online issn: 1307-6892

irc 2021
XV. international research conference
proceedings

open science index 15 2021

october 21-22, 2021 athens greece
international scholarly and scientific research & innovation

Open Science

Open Science Philosophy

Open science encompasses unrestricted access to scientific research articles, access to data from public research, and collaborative research enabled by information and communication technology tools, models, and incentives. Broadening access to scientific research publications and data is at the heart of open science. The objective of open science is to make research outputs and its potential benefits available to the entire world and in the hands of as many as possible:

- Open science promotes a more accurate verification of scientific research results. Scientific inquiry and discovery can be sped up by combining the tools of science and information technologies. Open science will benefit society and researchers by providing faster, easier, and more efficient availability of research outputs.
- Open science reduces duplication in collecting, creating, transferring, and re-using scientific material.
- Open science increases productivity in an era of tight budgets.
- Open science results in great innovation potential and increased consumer choice from public research.
- Open science promotes public trust in science. Greater citizen engagement leads to active participation in scientific experiments and data collection.

Open Science Index

The Open Science Index (OSI) currently provides access to over thirty thousand full-text journal articles and is working with member and non-member organizations to review policies to promote and assess open science. As part of the open science philosophy, and by making open science a reality; OSI is conducting an assessment of the impact of open science principles and restructuring the guidelines for access to scientific research. As digitalization continues to accelerate science, Open science and big data hold enormous promise and present new challenges for policymakers, scientific institutions, and individual researchers.

OSI is helping the global scientific research community discover, evaluate, and access high-quality research output. Renowned for its editorially curated and refereed collection of the highest-quality publications, OSI has always been and will remain free-of-charge.

OSI provides an efficient and thorough discovery process to the open science research database and provides links and free access to full-text articles. There are 50 open access journal categories that are curated and refereed by international scientific committees, the in-house editorial team, and trusted partners. Since its inception in 2007, OSI has made more than thirty thousand peer-reviewed open access full-text journal articles (PDF versions) freely available online without cost, barriers, or restrictions.

Open Science Access

With the Open Science Index, researchers can discover and access trusted peer-reviewed open access full-text scientific research articles with confidence. OSI helps researchers find appropriate non-profit open access journals to publish their work.

OSI gives one-click access to online full-text PDFs and expands the reach to global society by giving users free access from anywhere around the globe. Through cutting-edge open science collaboration, in an innovative public partnership, the non-profit OSI is devoted to making science open and reusable.

To learn more, visit online at waset.org

Open Science

Open Society

An open society allows individuals to change their roles and to benefit from corresponding changes in status. Open science depends to a greater or lesser extent on digital technologies and innovations in structural processes by an open society. When realized, open science research and innovation can create investment opportunities for new and better products and services and therefore increase competitiveness and employment. Open science research and innovation is a key component of thematic open science priorities. Central to the open science digital infrastructure is enabling industry to benefit from digital technology and to underpin scientific advances through the development of an open society. Open science research and innovation can also contribute to society as a global actor because scientific relations can flourish even where global relations are strained. Open science has a critical role across many areas of decision making in providing evidence that helps understand the risks and benefits of different open science choices. Digital technology is making the conduct of open science and innovation more collaborative, more global, and more open to global citizens. Open society must embrace these changes and reinforce its position as the leading power for science, for new ideas, and for investing sustainably in the future.

It is apparent in open society that the way science works is fundamentally changing, and an equally significant transformation is taking place in how organizations and societies innovate. The advent of digital technology is making research and innovation more open, collaborative, and global. These exchanges are leading open society to develop open science and to set goals for research and innovation priority. Open science goals are materializing in the development of scientific research and innovation platforms and greater acceptance of scientific data generated by open science research. Open science research and innovation do not need help from open society to come up with great ideas, but the level of success ideas ultimately reach is undoubtedly influenced by regulation, financing, public support, and market access. Open society is playing a crucial role in improving all these success factors.

Open Science

Open science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and collaborative tools. These innovations capture a systemic change to the way science and research have been carried out for the last fifty years. Science is shifting from the standard practice of publishing research results in scientific publications after the research and reviews are completed. The shift is towards sharing and using all available knowledge at an earlier stage in the research process. Open science is to science what digital technology is to social and economic transactions: allowing end users to be producers of ideas, relations, and services and in doing so, enabling new working models, new social relationships and leading to a new modus operandi for science. Open science is as important and disruptive as e-commerce has been for the retail industry. Just like e-commerce, the open science research paradigm shift affects the whole business cycle of doing science and research. From the selection of research subjects to the carrying out of research, to its use and re-use, to the role of universities, and that of publishers are all dramatically changed. Just as the internet and globalization have profoundly changed the way we do business, interact socially, consume culture, and buy goods, these changes are now profoundly impacting how one does research and science.

