, (d) potencijala za visoku preciznost i lakoće programiranja. Svaka od njih je najpogodnija za određeni skup problema. Neke paradigme su pogodnije da služe kao host računari, a druge kao akceleratori. Naravno, najbolje ih je iskoristiti kroz odgovarajuću vrstu sinergije.

Ovaj članak prvo predstavlja prednosti i nedostatke svakog od njih, a zatim raspravlja o mogućim načinima za njihovu sinergiju.

2. POREĐENJE ČETIRI RAČUNARSKE PARADIGME

Control Flow paradigma zasniva se na istraživanju Fon Nojmana. Ona je pogodna za transakcijsko računarstvo i mogla bi se efikasno koristiti kao host u hibridnim mašinama, koje kombinuju sve gore pomenute paradigme. U slučaju kada se Control Flow MultiCore mašina koristi kao host, transakcioni kôd je najbolje pokrenuti na Control Flow hostu, dok se druge vrste problema najbolje rešavaju na akceleratorima zasnovanim na drugim tipovima paradigmi. U slučaju kada kôd radi na podacima organizovanim u 2D, 3D ili n-dimenzionalnim strukturama, dobar nivo ubrzanja može se postići pomoću Control Flowa ManyCore akceleratora. Model programiranja je relativno lak za razumevanje. Brzina, snaga, veličina i potencijal za visoku preciznost mašina Control Flow su dobro poznati.

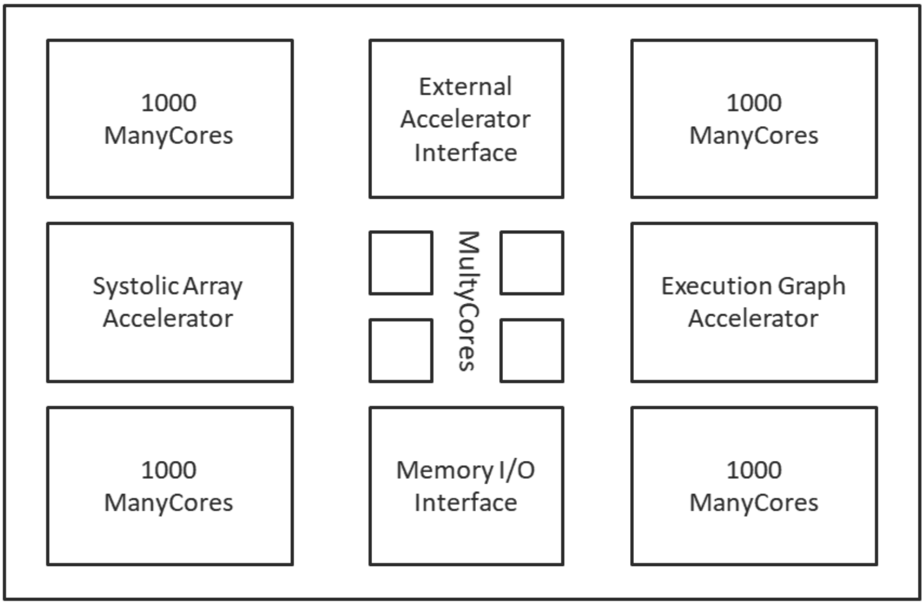
Data Flow paradigma inspirisana je istraživanjem Ričarda Fejnmana i drugih i insistira na činjenici da je računarstvo najefikasnije ako se podaci prenose, tokom računarskog procesa, na beskonačno male udaljenosti, kao u slučaju računarstva zasnovanog na grafu izvršenja (executon graph). U poređenju sa Control Flowom, ovaj pristup donosi ubrzanja, uštedu energije, manje dimenzije računarske opreme i veći potencijal za povećanu preciznost, ali koristi složeniji model programiranja, koji bi se mogao podići na više nivoe apstrakcije, u kom slučaju bi deo navedenih prednosti mogao nestati.

Diffusion Flow paradigma zasniva se na istraživanju masivnog paralelizma (IoT) moguće poboljšanog senzorskim mrežama (VSN). Jedna suštinska karakteristika ovog pristupa je velika oblast ili geografska pokrivenost, što znači da je teoretski nemoguće preneti podatke na male udaljenosti tokom računarske obrade. Ipak, neki nivo obrade je neophodan, možda u svrhu smanjenja podataka, ili neke vrste predobrade tokom „difuzije“ prikupljenih podataka prema hostu, za konačnu obradu tipa Big Data. Energetska efikasnost je visoka, dok je veličina zanemarljiva, kao i potencijal za najveću preciznost. S druge strane, model programiranja je evoluirao od početnog MIT PROTO pristupa i mora se pravilno savladati, što nije jednostavno.

Paradigma protoka energije Energy Flowa je namenjena samo za ubrzanje algoritama koji su najpogodniji za jednu od postojećih paradigmi. Bez obzira da li se koristi biomolekularni ili kvantnomehanički pristup, obrada se zasniva na energetskim transformacijama, a odgovarajući model programiranja mora da poštuje unutrašnju suštinu ako su žele najveće performanse. Za izvodljive algoritme ubrzanje je ogromno, potrebna snaga je minimalna, veličina je prihvatljiva, a potencijal za preciznost je nezamislivo veliki. Modeli programiranja su u porastu.

3. MOGUĆA ARHITEKTURA SUPERKOMPJUTERA NA ČIPU

U trenutnom stanju tehnologije, sa preko 100 milijardi tranzistora (BTr) na čipu, ili triliona tranzistora (TTr) na pločici (wafer), moguće je postaviti (na jedan čip) obe gore pomenute Control Flow mašine i obe gore pomenute Data Flow mašine. Međutim, mogući pojačivači (u obliku IoT-a ili VSN-a) i mogući akceleratori (u obliku biomolekularnog i/ili kvantnomehaničkog) moraju biti van čipa, ali lako dostupni preko odgovarajućih interfejsa.



Slika 1. Generička struktura budućeg superračunara na čipu
sa 100 milijardi tranzistora [10]

Naravno, memorija i klasični I/O moraju biti delimično na čipu, a delimično van čipa, povezivi odgovarajućim interfejsima.

Stoga, bez obzira da li se radi o 100 BTr ili 1 TTr strukturama, unutrašnja arhitektura na najvišem nivou apstrakcije treba da bude kao na Slici 1. Međutim, raspodela resursa bi mogla da se drastično razlikuje od jednog takvog čipa do drugog, zbog različitih zahteva aplikacija (orijentisani na transakcije ili orijentisani na „krckanje podataka“ — transactions-oriented or cranching-oriented), i zbog različitih zahteva za podacima (intenzivni zahtevi za memorijom zbog Big Data (masivnih podataka) statičkog tipa, ili intenzivni zahtevi za striming (prolazno) orijentisanih masivnih podataka dinamičkog tipa, koji dolaze i odlaze preko Interneta ili drugih protokola. Primeri koji slede pokrivaju simulacije problema velikih podataka (Big Data), gde je potrebno mašinsko učenje i odnose se na složene probleme u građevinarstvu ili srodnim oblastima:

(1) NBCE — građevinski inženjering zasnovan na prirodi;
(2) GNBE — genomska podrška za NBCE;
(3) EKIS — informacioni sistemi zemljotresa za predviđanje i alarm;
(4) NCEM — izrada novih građevinskih materijala za CO_2 i zemljotrese.

Algoritam koji se koristi u gore navedenim oblastima može biti:
(1) statistički i slučajni procesi koji oponašaju prirodu;
(2) NW ili SW[5] ili slično;
(3) PDE[6] tipa FE ili FD ili hibrid (FE = konačni element, FD = konačna razlika);
(4) tenzorski račun i matematička logika ili hibrid.

Za takav skup aplikacija pretpostavljamo da bi optimalna raspodela resursa bila kao u Tabeli 1.

4. NAPOMENA

U NBCE-u je bolje koristiti biološke strukture koje brzo rastu i naseljene su insektima koji stvaraju nanomaterijale, nego graditi betonske zidove, koji emituju CO_2 i koji su osetljivi na zemljotrese. Takođe, za zaštitu podvodnih objekata bolje je koristiti ribu i plankton nego metalne mreže. Pre svake investicije ove vrste, mora se uraditi studija izvodljivosti, zasnovana na simulaciji. Međutim, takve simulacije mogu biti dugotrajne i mogu trajati

[5] NW — Needleman Wunsch; SW — Smith Waterman
[6] PDE — Partial Differential Equasions — parcijalne diferencijalne jednačine

Tabela 1. Tip hardvera čipa i procenjeni broj tranzistora [10]

Tip Hardvera čipa	**Procenjeni broj tranzistora**
Jedan ManyCore sa memorijom	3,29 miliona
Jedan sistolni niz	< 1 milijarde
4000 ManyCore sa memorijom	11 800 miliona
Jedan vrhunski Data Flow, koji se može ponovno reprogramirati	< 69 milijardi
Jedan MultiCore sa memorijom	1 milijarda
Interfejs za I/O sa spoljnom memorijom	< 100 miliona
4 MultiCore sa memorijom	4 milijarde
Interfejs za spoljne akceleratore	< 100 miliona
TOTAL	< 100 milijardi

godinama. Rešenje je u prelasku sa Control Flowa na odgovarajuću kombinaciju ostale tri računarske paradigme.

U GNBE-u genetika vrsta i srodni procesi mogu u stvarnosti proći godine da se generišu željeni efekti za građevinarstvo. Međutim, kompjuterske simulacije na Control Flow mašinama, zasnovane na dovoljno detalja, mogle bi da potraju još više vremena. Dakle, rešenje je u odgovarajućoj sinergiji četiri paradigme.

U EKIS-u postoje modeli gradova zasnovani na cigli i cementu, ali simulacije zemljotresa sa ovim modelima kao ulazima mogu potrajati vek na najbržoj Control Flow mašini danas. Proces simulacije može se drastično ubrzati samo ako se koristi odgovarajući akcelerator protoka podataka. Pogodni su za PDE-ove tipa FE (Final Element), koji su potrebni za predviđanja i za PDE-ove tipa FD (Final Difference), koji su potrebni za alarmiranje u hitnim slučajevima.

U NCEM-u, novi materijali sa željenim svojstvima se najbolje nalaze ako se ML (Machine Learning — mašinsko učenje) algoritmi kombinuju sa klasičnim algoritmima, koji se koriste u istraživanju materijala. Ovakvi hibridni algoritmi su računarski intenzivni, pa je opet rešenje u sinergiji nekoliko paradigmi.

5. ZAKLJUČAK

Ovaj članak baca svetlo na potencijale koji dolaze iz sinergijskih interakcija četiri različite računarske paradigme.

Ovde primenjen pristup se razmatra u kontekstu građevinarstva, ali se lako može preneti u druge različite kontekste.

Test ovog tipa mogao bi se koristiti u obrazovne ili istraživačke svrhe u akademskoj zajednici i industriji.

Koncept koji se zagovara u ovom članku najbolje je primeniti na čipu koji uključuje neke od paradigmi koje se češće koriste i koji se efikasno povezuje sa drugim paradigmama koje se ređe koriste!

BIBLIOGRAFIJA

[1] J. L. Henning, „SPEC CPU2000: measuring CPU performance in the New Millennium," in *Computer*, vol. 33, no. 7, pp. 28–35, July 2000, DOI: 10.1109/2.869367.

[2] Mittal, Sparsh and Vetter, Jeffrey S, „A Survey of CPU-GPU Heterogeneous Computing Techniques", in *ACM Computing Surveys*, Volume 47, Issue 4, July 2015, Article No.: 69, pp. 1–35. https://doi.org/10.1145/2788396

[3] Kumar, Sameer et al., „Scale MLPerf-0.6 models on Google TPU-v3 Pods", in *Computer Science, Cornel University*, https://doi.org/10.48550/arXiv.1909.09756

[4] Wang, Yu Emma; Wei, Gu-Yeon and Brooksar, David, „Benchmarking TPU, GPU, and CPU Platforms for Deep Learning", in *Computer Science, Cornel University*, Oct 2019. https://doi.org/10.48550/arXiv.1907.10701

[5] Trifunovic, N.; Milutinovic, V.; Salom, J. et al., „Paradigm Shift in Big Data SuperComputing: DataFlow vs. ControlFlow", in *Journal of Big Data 2*, Springer, 4 (2015). https://doi.org/10.1186/s40537–014–0010-z

[6] Srivastava, Ankita and Mishra, Pramod Kumar, „A Survey on WSN Issues with its Heuristics and Meta-Heuristics Solutions", in *Wireless Personal Communication* 121, pp. 745–814 (2021). https://doi.org/10.1007/s11277–021–08659-x

[7] M. Centenaro; C. E. Costa; F. Granelli; C. Sacchi and L. Vangelista, „A Survey on Technologies, Standards and Open Challenges in Satellite IoT", in *IEEE Communications Surveys & Tutorials*, vol. 23, no. 3, pp. 1693–1720, 2021, DOI: 10.1109/COMST. 2021.3078433.

[8] Schlick T.; Portillo-Ledesma S.; Myers C. G. et al., „Biomolecular Modeling and Simulation: A Prospering Multidisciplinary Field", in *Annual Review of Biophysics*. 2021 May; 50: 267–301. DOI: 10.1146/annurev-biophys-091720–102019. PMID: 33606945; PMCID: PMC8105287.

[9] Baiardi, Alberto; Grimmel, Stephanie A.; Steiner, Miguel; et al., „Expansive Quantum Mechanical Exploration of Chemical Reaction Paths", *Laboratory of Physical Chemistry,* ETH Zurich, Acc. Chem. Res. 2022, 55, 1, 35–43 Publication Date: December 17, 2021. https://doi.org/10.1021/acs.accounts.1c00472[10]

[10] Milutinovic, Veljko; Azer, Erfan Sadeqi; Yoshimoto, Kristy et al., „The Ultimate DataFlow for Ultimate SuperComputers-on-a-Chip, for Scientific Computing, Geo Physics, Complex Mathematics, and Information Processing", 10th Mediterranean Conference on Embedded Computing (MECO), (2021), pp. 1–6, DOI: 10.1109/MECO52532.2021.9459725.

Vesna BENGIN*

THE MISSION OF THE BIOSENSE INSTITUTE: NEW TECHNOLOGIES IN AGRICULTURE

Abstract: Agriculture is facing enormous challenges. Not only does it have to provide enough safe food for the growing population, it also needs to leave a minimal environmental footprint and combat climate change. The available natural resources are limited and the conventional agricultural technologies do not provide sufficient means to address these challenges.

On the other hand, a number of key enabling technologies have matured in the recent decade, including nano and microelectronics, material science, remote sensing, communications, and artificial intelligence. However, although some of these technologies have already created a significant impact in various aspects of human life and wellbeing, they have not yet delivered their full impetus to the agrifood sector.

Agriculture needs to be optimized. More (higher quality) yields need to be produced with less inputs (water, fertilizers, labor, energy...) and with less risks (related to weather, pests, market conditions...). However, agriculture is a very complex biosystem, and its optimization presents a significant challenge as it requires full understanding of all underlying processes and their correlations. The first step in this process is to develop and deploy various sensors that will provide as much information as possible about the plant itself, the soil, atmospheric and meteorological conditions.

However, large amounts of sensing data *per se* do not offer necessary insights into complex processes in agriculture. To discover underlying interdependencies and to create actionable information from sensing data, multisensor data fusion, feature engineering, deep learning and big data analytics need to be deployed. As a results, sufficiently large sets of reliable data combined with AI algorithms have already proven their potential in optimizing agriculture.

Key words: *digital transformation, sensing technologies, artificial intelligence, agriculture*

---

* Vesna Bengin, BioSense Institute, University of Novi Sad

SENSING TECHNOLOGIES FOR AGRICULTURE

As the quality of any measure or action to be taken to optimize agriculture directly depends on the quality of input data, it is necessary to develop devices that will provide reliable, rapid and precise data of interest to agriculture, whether in-situ, proximally or remotely. However, no single technology or single operating frequency can provide the means to develop all needed sensors that will detect and measure various biological, physical and chemical properties and processes in the plant, the soil and its environment. To that end, a wide range of expertise needs to be combined, spanning from physics and biology to electrical engineering and materials science, with main research directions depicted in Fig. 1.

We develop molecular, optical, micro and nanofluidic, solid state, acoustic, photonic and nano and microelectronic sensing applications with the goal to enable sensing of previously unattainable parameters or provide solutions more accessible than the existing ones (e. g. faster, cheaper, smaller).

In addition, various remote sensing technologies, including thermal, microwave, terahertz, multispectral and hyperspectral imaging obtained from various sensing platform ranging from ground, through UAV to satellites,

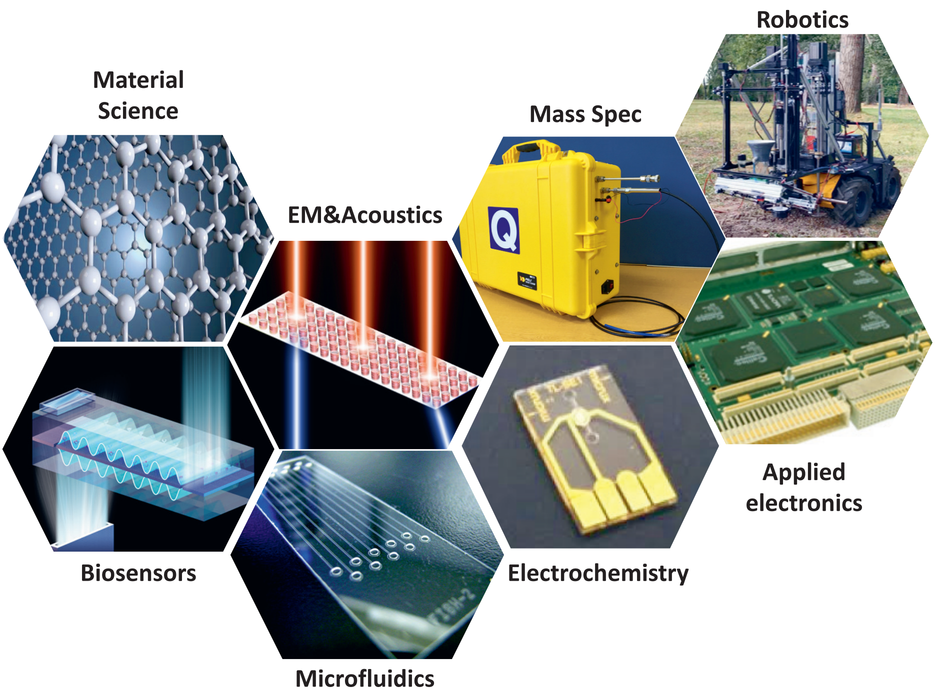


Figure 1. The core fields needed to develop necessary sensors for agriculture

enables us to discover qualitative and quantitative analytical data based on the interaction of electromagnetic radiation with living and non-living matter (e. g. NDVI, LAI, evapotranspiration, photosynthesis, etc.).

INFORMATION TECHNOLOGIES IN AGRICULTURE

To enable turning data into useful actionable information, research efforts across a number of directions in information technologies are needed that address challenges across the entire agri-food chain (from plant phenotyping, crop monitoring and yield prediction, to post harvest technologies, modelling, optimization and decision support). Fig. 2 depicts available data sources, technologies under development and final applications, all extremely relevant to agricultural applications.

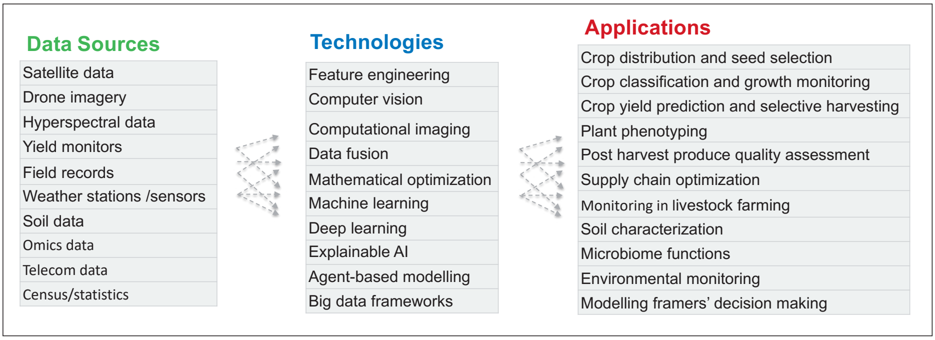


Figure 2. Interplay of data, information technologies and agricultural applications

For example, satellite technologies in combination with image denoising, CNNs and different algorithms such as Random Forest, are indispensable in application such as crop identification, Fig. 3, and crop monitoring, Fig. 4, including the detection of burnt farmland or irrigation optimization.

A number of challenges in agriculture require some sort of optimization, e. g. smart seed selection, optimsiation of seed distribution in retail, Fig. 5, vehicle routing problems etc.

The power of information technologies and the impact they have on the agricultural production can best be illustrated by the example of crop structure planning — one of the most common challenges in agricultural production. Namely, each year, a decision needs to be made on which crop to plant at which location to maximize profits, reduce yield and price risks, to obey to crop rotation, group crops to save fuel, and reduce the use of fertilizers and pesticides. On a farm of 6000 ha comprising of 70 fields grouped at 4 locations, where all five major crops are to be grown, the number of

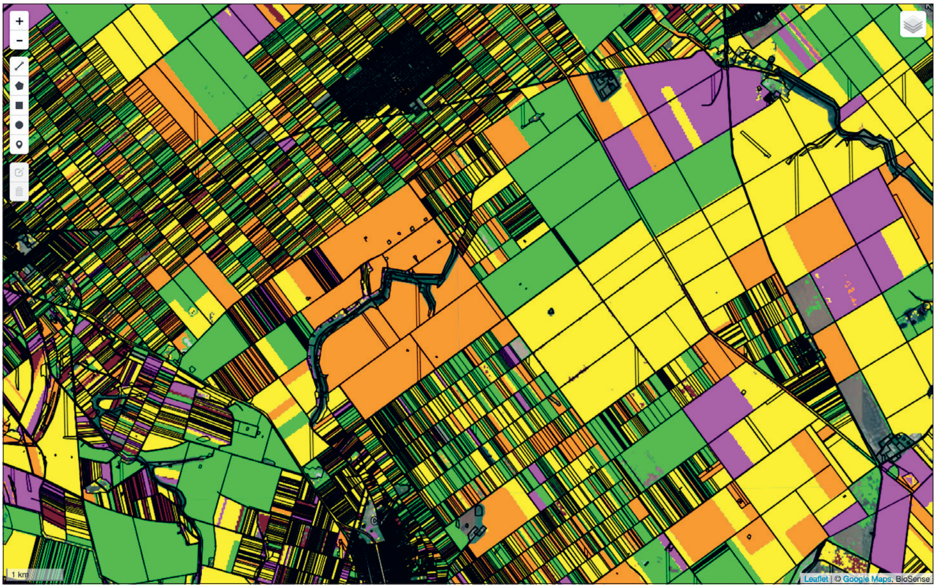


Figure 3. Satellite-based crop identification is West Balkans is burdened by very small size of plots (different colors denote different crop types identified with precision higher than 97%)

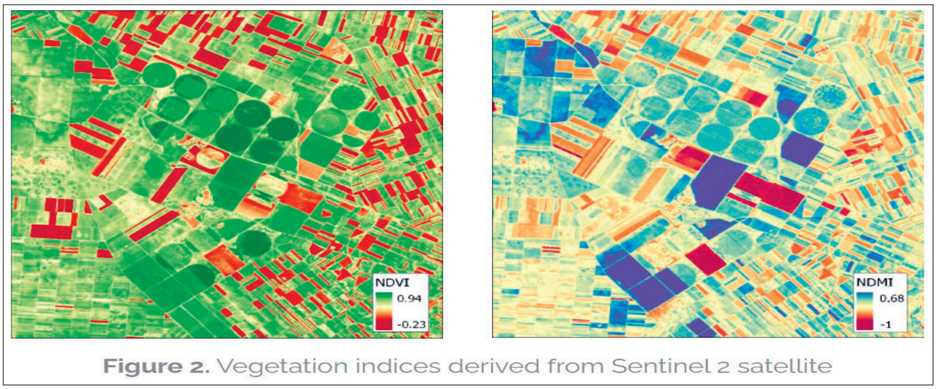


Figure 4. Satellite-based crop monitoring reveals health status of the plants (NDVI index).

crop combinations is around 10 quindecillion, far beyond the scope of any agriculture domain knowledge. By using AI-based crop structure planning solutions, with no additional investment whatsoever, increase of more than 60% in profits (with 10% reduction of compound risks) can be achieved, which, for a 6000 ha farm translates to additional profits of around 1.2 million € annually.

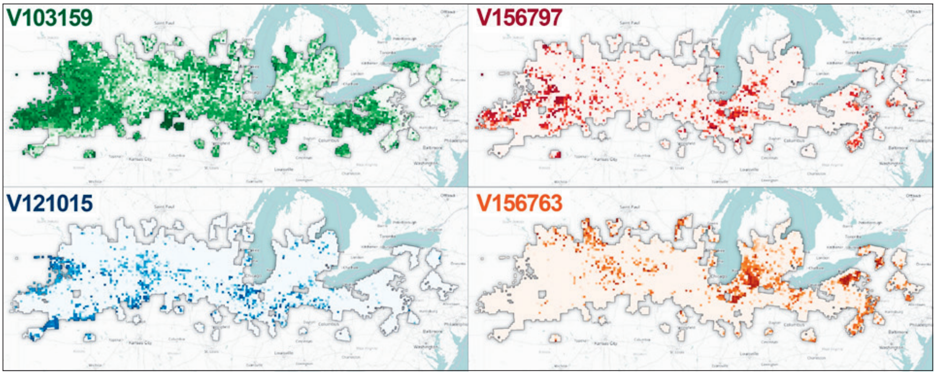


Figure 5. Optimal distribution of soy bean seeds in retail across the mid-west US

Vesna BENGIN

MISIJA INSTITUTA BIOSENS:
NOVE TEHNOLOGIJE U POLJOPRIVREDI

Sažetak

Poljoprivreda se suočava sa velikim izazovima. Ne samo da se od poljoprivrede očekuje da stvori dovoljno bezbedne hrane za rastuću svetsku populaciju, ona mora da ostavi minimalan ekološki otisak i izbori se sa uslovima klimatskih promena. Raspoloživi prirodni resursi su ograničeni i konvencionalne poljoprivredne tehnologije ne pružaju dovoljno mogućnosti za rešavanje ovih izazova.

Sa druge strane, u poslednjoj deceniji sazrele su brojne tehnologije, kao što su nano i mikroelektronika, nauka o materijalima, daljinska detekcija i veštačka inteligencija. Iako su neke od njih imale značajan uticaj na različite aspekte ljudskog života i blagostanja, ove tehnologije još uvek nisu ostvarile svoj puni potencijal u sektorima poljoprivrede i hrane.

Poljoprivredna proizvodnja se mora optimizovati, u smislu da je neohodno dostaviti veće i kvalitetnije prinose sa manje uloženih resursa (vode, đubriva, rada, energije...) i uz smanjenje rizika (vezanih za vremenske prilike, štetočine, tržišne uslove...). Međutim, poljoprivredna proizvodnja je u suštini veoma složen biosistem, a njena optimizacija predstavlja velik izazov, jer zahteva potpuno razumevanje svih fundamentalnih procesa i njihovih korelacija. Prvi korak u postupku optimizacije je stoga razvoj i upotreba različitih senzora koji će pružiti što više informacija o samoj biljci, zemljištu, atmosferskim i meteorološkim uslovima.

Ključne reči: digitalna transformacija, senzorske tehnologije, veštačka inteligencija, poljoprivreda

Mirjana KRANJAC[*]

STRATEGIJA SMART SPECIJALIZACIJE SRBIJE

Sažetak: Srbija je napravila svoju strategiju pametne specijalizacije (S4) za period od 2020. do 2027. godine. Ovaj dokument je osnova za razvoj sektora istraživanja i inovacija u evropskim zemljama i regijama. On se zasniva na podeli ovih aktivnosti prema specifičnostima regija i uvažavajući njihove komparativne prednosti. Stvaraju se regionalne niše istraživačkih delatnosti, koje teže da se ne preklapaju. Ovim se životni ciklus znanja i inovacija optimizuje, dajući veću dodatnu vrednost evropskom prostoru u celini. Nakon stvaranja ovakve strategije, neophodno je da se ona promoviše svim zainteresovanim stranama. Realizacija dokumenta mora neprekidno da se prati. On treba da se koriguje i unapređuje i u to uključi što više aktera. Da li je tako u Srbiji? Autori pokazuju da ovaj dokument nije dovoljno vidljiv, kao ni mnogi slični dokumenti. Nije prepoznata njegova vrednost niti mogućnost primene pri apliciranju za finansijska sredstva Evropske unije. U radu su date preporuke kako unaprediti životni ciklus S4 u cilju značajnih efekata za razvoj srpskog inovacionog ekosistema. Jedan od vidova daljeg razvoja S4 može biti izrada prekogranične strategije smart specijalizacije između Crne Gore i Srbije, kako bi se uvideli dobri načini saradnje dve zemlje u oblasti inovacija, istraživanja i razvoja. Autori identifikuju oblasti i načine saradnje dve zemlje.

Ključne reči: *smart specijalizacija, prekogranična smart specijalizacija, inovacije, istraživanje, razvoj*

1. UVOD

Koncept pametne specijalizacije je vizija Dominika Foreja i ekspertske grupe „Znanjem za rast“, koja je funkcionisala u okviru Evropskog istraživačkog prostora (European Reasearch Area — ERA). Problem koji je grupa uočila i analizirala je manja konkurentnost privrede u evropskim zemljama od ekonomije SAD, posebno u dostignućima njenih istraživačkih i

[*] Dr Mirjana Kranjac, Univerzitet u Novom Sadu, Pokrajinski sekretarijat za privredu i turizam, Vlada Vojvodine

inovacionih kapaciteta i u primeni inovacija i njihovom prenošenju na tržište. Analiza koju su uradili dala je zaključak da su istraživanja u Evropi previše fragmentirana, da ne postoji dovoljna koordinacija niti komunikacija među zemljama u ovoj oblasti. Naime, mnoge od zemalja ulažu u istraživanja sličnih oblasti, najviše u ono što je trenutno „moderno". Mnogo novca se troši bez ikakve kontrole i ikakvih rezultata. Oblasti, kao što su nanotehnologije i biotehnologije i informacione i komunikacione tehnologije (IKT) gotovo dominiraju investicijama. Oni preporučuju kreiranje savremenih istraživačkih i inovacijskih politika, zasnovanih na principima dobrog upravljanja. Takav pristup će doprineti razvoju onih oblasti koje najbolje koriste potencijale regiona vezane za njihove specifičnosti [1, 2]. Vojvodina je bila prva regija van EU, koja je razvila strategiju pametne specijalizacije (S3). Rezultat primene metodologije EU za razvoj S3 bio je da se Vojvodina, u oblasti istraživanja i razvoja, fokusira na poljoprivredu, proizvodnju hrane i IKT. Strategija Vojvodine bila je odličan primer strateškog dokumenta i EK je pozvala stručnjake koji su je izradili da je predstave trećim zemljama, kao TAIXS eksperte. Strategija je bila za period od 2015. do 2020. godine. Iako ju je prihvatila Skupština Vojvodine, realizovano je samo nekoliko definisanih aktivnosti. Nakon 5 godina svog životnog ciklusa, strategija nije ažurirana. Ostala je bez političkog razumevanja i završila svoj životni ciklus. Posledica je, između ostalog, da mnoge projektne prijave kod fondova EU iz Vojvodine ne dobijaju dodatne bodove tokom evaluacije, i ne finansiraju se. Ilustracija značaja posedovanja ovakve strategije je da je Institut za digitalnu poljoprivredu Biosens za svoj razvoj dobio oko 30 miliona evra jer je Vojvodina imala S3 [3, 4].