The discussion on broadening the footprint of science and on novel ways to produce and spread knowledge gradually evolved from two global trends: Open Access and Open Source. The former refers to online, peer-reviewed scholarly outputs, which are free to read, with limited or no copyright and licensing restrictions, while open source refers to software created without any proprietary restriction and which can be accessed and freely used. Although open access became primarily associated with a particular publishing

or scientific dissemination practice, open access already sought to induce a broader practice that includes the general re-use of all kinds of research products, not just publications or data. It is only more recently that open science has coalesced into the concept of a transformed scientific practice, shifting the focus of researchers' activity from publishing as fast as possible to sharing knowledge as early as possible. Open science is defined as the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process. As a result, the way science is done in the future will look significantly different from the way it is done now. Open science is the ongoing evolution in the modus operandi of doing research and organizing science. This evolution is enabled by digital technology and is driven by both the globalization of the scientific community and increasing public demand to address the societal challenges of our times. Open science entails the ongoing transitions in the way research is performed, researchers collaborate, knowledge is shared, and science is organized.

Open science impacts the entire research cycle, from the inception of research to its publication, and on how this cycle is organized. The outer circle reflects the new interconnected nature of open science, while the inner circle shows the entire scientific process, from the conceptualization of research ideas to publishing. Each step in the scientific process is linked to ongoing changes brought about by open science, including the emergence of alternative systems to establish a scientific reputation; changes in the way quality and impact of research are evaluated; the growing use of scientific blogs; open annotation; and open access to data and publications. All institutions involved in science are affected, including research organizations, research councils, and funding bodies. The trends are irreversible, and they have already grown well beyond individual projects. Theses changes predominantly result from a bottom-up process driven by a growing number of researchers who increasingly employ social media in their research and initiate globally coordinated research projects while sharing results at an early stage in the research process.

Open science is encompassed in five schools of thought:

- the infrastructure school, concerned with technological architecture
- the public school, concerned with the accessibility of knowledge creation
- the measurement school, concerned with alternative impact assessment
- the democratic school, concerned with access to knowledge
- the pragmatic school, concerned with collaborative research

According to the measurement school, the reputation and evaluation of individual researchers are still mainly based on citation-based metrics. The h-index is an author-level metric that attempts to measure both the productivity and citation impact of the publications of a scientist or scholar. The impact factor is a measure reflecting the average number of citations to articles published in an academic journal and is used as a proxy for the relative importance of a journal.

Numerous criticisms have been made of citation-based metrics, primarily when used, and often misused, to assess the performance of individual researchers. These metrics:

- are often not applicable at the individual level
- do not take into account the broader social and economic function of scientific research
- are not adapted to the increased scale of research
- cannot recognize new types of work that researchers are performing

Web-based metrics for measuring research output, popularized as altmetrics, have recently received much attention: some measure the impact at the article level, others make it possible to assess the many outcomes of research in addition to the number of scientific articles and references. The current reputation and evaluation system has to adapt to the new dynamics of open science and acknowledge and incentivize

engagement in open science. Researchers engaging in open science have growing expectations that their work, including intermediate products such as research data, will be better rewarded or taken into account in their career development. Vice-versa, the use, and reuse of open data will require appropriate codes of conduct requiring, for example, the proper acknowledgment of the original creator of the data.

These ongoing changes are progressively transforming scientific practices with innovative tools to facilitate communication, collaboration, and data analysis. Researchers that increasingly work together to create knowledge can employ online tools and create a shared space where creative conversation and collaboration can occur. As a result, the problem-solving process can be faster, and the range of problems that can be solved can be expanded. The ecosystem underpinning open science is evolving very rapidly. Social network platforms for researchers already attract millions of users and are being used to begin and validate more research projects.

Furthermore, the trends towards open access are redefining the framework conditions for science and thus have an impact on how open innovation is produced by encouraging a more dynamic circulation of knowledge. It can enable more science-based startups to emerge thanks to the exploitation of openly accessible research results. Open science, however, does not mean free science. It is essential to ensure that intellectual property is protected before making knowledge publicly available in order to subsequently attract investments that can help translate research results into innovation. If this is taken into account, fuller and broader access to scientific publications and research data can help to accelerate innovation. Investments that boost research and innovation in open science would benefit society with fewer barriers to knowledge transfer, open access to scientific research, and greater mobility of researchers. In this context, open access can help overcome the barriers that innovative organizations face in accessing the results of research funded by the public.

Open innovation

An open society is the largest producer of knowledge, but the phenomenon of open science is changing every aspect of the scientific method by becoming more open, inclusive, and interdisciplinary. Ensuring open society is at the forefront of open science means promoting open access to scientific data and publications alongside the highest standards of research integrity. There are few forces in this globe as engaging and unifying as science. The universal language of science maintains open channels of communication globally. Open society can maximize its gains through maintaining its presence at the highest level of scientific endeavor, and by promoting a competitive edge in the knowledge society of the information age. The ideas and initiatives described in this publication can stimulate anyone interested in open science research and innovation. It is designed to encourage debate and lead to new ideas on what and open society should do, should not do, or do differently.

An open society can lead to a research powerhouse; however, open society rarely succeeds in turning research into innovation and in getting research results to the global market. Open society must improve at making the most of its innovation talent, and that is where open innovation comes into play. The basic premise of open innovation is to open up the innovation process to all active players so that knowledge can circulate more freely and be transformed into products and services that create new markets while fostering a stronger culture of entrepreneurship. Open innovation is defined as the use of purposive inflows and outflows of knowledge to accelerate internal innovation. This original notion of open innovation was primarily based on transferring knowledge, expertise, and even resources from one company or research institution to another. This notion assumes that firms can and should use external ideas as well as internal ideas, and internal and external paths to market, as they seek to improve their performance. The concept of open innovation is continually evolving and is moving from linear, bilateral transactions and collaborations

towards dynamic, networked, multi-collaborative innovation ecosystems. This means that a specific innovation can no longer be seen as the result of predefined and isolated innovation activities but rather as the outcome of a complex co-creation process involving knowledge flows across the entire economic and social environment. This co-creation takes place in different parts of the innovation ecosystem and requires knowledge exchange and absorptive capacities from all the actors involved, whether businesses, academia, financial institutions, public authorities, or citizens.

Open innovation is a broad term, which encompasses several different nuances and approaches. Two main elements underpin the most recent conceptions of open innovation: the users are in the spotlight and invention becomes an innovation only if users become a part of the value creation process. Notions such as user innovation emphasize the role of citizens and users in the innovation processes as distributed' sources of knowledge. This kind of public engagement is one of the aims of open science research and innovation. The term 'open' in these contexts has also been used as a synonym for 'user-centric'; creating a well-functioning ecosystem that allows co-creation and becomes essential for open innovation. In this ecosystem, relevant stakeholders are collaborating along and across industry and sector-specific value chains to co-create solutions for socio-economic and business challenges. One important element to keep in mind when discussing open innovation is that it cannot be defined in absolutely precise terms. It may be better to think of it as a point on a continuum where there is a range of context-dependent innovation activities at different stages, from research to development through to commercialization, and where some activities are more open than others. Open innovation is gaining momentum thanks to new large-scale trends such as digitalization and the mass participation and collaboration in innovation that it enables. The speed and scale of digitalization are accelerating and transforming the way one designs, develops, and manufactures products, the way one delivers services, and the products and services themselves. It is enabling innovative processes and new ways of doing business, introducing new cross-sector value chains and infrastructures.

Open society must ensure that it capitalizes on the benefits that these developments promise for citizens in terms of tackling societal challenges and boosting business and industry. Drawing on these trends, and with the aim of helping build an open innovation ecosystem in open society, the open society's concept of open innovation is characterized by:

- combining the power of ideas and knowledge from different actors to co-create new products and find solutions to societal needs
- creating shared economic and social value, including a citizen and user-centric approach
- capitalizing on the implications of trends such as digitalization, mass participation, and collaboration

In order to encourage the transition from linear knowledge transfer towards more dynamic knowledge circulation, experts agree that it is essential to create and support an open innovation ecosystem that facilitates the translation of knowledge into socio-economic value. In addition to the formal supply-side elements such as research skills, excellent science, funding and intellectual property management, there is also a need to concentrate on the demand side aspects of knowledge circulation, making sure that scientific work corresponds to the needs of the users and that knowledge is findable, accessible, interpretable and reusable. Open access to research results aims to make science more reliable, efficient, and responsive and is the springboard for increased innovation opportunities, e.g. by enabling more science-based startups to emerge. Prioritizing open science does not, however, automatically ensure that research results and scientific knowledge are commercialized or transformed into socio-economic value. In order for this to happen, open innovation must help to connect and exploit the results of open science and facilitate the faster translation of discoveries into societal use and economic value.

Open Science

Collaborations with global partners represent important sources of knowledge circulation. The globalization of research and innovation is not a new phenomenon, but it has intensified in the last decade, particularly in terms of collaborative research, international technology production, and worldwide mobility of researchers and innovative entrepreneurs. Global collaboration plays a significant role both in improving the competitiveness of open innovation ecosystems and in fostering new knowledge production worldwide. It ensures access to a broader set of competencies, resources, and skills wherever they are located, and it yields positive impacts in terms of scientific quality and research results. Collaboration enables global standard-setting, allows global challenges to be tackled more effectively, and facilitates participation in global value chains and new and emerging markets.