2. STRATEGIJA SMART SPECIJALIZACIJE SRBIJE, ŽIVOTNI CIKLUS

Strategija pametne specijalizacije Srbije (S4) usvojena je početkom 2020. godine. Izrađena je prema metodologiji koju je predložila Evropska komisija. Tokom ovog procesa, koji je uključivao stručnjake iz različitih sektora, imala je snažnu podršku Vlade Srbije. Srbija je prošla sve korake, i to (Slika 1):

— Kvantitativno ispitivanje koje je identifikovalo neke potencijalne oblasti.

— Nakon niza sastanaka sa značajnim zainteresovanim stranama, sprovedena je kvalitativna analiza, koja je ukazala na potencijalne sektore sa konkurentskim prednostima. Ova analiza je ispravila greške koje su nastale nakon kvantitativne (iz statističkih podataka) analize.

— Proces preduzetničkog otkrivanja (Entrepreneurship Discovery Process — EDP).

— Napisana je prva verzija S4.

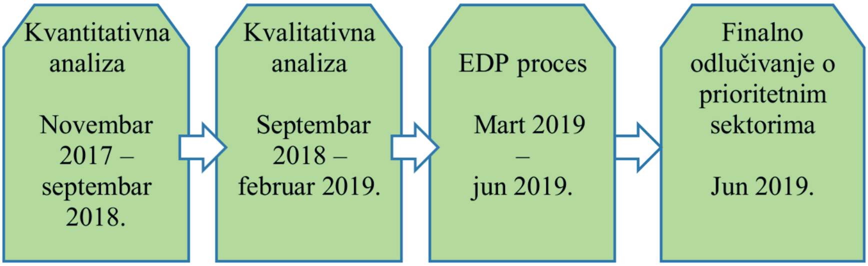
— Javna rasprava, koja je obuhvatila dosta ciljanih sastanaka i radionica svih predstavnika četvorostrukog heliksa (akademija, privreda, nevladin sektor, javni sektor), donela je listu prioritetnih tematskih oblasti za istraživanje, razvoj i inovacije.

Najvažniji podproces pri kreiranju S4 bio je tzv. proces preduzetničkog otkrivanja (EDP). „EDP je sveobuhvatan i interaktivan proces odozdo prema gore, u kojem članovi koji dolaze iz različitih okruženja (javna politika, biznis, akademska zajednica, nevladine organizacije, itd.) pronalaze i objavljuju informacije o potencijalnim novim aktivnostima i potencijalnim sposobnostima, a koje nastaju kroz razmenu informacija među tim stranama, dok kreatori politika istražuju rezultate i pristupe kako bi podstakli aktualizaciju ovog potencijala“ [5].

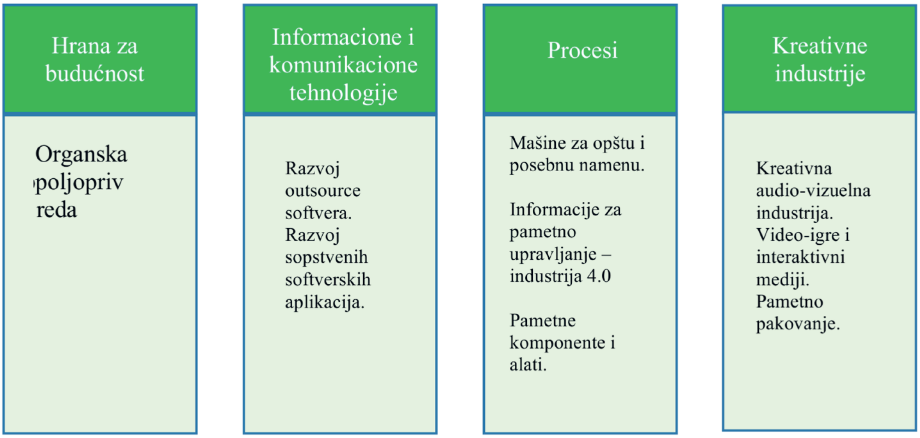
EDP stvara partnerstva čiji je cilj da integrišu preduzetnička znanja, koja su sada distribuirana na mnoge kompanije, organizacije, univerzitete, klijente i korisnike. Ovaj proces ima za cilj da uključi sve relevantne zainteresovane strane u proces stvaranja sektora od interesa za istraživanje, razvoj i inovacije, i ovo se dešava po prvi put. Proces preduzetničkog otkrivanja je interaktivan proces, zasnovan na ciljanom dijalogu, koji okuplja različite aktere u cilju identifikovanja prioritetnih oblasti pametne specijalizacije i razvoja odgovarajuće kombinacije politika za njihovu implementaciju [6, 7].

Procedura definiše vertikalne prioritetne oblasti identifikovane kao najvažniji sektori inovacionih aktivnosti u Srbiji. Oni su predstavljeni na Slici 2 i na sledećoj listi:

1. Informacione i komunikacione tehnologije:
   — veliki podaci (Big data) i poslovna analitika;
   — računarstvo u oblaku (Cloud computing);
   — internet stvari (Internet of Things);
   — razvoj softvera;
   — ugrađeni sistemi (Embedded systems).



Slika 1. Proces identifikacije potencijalno prioritetnih oblasti [3]



Slika 2. Prioritetni sektori [3]

2. Hrana za budućnost:
   — visokotehnološka poljoprivreda;
   — hrana sa dodatnom vrednošću;
   — održiva poljoprivreda i proizvodnja hrane.
3. Kreativne industrije:
   — kreativna digitalna audio-vizuelna produkcija;
   — industrija video-igara;
   — pametno i aktivno pakovanje.
4. Mašine i proizvodni procesi budućnosti:
   — mašine za posebne namene;
   — informacije u službi pametnog menadžmenta — industrija 4.0;
   — alati i delovi za proizvođače automobila, železničku opremu i avio-preduzeća;
   — pametna rešenja za ekologiju.

Pored vertikalnih prioriteta, pokazalo se da postoji nekoliko horizontalnih oblasti, koje bi morale da budu uključene u sve ostale, a to su:

1. — energetski efikasna i eko-pametna rešenja;
   — eko-pametni izvori energije.
2. Ključne omogućavajuće tehnologije (Key enabling technologies — KET):
   — fotonika;
   — napredni materijali;
   — napredne proizvodne tehnologije i elektronika;
   — biotehnologija;
   — blokčejn tehnologija;
   — autonomna vožnja, vazduhoplovni sistemi i inženjering.

Sa svim ovim zaključcima započeo je životni ciklus S4. Sledeća faza je implementacija i permanentno praćenje implementacije, evaluacija i korekcija. Da li se S4 implementira kako treba? Neke radne grupe su osnovane, ali su imale tek početne sastanke. Početna energija se raspršila, zainteresovane strane više nisu aktivne. Samo nekoliko novih strateških dokumenata Srbije pominje S4 [8].

3. SARADNJA CRNE GORE I SRBIJE U OBLASTI PAMETNE SPECIJALIZACIJE

Autor rada je bila učesnica prekograničnog projekta Mađarska—Srbija, koji je finansirala EU i u kome je urađena prekogranična strategija smart specijalizacije za regije: Velika južna dolina u Mađarskoj i Vojvodina u Srbiji. U tom projektu su definisane oblasti saradnje dve regije u oblasti istraživanja i razvoja. Ovakav pristup je vrlo originalan, a neophodan za razvoj oblasti, tj. klastera više regija. Autorka je bila predavač po pozivu na 19. Naučnostručnom skupu informacione tehnologije, koji je održan na Žabljaku, u organizaciji Elektrotehničkog fakulteta iz Podgorice, u februaru 2014. godine, upravo da predstavi stvaranje ovakvog dokumenta za regiju Vojvodine i time svojim znanjem i iskustvom pomogne izradu strategije smart specijalizacije Crne Gore. Crne Gora je 2019. godine uradila svoju strategiju sa važenjem do 2024. godine i time postala prva zemlja van EU, koja je uradila ovaj dokument. Time je definisala svoje prioritetne oblasti, a to su:

1. održiva poljoprivreda i lanci vrednosti hrane;
2. energija i održiva životna sredina;
3. održivi i zdravstveni turizam;
4. informacione i komunikacione tehnologije.

Ukrštajući prioritetne oblasti obe zemlje, nameće se mogućnost saradnje pre svega u oblasti poljoprivrede, odnosno proizvodnje organske hrane, funkcionalne hrane, odnosno hrane budućnosti. Ovo se može povezati sa seoskim, održivim, gastronomskim turizmom, stvarajući zajedničke turističke petlje sa potporom u dobroj hrani i piću.

Druga oblast saradnje može biti IKT oblast, koju su obe države obeležile kao važnu. Mogu da se prave zajedničke baze podataka, prave Big data obrade, koje bi dale bolje rezultate. Sve to bi dovelo do kvalitetnog mašinskog učenja i primene veštačke inteligencije. Dalje ova saradnja može da rezultuje stvaranjem i umrežavanjem pametnih gradova i sela. Klasteri ovakvih struktura treba da donesu potpuno nov kvalitet i jačanje saradnje i zajedništva, naročito što su i kulturnoistorijske osnove zajedničke i bliske. Ovaj rad treba da bude podstreh da se priđe izradi strategije prekogranične smart specijalizacije dve zemlje.

4. ZAKLJUČAK

Pametna specijalizacija je nov, veoma važan evropski koncept. Evropska komisija je uložila mnogo napora da razvije metodologiju za njen razvoj. Organizovali su peer review radionice za mnoge regione, kako bi se njihove strategije poboljšale. Uspostavljena je platforma S3 [10], na kojoj su sve urađene strategije pametne specijalizacije i niz korisnih alata za njihovo praćenje, upoređivanje, poboljšanje. Napisani su mnogi naučni članci i knjige. Zajednički istraživački centar EK (JRC) je glavno telo zaduženo za sve ove zadatke i ono ima na raspolaganju grupu stručnjaka za pružanje tehničke pomoći svima kojima je potrebna. Srbija je stigla do prvog cilja, urađen je S4. Ovo je ogroman rezultat, ali nije dovoljno. Izbijanje pandemije COVID-19 omelo je implementaciju S4. Sada je period posle pandemije, svet ima nove izazove, a Srbija ne sme da prekine aktivnosti oko S4, kao što je to sada slučaj. Ona ih mora dosledno sprovoditi. Strategija se mora promovisati, radionice moraju biti organizovane, zainteresovane strane moraju biti aktivno uključene. Svi koji su učestvovali u kreiranju Strategije moraju insistirati na njenom sprovođenju. To može da se uradi uz podršku EK, npr. preko već za S3 zaduženog JRC. Može se tražiti pomoć u organizovanju radionica, izradi izveštaja, poređenju ishoda sa drugim zemljama, kako bi se videlo dokle je Srbija stigla, a gde su ostale zemlje. Slučaj Vojvodine je primer kako se ne radi. Vojvodina je bila najbolja, prva regija van EU, koja je oblikovala Strategiju pametne specijalizacije, ali političari nisu prepoznali potrebu da se ovaj dokument ažurira i da mu se produži životni ciklus. Ovakvo stanje i stav idu na štetu svih građana, a posebno nauke Vojvodine.

Takođe, postoji potreba za izradama prekograničnih strategija smart specijalizacija i saradanji u bliskim prioritetnim oblastima što bi rezultovalo novim kvalitetom.

LITERATURA

[1] Foray D., Paul D., Bronwyn H.: „Smart specialization from academic idea to political instrument, the surprising career of a concept and the difficulties involved in its implementation“, MTEI Working paper-2011–001, Ecole polytechnic, Lausanne, 2011.

[2] McCann Ph, Ortega-Argilés, R.: „Smart specialisation, regional growth and applications to EU cohesion policy“, Working document from IEB 2011/14, Institute of Economy, Barcelona, 2011.

[3] Kranjac M., Salom J.: „The evaluation and implementation of a smart specialization strategy“, Cambridge Scholar, Publishing, 2022.

[4] BioSense: 2020. https: //biosens. rs/?page_
id=12564&lang=en Accessed October 11, 2021.
[5] Foray D.: „Smart Specialisation: Challenges and Opportunities for Regional Innovation Policies“, Routledge, 2015.
[6] Midtkandal I., Jens S.: „What is smart specialization“, Nordregio News, Issue 5, pp. 55–60, 2012.
[7] Martínez-López D., Palazuelos-Martínez M.: „Breaking with the past in smart specialisation: A new model of selection of business stakeholders within the entrepreneurial process of discovery“, Journal of the Knowledge Economy, 10(4), 2015.
[8] Kranjac M.: „How to ensure sustainable development of Serbia: supporting innovation and entrepreneurship?“, Etran, audio record, Belgrade, 2017.
[9] Kranjac M.: „Strategic development based on principles of smart specialization, case of Vojvodina“, Zbornik, 19. Naučnostručni skup informacione tehnologije, Žabljak: Elektrotehnički fakultet Podgorica, 24–28. februar, 2014, str. 1–5.
[10] S3 Platform: https: //s3platform. jrc. ec. europa. eu/search?q=how
+ many + s3&category=78605 Accessed October 11, 2021.

Mirjana KRANJAC

STRATEGY OF SMART SPECIALIZATION IN SERBIA

Abstract

Serbia created its own Smart Specialization Strategy (S4) for the period from 2020 to 2027. This document is the basis for the development of the research and innovation sector in European countries and regions. It is based on the division of these activities according to the specificities of the regions and taking into account their comparative advantages. Regional niches of research activities are being created that tend not to overlap. This optimizes the life cycle of knowledge and innovation, giving greater added value to the European space as a whole. After creating such a strategy, it is necessary to promote it to all interested parties. The implementation of the document must be continuously monitored, it should be corrected and improved and involve as many actors as possible. Is it like that in Serbia? The authors indicate that this document is not sufficiently visible, as well as many similar documents. Neither its value nor the possibility of application when applying for financial resources of the European Union was recognized. The paper gives recommendations on how to improve the life cycle of S4 in order to have significant effects for the development of the Serbian innovation ecosystem. One of the ways of further development of S4 can be the creation of a cross-border smart specialization strategy between Montenegro and Serbia in order to see good ways of cooperation between the two countries in the field of innovation, research and development. The authors identify areas and ways of cooperation between the two countries.

Key words: smart specialization, cross-border smart specialization, innovation, research, development

INFORMATIČKE NAUKE

INFORMATICAL SCIENCES

Miodrag GRBIĆ*

EXPLORING GENETIC DIVERSITY USING GENOMIC

Abstract: In this article, I will broadly address the application of genomic technologies in assessing biological diversity, including various organisms ranging from Chelicerates to plants. I will focus on the application of genomic technologies in the development of new bio-nano materials, and genomics of Montenegrin grapevine evolution.

INTRODUCTION

The nucleic acid sequence in the DNA molecule determines the hereditary and biochemical properties of one organism. The complete genome sequence of an organism contains all the information necessary for the development of the organism, its behavior, physiology, biochemical processes that can be carried out by the organism, its resistance to disease and hereditary properties that will be transmitted to the offspring. The sequence of the complete genome summarizes the „user manual" of a particular organism and represents the basis for understanding the life functions and the evolution of the organism, containing data on the history and origin of the organism. Sequencing and genomic analysis of the DNA sequences experienced an explosion after the publication of the *Homo sapiens* genome sequence in 2001. Bearing in mind the complexity of the sequencing process and the analysis of the genome, the question is why this complex analysis actually serves. First, this analysis gives us the „user's manual" for each organism and the genetic program for how individual organisms function. This genomic knowledge is the base knowledge-based economy and bioeconomy. This genomic knowledge is the basis for the development of new

* Miodrag Grbić, Department of Biology, University of Western Ontario, London, ONT, Canada

technologies including medicine and health, agriculture and food industry, veterinary medicine, industrial biotechnology, ecology, environmental protection, forensic medicine, and production safety. Genome sequencing technology has created a whole range of genomic disciplines integrating biology, informatics, and mathematics into computational biology and system biology. The three basic branches of genomics are structural genomics (which includes mapping and sequencing), comparative genomics (which researches diversity and genome evolution), and functional genomics which analyses the function of genes in biological systems. New „OMICS" technologies such as proteomics (which studies proteins), metabolomics (which studies metabolites), and transcriptomics (which explores RNA expression in various tissues) represent advances in our understanding of the genome. New applications include archaeological genomics, where our history is „read" through a history recorded in DNA, agricultural genomics, and environmental genomics.

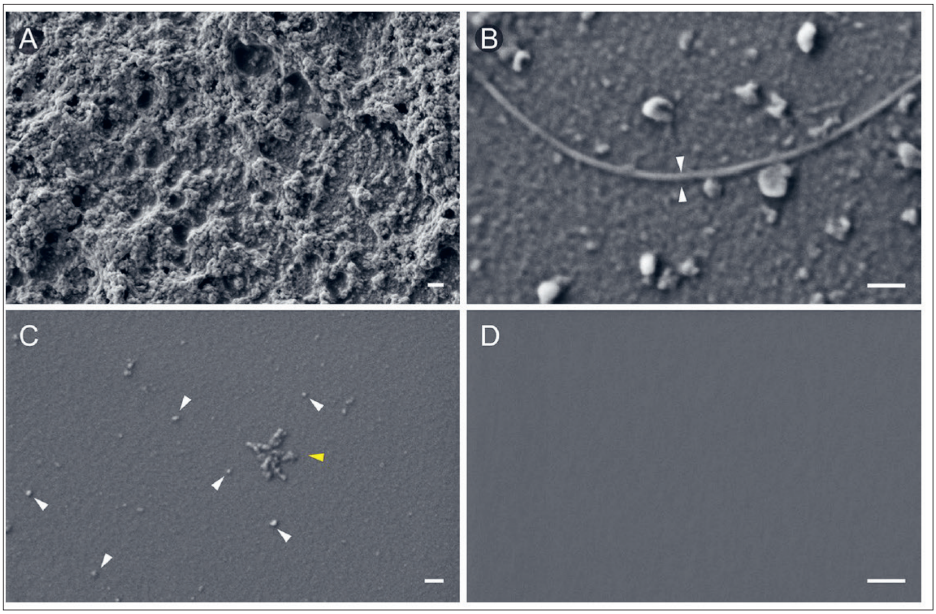
GENOMICS IN THE GENERATION OF NEW BIOMATERIALS

The development of modern technologies relies heavily on novel materials with applications in a wide range of fields, including medicine and pharmacology, food production, engineering, and catalysis. Specific characteristics of materials necessary for particular applications include properties such as mechanical strength, elasticity, biocompatibility, biodegradability, size, density, and a combination of biochemical and physical characteristics that are often found in biological materials. Materials such as spider and silkworm silk have served as a basis for the development of specific materials used in pharmacology, diagnostics, and regenerative medicine1–3, but have also inspired the development of biomimetic synthetic materials4–6. Nanomaterials are emerging as an important element for modern therapeutic treatments. For effective drug delivery, potential nano-carriers must be biocompatible, biodegradable, non-toxic, and non-immunogenic, and need to allow versatile conjugation of specific drug load and delivery in a particular cellular compartment 7–9. Currently, there are numerous nanomaterials on the market with often confusing nomenclature, such as „natural" biomaterials that can be of organic origin (spider or silkworm silk) or inorganic origin (titanium dioxide, nano-silver and many others), synthetic biomimetic materials (polymers, plastic), and materials produced by recombinant-technology. However, the application of nanomaterials in modern life has raised concerns about their environmental safety and potential for adverse effects on human health10.

Indeed, it has been shown that many engineered nanomaterials cause undesired effects in living organisms. For example, TiO2 affects circadian rhythm, immune and inflammatory response and basal metabolism, SiO2 affects immune response and genes involved in inflammation processes and polystyrene nanomaterials are associated with effects on apoptosis, inflammation, and basal metabolism11. Thus, biomaterials of natural origin are believed to have superior characteristics relative to synthetic materials. They are often biocompatible and biodegradable, and, as proteinaceous molecules, can be easily functionalized for specific applications. Silks are a family of proteinaceous materials secreted by many arthropods for different biological functions. The most studied examples of such organic materials are spider silk and silkworm silk. However, recent studies have uncovered „silk-like" materials from other arthropods including numerous insect species12, mites13 and even some marine species such as mussels14 and barnacles15.

All these silk proteins, encoded by fibroin genes, contain common structural motifs, such as the β-pleated sheet, generating unique silk properties depending on slight differences in individual sequences. The majority of these naturally produced silk threads are in the micron range of fibre thickness and are used for construction of cocoons, prey capturing, egg sac production, adherence to substrates or forming the pedicel of eggs16. Mechanical properties of these silk fibres can be expressed in terms their Young's modulus, which characterizes the deformation due to applied stress, and ranges from 7 GPa in silkworm to 13.5 GPa in the spider *Nephila clavipes* to a high Young's modulus of 28 GPa in the bugworm *Eumeta variegate*17. Another little-studied group that spins silk are phytophagous spider mites (Tetranychide). They belong to acari, in which the majority of species do not produce silk; however, spider mites from genus *Tetranychus* have evolved the production of versatile silk used for locomotion and dispersal from plant to plant18–20, protection from predators 18,20,21, sheltering from climatic conditions22, as a surface for egg laying23, and for communication via pheromone and social-clue deposition24,25. Silk fibers from Tetranychus urticae have a striking characteristic: they have diameters on the nanometre scale, representing the thinnest natural silk fiber produced by silk-spinning arthropods13,26,27.

This nano-silk displays an extraordinarily high Young's modulus that is almost double that of spider N. clavipes silk13,26 and is in the range of bagworm silk17, potentially representing a natural nano biomaterials with valuable characteristics for technology and medicine. The T. urticae genome sequencing project allowed the isolation of fibroin gene sequences; however, due to the limited amount of silk production by *T. urticae* and its fine structure, it was not possible to characterize this promising biomaterial. To overcome the limitation of lowsilk production rate, we established a culture of the related species, gorse spider mite, *Tetranychus lintearius*, that produces copious amounts of silk (Tl-S) (Fig. 1A), and using semi-industrial production generated a sufficient amount of silk for biochemical and physical characterization. We have shown that *T. lintearius* silk has thickness and physical characteristics similar to *T. urticae* silk. We characterized native Tl-S and produced nano-particles generated from Tl-S (Tl-SN) and compared then to *Bombyx mori* silkworm native silk fibroin (Bm-SF) and nanoparticles (Bm-SFN) as the standard in the field. Exposure of mouse fibroblasts to Tl-S-derived biofilm and nanoparticles demonstrated that this cell line can grow successfully in the culture at a comparable level to that seen in the presence of Bm-SF derived nanoparticles and biofilm. Finally, we fluorescently labeled Tl-SN and showed that they can enter the cytoplasm of cultured cells. These experiments suggest that Tl-S is a new cytocompatible material and a potential source of natural nanoparticles with potential for various applications, including pharmacology and biomedicine.

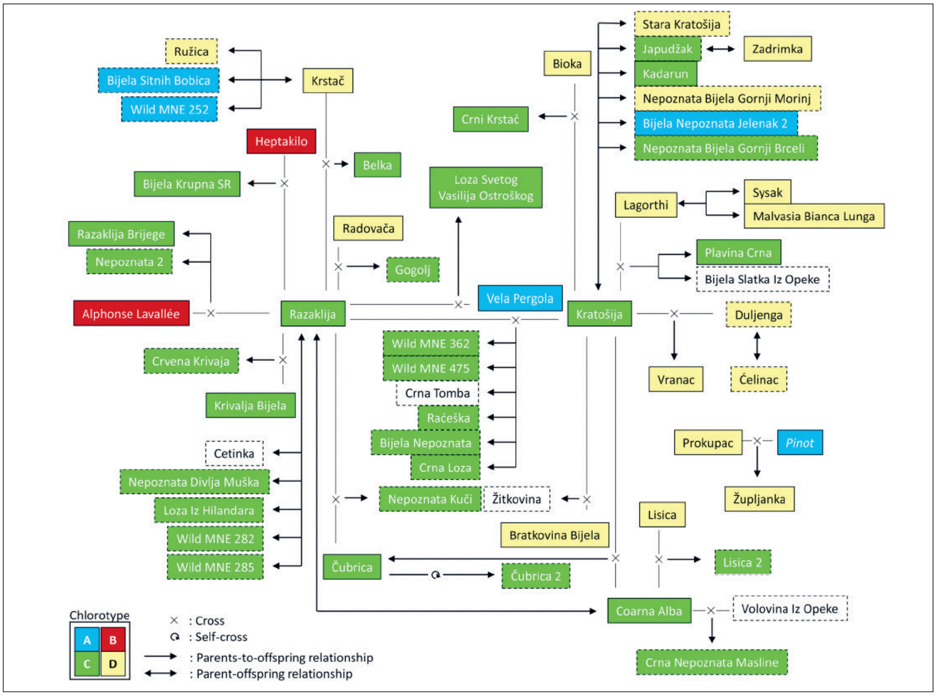


Focused Ion Beam milling combined with Scanning Electron Microscopy (FIB-SEM) of T. lintearius nanoparticles at different concentrations. (A) 10X dilution of T. lintearius nanoparticles showing dense conglomerates of particles. (B) 100X dilution shows spontaneous formation of silk fibers (arrowheads). (C) At 1000X dilution individual nanoparticles show diameter of ca. 20 nm (arrowheads) (D) Silicone wafer with no sample loaded (control). Scale bar: a—200 nm, b–d—100 nm.

IMPORTANCE OF GENOMICS FOR MONTENEGRO: AN EXAMPLE OF THE GRAPEVINE GENOMICS

Montenegro has incredible genetic diversity and it requires focused and strategic investment in genomics because without it the diversity and richness of species that Montenegro has could not be valorized or will be used by someone else in the development of new technologies. The peak of the iceberg and the potential in this area reveals the largest genomic research in Montenegro performed so far, the analysis of the genomic diversity of grapevine that was carried out in cooperation between Plantaze, the Ministry of Agriculture and the Ministry of Science of Montenegro and supported by CANU in cooperation with the world's leading institute for Grapevine Research (ICVV) from La Rioja, Spain. This research using the methods of genomic analysis 28, comparative genomics and genetic pedigrees analysis showed that Montenegro has 64 unknown grape varieties (so far the largest number in the surrounding) and their complete pedigree and their genetic relationships have been established. The central variety that is the parent of Montenegrin vine varieties is Kratosija, which shows great diversity and historical data from the Budva Statut from 1427 suggests that it is a variety cultivated in this area for centuries.

According to the pedigree structure of the Montenegrin grape varieties, this wine region has a similar structure with traditional wine regions such as Bordoux or Burgundy in France, where individual varieties such as Cabernet franc or Pinot noir are the basis of local pedigrees indicating the old traditional viticulture. This research, which is more extensively discussed in this Proceeding, is a starting point for further analysis that has the potential to place Montenegro in important international touristic wine destination but also to create a scientific and technological basis for the exploitation of these findings. The sequencing of the Vranac genome revealed its origin from the father Kratosija variety and the mother Duljenga and opened the pathway to genomics and metabolomics that should determine the structure of the Vranac genome and metabolites such as resveratrol and many complex polyphenols and tannins specific for this variety. The next step of this research is determination of the origin of the American variety Zinfandel by genomic technologies. Until now, the similarity of Zinfandel, Primitivo, Crljenak Kastelanski / Pribidgarag / Tribidraga was determined with a small number of molecular markers and there was no material analysed from Montenegro where Kratosija showed the highest biodiversity, being one of the prime candidates for the origin of Zinfandel. This great biodiversity suggests that Montenegro is the center of origin for this variety



First-order genetic relationships (trios and duos) detected for cultivated and wild grape varieties sampled in Montenegro. Chlorotypes (A, B, C or D) are indicated with different colors, according to the inserted code. If white, no information on chlorotype was available. Unidentified and unique genotypes in the ICVVSNP database are shown in boxes with broken borders. These genetic relationships were obtained with the likelihood-based method implemented in Cervus v. 3.0 for parentage analysis, on the basis of SNP genetic data.

that was later distributed world-wide. Thus, the next step in this project is the genomic analysis to determine from where Zinfandel was introduced to America. By using genomics, Montenegro has managed to preserve 64 unknown grapevine varieties from disappearing, determine their pedigrees and genetic relations and ensure the development of autochthonous viticulture and wine production for the next centuries. This pedigree of Montenegrin varieties has great significance for archeobotanics. The first paper on ancient DNA from grapevine seeds (Ramos-Madrigal et al. 2019) what published a while ago, showing that the grapeseed DNA discovered from the year 1050 corresponds to DNA of the modern variety Sauvignon Blanc, demonstrating that this variety has been grown in France for almost 1000 years. This discovery has an incredible significance for Montenegro, because the pedigree of Montenegrin grapevine varieties and the numerous archaeological findings from the Greek, Roman, and Medieval periods could allow a deeper understanding of the genetic history of Montenegrin varieties that would represent an incredible quantum leap for the history and genetics of the vines in Montenegro and the world. Research on the genetic diversity of the grapevine is the first serious step in the application of genomics in Montenegro, which could make it one of the backbones of the future societal and technological development.

REFERENCES:

[1] Omenetto, F. G. & Kaplan, D. L. New opportunities for an ancient material. Science https: //doi. org/10.1126/scien ce. 11889 36 (2010).
[2] Preda, R. C., Leisk, G., Omenetto, F. & Kaplan, D. L. Bioengineered silk proteins to control cell and tissue functions. Methods Mol. Biol. 996, 19–41 (2013).
[3] Huang, W., Ling, S., Li, C., Omenetto, F. G. & Kaplan, D. L. Silkworm silk-based materials and devices generated using bionanotechnology. Chem. Soc. Rev. https: //doi. org/10.1039/C8CS0 0187A (2018).
[4] Hubbell, J. A. Biomaterials in tissue engineering. Bio/Technology https: //doi. org/10.1038/nbt06 95–565 (1995).
[5] Lutolf, M. P. & Hubbell, J. A. Synthetic biomaterials as instructive extracellular microenvironments for morphogenesis in tissue engineering. Nat. Biotechnol. https: //doi. org/10.1038/nbt10 55 (2005).
[6] Lapidot, S. et al. Clues for biomimetics from natural composite materials. Nanomedicine https: //doi. org/10.2217/nnm. 12.107 (2012).
[7] Schacht, K. & Scheibel, T. Processing of recombinant spider silk proteins into tailor-made materials for biomaterials applications. Curr. Opin. Biotechnol. https: //doi. org/10.1016/j. copbi o. 2014.02.015 (2014).
[8] Schierling, M. B., Doblhofer, E. & Scheibel, T. Cellular uptake of drug loaded spider silk particles. Biomater. Sci. https: //doi.org/10.1039/c6bm0 0435k (2016).
[9] Holland, C., Numata, K., Rnjak-Kovacina, J. & Seib, F. P. The biomedical use of silk: past, present future. Adv. Healthc. Mater. https://doi. org/10.1002/adhm. 20180 0465 (2019).
[10] A Research Strategy for Environmental, Health, and Safety Aspects of Engineered Nanomaterials. A Research Strategy for Environmental, Health, and Safety Aspects of Engineered Nanomaterials. https: //doi. org/10.17226 /13347 (2012).
[11] Costa, P. M. & Fadeel, B. Emerging systems biology approaches in nanotoxicology: towards a mechanism-based understanding of nanomaterial hazard and risk. Toxicol. Appl. Pharmacol. https: //doi. org/10.1016/j. taap. 2015.12.014 (2016).
[12] Weisman, S. et al. Honeybee silk: recombinant protein production, assembly and fiber spinning. Biomaterials https: // doi.org/10.1016/j. bioma teria ls. 2009.12.021 (2010).
[13] Grbić, M. et al. The genome of Tetranychus urticae reveals herbivorous pest adaptations. Nature https: //doi. org/10.1038/natur e10640 (2011).