To learn more, visit online at waset.org

Open Science

Scholarly Research Review

The scholarly research review is a multidimensional evaluation procedure in which standard peer review models can be adapted in line with the ethos of scientific research, including accessible identities between reviewer and author, publishing review reports and enabling greater participation in the peer review process. Scholarly research review methods are employed to maintain standards of quality, improve performance, provide credibility, and determine suitability for publication. *Responsible Peer Review Procedure:* Responsible peer review ensures that scholarly research meets accepted disciplinary standards and ensures the dissemination of only relevant findings, free from bias, unwarranted claims, and unacceptable interpretations. Principles of responsible peer review:

• Honesty in all aspects of research
• Accountability in the conduct of research
• Professional courtesy and fairness in working with others
• Good stewardship of research on behalf of others

The responsibilities of peer review apply to scholarly researchers at all stages of peer review: Fairness, Transparency, Independence, Appropriateness and Balance, Participation, Confidentiality, Impartiality, Timeliness, Quality and Excellence, Professionalism, and Duty to Report.

Scholarly Research Review Traits:
• Scholarly Research Review Identities: Authors and reviewers are aware of each other's identity

• Scholarly Research Review Reports: Review reports are published alongside the relevant article

• Scholarly Research Review Participation: The wider academic community is able to contribute to the review process

• Scholarly Research Review Interaction: Direct reciprocal discussion between author(s) and reviewers, and/or between reviewers, is allowed and encouraged

• Scholarly Research Pre-review Manuscripts: Manuscripts are made immediately available in advance of any formal peer review procedures

• Scholarly Research Review Final-version Reviewing: Editorial revision of the language and format is conducted on the final version of the manuscript for publication

• Scholarly Research Review Platforms: The scholarly research review process is independent of the final publication of the manuscript and it is facilitated by a different organizational entity than the venue of publication

All submitted manuscripts are subject to the scholarly research review process, in which there are three stages of evaluation for consideration: pre-review manuscripts, chair-review presentation, and final-review manuscripts. All submitted full text papers, that may still be withstand the editorial review process, are presented in the conference proceedings. Manuscripts are tracked and all actions are logged by internal and external reviewers according to publication policy. External reviewers' editorial analysis consists of the evaluation reports of the conference session chairs and participants in addition to online internal and external reviewers' reports. Based on completion of the scholarly research review process, those manuscripts meeting the publication standards are published 10 days after the event date.

To learn more, visit online at waset.org

TABLE OF CONTENTS

1	ESW01: A Methodology for Approaching the Design of Interior Spaces *Eirini Krasaki*	1
2	Technological Characteristics of Coatings of Ottoman Mosques in Northern Greece *Maria Loukma, Maria Stefanidou*	2
3	2D Ti3C2@MoO3 Composite as an Efficient Anode Material for High-Performance Supercapacitors *Iffat Ashraf, Saba Ahmad, Muhammad Arslan Raza, Ghulam Ali, Syed Rizwan, Mudassir Iqbal*	9
4	Hardware Error Analysis and Severity Characterization in Linux-Based Server Systems *Nikolaos Georgoulopoulos, Alkis Hatzopoulos, Konstantinos Karamitsios, Konstantinos Kotrotsios, Alexandros I. Metsai*	10
5	Automatic ICD-10 Code Assignment of Clinical Narratives Text Using Bert in to Ethiopia National Classification of Diseases (NCoD) *Dereje Weyessa*	17
6	Donoho-Stark's and Hardy's Uncertainty Principles for the Short-Time Quaternion Offset Linear Canonical Transform *Mohammad Younus Bhat*	23
7	Model Predictive Control of a Vehicle Climate System *Karthik Poovendran*	24
8	Dimensioning and Fault Tolerance in an H-P2PSIP Architecture for IPv6-Based Mobility Managemen *Abel Diatta, Mandicou Ba, Bassirou Gueye, Ibrahima Niang*	25
9	Smart and Active Package Integrating Printed Electronics *Joana Pimenta, Lorena Coelho, José Silva, Vanessa Miranda, Jorge Laranjeira, Rui Soares*	33
10	Development Results of an Energy Supply Platform Based on Various Energy Harvesting Technologies *Matthias Fahland, Markus Tuomikoski, Bernd Fuchsbichler, Jan Hesse, Damian Hau, Katrina Morgan*	34
11	Evaluation of Packaging Elements from a Consumer'S Perspective; Empirical Findings from a Study for the Packaged Olive Oil *Angeliki Konstantoglou, Thomas Fotiadis, Kostas Rotsios, Dimitris Folinas*	35
12	Data-Driven Decision Making: A Reference Model for Organizational, Educational and Competency-Based Learning Systems *Emanuel Koseos*	36
13	Technological Applications in Automobile Manufacturing Sector - A Case Study Analysis *Raja Kannusamy*	37
14	Theorizing about the Determinants of Sustainable Entrepreneurship Intention and Behavior *Mariella Pinna*	42
15	Residual Stress Relaxation for the Assessment of Welded Structures *Hugo Tryla, Laurent Barrallier, Lorene Heraud, Philippe Bristiel*	43
16	Effect of Heat Input on the Transverse and Longitudinal Distortion of Multilayer Butt-Welded Joints on Submerged Arc Welding Process *Milad Bahrami, Michel Guillot*	44
17	Assessment of Different Additive Manufacturing Routes to Repair Titanium and Aluminium Alloys: Comparison of Power Bed Fusion and Solid-State Material Deposition Processes *Louise Toualbi, Cecile Davoine, Marc Thomas, Thibaut Bouilly, David Miot*	45

18	Thermo-mechanical Modelling of the Wire Arc Additive Manufacturing (WAAM) Process *Sami Hilal, Djamel Missoum-Benziane, Pierre Kerfriden, Sofiane Hendili, Matthieu Mazière*	46
19	Microstructural Evolution of an Interface Region in a Nickel-Based Superalloy Joint Produced by Direct Energy Deposition *Matthew Ferguson, Tatyana Konkova, Ioannis Violatos*	47
20	Streaming Communication Component for Multi-Robots *George Oliveira, Luana D. Fronza, Luiza Medeiros, Patricia D. M. Plentz*	54
21	On Quasi-LPV System Identification with (Deep) Recurrent Neural Networks *Alexander Rehmer, Andreas Kroll*	59
22	Performance Comparison of Deep Convolutional Neural Networks for Binary Classification of Fine-Grained Leaf Images *Kamal KC, Zhendong Yin, Dasen Li, Zhilu Wu*	66
23	Automatic Atlas Generation: Development of a Plugin for QGIS *Abdelkrim Bensaid, E. Slaoui, R. Nedjai, M. Al-Juhaishi*	74
24	Applying the Crystal Model Approach on Light Nuclei for Calculating Radii and Density Distribution *A. Amar*	75
25	A Real Time Monitoring System of the Supply Chain Conditions, Products and Means of Transport *Dimitris E. Kontaxis, George Litainas, Dimitris P. Ptochos*	81
26	Blue Whale Body Condition from Photographs Taken over a 14-Year Period in the North East Pacific: Annual Variations and Connection to Measures of Ocean Productivity *Rachel Wachtendonk, John Calambokidis, Kiirsten Flynn*	85
27	Fully Coupled Porous Media Model *Nia Mair Fry, Matthew Profit, Chenfeng Li*	86
28	Review of Carbon Materials: Application in Alternative Energy Sources and Catalysis *Marita Pigłowska, Beata Kurc, Maciej Galiński*	87
29	Application of Az Compound as a Bio-Desalination and a Bio-Treatment of Saline Wastewater *Eman Marei*	99
30	Comparative Transcriptome Profiling of Low Light Tolerant and Sensitive Rice Varieties Induced by Low Light Stress at Active Tillering Stage *Darshan Panda, Lambodar Behera, M. J. Baig, Sudhanshu Sekhar*	100
31	Pregnancy and Calving Rates following AI with Buffalo Sexed Semen through Modified Swim up Method *Asma Ul Husna, Shamim Akhter*	101
32	Effect of Heparin and Calcium Ionophore a23187 on Dromedary Camel Epididymal Sperm Functions and In - Vitro Fertilization *Hagar Elsharnoby, Nehal Abu-Elnaga, Omaima Kandil*	102
33	Overcome Effect of Heat Stress on Early Reproduction Success in Buffalo *Omaima Kandil, Ragab Dohreig , Ahmed Abdoon, Esraa Ismael, Hosam Tawfik, Ahmed Shaban, Pascal Mermillod*	103
34	Extracellular Vesicles: A New Insight into Endometritis in Arabian Mares *Sally Ibrahim, Mohamed Hedia, Mohamed Taqi, Mohamed Derbala, Karima Mahmoud, Youssef Ahmed*	104
35	Traces of the Past: The Croatian Nationalist Behavior *Elena Fagogeni*	105
36	Reviewing the Government Policy Response to Child Labour in Bangladesh: A Political Economy Analysis *M. Mahmudul Hoque*	106
37	Women in Nazi Germany: Were They Moulded to be Instruments of Destruction or They Willingly Did So? *Archi Biswas*	107