[14] Smeathers, J. E. & Vincent, J. F. V. Mechanical properties of mussel byssus threads. J. Molluscan Stud. https: //doi. org/10.1093/oxfordjour nals. mollu s. a0654 97 (1979).
[15] So, C. R. et al. Sequence basis of barnacle cement nanostructure is defined by proteins with silk homology. Sci. Rep. https: //doi.org/10.1038/srep3 6219 (2016).
[16] Sutherland, T. D. et al. Biocompatibility and immunogenic response to recombinant honeybee silk material. J. Biomed. Mater. Res. Part A https: //doi. org/10.1002/jbm. a. 36695 (2019).
[17] Yoshioka, T., Tsubota, T., Tashiro, K., Jouraku, A. & Kameda, T. A study of the extraordinarily strong and tough silk produced by bagworms. Nat. Commun. https: //doi. org/10.1038/s4146 7–019–09350 -3 (2019).
[18] Helle, W. & Sabelis, M. W. Spider Mites Their Biology, Natural Enemies and Control (1985).
[19] Clotuche, G. et al. The formation of collective silk balls in the spider mite tetranychus urticae Koch. PLoS ONE https:// doi.org/10.1371/journ al. pone. 00188 54 (2011).
[20] Sabelis, M. W. & Bakker, F. M. How predatory mites cope with the web of their tetranychid prey: a functional view on dorsal chaetotaxy in the Phytoseiidae. Exp. Appl. Acarol. https: //doi. org/10.1007/BF011 93804 (1992).
[21] Iwasa, T. & Osakabe, M. Effects of combination between web density and size of spider mite on predation by a generalist and a specialist phytoseiid mite. Exp. Appl. Acarol. https: //doi. org/10.1007/s1049 3–015–9902–7 (2015).
[22] Hazan, A., Gertler, A., Tahori, A. S. & Gerson, U. Spider mite webbing-III. Solubilization and amino acid composition of the silk protein. Comp. Biochem. Physiol. Part B Biochem. https://doi. org/10.1016/0305–0491(75)90038 -3 (1975).
[23] Clotuche, G., Yano, S., Akino, T. & Amano, H. Chemical investigation of aggregation behaviour in the two-spotted spider mite Tetranychus urticae. Exp. Appl. Acarol. https: //doi. org/10.1007/s1049 3–014–9779-x (2014).
[24] Cone, W. W., Predki, S. & Klostermeyer, E. C. Pheromone studies of the two spotted spider mite. 2. Behavioral response of males to quiescent deutonymphs. J. Econ. Entomol. 64, 379–382. https: //doi. org/10.1093/jee/64.2.379 (1971).
[25] Penman, D. R. & Cone, W. W. Role of web, tactile stimuli, and female sex pheromone in attraction of male twospotted spider mites to quiescent female deutonymphs. Ann. Entomol. Soc. Am. https: //doi. org/10.1093/aesa/67.2.179 (1974).
[26] Hudson, S. D., Zhurov, V., Grbić, V., Grbić, M. & Hutter, J. L. Measurement of the elastic modulus of spider mite silk fibers using atomic force microscopy. J. Appl. Phys. https: //doi. org/10.1063/1.48008 65 (2013).
[27] Clotuche, G., Mailleux, A. C., Deneubourg, J. L., Detrain, C. & Hance, T. The silk road of Tetranychus urticae: is it a single or a double lane? Exp. Appl. Acarol. https: //doi. org/10.1007/s1049 3–012–9520–6 (2012).
[28] Maraš, V., Tello, J., Gazivoda, A. et al. Population genetic analysis in old Montenegrin vineyards reveals ancient ways currently active to generate diversity in Vitis vinifera. Sci Rep 10, 15000 (2020). https://doi.org/10.1038/s41598–020–71918–7

Miroslav BOJOVIĆ[*], Veljko MILUTINOVIĆ[**]

AIHEAL: AN ARTIFICIAL INTELLIGENCE HEALTHCARE SYSTEM

Abstract: AIHeal healthcare system is a new solution creating explainable and causal artificial intelligence (AI) infrastructure by combining the benefits of the accuracy of deep-learning algorithms with visibility on the factors that are important to the algorithm's conclusion in a way that is accessible to physicians, solution in which they can trust, which open new routes to delivering better, faster, and more cost-effective medical care. The proposed AIHeal system defines a new structure of Electronic Health Record (EHR) that can connect longitudinal data providing insights across episodes of treatment and settings of care, and incorporating new types of data from wearables, sensors and genomics and other omics data. Another innovative contribution of this project is an extensible big data architecture which makes it possible to collect huge volumes and wide spectrum of data.

Key words: *Artificial Intelligence in Healthcare, Electronic Health Record, AI Clinical Decision Support*

INTRODUCTION

Advances in medicine, supported by innovation in technology, are accelerated dramatically in recent years. By the 1950s, medical knowledge had doubled in about 50 years. In 2020, the volume of medical knowledge will double in 73 days [1]. This evolution has significantly risen average life over the past century from less than 50 years to 78.9 years for the USA and to

[*] Miroslav Bojović, Professor of Computer Science, Department of Computer Science, School of Electrical Engineering, University of Belgrade, Belgrade, Serbia

[**] Veljko Milutinović, Professor of Computer Science, Department of Computer Science, University of Indiana in Bloomington, Luddy Hall 2062, IN 47408, Indiana, USA, e-mail: vmilutin@iu.edu

80.9 years for EU [2]. By 2050, 1 in 6 people will be over the age of 65, in Europe and North America, this will be 1 in 4. This demographic shift, combined with rapid urbanization, modernization, globalization and accompanying changes in risk factors and lifestyles, means chronic conditions will be more common, and an increasingly comorbid population's demand for healthcare will increase [3].

Managing patients with complex needs is typically more expensive, especially in conditions when health systems are already stretched. It also adds complexity to information flows, as large volumes of healthcare data no longer sit primarily in hospital. It also requires a different set of skills and a strong culture of collaboration between physicians across specialties. In this context, financial sustainability is a core challenge for European, and Montenegro also, healthcare systems. In 2018, healthcare expenditure ranged between 8.8 and 11.2 percent of the GDP (Italy 8.8%, Spain 8.9%, UK 9.8%, France 11.2%, Germany 11.2%) and are expected to continue rising [4]. Healthcare spending as a share of GDP has been growing since 1990, outpacing average wage growth and the growth of GDP itself [5].

Without major structural and transformational change, healthcare systems will struggle to find the funding needed to address growing demand, whilst maintaining or improving standards of care, access, and patient experience. Also, according to the World Health Organization, projected staff shortages of 9.9 million physicians and inevitable skill gaps will be limiting healthcare systems' ability to satisfy this increasing demands.

In recent decades, we are witnessing increased usage of artificial intelligence (AI) in various domains of science and technology. Healthcare systems are not an exception. With artificial intelligence in healthcare, a medical institution can significantly improve self-care, prevention, overall wellness, triage and diagnosis, clinical decision support, care delivery and management, etc.

In Montenegro, these challenges are even more serious. How many people, especially in rural areas, are far from satisfying standards of medical care? Medical staff shortages and skill are already not tolerable, with the trend to become worse.

The proposed AIHeal system has the potential to transform healthcare organizations and healthcare services and meet the current and future society needs not only in Montenegro.

Following chapters present the problem definition, existing healthcare systems and technologies, proposed solution, a new AIHeal healthcare system, analysis of the proposed AIHeal system, and conclusion. This work is described using the proposed method for presenting research results [6].

EXISTING HEALTHCARE SYSTEMS AND TECHNOLOGIES

Computers are utilized in healthcare for decades, ranging from relational database systems [7], over hardware acceleration of most commonly used operations [8] to complex simmulations of fluid dynamics [9]. Many efforts have been tried to exploit potentials of Artificial Intelligence to transform how care is delivered. Some of them are related to using robots for communication purposes and stimulating patients for activity [10, 11]. Others use sensors to extract data that will undergo the automated reasoning process [12, 13]. Effort has been made to exploit AI for advising medical personel [14].

While deep learning techniques produce state-of-the-art performance on a variety of tasks, one of its main criticisms is that the resulting models are difficult to naturally interpret. In this regard, many deep learning frameworks are often referred to as „black boxes", where only the input and output predictions convey meaning to a human observer. Since correct clinical decision-making can be the difference between life and death, many practitioners must be able to understand and trust the predictions and recommendations made by deep learning systems. There are authors attempts and approaches [15] to make clinical deep learning more interpretable include: Maximum Activation: A popular tactic in the image processing community is to examine the types of inputs that result in the maximum activation of each of a model's hidden units. This represents an attempt to examine what exactly the model has learned, and can be used to assign importance to the raw input features; Constraints: Others have imposed training constraints specifically aimed at increasing the interpretability of deep models. For example, Choi et al.'s Med2Vec framework [16] for learning concept and patient visit representations uses a non-negativity constraint enforced upon the learned code representations; Qualitative Clustering: In the case of EHR concept representation and phenotype studies, some studies point to a more indirect notion of interpretability by examining natural clusters of the resulting vectorized representations. This is most commonly performed using a visualization technique known as t-Distributed Stochastic Neighbor Embedding (t-SNE), a method for plotting pairwise similarities between high-dimensional data points in two dimensions [17]; Mimic Learning: Che et al. [18], [19] first train a deep neural network on raw patient data with associated class labels, which produces a vector of class probabilities for each sample. They train an additional gradient boosting tree (GBT) model on the raw patient data, but instead use the deep network's probability prediction as the target label. Since GBTs are interpretable linear models, they are able to assign feature importance to the raw input features while harnessing the power of deep networks. The approaches given and known from the open references cannot be recognized as full naturally interpretable, and are not widely accepted.

Companies around the world implemented numerous products where AI is having an impact in healthcare. Some of the prominent applications available today grouped around use cases in healthcare are [20]:

— Cronic Care Management: Sensely –virtual nurse, Karantis360–automated personal monitoring and alerting system, AICure–treatment adherence, Pill Pack–personalised presorted meds for repeat prescriptions.

— Care Delivery: Moxi–nurse assistant robot, Amelia–virtual health assistant, Bionic Pancreas–insulin/glucagon administration for Type-1 diabetes patients, EarlySense–contact-free patient monitoring.

— Clinical Decision Support: IBM Watson For Oncology, DeepMind–prediction of acute kidney injury.

— Improving population-health management: Mount Sinai Health Systems–risk prediction for emergency admissions, Sheba Medical Cancer–prediction of complications.

— Improving operations: Qventus–optimisation of operating room flow.

— Self-care/Prevention/Wellness: AliveCor CardiaMobile–personal ECG, Activity and sleep trackers.

— Triage and Diagnosis: Symptom checkers: Babylon, Mediktor, Ping An Good Doctor, Ada Health, K Health.

— Diagnostics: Sight Diagnostics–point of care blood testing, Arterys–medical image analysis, Idx –detection of diabetic retinopathy, Detection of eye diseases: DeepMind, UCL and Moorfields.

PROBLEM DEFINITION

Even available applications have many patients who uses its, why do they not widely accepted from patients and specially from physicians? What are common disadvantages for these applications? We would stress some of the problems which are important from the technological point of view and where technology could help:

— Applications are designed to cover only specific use cases.

— Applications focusing on areas that simplify routine processes are better accepted because they are more understandable for patients and physicians, and obviously free up physician time.

— Applications focusing on use cases such as Clinical Decision Support, Triage and Diagnosis, Diagnostics and other more complex segments of healthcare are less straightforward and are not understandable enough for physicians. Physicians cast doubt on the readiness of the technology and have not confidence to apply them in practice.

— The major challenge is data. Data can be incomplete or of poor quality and consequently AI applications cannot be better than the data used. Data are often poorly suited to AI due to quality issues, inconsistent formats, or the challenges of linking data and obtaining the necessary consent to use in different use cases.

The AIHeal healthcare system has an ambition to solve above mentioned disadvantages:

— The AIHeal can cover all medical use cases offering EHRs which may include structured (demographics, medications, diagnoses, laboratory tests, doctor's note, radiology documents, clinical information), semistructured (from various medical devices, wearables,...) and unstructured data (text, voice, biomedical images,..).

— The extensible big data architecture for such kind of data is proposed, capable of batch and stream processing.

— The most important aspect is that we also propose interpretable system, understandable for physicians. The decision making process can be based on deduction and known facts. Our goal is to create a solution which will be adopted from physicians recognizing that it will make or break the success of AI in healthcare

PROPOSED SOLUTION

The proposed AIHeal system (Figure1) offers three innovative contributions to the field of AI Healthcare systems.

Data

The major challenge is data. How does it possible to manage such quantity of data and reach the necessary quality and interchangeability. One of the most serious obstacles for AI in healthcare is that while enough data are available from pharmacy companies and healthcare institutions, they are not connected or interoperable. Two innovative contributions in a relationship with data are proposed.

The first innovative contribution is the new structure of Electronic Health Record (EHR). The proposed EHRs are medical records for patients with any information relating to the past, present, or future physical/mental health or condition of an individual and connect longitudinal data and new types of data from wearables, mobiles, NL data, sensors and

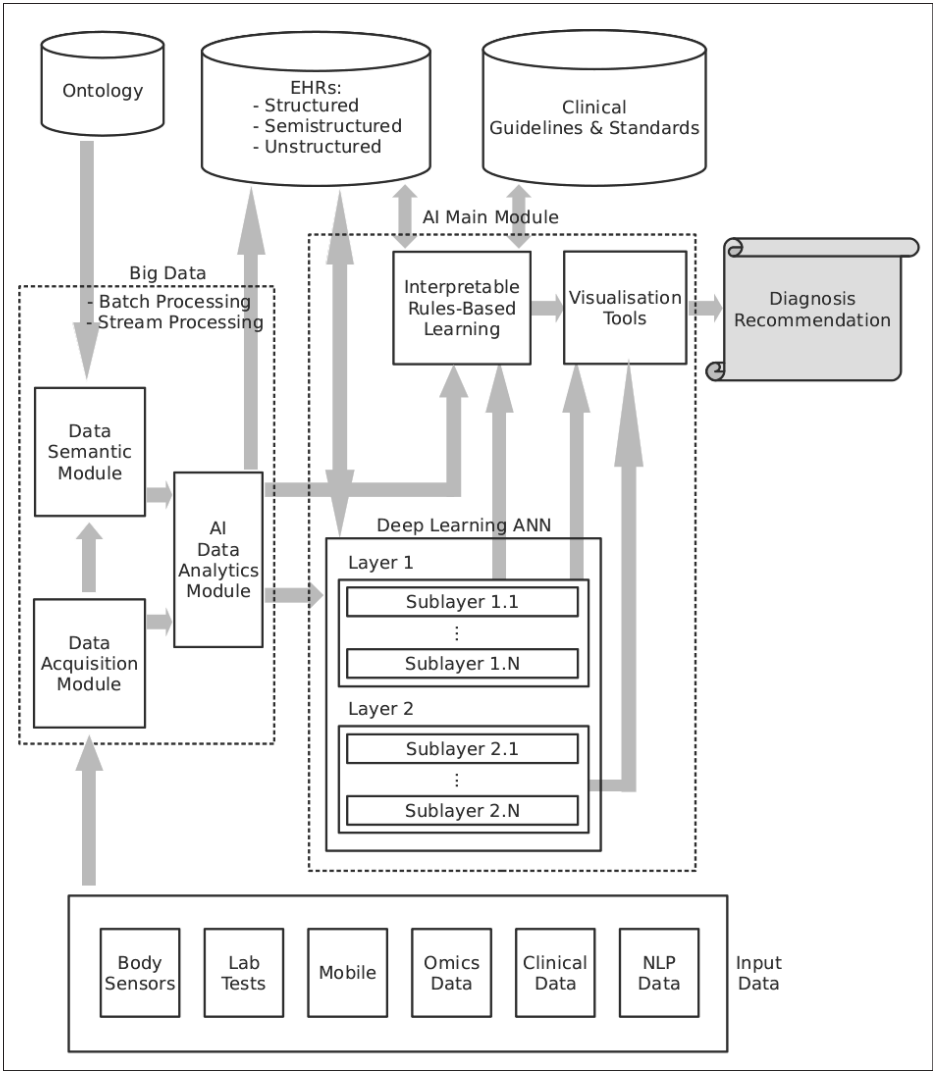


Figure 1. The structure of the proposed AIHeal system

genomics and other omics data. Genomics data are becoming more accessible as the costs of sequencing and bioinformatics techniques have significantly reduced. Healthcare data are typical big data which combine volume (high amounts of data), velocity (data is generated at a rapid pace), variety (data comes under different formats), veracity (data originates from trustable sources), and variability (variations in the data flow rates). Improvements in data, processing power and algorithms are rapidly changing what is feasible. The scope and quality of healthcare data produced and the potential to link datasets are opening new possibilities. Both the quality and consistency of data are improving as more data are machine generated. Focus is on developing consistent interconnected data infrastructure capable for data exchange and semantic interoperability of different medical fields and use cases. The proposed EHRs, unlike the traditional ones, may include structured (demographics, medications, diagnoses, laboratory tests, doctor's note, radiology documents, clinical information), semistructured (from various medical devices, wearables...) and unstructured data (text, voice, biomedical images,..). This data integration allows us to develop the „digital twin" of the patient and makes a prerequisite to design modular and flexible AI infrastructure capable to cover different use cases. This is multidisciplinary approach for medical and software development experts. The ambition of the project is not to cover every medical field, but the cardiology topic as a prototype and a proof of data infrastructure concept. The future multidisciplinary projects, based on this infrastructure, are expected to complete the puzzle.

Extensible big data architecture

The process of extracting from big data can be broken down into five stages [21]. These five stages form the two main sub-processes: data management and data analytics. Data management involves processes and supporting technologies to acquire and store data and to prepare and retrieve it for analysis: Acquisition and Recording; Extraction, Cleaning and Annotation; and Integration, Aggregation and Representation. Analytics, on the other hand, refers to techniques used to analyze and acquire intelligence from big data. The proposed extensible big data architecture consists of Data Acquisition, Data Semantic and AI Data Analytics modules (Figure 1). The Data Acquisition module is responsible to collect data from various sources. This component may be combination of HDFS, NoSQL such as MongoDB and SQL database. The Data Semantic module is mapping heterogeneous databases into common structure and semantics. The Web Ontology Language with XML syntax can be used as standard interchange format regarding ontology. The thorough analysis through the design and development of these modules is expecting to suggest the most suitable solution. The AI Data Analytics module is responsible for Extraction, Cleaning and Annotation, Integration, Aggregation and Representation, and is based on AI mechanisms proposed through the Main AI module proposal. The AI Data Analytics module compares every processed data to predefined user's threshold. If the value of a particular data exceeds alarming threshold value, it will be stored while an emergency alert is generated to the Main AI module. The Main AI module decides if it is value of data for final emergency alert, or regular storage in EHRs. The proposed extensible big data architecture is capable of batch and stream processing and be based on available frameworks. Various available frameworks will be analyzed. For batch processing mode Hadoop allows distributed big data on a cluster of machines but may not be appropriate for stream processing. When stream processing is required, Storm, S4, Apache Spark, Apache Flink may be a suitable choice. Apache Spark may be serious candidate. It is open source unified engine for distributed data processing that includes higher-level libraries for supporting SQL queries (Spark SQL) that restore data from many sources and manipulate them using SQL, streaming data (Spark Streaming), iterative machine learning algorithms through library mechanism (MLlib), provides efficient algorithms with high speed, structured data analysis using Hive, and graph processing based on GraphX. Further information extraction and processing from EHRs requires specialized toolsets for Natural Language Processing, Image Analytics (Visualization Toolkit, GIMIAS, Elastix, MITK), Machine Learning (Tensorflow, Keras, Theano, Torch, Caffe...) and analytics for „Omics" data (SparkSeq, SAMQA, Distmap, Hydra...).

Main AI module

Until the last few years, most of the techniques for analyzing rich EHRs data were based on traditional machine learning and statistical techniques such as logistic regression, support vector machines (SVM), and random forests. Recently, deep learning techniques have achieved great success in many domains through deep hierarchical feature construction and capturing long-range dependencies in data in an effective manner. Machine learning approaches can be broadly divided into two major categories: supervised and unsupervised learning. Supervised learning techniques involve inferring a mapping function y = f(x) from inputs x to outputs y. In contrast, the goal of unsupervised machine learning techniques is to learn interesting properties about the distribution of x itself. The representation of inputs is a fundamental issue spanning all types of machine learning frameworks. For each data point, sets of attributes known as features are extracted to be used as input to machine learning techniques. In traditional machine learning, these features are hand-crafted based on domain knowledge. One of the core principles of deep learning is the automatic data-oriented feature extraction. The vast majority of deep learning algorithms and architectures are built upon the framework of the artificial neural network (ANN). ANNs are composed of a number of interconnected nodes (neurons), arranged in layers. The most common deep learning architectures for analyzing EHR data differ in terms of their node types and the connection structure (e. g. fully connected versus locally connected). Several open source tools exist for working with deep learning algorithms in a variety of programming languages, including TensorFlow, Theano, Keras, Torch, PyTorch, Caffe, CNTK.

While deep learning techniques produce state-of-the-art performance on a variety of tasks, one of its main criticisms is that the resulting models are difficult to naturally interpret. In this regard, many deep learning frameworks are often referred to as „black boxes“, where only the input and output predictions convey meaning to a human observer. Since correct clinical decision-making can be the difference between life and death, many practitioners must be able to understand and trust the predictions and recommendations made by deep learning systems. There are authors attempts and approaches to make clinical deep learning more interpretable include: Maximum Activation: A popular tactic in the image processing community is to examine the types of inputs that result in the maximum activation of each of a model's hidden units. This represents an attempt to examine what exactly the model has learned and can be used to assign importance to the raw input features; Constraints: Others have imposed training constraints specifically aimed at increasing the interpretability of deep models. For example, Choi et al.'s Med2Vec framework for learning concept and patient visit representations uses a non-negativity constraint enforced upon the learned code representations; Qualitative Clustering: In the case of EHR concept representation and phenotype studies, some studies point to a more indirect notion of interpretability by examining natural clusters of the resulting vectorized representations. This is most commonly performed using a visualization technique known as t-Distributed Stochastic Neighbor Embedding (t-SNE), a method for plotting pairwise similarities between high-dimensional data points in two dimensions; Mimic Learning: first train a deep neural network on raw patient data with associated class labels, which produces a vector of class probabilities for each sample. They train an additional gradient boosting tree (GBT) model on the raw patient data, but instead use the deep network's probability prediction as the target label. Since GBTs are interpretable linear models, they are able to assign feature importance to the raw input features while harnessing the power of deep networks. The approaches given and known from the open references cannot be recognized as full naturally interpretable and are not widely accepted.

Our approach to reach naturally interpretable AI is synergistic AI concept. The main idea is to reach explainable and causal AI by combining the benefits of the accuracy of deep-learning algorithms with visibility on the factors that are important to the algorithm's conclusion, in a way that is accessible to physicians and other practitioners, as in rules-based systems, enabled by richer data and improved computing. Therefore, we propose AI in healthcare as a synergy of two concepts, where encoding clinical guidelines and/or existing clinical protocols provides a starting point, which then can be augmented by models that learn from data and demonstrate the distinctive properties of AI, the ability to perform tasks in complex environments without constant user guidance and improving performance by learning from experience. The proposed concept is realized through Main AI module (Figure 1) with two threads: supervised and unsupervised thread. Supervised thread is based on interpretable rules-based systems upgraded with visualization techniques. Unsupervised thread is deep learning ANN with two layers.

The first ANN layer, composed of more sub layers, analyses data to provide a mapping of types and features of disease, feeds Interpretable Rules-Based thread (IRB) allowing professionals to reach a clinical decision „independently" by supervised IRB thread. The second ANN layer, composed of more sub layers, analyses this map to present clinicians with a potential diagnosis and recommendation.

The solution proposed in this manuscript is based on the one previously described in a book chapter [22]. On top of the previous work, we've modeled a wrapper classes for parsing the data from various resources in an uniform manner, accompanied by a function that periodically adapts the artificial intelligence knowledge based on new data streamed over sensors and is received by medical profesionals. Implementation guidelines described in [22,23] are extended by modelling a process that automatically feeds data into the artificial intelligence model from various sources. The model is extendable and supports adding new types of input data to be combined with existing ones.

Actually, our approach is an attempt to follow the physicians thinking through the clinical decision procedure. Physicians first follow some rules and knowledge, then „call" experience (data) and make a decision. If we offer rules and knowledge in human understandable way physicians will believe and can check with their experience, but the deep learning ANN will extend that experience improving accuracy.

This transition from rules drawn up by experts to systems that learn from data would be exemplified and proven by applications to triage patients, applications to support clinical decision making in Cardiology Domain. Data from the WHO shows that cardiovascular diseases (CVDs) are a leading cause of deaths worldwide for both sexes and all ages. More deaths worldwide were caused by CVDs than all communicable, maternal, neonatal, and nutritional disorders combined, which is twice more than those caused by cancer.

The AIHeal system, as software that is potentially also a medical device, should be designed, implemented, tested, and documented using generally recognized quality assurance methods for software development used in the medical domain. The methodology is based on reference model for Clinical Decision Support Systems (CDSS) given in Figure 2.

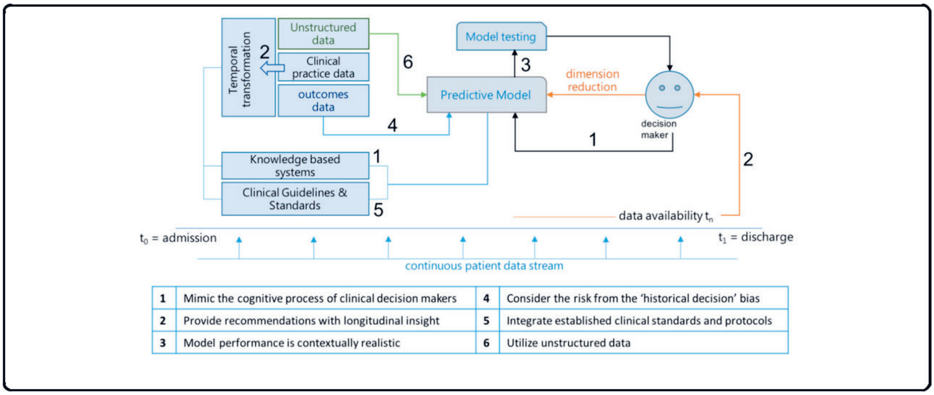


Figure 2. The reference model for Clinical Decision Support Systems (CDSS) [24]

The six elements/considerations of the reference model are: (1) Do CDSS mimic the cognitive process of clinical decision makers? (2) Do CDSS provide recommendations with longitudinal insight? (3) Is the model performance contextually realistic? (4) Is the 'Historical Decision' bias taken into consideration in CDSS design? (5) Do CDSS integrate established clinical standards and protocols? (6) Do CDSS utilize unstructured data?

Methodological steps to develop the AIHeal system are:

— Selection:

— Choose most appropriate data model for the new EHR to cover various medical use cases and clinical workflows.

— Choose most appropriate big data architecture to support batch and stream processing. In the case of AIHeal system based on rule based machine learning and deep learning artificial intelligence, an assessment of the quality of the data is necessary. The appropriate processes for anomaly detection, data cleansing, and handling of incomplete or missing data should have been applied to the dataset, and the existence of potential biases assessed and corrected.

— Choose most appropriate supervised thread, based on interpretable rules-based systems upgraded with visualization techniques and unsupervised thread based on deep learning ANN with two layers.

— Definition of acceptance criteria:

— Test that the AIHeal system satisfies security, privacy, and safety requirements applicable to medical devices

— Design:

— Implement the prototype covering workflow for cardiology domain. Design details are given in the concept description.

— Implementation:

— Tailor system to medical professionals expectations. System must offer interpretable results in which medical professionals can trust. The medical knowledge used in the construction of the AIHeal system cannot be proven clinically complete or objectively correct, but it must attempt to capture the current state of professional and scientific opinion. It must be possible to verify formally that the relevant medical knowledge satisfies certain requirements such as being unbiased, consistently interpreted, and reasonably completed.

— Quality assurance:

— Ensure that the quality of the AIHeal remains stable for internal and external updates and upgrades by new medical use cases.

ANALYSIS AND COMPARISON

Comparing to the existing solutions, the proposed one is more general. It not only exploits deep learning techniques for discovering hidden knowledge, but the recommendation to medical professionals are given in a set of directives based on commonly accepted reasoning, and the data feed from various resources including sensors is fed to the artificial intelligence model in a real time.

The additional innovative contribution of the proposed system is a combination of two AI concepts, where encoding clinical guidelines and/or existing clinical protocols, as in rules-based systems, provides a starting point, which then can be augmented by deep learning models that learn from data. The first AI concept makes that the decision explainable and the factors that are important to the algorithm's conclusion are visible which rise the confidence of the physicians and other practitioners. The second AI concept demonstrate the distinctive properties of AI, which also open new routes to delivering better, faster and more cost-effective care, and may have a greater focus on prevention and promoting wellness.

A patient wears specially designed mobile sensor, which is already available as a prototype, or implant. The data are transferred by mobile networks and collected by big data architecture in the proposed EHRs with any information relating to the past, present, or future physical/mental health of an individual. This data integration allows us to develop the „digital twin" of the patient.

The use case planned for implementation is an actual problem, ischemic heart disease, which affect majority of world and Montenegro population, more than 53% of all diseases.

It is intended to identify patients with highest risk and to react in proper time. AI based models can also screen people who are at low risk and do not have symptoms and help predict the progression of chronic coronary disease. The system can enable clinical pathways and protocols to be redesigned towards intervening proactively in the highest risk cases, even if they are asymptomatic. This could force the working patterns of practitioners from reactive care to proactive care.

How many lives could have been saved, if this kind of AI healthcare systems could have been used?

Physicians and other professionals who use AI will replace others who do not use AI. AI needs to serve as a decision-support tool. In the end, it is the decision of a doctor. AI is not going to substitute doctors in the foreseeable future. AI will be able to do some tasks that take a lot of time, need interconnection of many complex data, use mathematical formulas to determine the correct dosage of drugs, use numerous practical cases for data mining and analytics, what can free up physician time for concentration on diagnostic thinking and talking to patients.

Consequently, proposed system has the potential to transform healthcare organizations and healthcare services and meet the current and future society needs not only in Montenegro, but in many other countries.