38	Resistance Movements during the Holocaust *Urmimala Dev*	108
39	History of Holy Water and the Discovery of the Mount Zion Fountain: A Unifying Spring for People of All Faiths *Brandon WinsGaris*	109
40	Social Protection for Ex-Combatants as an Instrument to Reduce Poverty and Enhance Lasting Peace in the Post-Conflict of Colombia *Santiago Luna Castro*	131
41	Federal and State Immigration Policy in the U.S. on Un-Documented and Documented Immigrants: Implications for Economic Hardship, Mental Health, Community and Family Stability *Peter Kamps*	132
42	Machine Learning Methods for Flood Hazard Mapping *Stefano Zappacosta, Cristiano Bove, Maria Carmela Marinelli, Paola di Lauro, Katarina Spasenovic, Lorenzo Ostano, Giuseppe Aiello, Marco Pietrosanto*	147
43	Diagnosis and Structural Assessment of the Belice Bridge in Guatemala City *Jorge L. Campo, Rolando E. Torres, Silvia J. De Paz*	165
44	Food Market Policy Priorities in Georgia *Eteri Kharaishvili*	166
45	Educational Migration and Returnee Employment in Georgia's Labour Market *Manana Lobzhanidze*	167
46	Opportunities for Effective Conflict Management Caused by Global Crises *Marine Kobalava*	168
47	Economic Policy Promoting Economically Rational Behavior of Start-up Entrepreneurs in Georgia *Gulnaz Erkomaishvili*	169
48	Innovations and Agricultural Development Potential in Georgia *Tamar Lazariashvili*	170
49	Industrial Policy Directions in Georgia *Nino Grigolaia*	171
50	Natural Monopolies and Their Regulation in Georgia *Marina Chavleishvili*	172
51	Evaluating Factors Impacting Functioning Management Control Systems Becoming Dysfunctional Beyond Intra-Organizational Boundaries *Martin Kartomo*	173
52	Disclosure Deregulation of Quarterly Reporting *Jochen Zimmermann, Vanessa Behrmann, Lars Hornuf*	174
53	The Disposable Identities; Enabling Trust-by-Design to Build Sustainable Data-Driven Value *Lorna Goulden, Kai M. Hermsen, Jari Isohanni, Mirko Ross, Jef Vanbockryck*	215
54	Infertility Awareness: Knowledge and Attitude of Medical & Non-Medical Moroccan Young People *Sana El Adlani, Yassir Ait Ben Kaddour, Abdelhafid Benksim, Abderraouf Soummani, Mohamed Cherkaoui*	224
55	Turin, from Factory City to Talents Power Player: The Role of Private Philanthropy Agents of Innovation in the Revolution of Human Capital Market in the Contemporary Socio-Urban Scenario *Renato Roda*	225
56	Employee Engagement: The Key to Teaching Self Efficacy *Navya Jayachandran, Abey Edwin, H. Vipin*	232
57	A Mixed-method Study of Psychological Empowerment in Child Protection Practitioners *Amy Bromley*	233

58	Microfluidic Impedimetric Biochip and Related Methods for Measurement Chip Manufacture and Counting Cells *Amina Farooq, Nauman Zafar Butt*	234
59	Association of Hyperuricemia with Length of Stay, Hospital Re-admissions and in-Hospital Mortality in Patients Admitted with an Acute Heart Failure Exacerbation *John Leso, Wafic Itani, Miask Harutyunyan, Hiren Patel, Mikhail Torosoff*	235
60	Cross-Cultural Adaptation and Validation of the Child Engagement in Daily Life in Greek *Rigas Dimakopoulos, Marianna Papadopoulou, Roser Pons*	236
61	Ring Chromosome 14 Syndrome - Characterization, Dental Considerations, and Management *Chris S. Ivanoff, Athena E. Ivanoff, George P. Einstein, Orien L. Tulp, Timothy L. Hottel*	237
62	Gulcin's Method of Cessation Smoking *A.G. Sahan*	238
63	Functional Outcome and Quality of Life of Conservative versus Surgical Management of Adult Potts Disease: A Prospective Cohort Study *Mark Angelo Maranon, David Endriga*	239
64	Ustekinumab Associated with Vasculitis as Adverse Events - Signal Review *Abdulrahman Alomair*	240
65	Creation and Implementation of A New Palliative Care Drug Chart, via A Closed-Loop Audit *Asfa Hussain, Chee Tang, Mien Nguyen*	241

ESW01: A Methodology for Approaching the Design of Interior Spaces

Eirini Krasaki

***Abstract*—** This paper addresses the problem of designing spaces in a consistently changing environment. Space is considered as a totality of forces that coexist in the same place. Forces form the identity of space and characterize the entities that coexist within the same totality. Interior space is considered as a totality of forces which develop within an envelope. This research focuses on the formation of the tripole space-forces-totality and studies the relation of this tripole to the interior space. The point of departure for this investigation has been set the historic center of Athens, a city center where the majority of building mass is unused. The objective of the study is to connect the development of interior spaces to the alterations of the conceptions that form the built environment. The research focuses on Evripidou street, an axis around which expand both commercial and residential centers. Along Evripidou street, three case studies elaborate: a) In case study 01, Evripidou street is examined as a megastructure in which totalities of interior spaces develop. b) In case study 02, a particular group of entities (polykatoikia) that expand in Evripidou street is investigated. c) In case study 03, a particular group of entities (apartment) that derives from a specific envelope is investigated. Throughout the studies and comparisons of different scales, a design methodology that addresses the design of interior space in relation to the dynamics of the built environment is evolved.

***Keywords*—** methodology, research by design, interior, envelope, dynamics.

Eirini Krasaki is with the National Technical University of Athens, Greece (e-mail: eirini.krasaki@googlemail.com).

Technological characteristics of coatings of Ottoman mosques in Northern Greece

Maria Loukma, Maria Stefanidou

***Abstract*—** The paper presents the results of the analysis performed in 45 coatings samples from abandoned Ottoman mosques in Northern Greece. The aim of the paper is to record the technological characteristics of the materials of these buildings, which served as religious centers. Their specific stratigraphy, their components, their reinforcement using natural available fibers and their high porosity were key factors for their preservation. The characteristics reported are taken into account for designing compatible repair materials which will assist their maintenance and establish guidelines for specific problems**.**

***Keywords*—** coating mortars, construction materials, experimental process, Ottoman Balkan architecture.

I. Introduction

THE present paper focuses to the systematic analysis of a number of coating mortars, from various specific areas and accessible places of Ottoman mosques. The study of their technological evolution reveals a great source of information, regarding their application technology, as stratigraphy, layer's characteristics, their physic - mechanical properties, as well as their consistency and structure [3]. The analysis results of this paper, as recorded in the context of the laboratory experimental process, were statistically evaluated and led to their categorization according to their type (render or plaster). The purpose of the paper is to understand the current situation of the coating mortars, as they are multi-role materials, in order to clarify their components and the causes of deterioration as well as to classify their characteristics [4]. This way, the restoration will be based on scientific knowledge and compatible materials will be designed and applied assisting to their preservation [5].

The data collection, contributes to the study of Ottoman architecture in Greece, providing information that has not been recorded until now. This is the main step of the analysis method of the monuments, for the drafting of any future study of monuments restoration. It is also proposed for comparison with other buildings of the same historical period [6] as well as with monuments of another era or different area.

The research of the coating characteristics on each monument is one of the parameters that give information of the building technology [7]. Coatings are mortars which are comprised of the essential ingredients: binder, aggregate and water, with optional additions of various additives [8]. Covering the structure with coatings is a well-known technique in buildings since pre-historic period. Ottoman period is a continuity of the Byzantine period in plastering / rendering technique [9]. The technological characteristics of Ottoman period coatings are mostly based on lime and they usually consist of 2 or 3 layers, differing in the grain gradation and in the binder / aggregate ratio [10].

The main role of coatings is to protect the construction from environmental factors, while increasing its durability. Furthermore, the application of coatings produces indoor comfort by increasing thermal and waterproof insulation. The morphological configuration of the final surface [11], either facades (decoration with colors) or inside the monuments, is typical for the studied structures. The functional use of the coatings was the removal of humidity by allowing the masonry to 'breath' [12]. Their smooth external texture reduces the water penetration in order to restrict their hydrophilic nature. In this manner, the masons could control the porosity of the coatings which is a key factor of their use [13]. The equilibrium between reduced water penetration and high evaporation rate due to the high vapor permeability is due to the nature of the traditional binders combined ideally with the selected aggregates.

The purpose of the coatings of the studied Ottoman mosques, were to cover the exterior and interior spaces with a protective and decorative material. In all cases internally, the domes and the walls are not elaborated and the use of plaster gives an aesthetic covering. Plasters constitute the substrate for decoration either by painted geometric patterns or by prismatic stalactites. Externally, a variety of coatings was observed in the way the final surface is structured. There are cases where render is the final surface of the facades while in other cases finely cut stones and bricks are joined with mortar.

II. METHODOLOGY OF ANALYSIS

II.I Coating fragments sampling

The systematic approach and understanding of the monument, can lead to a focused sampling of building materials. A specific recording procedure was followed at each monument during sampling. It is important to record as much information as possible during the in-situ sampling before removing the mortar from their environment. Thus, each sample received a unique code per monument and cards for gathering the information were formed, the location of each sample was documented photographically and noted on the floor plan of each monument. The type of sample was recorded (plaster, render), and the construction part (wall, arch, dome, minaret, portico). The method of sampling (by hand, chisel, collecting) was noted.

Samples of coatings were taken from 9 Ottoman mosques that are preserved in region of Macedonia at northern Greece

Maria Loukma is with the Aristotle University of Thessaloniki, Greece (e-mail: mloukma@gmail.com).

and a total number of 45 samples were analyzed.

II.II Methodology

The tests performed during the analysis usually aim to identify the microstructure, physic-mechanical and chemical properties of the samples [14]. The analysis is based on a holistic approach in order to record the materials and determine their composition [15].