CONCLUSION

The ground-breaking objective of the AIHeal project is to create explainable and causal AI healthcare system by combining the benefits of the accuracy of deep-learning algorithms with visibility on the factors that are important to the algorithm's conclusion in a way that is accessible to physicians, as in rules-based systems, enabled by richer data and improved computing.

The new EHR concept, big data infrastructure and naturally interpretable AI decision system are proposed.

The proposed AI Healthcare System has the potential to transform healthcare organizations and healthcare services and meet the current and future society needs not only in Montenegro, but in many other countries.

BIBLIOGRAPHY

[1] Mesko B, *A Guide to Artificial Intelligence in Healthcare*, The Medical Futurist, 2019.
[2] Densen P, *Challenges and opportunities facing medical education*, Transactions of the American Clinical and Climatological Association, 122, 2011.
[3] *Life expectancy at birth in the EU*, Eurostat, July 2019. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/DDN-20190725–1.2019 Revision of World Population Prospects, UN.
[4] *OECD Statistics Health expenditure and financing*.
[5] Porter, M., *Value-based health care delivery: Core concepts,* Value-Based Innovation Summit, Cleveland, OH, 2019.
[6] Milutinovic V, „*The best method for presentation of research results*," IEEE TCCA Newsletter, 1–6, 1996.
[7] Bojovic M and Bojic D, *MobilePDR: A mobile medical information system featuring update via Internet*, IEEE Transactions on information Technology in Biomedicine, 9(1): 1–3, 2005.
[8] Velasevic DM and Bojovic MM, *Hardware support to operations of relational algebra*, Proceedings of the Twenty-Second Annual Hawaii International Conference on System Sciences, Architecture Track, IEEE Computer Society, 1: 191–192, January 1989.
[9] Korolija N, Djukic T, Milutinovic V and Filipovic N, *Accelerating Lattice-Boltzman method using Maxeler dataflow approach*, The IPSI BgD Transactions on Internet Research, 34, 2013.
[10] Gnjatović M, Tasevski J, Mišković D, Savić S, Borovac B, Mikov A and Krasnik R, *Pilot corpus of child-robot interaction in therapeutic settings*, 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom), IEEE, 253–258, September 2017.
[11] Tasevski J, Gnjatović M and Borovac B, *Assessing the Children's Receptivity to the Robot MARKO*, Acta Polytechnica Hungarica, 15(5): 47–66, 2018.
[12] Kos A, Tomažič S, Salom J, Trifunovic N, Valero M and Milutinovic V, *New benchmarking methodology and programming model for big data processing*, International Journal of Distributed Sensor Networks, 11(8), 2015.
[13] Trifunovic N, Milutinovic V, Salom J and Kos A, *Paradigm shift in big data supercomputing: dataflow vs. controlflow*, Journal of Big Data, 2(1): 1–9, 2015.
[14] Montani S and Striani M, *Artificial intelligence in clinical decision support: a focused literature survey*, Yearbook of medical informatics, 28(1): 120, 2019.
[15] Shickel B, Tighe PJ, Bihorac A and Rashidi P, *Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis*, IEEE journal of biomedical and health informatics, 22(5): 1589–1604, 2018.
[16] Choi E, Bahadori MT, Searles E, Coffey C, and Sun J, *Multi-layer Representation Learning for Medical Concepts*, arXiv, 1–20, 2016.
[17] Van Der Maaten L and Hinton G, *Visualizing Data using t-SNE*, Journal of Machine Learning Research, 9: 2579–2605, 2008.

[18] Che Z, Purushotham S, Khemani R, and Liu Y, *Interpretable deep models for icu outcome prediction*, AMIA Annual Symposium Proceedings, American Medical Informatics Association, p. 371, 2016.
[19] Che Z, Purushotham S, Khemani R and Liu Y, *Distilling knowledge from deep networks with applications to healthcare domain*, arXiv preprint arXiv: 1512.03542, 2015.
[20] *Transforming healthcare with AI: The impact on the workforce and organisations*, McKinsey&Company, March 2020.
[21] Labrinidis A and Jagadish HV, *Challenges and opportunities with big data*, Proceedings of the VLDB Endowment, 5(12), 2012.
[22] Bojović M, Milutinović V, Bojić D and Korolija N, *An Experimental Healthcare System: Essence and Challenges*, Handbook of Research on Methodologies and Applications of Supercomputing, IGI Global, 301–308, 2021.
[23] M. Bojović, D. Bojić, N. Korolija, V. Milutinovic, *H*ybrid *M*edical *R*ecords *S*ystem *f*or *A*rtificial *I*ntelligence*: E*ssence and *C*hallenges, 1st Serbian International Conference on Applied Artificial Intelligence (SICAAI), Kragujevac, Serbia, May 19–20, 2022.
[24] Zikos, D., DeLellis, N. *CDSS-RM: a clinical decision support system reference model*. BMC Med Res Methodol 18, 137. https://doi.org/10.1186/s12874–018–0587–6, 2018.

Jakob SALOM*, Veljko MILUTINOVIĆ*

DIGITALIZACIJA NASLEĐA I FORUM TEHNOLOGIJA

Sažetak: Savremeni digitalni bibliotečki sistemi se suočavaju sa izazovima vezanim za analitiku podataka, interoperabilnost, kontrolu pristupa i saradnju institucija i korisnika. Ovaj članak predstavlja rad koji je obavljen na dodavanju vrednosti konvencionalnim tokovima rada digitalne biblioteke u bibliotečkom sistemu sledeće generacije. Članak započinje kratkom istorijom projekta Austria-Forum, a i opisuje njegovo kreiranje i osnovne osobine. Potom se analiziraju karakteristike interaktivnosti velikih internet lokacija i porede sa karakteristikama austrijskog foruma. Naime, internet lokacije predstavljaju informacije ili usluge i one retko dozvoljavaju korisnicima da daju povratne informacije ili postavljaju pitanja. Konstatuje se da trenutno korišćene povratne informacije skoro isključivo pripadaju kategoriji koju nazivamo „pokrenuti od strane provajdera" i koji nemaju veoma poželjnu mogućnost povratnih informacija koje nazivamo „korisničkim". Zatim se predstavljaju pojedini primeri povratnih informacija koje daje korisnik i opisuju se napredne karakteristike novog NID (Net-Interactive Documents) sistema kao dela Austria-Foruma.

Ključne reči: *Forum tehnologija, NID, napomene, dijalozi, diskusije, povratne informacije, interaktivnost, primedbe, pitanja, Austria-Forum, Serbia-Forum*

1. UVOD

Definicija: Austria-Forum (www.austria-forum.org) je besplatno dostupna onlajn zbirka referentnih radova na nemačkom i engleskom jeziku prvenstveno o temama vezanim za Austriju. Što se sadržaja tiče, to je digitalna baza podataka austrijske kulture, nasleđa i znanja (Austria-Forum — das Wissensnetz aus Österreich — austrijska mreža znanja).

Kratka istorija: Sve je počelo 1994. godine zahtevom tadašnje savezne vlade da se za 1000-godišnjicu Ostarrichia realizuje dostojan projekat, koji

* IPSI DOO, Beograd

bi bio interesantan za javnost. Novo izdanje Leksikona Austrije, koje je austrijska državna izdavačka kuća već duže vremena bila planirala i čije je prethodno izdanje u dva toma tada bilo staro skoro 30 godina, idealno je odgovaralo tim zahtevima.

Profesor Hermann Maurer sa Politehnike u Gracu predložio je da se ovaj novi austrijski leksikon objavi ne samo kao štampano, već i kao digitalno izdanje. On je smatrao da treba koristiti Hiper-G sistem, koji može da obezbedi modernu multimedijalnu obradu i prezentiranje materijala u kombinaciji sa elegantnim metodama pretraživanja.

Pošto je „stari“ austrijski leksikon, koji je dugo bio dostupan isključivo kao antikvarna knjiga, nastao uz učešće mnogih interesnih grupa, uključujući i političke, postojala je šansa da ovaj projekat (nazvan AEIOU) bude široko prihvaćen. I bio je i to toliko ubedljivo da je od strane ministarstva ocenjen kao jedini projekat vredan finansiranja. Za elektronsku verziju, akronim AEIOU treba tumačiti kao „Annotatable Electronic Interactive Austria Universal Lexicon“ (Elektronski interaktivni univerzalni leksikon Austrije sa komentarima), i stoga je on već i imenom predvideo da korisnici treba da budu u mogućnosti da interaktivno daju komentare. Godine 1996. kada se, pored novog dvotomnog štampanog izdanja Leksikona Austrije, mogla koristiti i elektronska verzija AEIOU, to je bila mala senzacija.

Ovom istraživačkom poduhvatu su se kasnije pridružili Academia Europaea, Matematički institut Srpske akademije nauka i umetnosti (MISANU) i projekat u Nemačkoj, svaki entitet sa svojim forumom. Poslednjih godina projekat Foruma usvojilo je nekoliko zemalja u Aziji, uključujući Australiju i Novi Zeland. Na taj način se i Forum tehnologija proširila i sada te baze sadrže dela kultura, nasleđa i znanja tih zemalja (institucija). To je opisano u specijalnom izdanju časopisa IPSI Transactions on Internet Research (januar 2020, izdanje 16), posvećenom digitalnom očuvanju kulturne baštine.

Forum tehnologija poseduje niz prednosti koje nemaju drugi slični sistemi za digitalno očuvanje kulturnog nasleđa. Neki od njih su inherentni i zauvek će biti suštinska karakteristika Forum tehnologije. Ostale karakteristike će biti jedinstvene za ovu tehnologiju dok ostali veliki sajtovi ne usvoje iste ili slične karakteristike. Za više detalja pogledajte reference [7...10].

Inherentne karakteristike uključuju:

(1) zaštitu od kopiranja (copyrights) koja uključuje skoro sve moguće mehanizme, a ne samo kreativni zajednički mehanizam, koji koristi Vikipedija (Creative Commons license); vlasnik sadržaja odlučuje, iz široke lepeze mogućih zaštitnih mehanizama, koji će koristiti za svaki konkretan slučaj;

(2) najvažnijih 10 artefakata svake institucije, koji postaju odmah vidljivi, bez mnogo napora za one koji traže sadržaje.

Karakteristike koje će drugi, verovatno, usvojiti u budućnosti:

(1) Pretraga bi se mogla učiniti semantičkom. Na primer, ako neko specificira ključne reči „fizika“, „Austrija“ i „Italija“ semantička pretraga vraća „Bolcman“ — fizičar koji je rođen u Austriji, a umro u Italiji. Pretraga sintakse bi vratila, na primer, stavke koje se odnose na saradnju Austrije i Italije u fizici.

(2) Verzioniranje se koncentriše samo na izmene koje uzrokuju značajnije promene entropije sadržaja (za razliku od vikiverzioniranja).

Sa odgovarajućom tehnologijom dodataka lako se mogu poboljšati mogućnosti Forum tehnologije, koja je prvobitno izgrađena u Javi, ali prihvata dodatke u bilo kojoj modernijoj tehnologiji. Jedan primer je NID, koji omogućava da se knjige i ostali materijali lakše proučavaju i da se više uživa u njima.

Mnoge veb-stranice, uključujući neke od veoma velikih kompanija, pružaju mnogo tekstualnih i multimedijalnih informacija, ali ne dozvoljavaju korisnicima da prijave otkrivene greške ili propuste, ili da postavljaju pitanja i daju komentare. Često se ne navodi ni mejl adresa (obično zato što se strahuje od lavine mejlova), a ako se navede broj telefona, korisnici se obično „zabavljaju“ mnogo minuta muzikom dok čekaju da se odgovori, ili im se daje duga lista često postavljanih pitanja (FAQ), koja dosta često ne sadrži pitanje za koje je korisnik zainteresovan. Roboti za automatizovane odgovore obično reaguju samo na standardne fraze, odnosno često nisu mnogo bolji od sistema „helpa“ (pomoći) sa funkcijom pretraživanja. Pojedini sistemi dozvoljavaju razgovore sa stručnjacima, ali obično nisu besplatni.

Naravno, postoje veb-stranice koje dozvoljavaju, čak i insistiraju da korisnici unose podatke. Tipični slučajevi su sistemi e-učenja, koji zahtevaju od korisnika da odgovore na pitanja ili da popune kvizove sa višestrukim izborima, kompjuterske igre (za jednu ili više osoba), koje zahtevaju interakciju korisnika i eventualno komunikaciju između korisnika, i sistemi koji očekuju od korisnika da ocenjuju/prihvataju ponuđene proizvode/usluge (npr. sistemi za rezervaciju hotela ili restorana, onlajn prodavnice itd.). Ovakav unos podataka/odgovora koji se zahteva od korisnika nazivamo interakcijom „pokrenutom od provajdera“.

2. INTERAKTIVNOST AUSTRIA-FORUM SAJTA

Interakcije korisnika mogu se kategorisati u više nivoa, a najosnovniji oblik su jednostavne interakcije jednim klikom kao što su sviđanja ili nesviđanja. Sledeći nivo interakcije mogu biti nekolike reči korisnika za komentare ili povratne informacije. Drugi, progresivniji, tip interakcije uključuje mogućnost zajedničkog kreiranja sadržaja od strane više korisnika (pomenute igrice). Neki konvencionalni informativni portali na vebu dozvoljavaju generičke oblike osnovnih i sekundarnih interakcija korisnika, ali samo onih koji nisu povezani sa kontekstom sadržaja.

Veoma je važno da se korisnicima dozvoli da daju komentare o određenim informacionim entitetima dostupnim na veb-stranici. Interakcije korisnika ili podaci korisnika smatraju se gorivom za pokretanje analitike zasnovane na veštačkoj inteligenciji (AI) na veb-lokacijama sledeće generacije. Mehanizmi poslovne inteligencije koji rade u pozadini većine modernih veb-lokacija za potrošačke usluge prikupljaju korisničke podatke implicitno. Podjednako je važno korisnicima Sajta omogućiti da kontrolišu unose podataka koji se daju sistemu.

Austria-Forum je implementiran 2007. godine [1] i danas je to mreža sa oko 1,4 miliona multimedijalnih unosa i skoro 4.000 digitalizovanih knjiga. Na njemu je registrovanim korisnicima dozvoljeno da dodaju komentare (napomene, beleške) na dnu svake stranice, gde ostaju zapisana njihova korisnička imena i na taj način omogućavaju administratoru da menja ili briše komentare ili kontaktira autore kada je to potrebno.

Tokom godina ova karakteristika je nekoliko puta zloupotrebljena. Uglavnom su to bili programi koji šalju komentare i reklame na veliki broj stranica. Zbog zloupotreba ova mogućnost je sada ograničena na članove Redakcije Austria-Forum. Da bi svi korisnici mogli da šalju poruke, svaka stranica sadrži „dugme za povratne informacije“, koje omogućava anonimne povratne informacije Administraciji Austria-Foruma. Anonimno znači da su pisci napomena nepoznati, odnosno, ako postave pitanje, ne mogu dobiti odgovor, osim ako u svojoj poruci ne navedu način na koji bi se kontaktirali. Ova anonimnost je sigurno smanjila trud koji je potreban za slanje poruka, ali ju je i dalje koristilo manje od 1% čitalaca. Mogućnost zloupotrebe je bila otklonjena, ali činjenica da poruku može da vidi samo administrator, dovodila je do toga da istu vrstu poruke (ispravku, žalbu...) šalje više osoba. Ipak, čak i relativno mali broj dobijenih poruka stvorio je prilično opterećenje za tim Austria-Foruma, jer članovi tima nisu mogli da odgovore na neke napomene koje su zahtevale kontaktiranje stručnjaka.

Tipični primeri velikih sajtova su hiljade zbornika sa konferencija i drugih knjiga, koje izdavačke kompanije čine elektronski dostupnim. Govoreći o ovom pitanju, zašto čitaocima ne dozvoljavaju da šalju napomene (obično pitanja) u vezi sa određenim prilogom, odgovor izdavača ja standardan i očekivan, i to u sledećem smislu: „To nam je prevelik trud. Kada dobijemo odgovor na pitanje čitaoca, moramo da pronađemo kontakt parametre autora. U mnogim slučajevima to nije trivijalno, jer su autori bili u kontaktu sa urednicima knjiga, a ne sa nama. Čak i ako pronađemo parametre kontakta, oni možda više ne važe; a čak i da važe, može potrajati dosta vremena da dobijemo odgovor koji onda moramo da prosledimo. Sve u svemu, ne možemo sebi priuštiti takav napor."

Da bi se izbegli pomenuti problemi, neki izdavači naučnih časopisa, kao što je jedan od najčitanijih naučnih informatičkih časopisa — Communications of the ACM, dozvoljavaju napomene na priloge koji se objavljuju tek nakon što ih recenzenti rada odobre. Ovo raspoređuje napor odgovaranja na pitanja na malu ekspertsku grupu, koja je bila recenzent i prihvatila rad za objavljivanje. Problem sa ovakvom povratnom informacijom korisnika je što se ona nalazi na kraju rada, tako da nije lako odrediti tačno mesto na koje se komentar odnosi. Urednici CACM-a objašnjavaju da se ova funkcija koristi manje nego što se očekivalo, ali da oni smatraju da će „pasivno čitanje" sve više biti zamenjivano „aktivnim čitanjem", tj. da sigurno dolazi vreme interakcija koje iniciraju čitaoci.

3. NET-INTERACTIVE DOCUMENTS (NID)

Uverenje da bi čitaočeve napomene u papirima i knjigama, postavljene tačno na relevantnom mestu u tekstu, bile pravo rešenje, dovelo je u Austria-Forumu do razvoja softverskog paketa NID (Net-Interactive Documents). Takođe, bilo je jasno da se mora razviti i neki mehanizam koji će smanjiti teret rešavanja tih napomena. Čini se da NID efikasno rešava pomenute probleme, a, osim toga, pruža i dodatne mogućnosti.

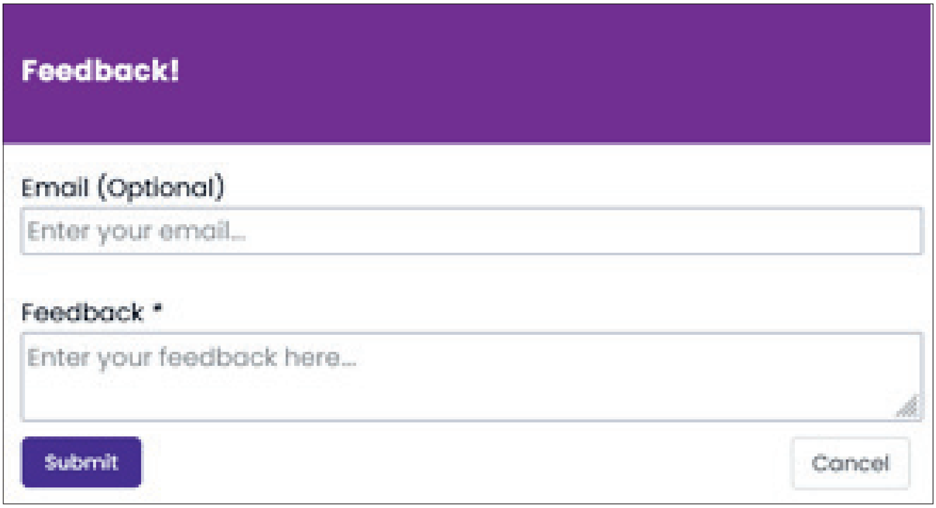
NID dokumenti su veličine od nekoliko listova do proizvoljno debelih knjiga. Svi oni se nalaze na NID serveru i sastoje se od stranica, predstavljenih kao slika prema standardu IIIF [2], koje se nižu jedna za drugom. To mogu biti skenirana dokumenta, koja se potom pretvaraju u PDF datoteke korišćenjem naprednih OCR tehnika. Shodno tome, ovo omogućava pretraživanje celog teksta u NID dokumentima, čak i teksta na slikama originalne knjige. NID knjige se takođe mogu proizvesti iz PDF ili Word datoteka ili PPT-a.

Što se tiče interakcije od strane korisnika, glavne tačke su:

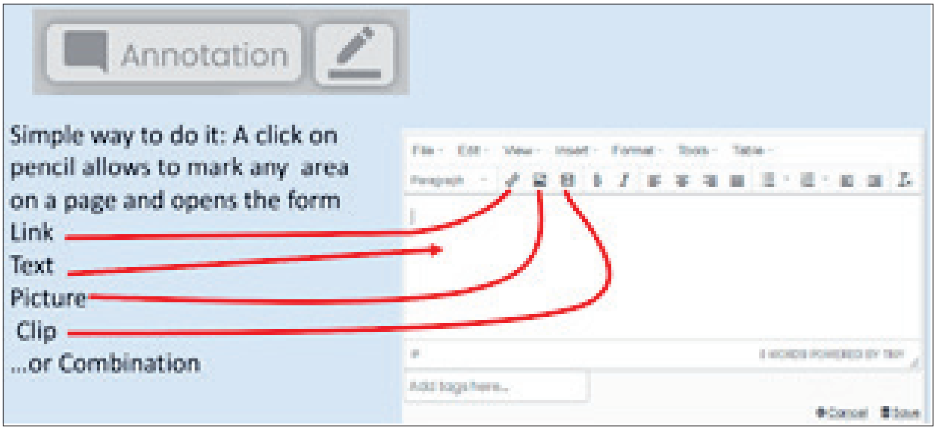
(1) Svaka stranica omogućava anonimnu povratnu informaciju (belešku na dnu stranice) korišćenjem obrasca prikazanog na Slici 1.

(2) Napomena se dozvoljava na bilo kojoj tački NID dokumenta. Potrebno je uzeti olovku sa linije menija i okružiti proizvoljnu (željenu) oblast na stranici. Ova radnja otvara obrazac koji omogućava umetanje teksta, linkova, slika ili drugih multimedijalnih objekata, korišćenjem forme prikazane na Slici 2.

Ključna tačka u NID-u je da se poruka sa novom korisnikovom beleškom/napomenom šalje jednoj ili više osoba, odgovornih za relevantni deo



Slika 1. Obrazac za povratne informacije dostupan je na svakoj stranici NID dokumenta

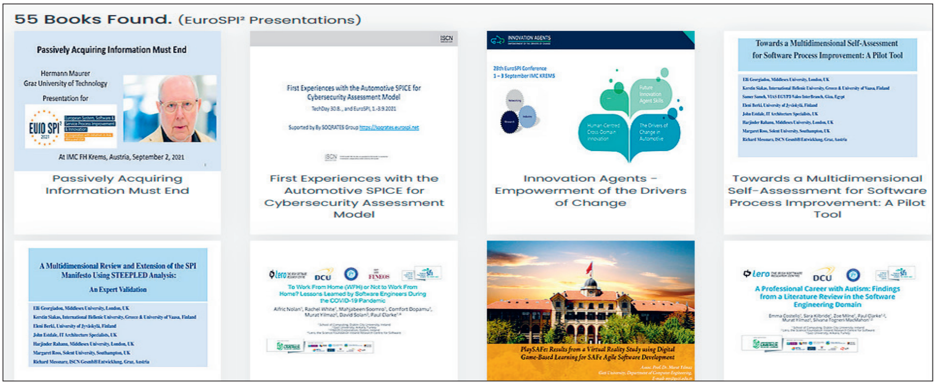


Slika 2. Ovaj obrazac omogućava umetanje različitih vrsta informacija na bilo koje mesto, na bilo kojoj stranici

dokumenta na koji beleška ukazuje. Time se ocenjivanje i modifikovanje napomene u potpunosti distribuira osobama odgovornim za predmetne stavke i ne smeta se urednicima ili izdavačima.

Evo kako je to urađeno za Zbornik EUROSPI 2021. konferencije [3]. Pedeset pet pojedinačnih radova sa konferencije (od kojih su pojedini prikazani na Slici 3) povezano je sa relevantnim autorima; pogledati: https://nid.iicm.tugraz.at/Home/Collections/31.

Prvu knjigu, čija je početna stranica prikazana na Slici 4, napisao je profesor Herman Maurer (pogledajte https://nid.iicm.tugraz.at/Home/



Slika 3. Osam od pedeset pet radova EUROSPI 2021. konferencije

Slika 4. Prva stranica NID dokumenta profesora Maurera sa EUROSPI 2021. konferencije

VievBook/411). Sve napomene koje su čitaoci dodavali u pokušaju da isprave ili dopune informacije na stranici, ili da postave pitanja išle su direktno autoru ovog rada, koji je njih hitno obrađivao. Na taj način su kasniji čitaoci odmah mogli da vide materijal koji je, reakcijama na ispravke i dopune čitalaca, dodatno uključen u originalni rad.

Ova knjiga je ovde izabrana jer daje i nezavisan pregled drugih karakteristika NID-a. U njoj su ukratko prikazane sve napredne mogućnosti koje ima NID sistem i preporučuje se svakome ko želi da za malo vremena sazna puno o ovom kvalitetnom i ne tako običnom softverskom paketu.

Funkcija beleške u NID-u nije važna samo za tehničke izveštaje. Ona se takođe uspešno koristi kod obuke studenata i srednjoškolaca svih starosnih nivoa. Oni se angažuju da pročitaju dokument sa revizijama, sprovedu zadato im istraživanje i dodaju nove informacije kao beleške. Čak su i učenici srednjih škola bili prilično animirani, što pokazuju stranice https://nid.kinderphilosophie.at/Home/VievBook/253. Primena NID-a može biti jako korisna za državnu upravu i sve vrste organizacija, delujući kao jednostavan sistem za upravljanje znanjem, ali i za mnogo više od toga, kao što je objašnjeno u [1].

NID nudi za svaki dokument mogućnost rukovanja beleškama na različite načine. Za NID dokument se može definisati da su one moguće i odmah dostupne javnosti čak i od strane anonimnih korisnika; ili da njih mogu da prave samo registrovani korisnici; ili ih može napraviti bilo ko, ali da su vidljive tek nakon potencijalnih izmena od strane autora testa; ili da su vidljive za sve, ili samo za određenu grupu korisnika, ili samo za onoga ko je napravio belešku, itd. NID nudi veoma fleksibilnu grupu struktura, koja omogućava npr. da različite grupe vide stranice sa različitim napomenama, ili im nije dozvoljeno da pogledaju neke stranice i sl.

Ugrađivanje eksplicitne mogućnosti interakcije korisnika na svim nivoima u platformu za objavljivanje sadržaja otvara pristup mnogim novim funkcijama. Takvi će sistemi u budućnosti koristiti sve vrste algoritama mašinskog učenja i veštačke inteligencije, kao i automatizovane alate za zaključivanje (donošenje zaključaka). Bitno je da se za potrebne održive tokove podataka za sistem koriste najbolje prakse od strane njegovih korisnika. Jedinstvena karakteristika NID-a je mogućnost kokreacije multimodalnog sadržaja u pravom kontekstu, u obliku beleški/napomena.

Još nekoliko činjenica u vezi sa korisničkim interfejsima. Prvo, kada se otvori NID dokument, može se izabrati jezik interfejsa (obično engleski ili nemački, ali su dostupni i drugi). Štaviše, složenost menija koji se prikazuje korisnicima može se prilagoditi željenoj grupi korisnika, pružajući ili pun meni, ili smanjenu verziju („pola meni"); ili veoma jednostavan meni samo za anonimne korisnike, koji mogu samo da okreću stranice, izvrše pretragu po celom tekstu, pogledaju sadržaj (ako je on dostupan, dostupan je klikom na svakoj stranici) i mali skup drugih akcija. Ako korisnik odluči da se prijavi, sistem će mu predstaviti sofisticiraniji meni.

4. SARADNJA I INTEROPERABILNOST

Digitalne biblioteke u interaktivnijim i međusobno povezanijim onlajn okruženjima suočavaju se sa dugoročnim izazovom saradnje korisnika i institucija. Dok se i bibliotekari i tehnolozi slažu oko činjenice da platforme za objavljivanje digitalnog sadržaja moraju imati intuitivna sredstva za promovisanje interaktivnosti korisnika i na nivou sistema i na različitim sistemima, najčešće postoje samo ograničene inherentne mogućnosti saradnje korisnika i interoperabilnosti podataka na različitim platformama. Istraživanja urađena da bi se pristupilo mogućnostima zajedničkih digitalnih bibliotečkih sistema otkrivaju da se ulažu napori da se promoviše standardizovano generisanje i potrošnja metapodataka za interoperabilnost i usluge udruživanja [4]. Većina sistema digitalnih biblioteka sve više koriste različite spoljne dodatke za društvene medije za promovisanje saradnje korisnika. Ove napore treba upotpuniti dodavanjem ugrađenih funkcija interakcije korisnika u digitalne biblioteke i standardnim interfejsima za programiranje aplikacija (API) za deljenje i korišćenje podataka.

Da bi se podržao razvoj sistema digitalne biblioteke u ovom pozitivnom pravcu, NID platformi su dodate sledeće mogućnosti saradnje i interoperabilnosti [5].

4. 1 Napomene, diskusije, kvizovi na javnom i grupnom nivou

Idući dalje od pasivnog potrošača informacija, korisnici NID biblioteke mogu da dodaju komentare ili da daju informacije u obliku napomena, komentara, povratnih informacija, kvizova, pa čak i da započnu diskusiju na nivou stranice dokumenta. Kokreacija sadržaja od strane korisničke zajednice takođe mora uključivati neophodne kontrole moderiranja. NID implementira ovu kontrolu, specificirajući ponašanje interakcije korisnika na osnovnom nivou, gde bibliotekar može odrediti da li objekat/dokument biblioteke dozvoljava anonimne interakcije ili samo prijavljeni registrovani korisnici mogu da doprinesu diskusijama i komentarima. Dodati sadržaji su dostupni opštim korisnicima bibliotečkog sistema, osim ako se moderiranje sadržaja bibliotečkih objekata ne sprovodi. Moderacija sadržaja se sprovodi navođenjem e-pošte moderatora bilo kog dokumenta u NID biblioteci. Ova kontrola moderiranja se može primeniti na nivou stranice, ali takođe omogućava dodavanje više moderatora sadržaja za određene delove dokumenta. Ova dodatna kontrola omogućava vidljivost doprinosa korisnika tek nakon odobrenja/aktivacije od strane moderatora ili urednika bibliotečkog sistema. Korisnici biblioteke mogu da kreiraju korisničke grupe i dodaju interakcije na dokumentima (napomene, kvizove i diskusije) koje su vidljive samo korisnicima predviđene grupe.

4. 2 Kompatibilnost

Besprekorna deljenja (Data sharing) podataka na nivou sistema su veoma kritična za održavanje robusnog i opsežnog informacionog ekosistema. Digitalne biblioteke moraju biti sposobne da dele i informacioni sadržaj pored postojeće razmene metapodataka. Bibliotečki sistem NID koristi standard IIIF da proširi besprekornu isporuku sadržaja i usluge udruživanja. NID predstavljanje podataka je zasnovano na jednostavnom principu dizajna, koji smanjuje upotrebu bilo kog vlasničkog formata podataka i specifičnih tehnologija. Koristi formate slika koji se mogu koristiti na bilo kojoj platformi i uređaju. Reprezentaciju slike IIIF dopunjuju podaci veza u JSON-LD formatu. Ovi standardni veb-podaci koriste otvoreni standardni format datoteke i format za razmenu podataka. Sadržaj originalnih podataka kao i korišćeni podaci za povezivanje (link data) predstavljeni su čitljivim tekstom.

NID sistem pruža načine za pristup, pregled, pretragu i deljenje objekata digitalne biblioteke u obliku slika, audio-zapisa i video-zapisa. Da bi podržao sinergiju digitalne biblioteke, NID sistem obezbeđuje više API koji omogućavaju deljenje informacija i agregaciju sadržaja. Ovo uključuje API za pretragu, API za prezentacije, API za slike i API za usluge podataka.

5. ZAKLJUČAK

Eksperimentalne studije [6] pokazuju da korisnici digitalnih biblioteka prepoznaju poboljšanja u njima koja su posledice modernih tehnologija. Svesni su poboljšanja mehanizama za pronalaženje informacija, upotrebljivosti i pristupnih interfejsa. Međutim, u isto vreme, korisnici osećaju potrebu za daljim poboljšanjima u skladu sa najsavremenijom naukom o podacima i informacionim tehnologijama. Uobičajene su tvrdnje o platformama digitalnih biblioteka otvorenog koda da one treba da budu proširive i da sve više dozvoljavaju upotrebu različitih tipova medija, standardnih metapodataka, sistema za upravljanje sadržajem i multimodalnog pristupa. Takođe se ulažu napori da se pokriju izazovna pitanja upravljanja pristupom autorskim pravima i implementacijom licenciranja. Brzo i pouzdano skladištenje i čuvanje podataka (hosting) su, takođe, vrlo visoko na agendama digitalnih biblioteka. Mnoge platforme za digitalno izdavaštvo počele su da nude sisteme za preporuke, zasnovane na filtriranju sadržaja. Takođe vidi se sve veća upotreba algoritama mašinskog učenja, koji nude različite analitike o bibliotečkim i korisničkim podacima. Iako sav ovaj razvoj pokazuje napredak, još uvek ima mnogo prostora za poboljšanja u domenima kao što su interoperabilnost, saradnja korisnika i institucija i upotreba veštačke inteligencije tokom kreiranja i potrošnje sadržaja.

Po svim ovim temama NID sistem Austria-Foruma je napravio ozbiljne iskorake. Ono čime se najviše bavi ovaj članak je potpuno nov način obezbeđivanja digitalnih knjiga ili dokumenata, koji omogućava korisnicima interakciju sa materijalom na nove načine. Tako je „pasivno proučavanje" nekog materijala zamenjeno u NID-u „proučavanjem informacija uz aktivno učešće“, a kao nusproizvod povećava se sadržaj znanja u originalnom dokumentu. Mnogi IT istraživači to vide kao budućnost konzumiranja digitalnog materijala.

Očekuje se da će razvoj NID-a duboko uticati na to kako će izgledati buduće digitalne biblioteke, ali inercija da se postojeći sistemi promene u nove zahtevaće nekoliko godina strpljenja.

BIBLIOGRAFIJA

[1] Maurer, Hermann: „Bericht zum Austria-Forum“, 2022, https://nid.iicm.tugraz.at/Home/ViewBook/562

[2] International Image Interoperability Framework, https://iiif.io/

[3] Maurer, Hermann; Zaka, Bilal; Eisenberger, Sonja: „Passively Acquiring Information Must End“, Proceedings Eurospi 2021, Springer CCIS 1442, 2021, 151–163.

[4] Verma, L., and Kumar, N.: „Comparative Analysis of OpenSource Digital Library Softwares: A Case Study“, DESIDOC Journal of Library & Information Technology, 38(5), 361–368, 2018. https://doi.org/10.14429/djlit.38.5.12425

[5] Zaka, Bilal: „Emerging Technologies in Digital Libraries: Net Interactive Document (NID) Experiences and Prospects“, IPSI Bgd Transactions on Internet Research, Vol. 18, No. 2, July 2022, pp. 70–76.

[6] Alokluk, J. A., & Al-Amri, A.: „Evaluation of a Digital Library: An Experimental Study“, Journal of Service Science and Management, 14, 2021, 96–114. https://doi.org/10.4236/jssm.2021.141007

[7] Zaka, B., Maurer, H., and Delilovic, N.: „Investigating Interaction Activities in Digital Libraries: The Networked Interactive Digital Books Project“, IPSI Transactions on Internet Research, Vol. 16, No. 1, January 2020, pp. 75–82.

[8] Delilovic, N., Ebner, M., Maurer, H., and Zaka, B.: „Experiences Based on a Major Information Server IPSI Transactions on Internet Research, Vol. 16, No. 1, January 2020, pp. 68–74.

[9] Eisenberger, Sonja: „The Incorporation of NID into Digital Libraries and Its Benefits“, IPSI Bgd Transactions on Internet Research, Vol. 18, No. 2, July 2022, pp. 77–81.

[10] Zaka, Bilal and Maurer, Hermann: „Web Sites Need Much More User-Driven Interactivity“, IPSI Bgd Transactions on Internet Research, Vol. 18, No. 2, July 2022, pp. 64–69.

Miodrag J. MIHALJEVIĆ[*]

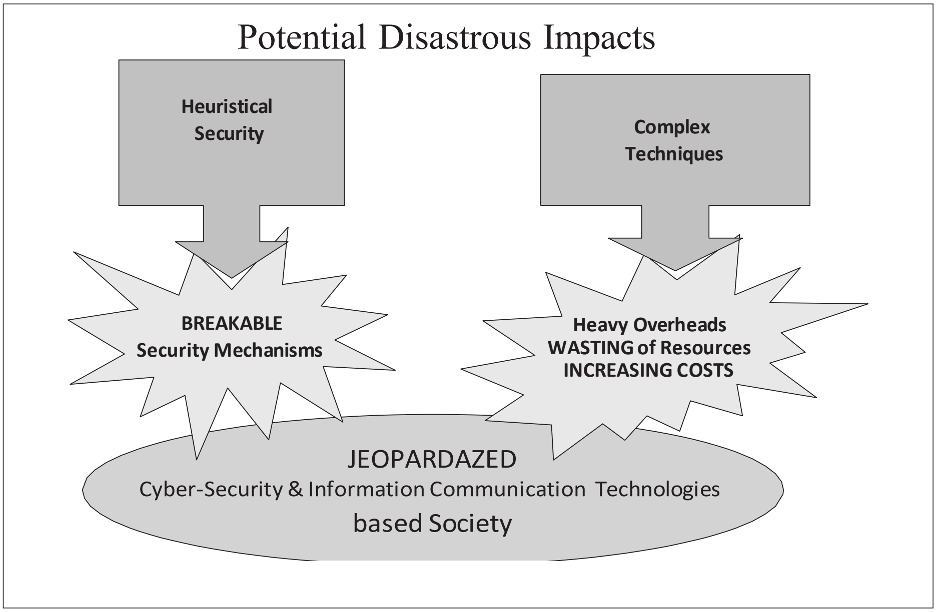
ILUSTRATIVNI NAPRECI U TEHNIKAMA KRIPTOLOGIJE I BLOKČEJN TEHNOLOGIJE

Sažetak: Cilj ovog rada je da ilustrativno ukaže na neke otvorene probleme i neke napredne tehnike u domenu informacione bezbednosti i blokčejn tehnologije. Kao ilustrativni problemi i povezana napredna rešenja razmatraju se domeni za zaštitu tajnosti i za redukciju energetske potrošnje u javnim blokčejn sistemima. Prikazuje se pristup za ojačavanje kriptografske sigurnosti postupaka šifrovanja niske složenosti i primena trgovine između potrebnih energetskih i memorijskih resursa u klasi blokčejn konsenzus protokola zasnovanih na tzv. dokazu o radu.

1. UVOD

Informaciono-komunikacione tehnologije (IKT) i digitalni (sajber) prostor su neraskidivo isprepletani sa našim fizičkim trodimenzionalnim prostorom. IKT i sajber prostor se stalno proširuju i pružaju nam nove pogodnosti. Blokčejn tehnologija i informaciona bezbednost su izuzetno bitne komponente i za IKT-e i u digitalnom prostoru da sve njihove pogodnosti ne bi postale i ulazna vrata za zlonamerne aktivnosti sa potencijalno katastrofalnim posledicama. Osnova za informacionu bezbednost je kriptologija, naučna disciplina formirana sredinom dvadesetog veka, a blokčejn tehnologija je desetak godina stara tehnologija zasnovana na nekim rezultatima kriptologije. Kriptologija daje osnove za zaštitu tajnosti i kontrolu integriteta, autentičnosti i neporecivosti. Kriptologija se razvila nad vekovima starim problemima šifrovanja i „razbijanja šifara". Interes za blokčejn tehnologiju počinje sa njenom prvom velikom primenom — bitkoin kriptovalutom. Cilj ovog rada je da ukaže na neke otvorene probleme i neke

[*] Dr Miodrag Mihaljević, naučni savjetnik, dopisni član Srpske akademije nauka i umetnosti (SANU), zamjenik direktora Matematičkog instituta SANU



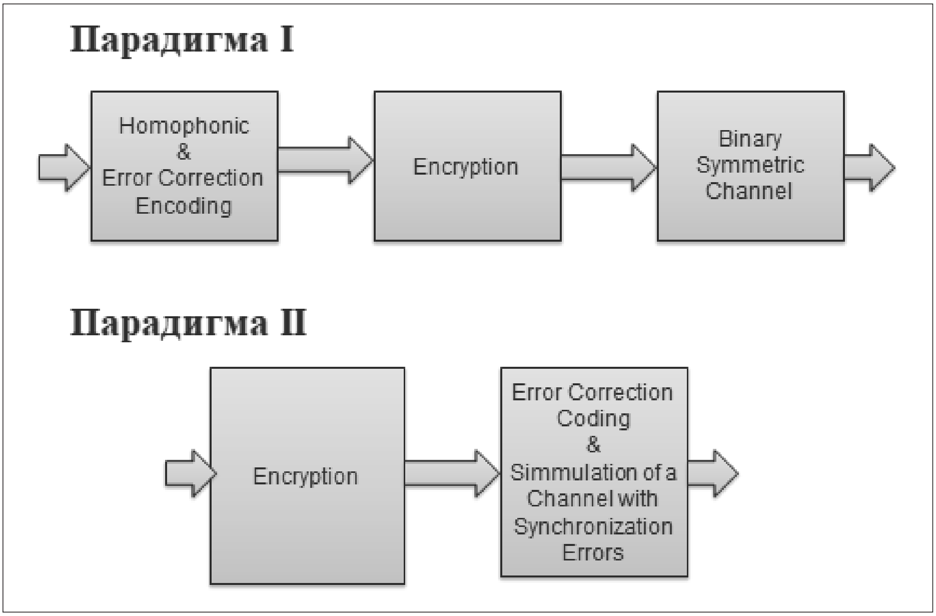
Slika 1. Ilustracija implikacija primene tehnika koje su samo heuristički sigurne i/ili koje imaju visoku implementacionu složenost

napredne tehnike u domenu informacione bezbednosti i blokčejn tehnologije. Kao na osnovni otvoreni problem ukazuje se na potrebu da tehnike koje podržavaju informacionu bezbednost i blokčejn istovremeno treba da obezbede željenu visoku sigurnost, ali i malo dodatno opterećenje funkcionalnosti sistema. Slika 1 ilustruje posledice neadekvatne sigurnosti i/ili složenosti tehnika koje se koriste u sajber prostoru.

Kao ilustrativni problemi i povezana napredna rešenja razmatraju se domeni za zaštitu tajnosti i za redukciju energetske potrošnje u javnim blokčejn sistemima. Prikazuje se pristup za ojačavanje kriptografske sigurnosti postupaka šifrovanja (enkripcije) niske složenosti i primena trgovine između potrebnih energetskih i memorijskih resursa u klasi blokčejn konsenzus protokola, zasnovanih na tzv. dokazu o radu.

2. INFORMACIONA BEZBEDNOST — OJAČANA KRIPTOGRAFSKA SIGURNOST ENKRIPCIJE PRIMENOM KODOVA ZA ISPRAVLJANJE GREŠAKA

Kao napredne tehnike enkripcije koje imaju visok potencijal primenljivosti i u blokčejn sistemima, u okviru ovog odeljka ukazuje se na pristup zasnovan na korišćenju kodova za ispravljanje grešaka za ojačavanje kriptografske



Slika 2. Dve paradigme za ojačavanje kriptografske sigurnosti date tehnike enkripcije primenom rezultata iz oblasti kodova

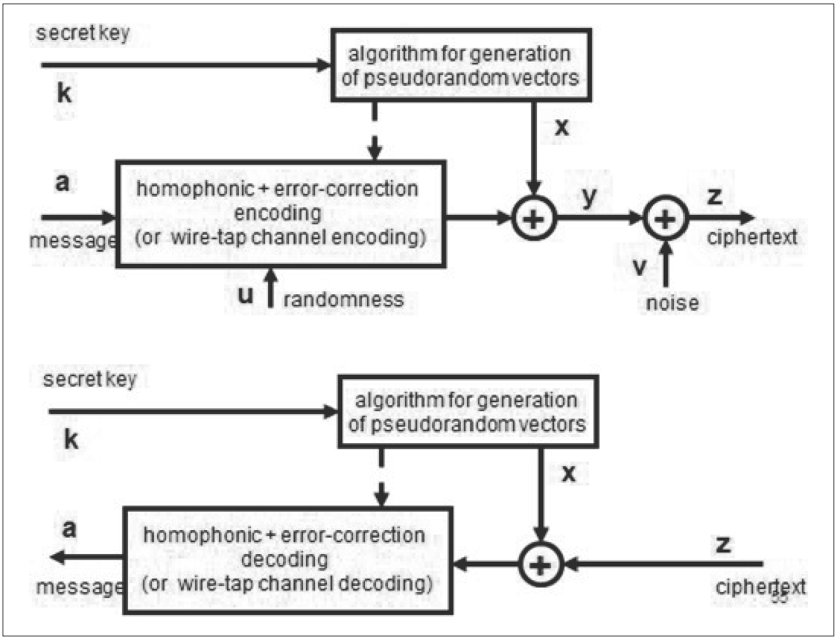
sigurnosti algoritama šifrovanja niske implementacione složenosti. Dva oblika ovog pristupa prikazana na Slici 2 su razmatrana u [1]–[5]. Paradigma I razmatrana u [1]–[3] zasniva se na primeni homofonskih i kodova za ispravljanje grešaka u binarnim simetričnim komunikacionim kanalima. Ovako kodovan ulazni vektor je predmet enkripcije, a dobijeni šifrat predmet degradacije aditivnim šumom, koji simulira binarni simetrični kanal u kome se svaki ulazni bit komplementira sa unapred zadatom verovatnoćom. Paradigma II razmatrana u [4] i [5] se zasniva na specijalnoj degradaciji inicijalno dobijenog šifrata, primenom simulatora komunikacionog kanala sa sinhronizacionim greškama u kome može da nastupi brisanje ili umetanje bita u inicijalno formirani šifrat.

Partikularni oblici Paradigmi I i II prikazani su na slikama 3 i 4.

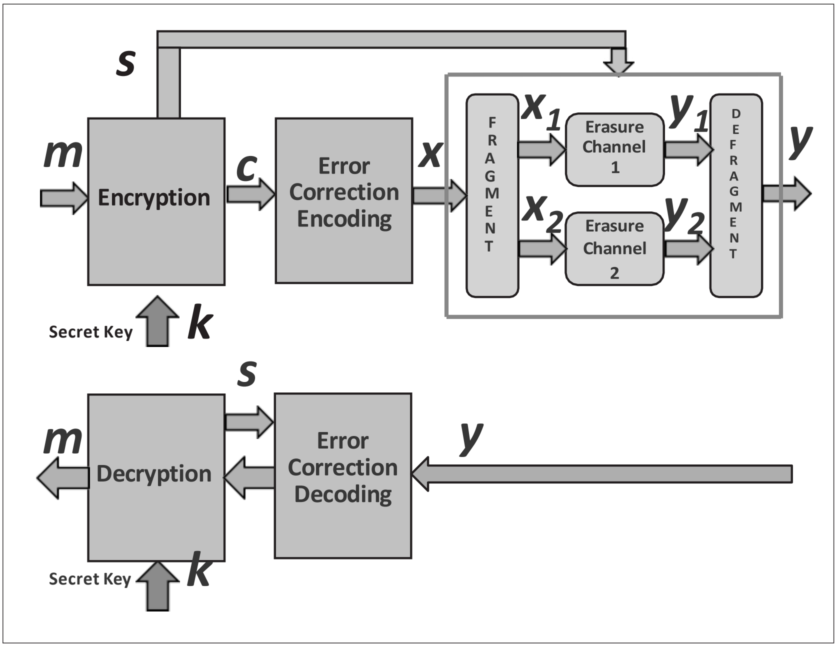
Na Slici 3 je data partikularna forma kriptografski ojačane sekvencijalne enkripcije saglasno Paradigmi I, koja je predložena i analizirana u [1]–[3].

Slika 4 prikazuje partikularnu formu kriptografski ojačane blok enkripcije saglasno Paradigmi II, koja je predložena i analizirana u [4] i [5].

Za detaljna tumačenja i razmatranja ovih predloga sugeriše izvorno upoznavanje na osnovu [1]–[4].



Slika 3. Partikularna forma kriptografski ojačane sekvencijalne enkripcije saglasno Paradigmi I prikazana i analizirana u [1]–[3]



Slika 4. Partikularna forma kriptografski ojačane blok enkripcije saglasno Paradigmi II prikazana i analizirana u [4]–[5]

3. BLOKČEJN TEHNOLOGIJA — KONSENZUS PROTOKOL SA SUŠTINSKI REDUKOVANOM POTROŠNJOM ENERGIJE

Blokčejn je distribuirani način dodavanja i čuvanja velikih količina podataka iz raznih domena tako da se dodaju samo verifikovani podaci, i jednom dodati podaci više ne mogu da budu menjani.

Suštinska svojstva su:

— da se verifikacija podataka koji se dodaju vrši bez postojanja tzv „treće strane od apsolutnog poverenja";

— da se nepromenjivost prethodno dodatih podataka ostvaruje primenom tehnika kriptologije.

Ukazuje se da su dve suštinske komponente svakog blokčejn sistema digitalna knjiga — blokčejn i konsenzsus protokol. Takođe, bitno svojstvo većine blokčejn sistema je da omogućavaju postojanje pametnih ugovora. U nastavku se sumiraju osnovna svojstva navedene tri komponente.

Digitalna knjiga — lanac blokova. Osnovna komponenta blokčejn tehnologije je distribuirana baza podataka, koja može da se modeluje kao digitalna knjiga (Ledger), u koju mogu da se dodaju strane, i svaka dodata strana je neraskidivo vezana sa svim prethodno dodatim stranama, koje ne mogu da se menjaju. Alternativno, ova knjiga se može posmatrati kao niz ulančanih blokova, odakle i naziv blokčejn.

Konsenzus protokol. Kao kompenzacija sa nepostojanje treće strane od arbitražnog poverenja, blokčejn tehnologija koristi konsenzus protokol za verifkovano ažuriranje digitalne knjige. Blokčejn konsenzus protokol je osnova za distribuirano ažuriranje.

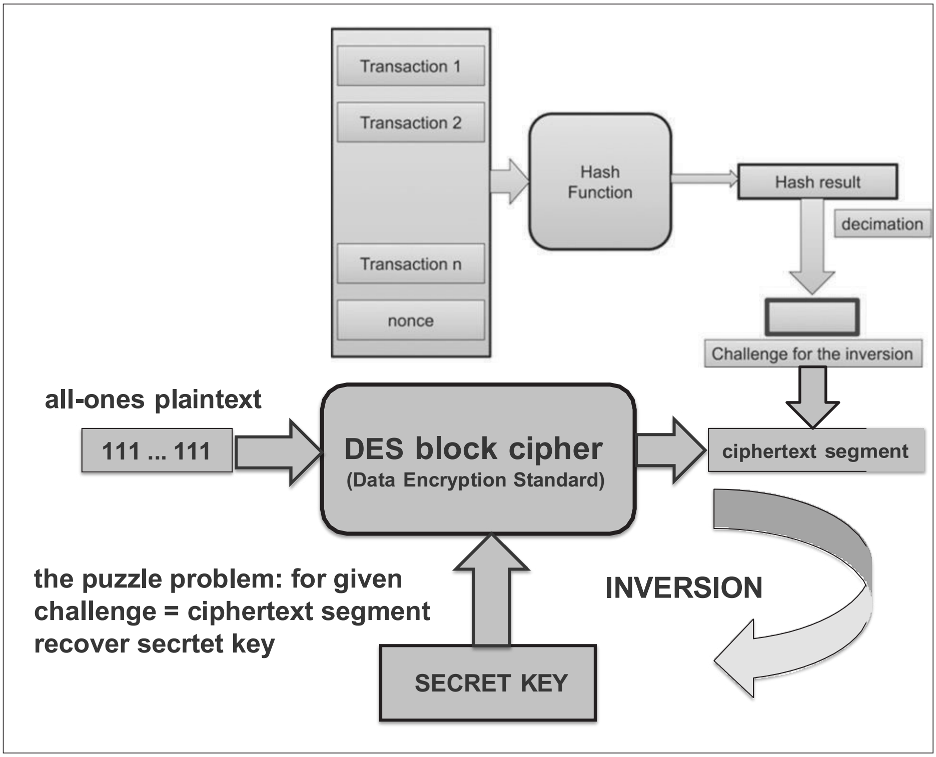
Pametni ugovori (Smart contracts) su programi kojima se vrši automatsko ugovaranje (mašina — mašina, bez direktnog ljudskog učešća) na osnovu specificiranih preferenci ugovornih strana. Pametni ugovori omogućuju postojanje veoma kratkotrajnih ugovornih obaveza i uspostavljanje ogromnog broja ugovora, a blokčejn paradigma obezbeđuje njihovu „overu" bez angažovanja treće strane od poverenja (notara).

Bitna klasa blokčejn konsenzus protokola je zasnovana na nekom kriptografskom problemu koji se naziva kriptografska slagalica.

Slika 5 pokazuje primer problema slagalice za konsenzus protokol predložen u [6]–[7]. Kao problem koji entiteti-verifikatori (rudai) rešavaju tokom izvršenja blokčejn konsenzus protokola postavljeno je sledeće:

— nepoznatim tajnim ključem primenom poznatog algoritma enkripcije generisati šifrat koji odgovara otvorenom tekstu koji se sastoji od vektora jedinica;

— na osnovu dobijenog šifrata (kao izazova) rekonstruisati tajni ključ kojim je šifrat generisan.



Slika 5. Konsenzus slagalica predložena i razmatrana u [6]–[7]

4. ZAKLJUČAK

Dat je jedan kompaktan rezime nekih naprednih pristupa od interesa za informacionu bezbednost i blokčejn tehnologiju objavljenih u [1]–[5] i [6]–[7], respektivno. Cilj je bio da se potencijalno zainteresovanim čitaocima da objedinjeni rezime izabranih tehnika: (a) za ojačavanje kriptografske sigurnosti primenom kodova za ispravljanje grešaka; (b) blokčejn konsenzus protokola koji omogućavaju redukciju energetskih resursa, koji se koriste u procesu verifikacije blokčejn transakcija.

LITERATURA

[1] F. Oggier and M. J. Mihaljevic, „An Information-Theoretic Security Evaluation of a Class of Randomized Encryption Schemes“, *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 2, Feb. 2014, pp. 158–168.

[2] M. J. Mihaljevic and F. Oggier, „Security Evaluation and Design Elements for a Class of Randomized Encryptions“, *IET Information Security*, vol. 13, no. 1, Jan. 2019, pp. 36–47.

[3] V. Mikhalev, M. J. Mihaljevic, O. Kara and F. Armknecht, „Selected Design and Analysis Techniques of Contemporary Symmetric Encryption“, in Security of Ubiquitous Computing Systems, Eds. G. Avoine and J. Hernandez-Castro, Springer, 2021, pp. 49–62.

[4] M. J. Mihaljevic, „A Security Enhanced Encryption Scheme and Evaluation of Its Cryptographic Security“, *Entropy*, vol. 21(7), July 2019 (11 pages).

[5] M. J. Mihaljevic, L. Wang and S. Xu, „An Approach for Security Enhancement of Certain Encryption Schemes Employing Error Correction Coding and Simulated Synchronization Errors“, *Entropy*, vol. 24(3), 406; March 2022 (10 pages).

[6] M. J. Mihaljevic, „A Blockchain Consensus Protocol Based on Dedicated Time-Memory— Data Trade-Off“, *IEEE Access*, vol. 8, Aug. 2020, pp. 141258–141268.

[7] M. J. Mihaljevic, L. Wang, S. Xu and M. Todorovic, „An Approach for Blockchain Pool Mining Employing the Consensus Protocol Robust against Block Withholding and Selfish Mining Attacks“, *Symmetry*, vol. 14 (8), 1711, Aug. 2022 (28 pages).

Miodrag J. MIHALJEVIĆ

ILLUSTRATIVE ADVANCES IN CRYPTOLOGY AND BLOCKCHAIN TECHNOLOGY TECHNIQUES

Abstract

The aim of this paper is to provide an illustrative address on some open problems and some advanced techniques in the fields of information security and blocking technology. The domains of secrecy protection and reduction of energy consumption in permissionless blockchain systems are considered. The following is pointed out: (a) An approach to strengthening the cryptographic security of lightweight encryption employing error-correction coding, and (b) An approach that employs trade-off between necessary energy and memory resources in the Proof-of-Work (PoW) based blockchain consensus protocols.

PRIMJERI DOBRE PRAKSE

EXAMPLES OF GOOD PRACTICE

Vesna MARAŠ*

VJEKOVI ZA BUDUĆNOST: CRNOGORSKO VINOGRADARSTVO I VINARSTVO

Sažetak: U periodu 2014–2017. godine multidisciplinarni tim, predvođen crnogorskim i španskim istraživačima, realizovao je najobimnije istraživanje genetičkih resursa vinove loze u Crnoj Gori. Istraživanje je obuhvatilo genetičku karakterizaciju postojećeg diverziteta vinove loze u gajenih i divljih populacija, uključujući i njihovu pedigre analizu.

U radu je predstavljeno bogatstvo genetskog nasljeđa vinove loze u Crnoj Gori koje i dalje stvara veoma bogat diverzitet vrste *Vitis vinifera*. Od 476 genotipiziranih uzoraka, pomoću SSR i SNP markera izdvojen je 101 genotip koji odgovara gajenim sortama (*Vitis vinifera* ssp. *sativa*), 43 genotipa su odgovarala divljim formama (*Vitis silvestris*). U klasteru gajenih sorti identifikovano je 50 poznatih sorti, za 51 neidentifikovani genotip smatra se da su novootkrivene, potencijalno autohtone crnogorske sorte. Rezultati istraživanja su potvrdili da se crnogorsko vinogradarstvo bazira na porodici genetički povezanih sorti, što je karakteristično i u drugim čuvenim tradicionalnim regionima za gajenje vinove loze i proizvodnju vina. U Crnoj Gori ova porodica je stvorena oko sorti kratošija i razaklija. U istraživanjima je otkriveno 25 punih pedigre sorti (triosi) i 27 duosa. Pronađen je puni pedigre vranca, koji je potomak kratošije i duljenge. Proces domestifikacije vinove loze započet još prije 8000 godina, u doba neolita, i dalje se nastavlja u vinogradarskim područjima Crne Gore. Pronađeni su predstavnici svih stepena domestifikacije vinove loze. Istorija crnogorskog vinogradarstva sačuvana je u njenim sortama, koje su kroz vjekove gajenja na ovim prostorima opstale i osigurale budućnost crnogorskog vinogradarstva. To nas obavezuje na njihovo očuvanje i potrebu da crnogorsko vinogradarstvo i vinarstvo sačuva svoju autentičnost.

Ključne riječi: *vinova loza, genetički resursi, diverzitet, sorta, autohtone sorte, kratošija, razaklija, vranac*

* Vesna Maraš, Kompanija „13. jul — Plantaže", Podgorica

UVOD

Crna Gora je odigrala važnu ulogu u dugoj istoriji gajenja vinove loze na zapadnom Balkanu. Njeni veoma različiti klimatski uslovi, tipovi zemljišta i orografija stvarali su različite uslove okruženja, koje je zajedno sa svojim istorijskim i geografskim kontekstom promovisalo stvaranje bogatog genetičkog diverziteta vinove loze. Potreba za karakterizacijom i očuvanjem genetičkih resursa vinove loze u tradicionalnim vinogradarskim područjima, favorizujući gajenje autohtonih i lokalnih sorti posljednjih decenija, intezivira se usljed globalizacije tržišta vina i klimatskih promjena. Grožđe od jedinstvenih domaćih sorti je osnova za proizvodnju vrlo prepoznatljivih vina sa stvarnim potencijalom da revitalizuju lokalnu vinsku industriju i istaknu se na tržištu, što je već postao trend u zemljama Starog svijeta. Pored toga, genotipizacija i analiza autohtonih i tradicionalnih sorti koristi se u utvrđivanju roditeljstva sorti[8–10], porijekla sorti, kao i načina širenja nekih sorti vinove loze u tradicionalnim vinogradarskim regionima.[20, 26] Dominantno mjesto u sortimentu Crne Gore imaju autohtone sorte vinove loze, koje se pominju još u XV vijeku (Srednjovjekovni statut Budve).[1] U Staroj Crnoj Gori kratošija je bila najrasprostranjenija sorta vinove loze sa arealom gajenja 100–150 km. Zbog heterogenosti njene populacije, koja je rezultat dugog gajenja na našim prostorima u novijim vinogradima, sortu kratošiju potisnula je sorta vranac, koja je, za razliku od nje, stabilna u ekspresiji svojih najboljih osobina. Kao rezultat toga, zasadi pod sortom kratošija su se smanjili, dok je vranac tokom XX vijeka postao najčešće gajena i najreprezentativnija sorta za proizvodnju crvenih vina u Crnoj Gori[4], i regionu. Pored ove dvije glavne vinske sorte, u Crnoj Gori gaje se i druge manje poznate autohtone sorte (krstač, žižak, bioka, čubrica), a prisutan je i veliki broj domaćih, odomaćenih i autohtonih sorti vinove loze, koje su nedovoljno poznate, manje zastupljene i nijesu u potpunosti identifikovane.[5] Za neke postoji određena identifikacija (ampelografski opis) pomoću kodova OIV-a, ali to često nije dovoljno za njihovu potpunu i sigurnu identifikaciju. Zato se pored ampelografsko-ampelometrijskih metoda, kao najpouzdanijih, koriste molekularno-genetičke metode, kojima utvrđujemo strukturu DNK. Genetička identifikacija sorte vinove loze se radi u cilju sigurne identifikacije sorte kako bi se pouzdano utvrdili sinonimi i homonimi sorte, njihovo porijeklo i nastanak — pedigre (utvrđivanje njihovog srodstva, roditelja sorti i njihove povezanosti). Zahvaljujući ranijim rezultatima rada na genetičkoj identifikaciji sorti, i nakon uspostavljene saradnje sa ICVV-om (Institut za vinovu lozu i vino, Logrono, La Rioha, Španija), nastavili smo sa daljim istraživanjima i proučavanjem sorti vinove loze u Crnoj Gori. Cilj naših zajedničkih istraživanja bio je usmjeren na genetičku identifikaciju i karakterizaciju postojećeg diverziteta vinove loze, gajenih i divljih populacija i definisanje značaja autohtonih i drugih odomaćenih sorti u crnogorskom vinogradarstvu, u cilju njihovog očuvanja, zaštite i bolje valorizacije. Kompletiranjem istraživanja crnogorskih sorti vinove loze, predstavljena je genetska istorija, sortiment i originalnost crnogorskog vinogradarstva, odnosno genetsko nasljeđe vinove loze u Crnoj Gori.

U radu su predstavljeni rezultati širokog istraživanja ugroženih genetičkih resursa u Crnoj Gori, uključujući gajene čokote iz starih vinograda i čokote koje rastu u divljini. Odabrani čokoti sorti su uzorkovani i genotipizovani pomoću SNP-a (polimorfizama jednostrukih nukleotida) i SSR markera (jednostavnih ponavljanja sekvenci), a jedinstveni genetski profili upoređivani su sa međunarodnim bazama podataka za pravilnu identifikaciju sorti i otkrivanje sinonima (različita imena za jednu sortu) i homonima (isto ime za različite sorte). Pored toga, analizirali smo genetsku raznolikost i strukturu vinove loze u Crnoj Gori i proučavali njihove genetske veze (pedigre analiza). Pomoću nekoliko hloroplastnih DNK markera utvrđen je tip hlorotipa u sorte (postoje četiri različita glavna tipa, A, B, C, D[14]), koji se nasljeđuju po majci. Hlorotipovi su korisni za određivanje ženskog roditelja u pedigreu i za prikaz višestrukog porijekla gajenih sorti vinove loze, analizom distribucije hlorotipova u lokalnom sylvestrisu i gajenoj vinovoj lozi.[15] Rezultati su otkrili porijeklo savremenih crnogorskih sorti od autohtonih i/ili introdukovanih sorti; postojanje anahronih prasorti u ovom regionu; događaje introgresije iz populacija divlje vinove loze, koji oblikuju genetsku strukturu pripitomljenih sorti vinove loze.

Ovo istraživanje je rezultat dugogodišnjeg rada kompanije „13. jul — Plantaže" i njenog sektora za razvoj, na čelu sa prof. dr Vesnom Maraš, i saradnje sa vodećim svjetskim naučnim centrima u Španiji, Kanadi i Sloveniji. Crnogorski istraživači su sa naučnim timom iz Instituta za vinovu lozu i vino iz Španije — ICVV (Logronjo, La Rioha), predvođenim prof. dr Joseom Miguelom Martínezom Zapaterom, uz učešće prof. Miodraga Grbića sa Univerziteta u Kanadi, prof. dr Nataše Štajner sa Univerziteta u Ljubljani, članova Nacionalnog udruženja vinogradara i vinara Crne Gore, proizvođača grožđa i vina u Crnoj Gori, realizovali ovo, po vinogradarstvo i vinarstvo Crne Gore, izuzetno značajno istraživanje.

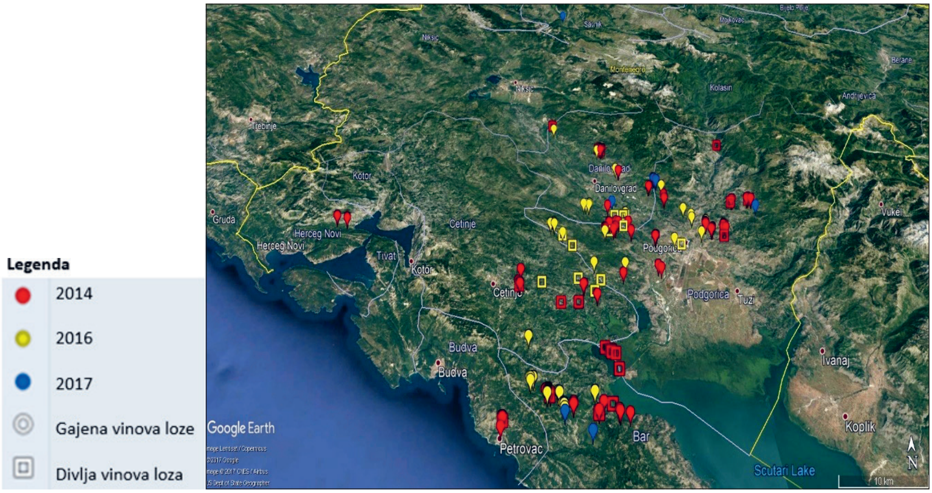
Kroz ovo istraživanje ostvarena je snažna saradnja privrede sa akademskim institucijama, Ministarstvom nauke i tehnološkog razvoja i Ministarstvom poljoprivrede, šumarstva i vodoprivrede Crne Gore. U projektu su bile uključene najznačajnije institucije iz oblasti nauke i visokog školstva u našoj zemlji, Crnogorska akademija nauka i umjetnosti (CANU), Univerzitet Crne Gore i Univerzitet Donja Gorica.

Multidisciplinarni tim predvođen crnogorskim i španskim istraživačima objavio je rezultate ovog najobimnijeg istraživanja vinove loze u Crnoj Gori. U radu[2] je predstavljeno bogatstvo genetskog nasljeđa vinove loze u Crnoj Gori, koje nesumnjivo stvara veoma bogat diverzitet u okviru vrste *Vitis vinifera*, obogaćujući domaće i evropsko vinogradarstvo i vinarstvo (https: //www. nature. com/articles/s41598–020–71918–7).

Ovaj rad predstavlja prvi korak u promociji šire upotrebe autohtonih i tradicionalnih sorti vinove loze u savremenoj crnogorskoj vinskoj industriji, doprinoseći diverzifikaciji i obogaćivanju evropskog vinskog sektora. Imajući u vidu koliki je značaj autohtonih sorti u vinogradarstvu i vinarstvu naše zemlje, ovo je istraživanje doprinijelo kompletnijem upoznavanju, valorizaciji i još boljoj komercijalizaciji crnogorskih sorti, uz snažnije promovisanje i pozicioniranje Crne Gore kao značajne i kvalitetne vinske destinacije na vinskoj mapi svijeta.

MATERIJALI I METODE

Kolekcija biljnog materijala: Analizirano je ukupno 476 čokota gajene i divlje vinove loze, starosti 50–300 godina sa afirmisanih vinogradarskih područja Crne Gore. Svaki analizirani čokot posjeduje svoj originalni kôd pod kojim se vodi, fotografije, lokalno ime (ime na terenu pod kojim je pronađen), a evidentirani su i ime vlasnika čokota kao i podaci za lokaciju (region, nadmorska visina, GPS koordinate). Tokom terenskih istraživanja,



Slika 1. Mjesta uzorkovanja gajene i divlje vinove loze

uzorkovano je 419 čokota (379 čokota gajenih u starim vinogradima — *Vitis vinifera* ssp. *sativa* i 45 čokota koji samoniklo rastu u prirodi, pod pretpostavkom da pripadaju divljoj lozi — podvrsti *Sylvestris*). U istraživanju je bilo uključeno i 57 uzoraka iz kolekcije vinove loze (*ex situ Vitis* kolekcija) sa Oglednog imanja Biotehničkog fakulteta, koja je podignuta 1956/60. godine od strane akademika Marka Ulićevića i njegovih saradnika.

Uzorci su prikupljeni u nekoliko terenskih ekspedicija u 2013, 2014, 2016. i 2017. godini. Mladi listovi sakupljani su na licu mjesta za svaki uzorak i čuvani u ledu do skladištenja na -80°C za DNK ekstrakciju i genotipizaciju. Pored toga, set od 57 uzoraka iz *ex situ Vitis* kolekcije na Biotehničkom fakultetu Univerziteta Crne Gore (BTF kolekcija) uključen je kao referenca u cilju identifikacije autohtonih sorti.

DNK izolacija, genotipizacija i sortna identifikacija. DNK je izolovana iz smrznutog lišća. Uzorci su u početku genotipizirani za osnovni set od 48 nuklearnih SNP markera[12], koristeći usluge genotipizacije SNP-a od strane Španskog nacionalnog centra za genotipizaciju (CEGEN), ili Jedinice za sekvenciranje i genotipizaciju Univerziteta Baskije, koristeći Fluidigm tehnologiju. Tri hloroplasta SNP-a, koja omogućavaju razlikovanje glavnih haplotipova hloroplasta vinove loze (A, B, C i D)[14] korišćena su za određivanje hlorotipa uzorka. Izdvojeni genetički profili (144) za 48 SNP-a upoređivani su sa SNP bazom podataka Instituta za vinovu lozu i vino (ICVV-SNP) za identifikaciju sorti, koja uključuje genetske profile za više od 2800 genotipova za 48 SNP-a. U onim slučajevima u kojima se SNP profil nije podudarao sa identifikovanom sortom u bazi podataka ICVV-SNP, uzorci su dodatno genotipizirani za 9 SSR markera (*VVS2, VVMD5, VVMD7, VVMD25, VVMD27, VVMD28, VVMD32, VrZAG62* i *VrZAG79*) u platformi za genotipizaciju Centra za biomedicinska istraživanja iz La Rioje (CIBIR). SSR profili su upoređeni sa SSR profilima dostupnim u VIVC[17] i u evropskoj *Vitis* bazi podataka[16]. Svih 144 genotipa genotipizovani su za dodatni set od 192 SNP markera[19], koristeći prethodno navedene platforme za genotipizaciju. Nažalost, nijesu postignuti zadovoljavajući rezultati za 13 genotipova, pa je 131 genetički profil vinove loze genotipizovan na 240 lokusa, korišćen za analizu strukture populacije i genetičkog diverziteta.

Analiza strukture populacije i parametara genetičkog diverziteta. Stablo razdvajanja sa neponderisanim susjedom (UwNJ) i analiza glavne koordinate (PCoA) izračunati su kako bi se istražila veza između gajenih i divljih genotipova vinove loze prikupljenih širom Crne Gore. U tu svrhu izračunata je matrica različitosti sa 10.000 'bootstrap' koraka pomoću softverskog paketa DARwin v. 6.0.21[20] za 131 genotip, uzimajući u obzir 194 SNP-a (SNP-ovi koji nijesu genotipizirani u najmanje 10 genotipova su odbačeni). Ova matrica različitosti korišćena je za analize PCoA i UwNJ, koja je napravljena na osnovu 1.000 'bootstrap' replika.

Zatim je korišćena Bayesova klaster metoda primenjena u STRUCTURE v. 2.3.4[21] softveru da bi se utvrdio broj genetskih grupa prisutnih u skupu od 131 genotipa vinove loze. Tako smo testirali postojanje genetske strukture, uzimajući u obzir brojne hipotetičke genetske grupe (K) u rasponu od 1 do 10, koristeći ciklus od 100.000 'burn-in' koraka, praćenih sa 150.000 Markov Chain Monte Carlo ponavljanja (MCMC metoda). Da bi se procijenila konzistentnost rezultata, izvedeno je 10 ponavljanja po K vrijednosti, pri čemu je svaki uzimao u obzir model sa koreliranim frekvencijama alela među populacijama. Najvjerovatniji broj genetičkih grupa je dobijen korišćenjem ΔK kriterijuma primijenjenih u STRUKTURI HARVESTER v. 0.6.94[23], a CLUMPP v. 1.1.2[24] je korišćen za svrstavanje 10 različitih ponavljanja. Genotipovi su dodijeljeni genetskoj grupi, s obzirom na prag $q \geq$ 0,50. Rezultati su grafički predstavljeni pomoću veb-softvera STRUCTURE PLOT v. 2.0.[25]

Parametri genetskog diverziteta (broj efektivnih alela (N_e), indeks informacija (I), uočena heterozigotnost (*Ho*), stvarna očekivana heterozigotnost (*UHe*) i koeficijent fiksacije (F)) izračunati su za genotipove dodijeljene dvjema genetičkim grupama identifikovanim sa STRUCTURE, odbacujući one identifikovane kao *Vitis* spp. interspecifične hibride zbog njihove *nevinifera* genetske pozadine. Da bi izbjegli efekat jedinki sa sličnim vrijednostima q predaka u obje genetske grupe, postavili smo prag vrijednosti q od 0,60 za grupno dodjeljivanje. Tako su gajene i divlje genetske grupe obuhvatale 82, odnosno 38 nesuvišnih genotipova. F_{ST} statistika je korišćena za analizu genetske distance između gajenih i divljih genotipova. Nepolimorfni markeri (5 i 26 za gajene, odnosno divlje podskupove) isključeni su za procjenu parametara raznolikosti. Proračuni su izvedeni pomoću GenAlEx v. 6.5.[26] Kao što je ranije naznačeno[19], srednje vrednosti ovih pokazatelja podvrgnute su analizi *t*-testa, kako bi se otkrile značajne razlike između obje grupe, koje se smatraju značajnim pri $p < 0.05$. Ova analiza je izvršena korišćenjem IBM SPSS Statistics v. 25.0 (Chicago, IL, USA).

Konačno, izvršili smo drugi krug hijerarhijske analize STRUCTURE u cilju procjene dodatnih nivoa genetske stratifikacije[27], fokusirajući se na konačnu genetičku grupu gajenih sorti vinove loze (91 genotip koji se ne ponavlja). Korišćen je isti postupak opisan ranije, ali kao prag pripadnosti grupi smo izabrali strožiju q vrijednost (0,70). Konzistentnost rezultata genetske strukture, dobijena u ovom drugom krugu, procijenjena je PCoA i UwNJ analizom, obavljenom kao što je prethodno detaljno opisano.

Analiza roditeljstva. Izdvojeni genotipovi vinove loze (144) su objedinjeni sa onima iz baze podataka ICVV-SNP da bi se upotpunili 1921 genotip podacima za 240 SNP-a i analizirani kako bi se otkrili mogući srodnički odnosi prvog reda (trio i parovi roditelj — potomstvo), primijenjenom metodom zasnovanom na vjerovatnoći u Cervus v. 3.0[71], kao što je prethodno detaljno objašnjeno.[15, 19] Vjerovatnoća svakog pronađenog trija i para roditelj — potomstvo (dvojac) procijenjena je, uzimajući u obzir prirodni logaritam ukupnog rezultata vjerovatnoće (LOD), razmatrajući maksimalni broj neusklađenih lokusa od 1 SNP-a za duose, odnosno 2 SNP-a za triose. Za svaki trios, hlorotipovi su korišćeni da bi se utvrdilo ko se od pretpostavljenih roditelja ponaša kao majka, prema majčinom prenosu hloroplasta u vinovoj lozi.[14]

REZULTATI I DISKUSIJA

1. GENETIČKA IDENTIFIKACIJA GAJENIH SORTI VINOVE LOZE I DIVLJE LOZE U CRNOJ GORI

SNP genotipizacijom svih 476 uzoraka na 48 lokusa potvrđena su 144 različita genetska profila, među kojima 101 genotip (68 uzoraka sa terena i 33 iz kolekcije BTF) odgovara gajenim sortama (*Vitis vinifera* ssp. *sativa*), a 43 genotipa su odgovarala divljim formama (*Vitis silvestris).* Od 101 genotipa koji odgovara gajenim sortama (*Vitis vinifera* ssp. *sativa*), 50 genotipova je bilo moguće identifikovati u bazama podaka, dok 51 genotip nije bilo moguće povezati sa drugim poznatim sortama. Od 50 identifikovanih sorti, 29 su poznate vinske sorte, 6 stonih sorti i 15 sorti sa dvostrukom namjenom vinka / stona sorta, prema Vitis internacionalnom katalogu sorti[17] (VIVC, www.vivc.de). Od 419 uzoraka/čokota sa terena najčešće pronađeni genotip odgovarao je sorti kratošija, koja je pronađena 107 puta, zatim vranac 76, lisica 35, razaklija 27, krstač 22 i bioka 8. Dvije najvažnije sorte za crnogorsko vinogradarstvo (kratošija i vranac)[4, 6] su dva najčešće pronađena genotipa u starim crnogorskim vinogradima, što potvrđuje njihovu važnost i značaj kroz istoriju za vinogradarstvo i vinarstvo Crne Gore. Važno je istaći da je od 45 identifikovanih sorti u kolekciji Biotehničkog fakulteta, veliki broj, 33 genotipa, sačuvan samo u kolekciji što ukazuje na brz nestanak starih sorti i ističe važnost ovih istraživanja u cilju očuvanja lokalnih genetičkih resursa od izumiranja. Veliki broj od 50 identifikovanih sorti trenutno se smatra autohtonim sortama sa zapadnog Balkana, kao što su bratkovina bijela, coarna alba, hrvatica, kratošija, prokupac, vranac ili žilavka, dok se druge sorte smatraju izvornim iz istočnih zemalja poput Grčke (Heptakilo, Karistino ili Muscat a aetits grains), Turske (Chaouch blanc, Kadarun ili razaklija), Libana (Afus ali), Jermenije (Krivalja bijela), Azerbejdžana (Sysak) i

Gruzije (Rkatsiteli). Pored toga, identifikovane su i sorte iz zapadnih zemalja poput Francuske (npr. Bicane, Cabernet franc i Merlot), Njemačke (npr. Mueller thurgau), Italije (npr. Malvasia bianca lunga) i Austrije (npr. Silvaner gruen), kao i oplemenjene sorte poput Angelo pirovano i Perlona i *Vitis* spp. interspecifični hibridi poput Isabelle i Varousseta. Ova identifikacija značajanog broja stranih sorti potvrđuje istorijske izvore, koji ukazuju na turbulentnu istoriju i događaje u Crnoj Gori, koji su omogućili unos sorti u zemlju iz različitih regiona u različito vrijeme i sa različitim ciljevima. Uvođenje sa istoka vjerovatno je posljedica širenja vinogradarstva iz regiona Bliskog istoka u zapadnoevropske zemlje, dok su novija uvođenja uglavnom pratila intenzivnu razmjenu sorti tokom druge polovine XIX vijeka u borbi protiv filoksere i drugih štetočina i bolesti u Evropi.

Rezultati ovog istraživanja otkrivaju i 51 neidentifikovani (nepoznat) genotip vinove loze, koji predstavljaju potencijalno nove sorte, i autohton crnogorski genetski materijal vinove loze. Devet od njih (crni krstač, belka, loza Svetog Vasilija Ostroškog, nepoznata bijela brijestovo, bijela krupna SR i dr.) je pronađeno na najmanje dvije lokacije, što potvrđuje da su one vegetativno razmnožavane (reznicama), da su gajene kao lokalne sorte i da su na ivici izumiranja. Zanimljivo je da je utvrđeno da su neki od neidentifikovanih genotipova potomci drugih poznatih lokalnih sorti (npr. belka je potomak razaklije i krstača, crni krstač je potomak kratošije i bioke), što ukazuje na njihovo crnogorsko porijeklo. Preostala 42 neidentifikovana genotipa pojavila su se samo jednom u uzorkovanju gajenih sorti i to 28 na vinogradarskim terenima Crne Gore (bijela nepoznata, bijela nepoznata Jelenak; crna loza; crna nepoznata, Crmnica; crna nepoznata, Masline; nepoznata bijela, Gornji Brčeli; nepoznata bijela, Gornji Morinj; nepoznata, Brijege; nepoznata, Kosić; nepoznata, Kuči; nepoznata, Piperi; nepoznata, Žabljak i dr.), a 14 je nađeno samo u kolekciji Biotehničkog fakulteta pod imenima bijela bezimena, bijela sitnih bobica, bijela slatka iz opeke, cetinka, crna tomba, ćelinac, duljenga, volovina iz Opeke, jasenka i dr. Neke od njih mogu predstavljati i autohtone sorte, na šta ukazuje genetička identifikacija potomaka za neke od ovih sorti (npr. crna nepoznata (Masline) potomak je sorte volovina iz Opeke). Dodatna posebnost crnogorskog vinogradarstva je trenutno prisustvo značajnog broja navodnih „prasorti“. To su čokoti koje su gajili samo lokalni proizvođači grožđa, tj. biljke koje direktno rastu iz sjemena ili su samo jednom razmnožene reznicama, od mjesta klijanja sjemena. One se mogu dalje umnožiti i distribuirati, i na taj način postati sorte. Na ovaj način je nastala većina sorti u prošlosti, ali ovaj proces više nije aktivan u zapadnoevropskim regionima.

Kada je u pitanju analiza populacije divlje loze *Vitis sylvestris*, svi uzorci (43) nose jedinstvene genotipove. Uzorci divlje MNE 273 pronađeni su na dvije strane puta (lokalitet Crmnica), što ukazuje na njihovo vegetativno razmnožavanje.

Na kraju, analize hlorotipa otkrile su većinu sorti koje nose hlorotipove C (43 genotipa, 42,6%) i D (33, 32,7%), zatim hlorotipove A (13, 12,9%) i B (3, 3,0%). Suprotno tome, većina biljaka uzorkovanih u divljini nosi hlorotip A (30 genotipova, 69,8%), zatim hlorotipove D (7, 16,7%) i C (6, 13,9%).

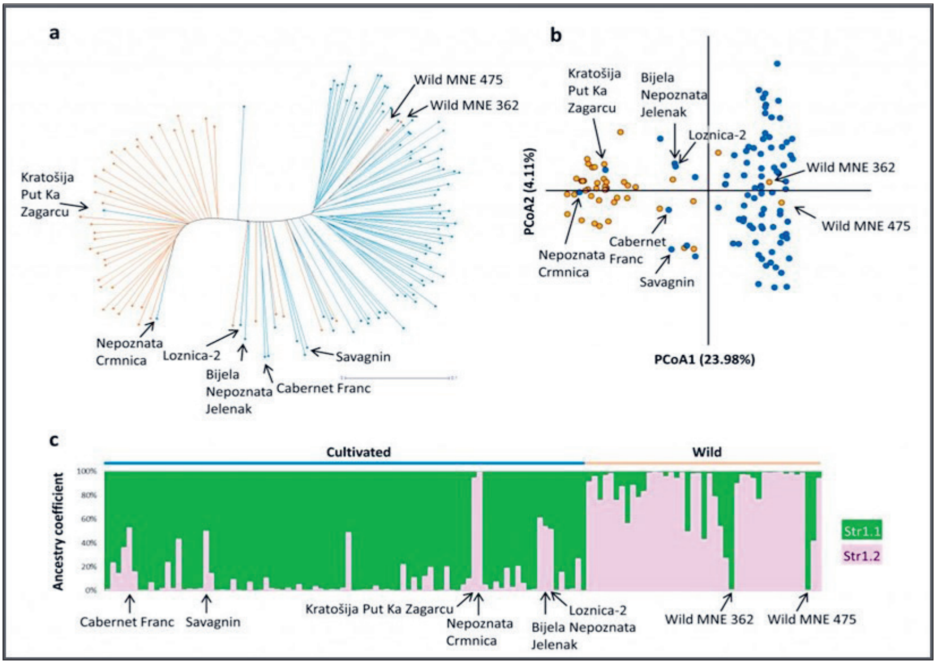
2. ANALIZA STRUKTURE POPULACIJE GAJENE I DIVLJE VINOVE LOZE U CRNOJ GORI

Izdvojenia 144 genotipa (101 gajena i 43 divlje loze) genotipizovano je na 192 dodatna SNP lokusa. Dobijeni su zadovoljavajući rezultati za 131 genotip, koji su korišćeni za dalju strukturu populacije, genetički diverzitet i pedigre analize. Stablo udaljenosti UwNJ (razdvajanja) generisalo je dva glavna klastera (Slika 2a), koja su uglavnom povezana sa porijeklom uzorkovanja genotipova obuhvaćenih analizom. Tako je jedan od klastera obuhvatio većinu genotipova prikupljenih kao divlja loza, dok je drugi obuhvatio većinu genotipova prikupljenih kao gajene biljke.

Zanimljivo je da su dva genotipa iz čokota gajenih u starim vinogradima (kratošija, put ka Zagaraču i nepoznata, Crmnica) smještena u klasteru divlje vinove loze, dok su dva genotipa sakupljena u divljini (divlja MNE 362 i divlja MNE 475) grupisani u klasteru gajene vinove loze. Rezultati su potvrdili ovu glavnu genetičku diferencijaciju, s tim što je PCoA1 jasno razdvajao genotipove prikupljene u divljini od onih prikupljenih kao gajene biljke (Slika 2b), potvrđujući i rezultate grupisanja za kratošiju (put ka Zagaraču), nepoznatu (Crmnica), divlju MNE 362 i divlju MNE 475.

Dva genotipa iz čokota uzorkovana kao gajene biljke (*sativa*) u starim vinogradima su sa genetičkog stanovišta potvrđene kao potpuno divlje — *sylvestris*, kratošija, put ka Zagaraču, i nepoznata, Crmnica. Ove biljke nose hlorotip A i jedinstvene genotipove bez podudaranja u konsultovanim bazama podataka, što ukazuje da potiču iz *Sylvestris* loze koja je vegetativno razmnožena.

Pronađen je drugi set gajenih biljaka sa sličnim procentom porijekla sa *Sativa* i *Sylvestris* genetskim podgrupama (poput bijela nepoznata, Jelenak, Loznica–2) i one su po svom postanku primjeri aktivnih mehanizama introgresije (prijelaz gena iz jedne vrste u drugu) genomskih regiona tj. iz *Sylvestrisa* u nove sorte. Ovi hibridi između dvije podvrste (*Sylvestris/Sativa*) vjerovatno su bili veoma važni u mnogim vinogradarskim oblastima, gdje su divlje loze dale važan doprinos genetskom fondu gajenih sorti. Svi ovi rezultati predstavljaju tragove prvih koraka u stvaranju novih sorti ili čak sekundarnu domestifikaciju divlje loze.



Slika 2: Analize različite genetske strukture vinove loze uzorkovane u Crnoj Gori kao gajena (plava) ili u divljini (narandžasta); a i b. strukturna analiza; c. svaki genotip (131) prikazan je kao vertikalna linija, sa segmentima boja čije su dužine proporcionalne njihovom pretpostavljenom pretku sa Str 1.1 i Str 1.2 (prikazane zelenom i rozom bojom).

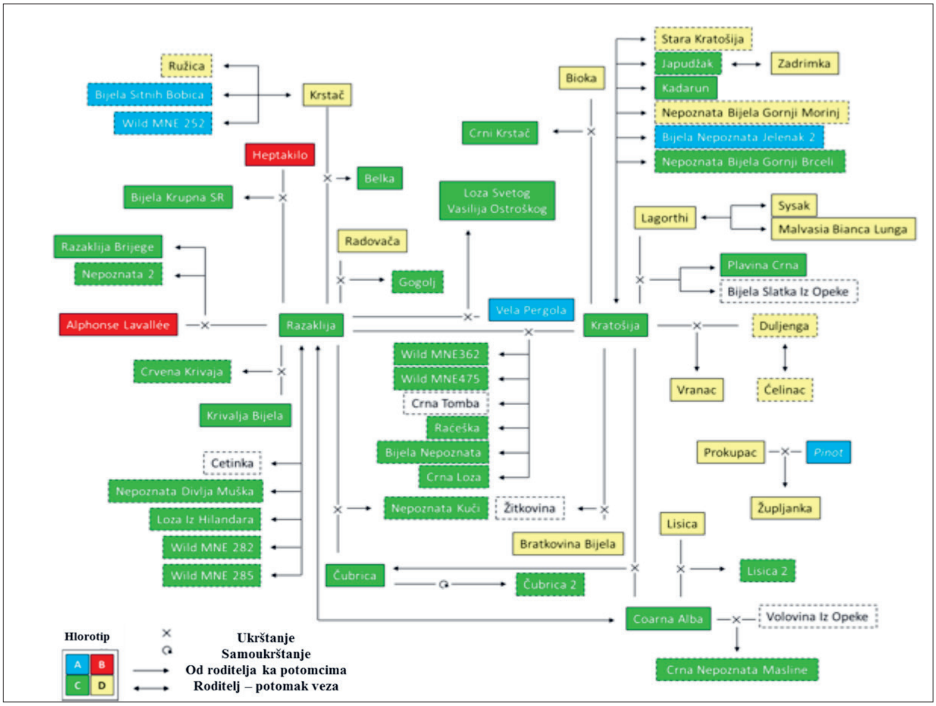
Pored toga, otkriveno je šest jedinki koje su uzorkovane kao divlja vinova loza sa većim *Sativa*, nego *Sylvestris* genetskim sastavom (četiri sa hlorotipom C). Dva genotipa čiji su uzorci sakupljeni u divljini, divlja MNE 362 (lokalitet Orasi) i divlja MNE 475, (lokalitet Meterizi), pokazuju izuzetno visoke koeficijente porijekla u genetskoj podgrupi gajenih sorti i kompatibilni su potomci kratošije i razaklije, što ih svrstava u podvrstu *Sativa*. Moguće je da potiču od sjemena proizvedenog u vinogradima, i raširenog u prirodnom okruženju od strane ljudi i životinja. Ostali genotipovi koji pokazuju visok procenat *Sativa* genetskog sastava (pripadaju genetskoj grupi Str 1.1) su: divlja MNE 361 (lokalitet Orasi), divlja MNE 476 (Podgorica), divlja MNE 285 (Rijeka Crnojevića) i divlja MNE 282 (Kamenik).

3. ANALIZA RODITELJSTVA GAJENE I DIVLJE VINOVE LOZE U CRNOJ GORI

U analizi roditeljstva 131 genetski profil vinove loze objedinjen je sa onima koji su sačuvani u bazi podataka ICVV-NP, za široku pretragu mogućih srodničkih odnosa prvog reda, koristeći 240 SNP profila. Dobijeno je 25 kompatibilnih trojki (majka + otac + potomak) i 27 duosa (roditelj — potomci)[2]. Pronađeni triosi (razaklija x heptakilo = bijela krupna SR; razaklija x Alphonse lavale = razaklija, Brijege; Razaklija x Alphonse lavale = nepoznata 2; razaklija x Krivalja bijela = Crvena krivaja; razaklija x krstač = belka; razaklija x radovača = gogolj; razaklija x Vela pergola = loza Svetog Vasilija Ostroškog; razaklija x kratošija = divlja MNE 362* lokalitet Orasi; razaklija x kratošija = divlja MNE 475* lokalitet Meterizi; razaklija x kratošija = crna tomba; razaklija x kratošija = raćeška; razaklija x kratošija = bijela nepoznata; razaklija x kratošija = crna loza; razaklija x čubrica = nepoznata, Kuči; kratošija x bratkovina bijela = žitkovina; kratošija x Coarna alba = čubrica; kratošija x duljenga = vranac; kratošija x Lagorthi = plavina crna; kratošija x Lagorthi = bijela slatka iz opeke; kratošija x bioka = crni krstač; Coarna alba x lisica = lisica 2; Coarna alba x volovina iz opeke = crna nepoznata, Masline; Prokupacx pinot = župljanka; Coarna alba x Chasselas = Clairette mazel ; čubrica 2 x čubrica = čubrica).

Pronađeno je 27 duosa (Wild MNE 275 lokalitet Lesendo — Wild MNE 276, lokalitet Lesendro; Wild MNE 363, lokalitet Orasi — Wild MNE 364, lokalitet Orasi; kratošija — stara kratošija; Chaouch blanc — čauš crveni; Muscat a petits grains — Muscat rouge de madere; kratošija — nepoznata bijela, Gornji Morinj; zadrimka — japudžak; kratošija — kadarun; Koenigin der weingaerten — Afus ali; kratošija — bijela nepoznata, Jelenak 2; krstac — Wild MNE 252, lokalitet Danilovgrad; ćelinac — duljenga; kratošija — nepoznata bijela, Gornji Brceli; razaklija — loza iz Hilandara; razaklija — cetinka; kratošija — japudžak; Angelo pirovano — Muscat hamburg; Silvaner gruen — Savagnin; Malvasia bianca lunga — Lagorthi; krstac –ružica; razaklija — Coarna alba; Lagorthi — Sysak; điritkinja — Karystino; bijela sitnih bobica — krstac; razaklija — Wild MNE 285, lokalitet R. Crnojevića; razaklija — nepoznata divlja muška; razaklija — Wild MNE 282, lokalitet Kamenik.

Rezultati istraživanja ukazuju na vodeću ulogu razaklije i kratošije u stvaranju crnogorskog genetičkog diverziteta vinove loze, koji su uključeni kao roditelji u 14 (razaklija), odnosno 12 (kratošija) pedigrea. Zajedno imaju šest potomaka, uključujući dva genotipa uzorkovana kao divlja vinova loza (divlja MNE 362 i divlja MNE 475). Pored toga, od 27 duosa potvrđeno je šest kompatibilnih veza roditelja i potomaka za kratošiju i šest za razakliju od kojih su dva genotipa uzorkovana kao divlje vinove loze (divlja MNE 282, i divlja MNE 285). Pored toga, identifikovan je jedan gajeni genotip (čubrica-2) kao kompatibilni rezultat samoukrštanja sorte čubrica. Među divljom lozom nijesu pronađeni puni trojci, ali su otkrivena dva pouzdana



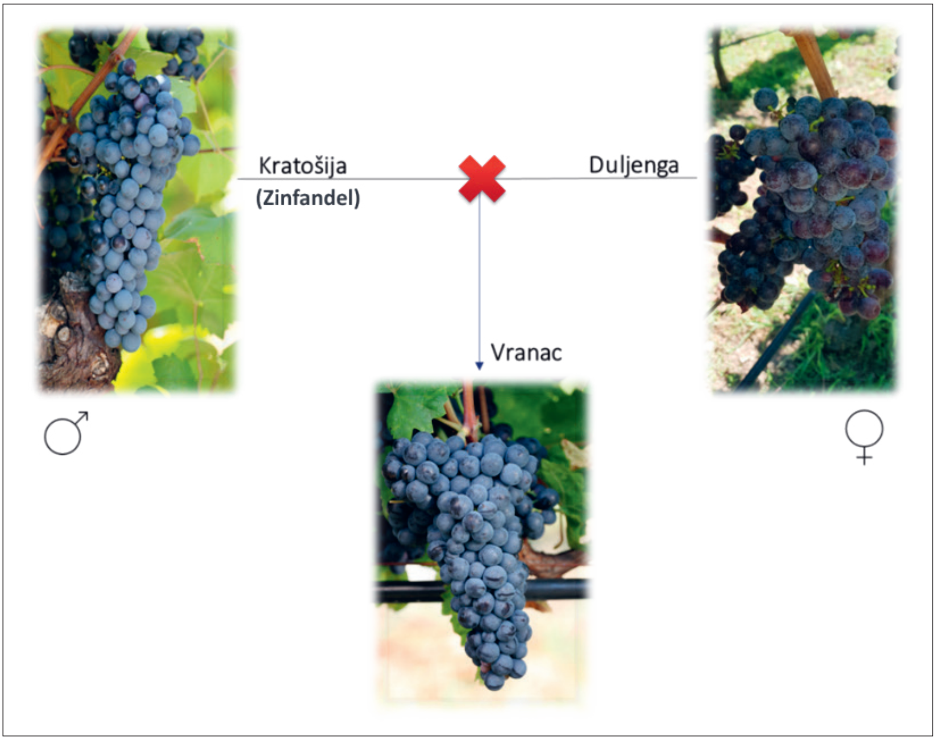
Slika 3. Prikaz pedigrea i genetičke povezanosti ispitivanih gajenih i divljih formi vinove loze u Crnoj Gori. Hlorotipovi (A, B, C ili D) su označeni različitim bojama, prema kodu. U slučaju bijele boje, informacije o hlorotipu nijesu dostupne. Neidentifikovani i jedinstveni genotipovi u bazi podataka ICVV-SNP prikazani su u okvirima s izlomljenim granicama.

duosa u dvije različite populacije, u tvrđavi Lesendro (koja uključuje divlju MNE 275 i divlju MNE 276) i Orasima (uključujući divlju MNE 363 i divlju MNE 364).

Rezultati istraživanja su potvrdili da se crnogorsko vinogradarstvo bazira na porodici genetski povezanih sorti što je karakteristično i u drugim čuvenim tradicionalnim regionima za gajenje vinove loze i proizvodnju vina. U Crnoj Gori ova porodica je uglavnom stvorena oko sorte kratošija i razaklija (Slika 3).

Kratošija je u središtu crnogorskog vinogradarstva kao što su Cabernet franc i Pinot noir u središtu Bordoa i Burgundije, čuvenih vinogradarskih regiona u Francuskoj. Kratošija je i najstarija sorta, koja se gajila u Crnoj Gori, ima skoro 20 potomaka u Crnoj Gori i roditelj (otac) je sorte vranac. Pored vranca, neki od potomaka kratošije u Crnoj Gori su čubrica, raćeška, žitkovina, crni krstač, crna loza, crna tomba i dr.

Sorta vranac se danas smatra starom autohtonom crnogorskom sortom vinove loze. Od ranije je poznato postojanje odnosa prvog stepena između kratošije i vranca.[13] Pedigre analiza omogućila je i otkrivanje genetskog porijekla vranca.[2] Rezultati pouzdano podržavaju da je vranac nastao hibridizacijom između sorti duljenga i kratošija. Ovaj pedigre je takođe potpomognut dodatnim setom od 20 SSR markera, a analize hlorotipa identifikovale su duljengu kao ženskog roditelja vranca i utvrdili da je potomak duljenge i kratošije, te na najbolji način potvrdili njegovu autohtonost i porijeklo (Slika 4).



Slika 4. Pedigre analiza sorte vranac (roditelji sorte vranac: kratošija i duljenga)

Značajno je navesti i utvrđivanje genetike loze iz Manastira Ostrog (loza Svetog Vasilija Ostroškog), koja datira prema pisanim podacima iz 1672. godine, što je jedan od najstarijih pisanih podataka o vinovoj lozi u Crnoj Gori i na Balkanu. Istraživanje je potvrdilo da je ova loza rezultat ukrštanja sorte razaklije sa sortom vinove loze pod imenom kosoranka (nađenom u blizini manastira), koja je ko Vela pergola registrovana u Vitis katalogu.

4. FORMIRANJE NACIONALNE KOLEKCIJE CRNOGORSKIH AUTOHTONIH I ODOMAĆENIH SORTI VINOVE LOZE

U cilju očuvanja i valorizacije genetskih resursa vinove loze u Crnoj Gori, sa potvrđenim genotipovima — sortama, podignuta je nacionalna kolekcija crnogorskih autohtonih i odomaćenih sorti vinove loze na Ćemovskom polju. Kroz učešće u Evropskom kooperativnom programu za očuvanje biljnih genetiških resursa (ECPGR), za rod Vitis sorte su predstavljene i uključene u evropske Vitis baze: http: //www. eu-vitis. de/index. php i VVIC (Vitis International Variety Catalogue) — https: //www.vivc.de/. Dalje ispitivanje i karakterizacija njihovog agrobiološkog i enološkog potencijala će doprinijeti razvoju specifičnog crnogorskog vinogradarstva, zasnovanog na autohtonim sortama, obogaćujući asortiman crnogorskih vina u budućnosti. U cilju predstavljanja dobijenih rezulata domaćoj i međunarodnoj naučnoj i stručnoj javnosti, u saradnji sa španskim partnerima u 2017. godini, organizovali smo Prvu međunarodnu konferenciju o vrancu i drugim crnogorskim autohtonim sortama vinove loze u Podgorici. Rezultati koji su iznijeti na ovoj eminentnoj konferenciji, uz učešće naučnika iz 17 država, definisali su Crnu Goru kao jedan od centara diverziteta vinove loze na Balkanu, koja može dati doprinos globalnom diverzitetu vinove loze, istovremeno je pozicionirajući kao specifičnu i originalnu kvalitetnu vinsku destinaciju. Suorganizator konferencije je bila i Crnogorska akademija nauka i umjetnosti koja je štampala zbornik radova sa konferencije.[22]

ZAKLJUČAK

Ovo genomsko istraživanje omogućilo je evidenciju i katalogizaciju genetskog diverziteta vinove loze u Crnoj Gori i predstavilo detaljan pedigre — porijeklo sorti i njiihove međusobne genetske povezanosti. Višegodišnje istraživanje je pokazalo da se proces pripitomljavanja vinove loze, koji je započet još prije 8000 godina, u doba neolita, i dalje nastavlja u vinogradarskim područjima Crne Gore. Pronađeni su predstavnici svih stepena domestifikacije vinove loze. Istraživanje je otkrilo sortnu strukturu i vinogradarsku istoriju Crne Gore, potvrđujući da je diverzitet crnogorskih sorti mnogo bogatiji u poređenju sa okolnim balkanskim zemljama. Dio ovako velikog diverziteta ukazuje i potvrđuje i samu lokalnu istoriju, što ukazuje i na višestruka i intenzivna unošenja vinove loze, iz različitih vinskih regija, u različitim istorijskim periodima. Otkriven genetski fond uključuje mnoge autohtone sorte, neke na ivici izumiranja, povezane u složenu rodbinsku mrežu, gdje su dvije sorte kratošija i razaklija igrale glavnu ulogu u stvaranju autohtonih sorti. Od ispitivanih čokota/uzoraka, pomoću SSR i

SNP markera identifikovano je 144 različitih genetskih profila, među kojima je 101 genotip odgovarao gajenim sortama (*Vitis vinifera* ssp. *sativa*), a 43 genotipa su odgovarala divljim formama (*Vitis silvestris).* U klasteru gajenih sorti (101) identifikovano je 50 poznatih sorti. Pronađen je 51 neidentifikovani genotip vinove loze, koji predstavlja nepoznate sorte za vinogradarsku nauku. Smatra se da se radi o novootkrivenim, potencijalno autohtonim crnogorskim sortama vinove loze.
Rezultati istraživanja su potvrdili da se crnogorsko vinogradarstvo bazira na porodici genetički povezanih sorti, što je karakteristično i u drugim čuvenim tradicionalnim vinogradarskim regionima. U Crnoj Gori, ova porodica uglavnom je stvorena oko vinske sorte kratošija i stone sorte razaklije, koje su roditelji mnogobrojnih crnogorskih autohtonih sorti. Pronađeno je 25 punih pedigre sorti (triosi) i 27 duosa. Razaklija je ženski roditelj u 14 punih pedigrea, dok je kratošija u 12, a zajedno imaju 6 potomaka. Kratošija je roditelj (trio i duo odnosi) u skoro 20 genotipova vinove loze gajenih u Crnoj Gori. Pored vranca, neki od potomaka kratošije u Crnoj Gori su čubrica, raćeška, žitkovina, crni krstač, crna loza, crna tomba. Sorta kratošija je u ovim istraživanjima najčešće pronađen genotip u starim crnogorskim vinogradima (106 puta) što, pored ostalog, potvrđuje njenu važnost u istoriji crnogorskog vinogradarstva. To je najstarija sorta vinove loze, koja se gaji u Crnoj Gori i vjerovatno je nastala na ovim prostorima. Kratošija je najstarije ime za sortu vinove loze, koja se gaji u inostranstvu pod imenom Primitivo/Zinfandel — crljenak kaštelanski.

U istraživanjima je pronađen puni pedigre vranca, koji je potomak kratošije i duljenge. Pronađen je i puni pedigre loze iz Manastira Sveti Vasilije Ostroški, to su razaklija i sorta volovnik/kosoranka, pronađena u okolini manastira, koja je u Vitis katalogu registrovana pod imenom Vela pergola.

Potvrđeni veliki diverzitet vinove loze u Crnoj Gori, bogatstvo sortimenta i genetskog nasljeđa ukazuje na dugu tradiciju gajenja vinove loze u našoj zemlji kao i na potrebu njegovog očuvanja. Ova istraživanja su omogućila valorizaciju genetskog nasljeđa vinove loze u Crnoj Gori, pronalaskom 51 nepoznatog genotipa za nauku, utvrđujući nove sorte, determinisanjem njihovog pedigrea, ali i spasavanje od njihovog nestajanja, kao fundamentalnog razvojnog resursa za dalji razvoj vinogradarstva u Crnoj Gori. U cilju njihovog očuvanja i valorizacije, potvrđeni genotipovi posađeni su u nacionalnoj kolekciji crnogorskih autohtonih i odomaćenih sorti vinove loze na Ćemovskom polju, čime je stvorena osnova za dalji razvoj crnogorskog vinogradarstva i vinarstva. To predstavlja resurs i temelj za razvoj i konkurentnost novih crnogorskih vina na tržištu za decenije koje dolaze.

Na osnovu predstavljenih rezultata može se zaključiti da je istorija crnogorskog vinogradarstva sačuvana u njenim sortmama, koje su kroz vjekove gajenja na ovim prostorima ispisale i osigurale uspješnu budućnost crnogorskog vinogradarstva. Zahvaljujući rezultatima ovih istraživanja, Crna Gora ima najmodernije genetske i genomske naučne činjenice — podatke za svoje specifično autohtono vinogradarstvo, što je presudno za naučnu, stručnu, razvojnu i marketinšku promociju ove strateške grane crnogorske poljoprivrede.

LITERATURA

[1] Ljubić, Š.: Statuta et leges civitatis Buduae, civitatis Scardonae, et civitatis et insulae Lesinae. Monumenta historico iuridica Slavorum meridional lum. Jugoslovenska akademija znanaosti i umjetnosti JAZU. Zagreb (1882); Srednjovjekovni Statut Budve. Bibiloteka — anali Budve, urednik dr M. Luketić, Istorijski arhiv, Budva 1988.

[2] V. Maraš, J. Tello, A. Gazivoda, M. Mugoša, M. Perišić, J. Raičević, N. Štajner, R. Ocete, V. Božović, T. Popović, E. García-Escudero, M. Grbić, J. M. Martínez-Zapater, J. Ibáñez: Population genetic analysis in old Montenegrin vineyards reveals ancient ways currently active to generate diversity in *Vitis* vinifera. Scientific Report, 1015000. (https://www.nature.com/articles/s41598–020–71918–7), 2020.

[3] Maraš, V. et al.: Origin and characterization of Montenegrin grapevine varieties. *Vitis* 54, 2015, 135–137.

[4] Maraš, V, Bozovic, V, Giannetto, S. & Crespan, M.: SSR molecular marker analysis of the grapevine germplasm of Montenegro. *J. Int. Sci. Vigne Vin.* 48, 2014, 87–97.

[5] Maraš, V.: Ampelographic and genetic characterization of Montenegrin grapevine varieties, in: Advances in Grape and Wine Biotechnology. DOI: https://doi.org/10.5772/intechopen.85676, 2019.

[6] Pejović, Lj.: Ampelografska proučavanja kratošije. Jugoslovensko vinogradarstvo i vinarstvo. Broj 3–4, 1988.

[7] Drori, E. et al.: Collection and characterization of grapevine genetic resources (*Vitis vinifera*) in the Holy Land, towards the renewal of ancient winemaking practices. *Sci Rep* 7, 2017.

[8] Štajner, N., et al.: Genetic clustering and parentage analysis of Western Balkan grapevines (*Vitis vinifera* L.). *Vitis* 54, 2015, 67–72.

[9] Boccacci, P, Torello-Marinoni, D, Gambino, G, Botta, R. & Schneider, A.: Genetic characterization of endangered grape cultivars of Reggio Emilia province. *Am. J. Enol. Vitic.* 56, 2005, 411–416.

[10] Vouillamoz, J. F. et al.: Genetic characterization and relationships of traditional grape cultivars from Transcaucasia and Anatolia. *Plant Genetic Resources* 4, 2007, 144–158, doi: doi: 10.1079/PGR2006114.

[11] De Lorenzis, G., et al.: SNP genotyping elucidates the genetic diversity of *Magna Graecia* grapevine germplasm and its historical origin and dissemination. *BMC Plant Biol.* 19, 2019.
[12] Cabezas, J. A. et al.: A 48 SNP set for grapevine cultivar identification. BMC Plant Biol. 11, 153, 2011.
[13] Calo, A., Costacurta, A., Maraš, V., Meneghetti, S. & Crespan, M.: Molecular correlation of Zinfandel (Primitivo) with Austrian, Croatian, and Hungarian cultivars and Kratošija, an additional synonym. Am. J. Enol. Vitic. 59, 2008, 205–209.
[14] Arroyo-Garcia, R. et al.: Chloroplast microsatellite polymorphisms in Vitis species. Genome 45, 2002,1142–1149.
[15] Cunha, J. et al.: Genetic relationships among Portugueses cultivated and wild Vitis vinifera L. germplasm. Front Plant Sci. 11, 127, 2020.
[16] Maul, E. et al.: The European *Vitis* Database (www.eu-vitis.de) — a technical innovation through an online uploading and interactive modification system. *Vitis* 51, 2012,79–85.
[17] Maul, E.: Vitis International Variety Catalogue — www.vivc.de — (accessed February 2020), 2020.
[18] Lijavetzky, D., Cabezas, J. A., Ibáñez, A., Rodriguez, V. & Martínez-Zapater, J. M.: High throughput SNP discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics* 8, 424, 2007.
[19] Ghaffari, S. et al.: Genetic diversity and parentage of Tunisian wild and cultivated grapevines (*Vitis vinifera* L.) as revealed by single nucleotide polymorphism (SNP) markers. *Tree Genet. Genomes* 10, 2014,1103–1113.
[20] Perrier, X. & Jacquemond-Collet, J. P.: DARwin software. http://darwin.cirad.fr (2006).
[21] Pritchard, J. K., Stephens, M. & Donnely, P.: Inference of population structure using multilocus genotype data. *Genetics* 155, 2000, 945–959.
[22] Prva međunarodna konferencija o vrancu i drugim crnogorskim autohtonim sortama vinove loze. Radovi sa naučnog skupa. Crnogorska akademija nauka i umjetnosti, 2021. ISBN 978–86–7215–491–7.
[23] Earl, D. & vonHoldt, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetic Resources* 4, 2012, 359–361.
[24] Jakobsson, M. & Rosenberg, N. A. CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics* 23, 2007, 1801–1806.
[25] Ramasamy, R. K., Ramasamy, S., Bindroo, B. B. & Naik, V. G.: STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus* 3, 2014.
[26] Peakall, R. & Smouse, P. E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research — an update. *Bioinformatics* 28, 2012, 2537–2539.
[27] Vähä, J. P., Erkinaro, J., Niemelä, E. & Primmer, C. R.: Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Mol. Ecol.* 16, 2007, 2638–2654.

[28] Kalinowski, S. T., Taper, M. L. & Marshall, T. C.: Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol. Ecol.* 16, 1099–1106, doi: 10.1111/j. 1365–294X. 2007.03089. x, 2007.

Vesna MARAŠ

CENTURIES FOR THE FUTURE: MONTENEGRIN VITICULTURE AND WINEMAKING — AN EXAMPLE OF COOPERATION BETWEEN THE SCIENTIFIC MOTHERLAND AND DIASPORA

Abstract

In the period 2014–2017, a multidisciplinary team led by Montenegrin and Spanish researchers carried out the most extensive research of grapevine genetic resources in Montenegro. The research included the genetic characterization of the existing grapevines diversity in cultivated and wild populations, encompassing their pedigree analysis.

The paper presents the richness of the grapevine genetic heritage in Montenegro, which still creates a very rich diversity of the *Vitis vinifera* species.

Out of 476 genotyped samples, using SSR and SNP markers, 101 genotypes corresponding to cultivated varieties (*Vitis vinifera* ssp. *sativa*) were identified, 43 genotypes corresponded to wild forms (*Vitis sylvestris*). In the cluster of cultivated varieties, 50 known varieties were identified, 51 unidentified genotypes are considered to be newly discovered, potentially autochthonous Montenegrin varieties. The results of the research confirmed that Montenegrin viticulture is based on a family of genetically related varieties, which is also characteristic of other famous traditional regions for the cultivation of vines and wine production. In Montenegro, this family was created around the Kratosija and Razaklija varieties. The research revealed 25 full pedigree varieties (trios) and 27 duos. The full pedigree of Vranac, who is a descendant of Kratošija and Duljenga, was found. The process of grapevine domestication, that began 8,000 years ago, in the Neolithic Age, continues in the wine-growing areas of Montenegro. All levels representatives of grapevine domestication were found.

The history of Montenegrin viticulture is preserved in its varieties that have survived through cultivation centuries in these areas and ensured the future of Montenegrin viticulture. This obliges us to preserve them and the need for Montenegrin viticulture and winemaking to preserve its authenticity.

Key words: vine, genetic resources, diversity, variety, autochthonous varieties, Kratošija, Razaklija, Vranac

Karolj SKALA[*]

MIPRO AS A HUB FOR REGIONAL COOPERATION AND INTEGRATION OF THE DIASPORA

Lampadem tradere — Innovative promotion of partnership!

INTRODUCTION

MIPRO is an international meeting and cooperation platform (hub) for experts from the fields of science, education, business and governments. MIPRO is a scientific, technological, professional and business meeting place, collaboration interface and acceleration platform. MIPRO dates back to 1978 and started with tutorials for lifelong learning of engineers in the field of microprocessor architecture and applications. Over the last decades, MIPRO has become one of the largest and most important congresses in Central and Eastern Europe, based on innovation and new technologies such as ICT. The core of the congress is formed by conferences covering a large part of ICT (such as microelectronics, electronics, data science, biomedicine, telecommunications, smart and security systems, digital economy, smart cities and environment, robotics and many others). Other indispensable segments include panels, workshops, tutorials, keynote speeches, plenary sessions, project meetings and the exhibition of EU projects, technical solutions and services. The annual MIPRO Congresses are regularly attended by more than a thousand participants from thirty to forty countries from all over the

[*] Karolj Skala, General Chair of MIPRO Convention, Member of Hungarian Academy of Science, Member of Croatian Academy of Technical Sciences

world. In the 46 years of its existence, more than 8900 scientific and technical papers have been published (regularly indexed in IEEE Xplore Digital Library). In addition, more than 200 panel discussions, 150 workshops and 120 tutorials were held and more than 600 exhibitors participated in the exhibitions. The mission of MIPRO is Lampadem tradere, to create, share and disseminate knowledge. It follows this mission by adapting to the demands of the world market and national interests in times of digital transformation, by networking and contributing strategies that follow EU technology policies and development directions. MIPRO aims to become a hub and a major dissemination point STEM by joining forces with other ICT associations in Europe to provide a well-developed infrastructure for the promotion of knowledge and cutting-edge technologies.

DESCRIPTION OF THE ORGANISATION

MIPRO is a public professional association with an office in Rijeka and a staff. It has a large number of industry members (65) and individual members (160) with a current base of 12,500 members. MIPRO relies on an already established network of direct collaborators at various faculties, institutions, government agencies and public administrations. More than 100 experts with different profiles have relevant expertise and experience. They are members and collaborators of MIPRO in ICT and in all other fields, including social sciences, humanities and technical design arts. Experience in logistics and expertise in organising large events with more than 1,000 participants. MIPRO has an extensive network of staff in HR, SE and the EU, technical support and service from professionals and stakeholders. With a tradition of 46 years, MIPRO has become a respected organisation for the promotion and knowledge transfer of new innovative visions, ideas and technologies in the field of ICT and technological-social innovation. MIPRO has 18 friends in 12 European countries (Austria, Croatia, Estonia, Finland, Italy, Norway, Germany, Hungary, Macedonia, Slovenia, Serbia and Ukraine). They are national activists, ambassadors of MIPRO and participate in the activity through experience and implementation of tasks. MIPRO is equipped with adequate multimedia communication and ICT infrastructure and will further develop and adapt platforms, tools and methods to fulfil its mission.

Organisational development activities

— Campaigns to raise public awareness of ICT and advocacy initiatives and models of good vision and best practise.

— Campaigns to promote efficiency, reliability and transparency in accountability and good governance of innovative technological solutions and services.

— Activities that directly and spontaneously strengthen civic influence and engagement in public advocacy and decision-making.

— Initiatives that influence distributed and parallel interaction of awareness based on information and communication processes.

— Public awareness and capacity building activities related to the protection of digital, social and natural ecosystems to enhance sustainability.

Added value of the activity

An integration approach will enable local communities and the hierarchical and wider society to significantly increase efficiency in a socio-economic sense and significantly increase awareness, need and opportunity for direct participation of all people in science, technology and society activities at all levels. Innovativeness In particular, the innovative use of cyber principles, information and communication technologies and informatics based on the hierarchy of Dew-Fog cloud computing services in line with the latest trends (e. g. smart product/service, smart city, smart island...), refined by integrating the basic needs of future exchange of goods and services after a pandemic, civil society proposals and joint decision-making, and permanent educational activities. In addition, Dew Computing solutions can significantly contribute to reducing the burden on the natural environment (intelligent management of processes and services). The basis of the planned system is an automated, intelligent organisation to optimise communication and cooperation (timely detection of problems and their joint solution).

CAPACITY AND POTENTIAL

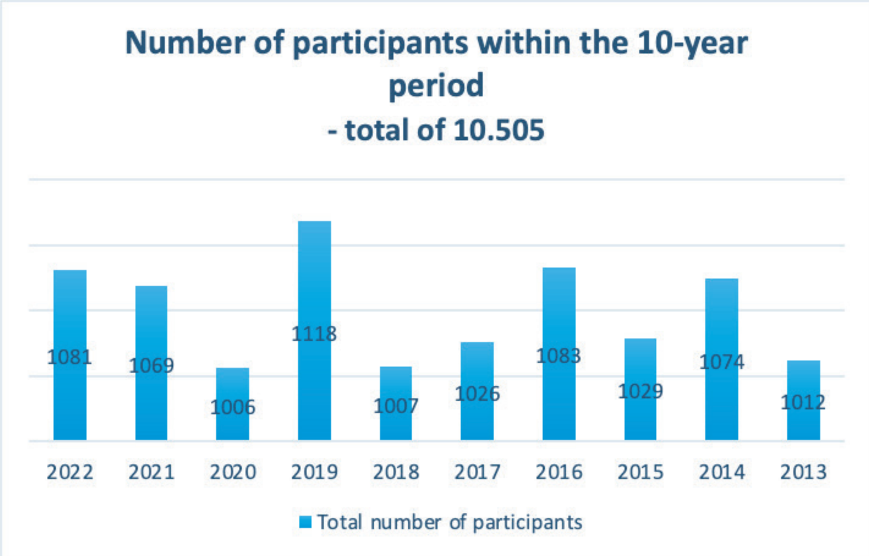
MIPRO has been active since 1978, initially as an organisation for the lifelong training of engineers in the field of microprocessor architecture and applications. Over the last decades, the annual MIPRO Congress has become one of the largest and most important congresses in this part of Europe dealing with innovations and new technologies in ICT (Information and Communication Technology). The congress offers conferences covering a wide spectrum of ICT (such as microelectronics, electronics, data science, biomedicine, telecommunications, smart and security systems, digital economy, smart cities and environment, robotics and many others). Other indispensable parts of the programme are panels, round tables, workshops, tutorials, outstanding lectures, plenary sessions, project meetings and an exhibition of EU projects, technical solutions and business services. MIPRO has become an international meeting place for experts from the fields of science, education, business, banking, government and local government. MIPRO is simultaneously a scientific, technological, professional and business communication and cooperation platform. Experience and expertise MIPRO's mission is to develop innovative solutions, share and disseminate knowledge. In this time of digital transformation, we follow this mission by adapting to global market demands and national interests, by networking and contributing to strategies that follow EU and HR technology policies and development guidelines. MIPRO aims to become a centre and focal point for communication and cooperation by joining forces with other ICT associations in the EU. MIPRO will be the coordinator for all aspects of project implementation in terms of leadership, collaboration, communication and dissemination of research results as an integral part of the project. This includes applying an innovative vision, increasing the visibility of research results and public engagement with science, technology and innovation in society. The aim is to ensure that the project carried out has a social, political and economic impact. MIPRO ensures the transfer of knowledge and the use of technical solutions, results and initiatives to relevant decision-makers at local, regional, national and European level. MIPRO takes care of the implementation of policy dissemination activities within these project initiatives, focusing on dissemination in the public press and social media.

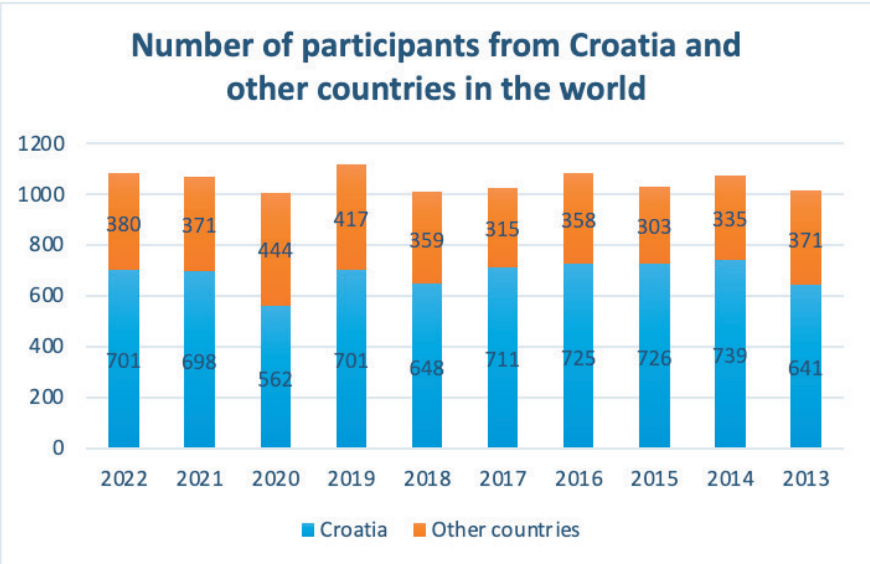
STATISTICS WITHIN THE LAST 10 YEARS (2013–2022)

Number of participants

		2022	2021	2020	2019	2018	2017	2016	2015	2014	2013	TOTAL
Total number of participants		1081	1069	1006	1118	1007	1026	1083	1029	1074	1012	**10.505**
From Croatia	NUMBER	701	698	562	701	648	711	725	726	739	641	**6.852**
	%	**64,88**	**65,32**	**55,82**	**62,74**	**64,31**	**69,26**	**66,90**	**70,58**	**68,85**	63,38	**65,23**
* From other countries	NUMBER	380	371	444	417	359	315	358	303	335	371	**3.653**
	%	**35,12**	**34,68**	**44,18**	**37,26**	**35,69**	**30,74**	**33,10**	**29,42**	**31,15**	36,62	**34,77**
From the former Yugoslavia (except Croatia)	NUMBER	121	149	140	107	103	124	126	119	129	155	**1.273**
	%	**11,19**	**13,98**	**13,96**	**9,56**	**10,23**	**12,09**	**11,67**	**11,54**	**12,05**	15,33	**12,12**

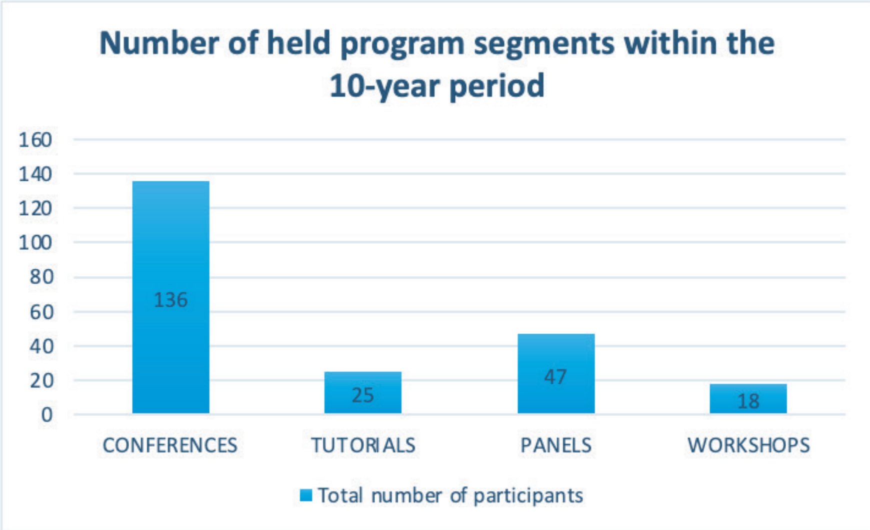
* *Participants from other countries include participants from the ex Yugoslavia*





Number of program segments that were held within MIPRO conventions

	2022	2021	2020	2019	2018	2017	2016	2015	2014	2013	TOTAL
Conferences	15	15	16	14	13	13	13	12	13	12	**136**
Tutorials	2	1	0	3	3	3	3	3	3	4	**25**
Panels	1	1	1	6	4	8	7	7	6	6	**47**
Workshops	5	3	0	4	0	1	0	3	1	1	**18**
TOTAL	23	20	17	27	20	25	23	25	23	23	**226**



Exhibition demonstrations and appearances

	2022	2021	2020	2019	2018	2017	2016	2015	2014	2013	TOTAL
Number of exhibitors	8	8	1	14	11	17	20	26	31	25	**161**

CONCLUSION

In 46 years of continuous work, from microprocessors to grid and cloud systems, MIPRO has participated in knowledge and technology transfer in the region and beyond. It provides an integration and development platform in the field of ICT. As an interface for collaboration and communication, it creates a symbiosis of professions and accelerates the development and application of advanced technical solutions to develop smart digital ecosystems for the future industrial revolution.

With its tradition and best practises, MIPRO can be an integrating factor in the region and beyond. It works to reintegrate intellectual capital through networking and building collaborations with the scientific diaspora.

REFERENCE AND INVITATION

For more information, please visit the website http://www.mipro.hr or direct mail contact mipro@mipro.hr

We invite all interested parties to register for the 46th MIPRO 2023, which will be held in Opatija from 22 to 26 May 2023 under the theme: ICT for a smart and green present and future.

Pavle R. ANĐUS[*]

AN EXAMPLE OF A GOOD AND SUSTAINABLE INTERNATIONAL COOPERATION — BIOPHYSICS SCHOOL „ACADEMICIAN RADOSLAV K. ANDJUS“

Abstract: After organizing a successful international Biophysics congress at Sveti Stefan dedicated to Radoslav K. Andjus (1926–2003), the founder of Serbian Biophysics and the leader of the Belgrade School of Physiology, in 2006 the members of the Yugoslav Biophysical Society decided to honor the academician Andjus (he was member of both Serbian and Montenegrin Academies) by establishing a biennial series of international Biophysics schools & workshops dedicated to his memory. Since then 8 Schools have been organized. The main concept was that this is an international event organized at Montenegro coast with the mutual core organization by colleagues from Serbia (Biophysical Society and Faculty of Biology University of Belgrade) and Montenegro (Institute of Marine Biology, Kotor). Co-organizers were usually distinguished scientists from abroad. Since the topics of the School were often tackling Neurobiophysics the School coined its acronym „NERKA“ (**NE**uro **R**adoslav **K** **A**ndjus). The School remains a lighthouse of collaboration between Serbia and Montenegro attracting international peers, experts as well as young participants.

INTRODUCTION

Radoslav Krstov Andjus (1926–2003) was one of the main proponents of the Belgrade School of Physiology and the founder of the Serbian Biophysics. He was member of the Serbian (elected age 33) and Montenegrin academies of sciences and arts. He was one of the founders of the Institute for Marine Biology in Kotor. His roots are in his fatherland — Paštrovići and Sveti Stefan of the Montenegrin coast. He is buried at the graveyard

[*] Pavle R. Anđus, Full professor, PhD; Center for laser microscopy, Faculty of Biology University of Belgrade, Serbia

of monastery Praskvica, above the Miločer park with a beautiful view of Sveti Stefan.

After organizing a successful international Biophysics congress at Sveti Stefan (Proceedings published in the Annals N. Y. Acad. Sci. Vol 1048 2005) dedicated to Radoslav K. Andjus, in 2006 we from the Yugoslav Biophysical Society decided to honor academician Andjus by organizing an international Biophysics school & workshop dedicated to his memory, with the intention to establish a biennial series. Since then 8 Schools have been organized. The main concept was that this is an international event organized at Montenegro coast with the mutual core organization by colleagues from Serbia (Biophysical Society and Faculty of Biology University of Belgrade) and Montenegro (Institute of Marine Biology, Kotor). Co-organizers were usually distinguished scientists from abroad. Since the topics of the School were often tackling Neurobiophysics the School coined its acronym NERKA (**NE**uro **R**adoslav **K** **A**ndjus).

NERKA 1 (2006)

The International Workshop „Imaging in Neurosciences and Beyond“ was held at Sveti Stefan (Montenegro) September 23–30, 2006. The Workshop was organized by the Faculty of Biology, University of Belgrade and the Yugoslav Biopohysical Society. There were 30 participants

NERKA 1. Organizers and lecturers of the First NERKA in a visit to Budva old town.

— 10 students with fellowships, and 7 published papers in the workshop proceedings on a CD. Five papers were chosen by the program committee for oral section presentations. There were 12 invited speakers who also prepared handouts for the participants some of which could be obtained prior to the meeting at an internet site. Participants were from Belarus, Bosnia and Herzegovina, Croatia, Czech Republic, Estonia, Russia, Serbia, Slovenia, and USA.

The program consisted of three sections dealing with MRI, advanced microscopy, and imaging in electrophysiology. Each working day was summed up in the evening at a panel discussion with speakers and participants. Finally, a written exam was given to the interested participants and a diploma for 3 ECTS credits verified by the Faculty of Biology, University of Belgrade was given to the ones that past. Six students applied and past the exam.

NERKA 2 (2008)

In 2006 Montenegro got its independence, and the organization of NERKA had to be transferred to Belgrade. However, this was considered only a temporary solution with the remaining initial goal to sustain the collaboration between the two academic environments.

The Workshop „Neuroimaging and complementary techniques“ was organized by twining with the Training School of the EC COST Action B30 („Neural Repair and Plasticity“). It was also organized in cooperation with the Society for Neurosciences of Serbia (that also recruited speakers) and with two academic institutions in the country, School of Biology, and Institute for Biological Research „Siniša Stanković“ both part of University of Belgrade (that offered their labs and lecture halls).

In addition to the main approaches in neuroimaging (as established at the 2006 School) this time in addition to high quality lectures emphasis was given also to practical courses. Students entered MRI clinical facilities, histology and electrophysiology labs. They were then given practical courses at these facilities and later the students went through hands-on courses in any of the offered facilities.

NERKA 3 (2010)

In December 2010 NERKA returned to Montenegro. With the help of Dr Veljko Milutinović and his IPSI team a workshop on „Imaging and Modern Biophysical Approaches“ was organized at Pržno-Miločer. Several distinguished lecturers from over the world gathered in the hotel „Residence“ and

NERKA 2. Clasroom (Faculty of Biology, Belgrade), and groop photo in front of the venue in Belgrade.

TIR

The IPSI BgD Transactions on Internet Research

Multi-, Inter-, and Trans-disciplinary Issues in Computer Science and Engineering

A publication of

IPSI Bgd Internet Research Society
New York, Frankfurt, Tokyo, Belgrade
January 2012 Volume 8 Number 1 (ISSN 1820-4503)

Table of Contents:

NERKA 3. Front page of IPSI BgD Transaction dedicated to the talks from the Workshop.

as a special dinner social celebrated the coming New Year 2011. After the workshop a special volume of IPSI BgD Transactions was issued (see abow).

NERKA 4 (2012)

The International Workshop: „Methods on the interface of Neurochemistry and Electrophysiology" was again organized in Belgrade (August/September, 2012) since it was considered convenient to twin with the *Regional Biophysical Conference 2012* in Kladovo, Serbia (September, 2012). It was also be organized in cooperation with the Biophysical Society of Serbia and with the Faculty of Biology, and Institute for Biological Research „Siniša Stanković" both part of University of Belgrade (that offered their labs and lecture halls). Also included in the organization of the Workshop there were the Military Medical Academy and the Faculty of Physical Chemistry as well as the Institute of General and Physical Chemistry.

This Workshop should deal with techniques (lectures and practicals) on the interface of Neurochemistry and Electrophysiology and their applications in Neurosciences:

1) Magnetic resonance spectroscopy and transcranial magnetic stimulation
2) Imaging of ions and molecules in electrophysiology
3) Neuroelectrochemistry on tissue samples
4) Chemical tracing of transmitter electrophysiological responses and their modulation by ecto- and endo-phosphatases.

The School as well as the Regional Conference were supported by European and International Biophysics societies (EBSA and IUPAB) and the Ministry of Education and Science Rep. of Serbia.

NERKA 5 (2014)

This School was meant to be organized on the occasion of the International Year of Light and was entitled „NEUROPHOTONICS — Towards the International Year of Light and Light-based Technologies 2015" and it was coorganized with the Photonics Center of the Institute of Physics, University of Belgrade. Foreign organizer was Ivan Milenković from Faculty of Medicine, University of Leipzig.

After a selection process 25 students were invited to the School. 13 where from abroad (Argentina, India, Italy, Netherlands, New Zeeland, Poland, and Turkey) and 12 from Serbia. Other local students were also invited for lectures only. There were 14 2–3h lectures by experts from Croatia, Germany, Italy, Serbia, Spain, Sweden, and USA. Lectures were intertwined with hands-on and practical sessions. There was a special lecture with discussion on Ethical use of animals in neuroscience given by the head of the Ethics Council of Republic of Serbia.

NERKA 5. Prof. Andjus with some international students at a get-together (upper left). Hands-on session at Institute of Physics (upper right). Group photo in front of the venue.

There were many get-together occasions to discuss and meet the speakers. A farewell dinner was organized at the end.

Finally the students were given certificates and filled in an anonymous evaluation test. The evaluation test showed that the students strongly agree (70%) or just agree (30%) that the School was a success.

In addition to the International Brain Research Organization (IBRO) funds the School was also supported by the Ministry of education, science and technological development Republic of Serbia.

NERKA 6 (2016)

For the Summer School „Imaging neuroinflammation“ organized in June, 2016 NERKA finally returns to the in Kotor. This School was, organized traditionally by the Biophysical Society of Serbia, co-organized by the Faculty of Biology, University of Belgrade, Institute of Marine Biology as the local organizer and with the fellowships sponsored by the International Union for Pure and Applied Biophysics (IUPAB).

The topic of the School was to introduce modern methods of biophysical imaging and image analysis in studies of inflammatory phenomena and markers of neurodegenerative diseases. It was also envisaged to raise discussion among the participants on the translational value of these techniques and the clinical relevance of presented experimental markers. The techniques covered were MRI, MRS, PET, EPRI, and advanced microscopy (non-linear microscopy- dual photon fluorescence, Fluorescence correlation spectroscopy & microscopy, PALM, STED, CARS, digital holographic microscopy). A multidisciplinary faculty was chosen ranging from fields of physics (applications of laser technology and advanced microscopy), physical chemistry (magnetic resonance techniques and applications), biology (experimental models of neurodegenerative diseases), and medicine (clinical studies and markers of neuroinflammation).

There was a lineup of 13 expert speakers from Australia, Austria, Canada, Italy, Serbia, UK, and USA. The students were from 8 countries: Armenia, Bulgaria, Croatia, Hungary, Italy, Macedonia, Serbia and Slovakia. Altogether 33 students were registered. 15 students were awarded a fellowship on competitive basis.

NERKA 6. Group photo at the terrace of the Institute of Marine Biology, Kotor.

NERKA 7 (2018)

The topic of this School was „Mechanobiology" and it was organized with the great help pf Dr Boris Martinac (who participated at NERKA 3) from Victor Chang Cardiac Research Institute, Sydney Australia. This School was organized in the wake of the 2021 Nobel prize in Physiology or Medicine given for the discovery of mechanosensitive Piezo channels to which Dr Martinac contributed significantly.

After a selection process, 24 students were invited to the School. Seven of these students where from abroad (Austria, Egypt, France, Slovakia, Slovenia

NERKA 7. A group photo at the terrace of the Institute of Marine Biology, Kotor (left). The faculty (Dr Boris Martinac sitting) at the famous church site behind the Institute (right).

and Turkey) and 13 from Serbia. Other local students were also invited for lectures. There were 12 1.5h lectures by experts from France, Greece, Hungary, Italy, Japan, Serbia, Spain, UK and USA. Lectures were intertwined with time dedicated to discussions and a discussion panel has also been organized on the Translational value of Mechanobiology (moderated by P. R. Andjus, B. Martinac, K. Radotić and K. Naruse).

There were two get-together occasions in the evening to discuss and meet the speakers and a farewell dinner preceded by a boat excursion at the end.

At the closing of the School the students were given certificates of participation and took an anonymous evaluation survey in addition to a personalized questionnaire.

In addition to the Biophysical Society the School was also supported by IBRO, IUPAB, EBSA, and COST.

NERKA 8 (2021)

The Covid19 pandemic prevented us to organize a School in 2020, however as soon as the situation bettered we boldly embarked on the organization of a School in 2021. It was a success.

The Summer Course was enttled „Ion Channels and Neuronal Excitability". It was organized by the Faculty of Biology University of Belgrade with the support of the European FENS-NENS PhD program on Neurobiology. This school covered the topic of Ion Channels and Neuronal Excitability — techniques, modeling, physiology & pathology. The foreign organizer of the School was Dr Marco Canepari from University of Grenoble, France.

We had 27 students from Serbia, N. Macedonia, Romania, Hungary, Turkey, Italy, Greece, Germany and Russia. Participants presented posters during the School. We made sure that students had enough time during breaks to get to know each other better and discuss with the lecturers the topics they covered in their lectures. The School Course was supported by FENS-IBRO, European Society of Neurochemistry, and companies CAIRN and LabEx ICST. Most of the participants received fellowships of different kinds.

There were 13 lectures by experts from France, Hungary, N. Macedonia, Italy, Serbia, UK and USA. Lectures were intertwined with the time dedicated to discussion. The final panel discussion treated some relevant questions of career development and stimulated a fruitful interaction of the participants and faculty.

NERKA 8. The Faculty at the church site (above; Dr M. Canepari, 2nd from the right). A group photo in front of the Institute of Marine Biology, Kotor.

CONCLUSION

One of the main achievements of this series of Schools is the sustained collaboration between Belgrade and Kotor and the revival of the intense academic exchange that was initiated and strongly developed from the very beginning of the foundation of the Institute of Marine Biology. This however with the ever present international component within instructors as well as students, with the ultimate goal to develop an international center for multidisciplinary research deeply embedded in the beautiful Bay of Kotor. Our endeavor was always fully supported by the staff and colleagues from the Institute of Marine Biology, particularly we would like to highlight here the unreserved collaboration of Drs Aleksandar Joksimović and Mirko Đurović.

Pavle R. ANĐUS

PRIMER DOBRE I ODRŽIVE MEĐUNARODNE SARADNJE — BIOFIZIČKA ŠKOLA „AKADEMIK RADOSLAV K. ANĐUS"

Sažetak

Nakon organizovanja uspešnog međunarodnog biofizičkog kongresa na Svetom Stefanu posvećenog Radoslavu K. Anđusu (1926–2003), osnivaču srpske biofizike i rukovodiocu Beogradske fiziološke škole, članovi Biofizičkog društva Jugoslavije su 2006. godine odlučili da odaju priznanje akademiku Anđusu (bio je član i SANU i CANU) osnivanjem bijenalne serije međunarodne biofizičke škole i radionice njemu posvećene. Od tada je organizovano 8 škola. Osnovna koncepcija je bila organizovanje međunarodnog događaja koji na Crnogorskom primorju, kao zajedničkom jezgru, organizuju kolege iz Srbije (Društvo za biofiziku i Biološki fakultet Univerziteta u Beogradu) i Crne Gore (Institut za biologiju mora, Kotor). Suorganizatori su obično ugledni naučnici iz inostranstva. Pošto su se teme škole često bavile neurobiofizikom, škola je dobila akronim „NERKA" (**NE**uro **R**adoslav **K. A**nđus). Škola ostaje svetionik saradnje Srbije i Crne Gore, koji okuplja međunarodne saradnike, stručnjake kao i mlađe polaznike.

Crnogorska akademija
nauka i umjetnosti

UCG
Univerzitet Crne Gore

NAUČNI SIMPOZIJUM

NAUKA I MALE ZEMLJE: SINERGIJA DIJASPORE, MATICE I PRIJATELJA CRNE GORE

2. i 3. novembar 2022, Podgorica

SCIENTIFIC SYMPOSIUM

SCIENCE AND SMALL COUNTRIES: SYNERGY OF THE DIASPORA, THE MOTHERLAND & FRIENDS OF MONTENEGRO

November 2 & 3, 2022, Podgorica

SPISAK UČESNIKA
/ LIST OF THE PARTICIPANTS
(po abecednom redu / by alphabetical order)

Ime / Name	Zaposlenje / Affiliation	e-mail
Pavle R. Anđus	▪ Center for Laser Microscopy, Institute for Physiology & Biochemistry “Ivan Djaja” ▪ Faculty of Biology, University of Belgrade	pandjus@ bio.bg.ac.rs
Vaso Antunović	▪ Associate Member, CANU	ivanka@ffh.bg.ac.rs
Dušica Babović-Vuksanović	▪ Professor of Medical Genetics and Pediatrics, Consultant ▪ Department of Clinical Genomics, Mayo Clinic, Rochester, MN, USA ▪ Director of Neurofibromatosis Comprehensive Program, Mayo Clinic	dbabovic@mayo.edu
Slobodan Backović	▪ Academician, CANU	backovic@yandex.ru
Adis Balota	▪ University Mediterranean, Podgorica	adisb@t-com.me
Vesna Bengin	▪ BioSense institute, Department of Physics, Faculty of Sciences, University of Novi Sad	vesna.bengin@ gmail.com bengin@biosense.rs
Miroslav Bojović	▪ Faculty of Electrical Engineering, University of Belgrade	mbojovic@ etf.bg.ac.rs mbojovic@etf.rs
Ranislav Bulatović	▪ Academician, CANU	ranislav@ucg.ac.me
Ivan Božović	▪ Brookhaven National Laboratory, NY, USA, and Yale University, CT, USA ▪ Member of Academia Europaea, London, UK ▪ Foreign Member of the Serbian Academy of Sciences and Arts ▪ Fellow of the American Physical Society (APS) ▪ Fellow of the International Society for Optics and Photonics (SPIE)	bozovic@bnl.gov
Vladimir Božović	▪ Rector of the University of Montenegro	rector@ucg.ac.me
Ilir Capuni	▪ Massachusetts Institute of Technology (MIT), Boston, USA ▪ Boston University, USA ▪ Atlas Imaging, San Francisco, CA, USA ▪ Marin Barleti University, Tirana, Albania	ilir.capuni@gmail. com lir@bu.edu

Ime / Name	Zaposlenje / Affiliation	e-mail
Anđela Draganić	▪ Assistant Professor, University of Montenegro	andjelad@ucg.ac.me
Vladimir Dragović	▪ Department of Mathematical Sciences, The University of Texas at Dallas, USA, USA/MISANU, Serbia	Vladimir.Dragovic @utdallas.edu
Vladan Đokić	▪ Faculty of Architecture, University of Belgrade, Rector of the University of Belgrade	vdjokic@arh.bg.ac.rs
Igor Đurović	▪ Academician, CANU	idjurovic@ucg.ac.me
Siniša Đurović	▪ Department of Electrical and Electronic Engineering, University of Manchester, UK	Sinisa.Durovic@ manchester.ac.uk
Nenad Filipović	▪ Faculty of Mechanical Engineering, University of Kragujevac ▪ Rector of the University of Kragujevac ▪ Harvard University, Boston, USA	fica@kg.ac.rs
Vladimir Filipović	▪ Faculty of Mathematics, University of Belgrade	vladaf@matf.bg.ac.rs
Miodrag Grbić	▪ Department of Biology, University of Western Ontario, London, ONT, Canada	mgrbic@uwo.ca
Marija Ilić	▪ Massachusetts Institute of Technology (MIT), Boston, USA ▪ Member of the USA National Academy of Engineering ▪ Member of Academia Europaea, London, UK	ilic@mit.edu
Ayhan Irfanoglu	▪ Vice Head of the School of Civil Engineering, Purdue University, IND, USA	ayhan@purdue.edu airfanoglu@ gmail.com
Milojica Jaćimović	▪ Academician, CANU	milojica@ jacimovic.me
Branislav Jelenković	▪ Faculty of Physics, University of Belgrade, Institute of Physics, Belgrade ▪ Corresponding Member of the Serbian Academy of Sciences and Arts	branaj@ipb.ac.rs
Gordan Karaman	▪ Academician, CANU	karaman@t-com.me
Miljan Knežević	▪ Faculty of Mathematics, University of Belgrade	kmiljan@ matf.bg.ac.rs kmiljan@gmail.com
Mirjana Kranjac	▪ Faculty of Technical Sciences, University of Novi Sad	mirjana.kranjac@ vojvodina.gov.rs mirjana.kranjac@ hotmail.com
Antonija Krstačić	▪ Faculty of Dental Medicine and Health Osijek, Medical Faculty ▪ University J. J. Strossmayer, Osijek; CHC „Sestre milosrdnice“ ▪ Traumatology Clinic, Zagreb	akrstacic@ gmail.com
Jelena Latinović	▪ Biotechnical Faculty, University of Montenegro, Podgorica	jelenalat@ucg.ac.me
Nedeljko Latinović	▪ Biotechnical Faculty, University of Montenegro, Podgorica	nlatin@ucg.ac.me
Vlado Lubarda	▪ Academician, CANU	vlubarda@ eng.ucsd.edu

Ime / Name	Zaposlenje / Affiliation	e-mail
Vesna Maraš	▪ Company „13. jul Plantaže" — Podgorica	vesnam@t-com.me
Slobodan Marković	▪ Department of Geography, Tourism and Hotel Management ▪ Faculty of Natural Sciences and Mathematics, University of Novi Sad ▪ Corresponding member of the Serbian Academy of Sciences and Arts	baca.markovic@gmail.com slobodan.markovic@dgt.uns.ac.rs
Miodrag Mateljević	▪ Faculty of Mathematics, University of Belgrade ▪ Member of the Serbian Academy of Sciences and Arts	miodrag@matf.bg.ac.rs
Nenad Medvidović	▪ Professor and Chair, Computer Science Department, Viterbi School of Engineering, University of Southern California, Los Angeles, USA	neno@usc.edu
Fedor Mesinger	▪ Serbian Academy of Sciences and Arts	fedor.mesinger@gmail.com
Miodrag Mihaljević	▪ Corresponding Member, Serbian Academy of Sciences and Arts ▪ Member of Academia Europaea ▪ Deputy Director, Mathematical Institute of the Serbian Academy of Sciences and Arts	miodragm@turing.mi.sanu.ac.rs
Gradimir Milovanović	▪ Faculty of Electrical Engineering, University of Niš ▪ Member of the Serbian Academy of Sciences and Arts	gradimir.milovanovic@mac.com
Veljko Milutinović	▪ Indiana University, Bloomington, USA ▪ Life Fellow of the Institute of Electrical and Electronics Engineers (IEEE), Washington DC, USA, Member of Academia Europaea, exMember of the Board, London, UK ▪ Founding Member of the Serbian Academy of Engineering, Belgrade ▪ Foreign Member of the Montenegrin Academy of Sciences and Arts (CANU)	vm@etf.rs veljko.milutinovic.1951@gmail.com
Ognjen Š. Miljanić	▪ Department of Chemistry, University of Houston, TX, USA	ognjen.miljanic@gmail.com
Predrag Miranović	▪ Academician, CANU	pedjam@ucg.ac.me
Nikola Mišković	▪ Department of Control and Computer Engineering ▪ Faculty of Electrical Engineering and Computing (FER), University of Zagreb ▪ LABUST — Laboratory for Underwater Systems and Technologies, University of Zagreb	nikola.miskovic@fer.hr
Branislav Mitrović	▪ Prof. Emeritus, Faculty of Architecture, University of Belgrade ▪ Member of the Serbian Academy of Sciences and Arts	office@mitarh.rs
Onur Mutlu	▪ ETH Zürich, Distinguished Professor ▪ Carnegie Mellon University (CMU), Pittsburgh, PA, USA, Distinguished Professor ▪ Fellow of the Institute of Electrical and Electronics Engineers (IEEE) ▪ Fellow of the Association for Computing Machinery (ACM)	omutlu@gmail.com

Ime / Name	Zaposlenje / Affiliation	e-mail
Angelia Nedich	▪ Moscow State University, Russia ▪ Massachusetts Institute of Technology (MIT), Boston, USA ▪ Arizona State University, Tempe, Arizona, USA	Angelia.Nedich@asu.edu
Goran Nikolić	▪ Academician, CANU	nikolicg@ucg.ac.me
Irena Orović	▪ Vice-Rector, University of Montenegro	irenao@ucg.ac.me
Vukić Pulević	▪ Academician, CANU	piskovina@gmail.com
Velimir Radmilović	▪ Serbian Academy of Sciences and Arts ▪ Honorary Member of the Academy of Engineering Sciences of Serbia ▪ Faculty of Technology and Metallurgy, Nanotechnology and Functional Materials Center, University of Belgrade	vrradmilovic@tmf.bg.ac.rs
Saša V. Raković	▪ Beijing Institute of Technology (BIT)	sasa.v.rakovic@gmail.com
Jakob Salom	▪ IPSI Belgrade, Serbia, Department for Digital Preservation of National Heritage	jasasalom@yahoo.com jasasalom@gmail.com
Marija Santini	▪ School of Medicine, University of Zagreb ▪ University Hospital for Infectious Diseases "Dr. Fran Mihaljević", Zagreb ▪ President of the Croatian Society for Infectious Diseases of the Croatian Medical Association	marijasantini.ms@gmail.com
Petar Seferović	▪ Academician, Serbian Academy of Sciences and Arts, Faculty of Medicine, University of Belgrade, and Heart Failure Center, Belgrade University Medical Center ▪ Vice-President, European Society of Cardiology ▪ President, Heart failure Association of the ESC (2018-2020) ▪ President, Heart Failure Society of Serbia	seferovic.petar@gmail.com
Dan Shechtman	▪ Technion — Israel Institute of Technology, Haifa, Israel ▪ The Nobel Lauerate in Chemistry, 2011.	danny.shechtman@gmail.com
Karolj Skala	▪ Ruđer Bošković Institute, Centre for Informatics and Computing, Zagreb, Croatia ▪ Member of the Hungarian Academy of Science ▪ Member of the Croatian Academy of Engineering	skala@irb.hr
Michal Goran Stanišić	▪ Poznan University of Medical Sciences, Poland ▪ Member of the Board of the Polish Society for Vascular Surgery ▪ Member of the Board of the Polish Society of Phlebology	michal.stanisic@angiodiabetica.pl
Isidora Stanković	▪ Assistant Professor, University of Montenegro	isidoras@ucg.ac.me
Ljubiša Stanković	▪ Academician, Vice-President CANU	ljubisa@ucg.ac.me
Bratislav Stipanić	▪ Faculty of Civil Engineering, University of Belgrade and PBU, Novi Sad, Serbia	stipanic.bratislav@gmail.com
Radovan Stojanović	▪ Faculty of Electrical Engineering, University of Montenegro, Podgorica	stox@ucg.ac.me

Ime / Name	Zaposlenje / Affiliation	e-mail
Biljana Šćepanović	▪ Minister of Science and Technology Development in the Government of Montenegro	kabinet@mntr.gov.me
Ramo Šendelj	▪ University Donja Gorica, Podgorica	ramo.sendelj@gmail.com
Zhilbert Tafa	▪ Indiana University, Bloomington, USA and UBT College, Priština, Kosovo	tafaul@t-com.me
Svjetlana Terzić	▪ Academician, CANU	sterzic@ucg.ac.me
Željko Torbica	▪ Head, School of Construction Management Technology, Purdue University, USA	ztorbica@purdue.edu
Danilo Vojvodić	▪ Medical Faculty, University of Defense, MMA Belgrade	vojvodic.danilo@gmail.com
Dragan K. Vukčević	▪ President of the Montenegrin Academy of Sciences and Arts	canu@canu.ac.me dragan.vukcevic@udg.edu.me
Petar Vukoslavčević	▪ Academician, CANU	petarvuk@ucg.ac.me
Perko Vukotić	▪ Academician, CANU	pvukotic@ucg.ac.me
Gordana Vunjak-Novaković	▪ University Professor, The Mikati Foundation Professor of Biomedical Engineering and Medicine, Columbia University, NY, USA ▪ Member of the USA National Academy of Medicine ▪ Member of the USA National Academy of Engineering ▪ Member of the American Academy of Arts and Sciences ▪ Member of the USA National Academy of Inventors ▪ Member of Academia Europaea, London, UK ▪ Foreign Member of SANU, Belgrade, RS ▪ Member of the International Academy for Medical and Biological Engineering	gv2131@columbia.edu

SADRŽAJ

OTVARANJE

MEDICINSKE NAUKE

MATEMATIČKE NAUKE

PRIRODNE NAUKE

INŽENJERSKE NAUKE I

INŽENJERSKE NAUKE II

INFORMATIČKE NAUKE

PRIMJERI DOBRE PRAKSE

CONTENTS

OPENING

MEDICAL SCIENCES

MATHEMATICAL SCIENCES

ENGINEERING SCIENCES II

INFORMATICAL SCIENCES

EXAMPLES OF GOOD PRACTICE

CRNOGORSKA AKADEMIJA NAUKA I UMJETNOSTI
MONTENEGRIN ACADEMY OF SCIENCES AND ARTS

Naučni simpozijum

NAUKA I MALE ZEMLJE: SINERGIJA DIJASPORE, MATICE I PRIJATELJA CRNE GORE

Scientific Symposium

SCIENCE AND SMALL COUNTRIES: SYNERGY OF THE DIASPORA, THE MOTHERLAND & FRIENDS OF MONTENEGRO

Lektor / Lector
Anita Labović

Priprema za štampu / Prepress
Bojan R. Popović
Medeon — Podgorica

Tiraž / Circulation
150

Štampa / Press
Golbi Print — Podgorica
2023.

CIP — Каталогизација у публикацији
Национална библиотека Црне Горе, Цетиње

ISBN 978-86-7215-533-4
COBISS.CG-ID 27343620

ISBN 978-86-7215-533-4