The experimental procedures were carried out on all the samples from different Ottoman mosques. Initially, each sample was photographed, measured and its color was characterized according to the Munsell Chart. The quantity and preservation state of each sample was also defined. In the majority of samples, microscopical observation was made with the technique of Image Analysis, which gives information on the structure and pathology of the materials. Inclusions, aggregate type and size as well as the present of salts are recorded in the structure and generally the pathology status of the sample (cracks, gaps, adhesion of materials) could be seen.

The physic-mechanical properties of coating samples were determined with the following techniques:

1. Sieve analysis. The sample is scrubbed and sieved according to European Standard (EN1015-1:1998). The following step is the calculation of the percentage of the aggregates passing through the sieves and the design of the aggregates' gradation curve.

2. Determination of open porosity and apparent specific gravity with the indirect method of water absorption under vacuum according to RILEM CPC 11.3 [16].

3. The determination of the compressive strength, is achieved with the method of the direct crushing value of the irregular shaped samples after completing and leveling them with a soft plaster (hemi hydrate gypsum) [17].

The chemical properties of coating were identified with:

1. Atomic Absorption Spectrometry (AAS) for the determination of metal oxides %wt,

2. Ionic Chromatography (IC) for the determination of water-soluble salts %wt.

3. Determination of the aggregates' percentage.

The protocols followed [18], made clear most of the technological characteristics and mechanical properties of the coatings for each monument that was studied. The results in each monument are related mainly to the combination of deterioration factors and the building techniques and materials. In this way the data produced from each monument are available for comparisons and evaluation.

According to their stratigraphy, were divided into two categories. The majority were those of two-layers (a total of 38 samples), and only 7 samples were of one-layer.

III. RESULTS OF ANALYSIS

III.I Analysis of two layer coatings

Samples consisting of two layers mainly cover the interior surfaces of the monuments (walls and domes), with the internal one coming in contact with the masonry and the external layer being the substrate for the final painting.

The chemical composition of binders from both layers showed that they are lime-based mortars with a variety in the percentage of natural pozzolan compounds (Table I-II). The chemical analysis assisted towards the calculation of the Cementation Indices (C.I.) [19]. In most cases, the values of C. I. [20] of the internal layer were higher than those of the external. In one case, the values were similar in both layers. Contrary, only in two cases the C.I. values from the external layer were higher than the internal. (Table I-II).

TABLE I Chemical properties of internal layer binder

monument / sample location	sample code	total soluble oxide (%ww)			C.I.
		Ca0	MgO	$SiO_2+Al_2O_3+Fe_2O_3$	
Koursoum mosque/ interior dome	KK 3.1	36,38	3,15	30,16	1,22
Koursoum mosque/ interior masonry	KK 9α	37,76	1,08	23,27	1,62
Koursoum mosque/ interior masonry	KK 13	34,98	3,55	23,99	1,56
Geni mosque/ interior masonry	EG 5.2	30,78	4,54	56,33	3,84
Orta mosque/ interior masonry	VO 7	26,61	2,35	40,47	3,47
Ahmet Bey mosque/ interior masonry	GA 5	36,4	0,32	42,37	3,16
Iskender Bey mosque/ interior masonry	GI 2.1	36,37	1,66	16,98	1,17
Iskender Bey mosque/ interior dome	GI 3	27,98	2,77	49,53	4,06
Moustafa Bey mosque/ interior masonry	SM 4	33,58	2,65	43,89	3,04
Moustafa Bey mosque/ interior dome	SM 9	39,18	0,53	22,71	1,55
Mehmet Bey mosque/ interior masonry	S.S. 2.1	38,44	3,25	25,86	1,55
Echoxi mosque/ interior masonry	DE 2α	23,79	8,62	14,2	0,9
Echoxi mosque/ interior masonry	DE 4	23,79	6,48	26,43	1,9
Echoxi mosque/ interior masonry	DE 5	40,58	12,14	3,87	0,1

TABLE II Chemical properties of external layer binder

monument name / sample location	sample code	total soluble oxide (%ww)			C.I.
		Ca0	MgO	$SiO_2+Al_2O_3+Fe_2O_3$	
Koursoum mosque/ interior dome	KK 3.2	39,18	2,22	11,96	1,98
Geni mosque/ interior dome	EG 4	54,57	3,73	19,03	0,76
Geni mosque/ interior masonry	EG 5	37,78	0,79	22,03	1,52
Orta mosque/ interior masonry	VO 3.2	33,59	1,26	27,52	2,12
Orta mosque/ interior minaret	VO 8	28	1,56	37,78	2,25
Ahmet Bey mosque/ interior masonry	GA 5.1	33,6	1,14	41,83	3,11
Moustafa Bey mosque/ interior masonry	SM 3.2	39,2	1,51	17,44	1,14
Mehmet Bey mosque/ interior masonry	S.S. 2.2	40,6	0,77	17,67	1,19
Echoxi mosque/ interior masonry	DE 2β	20,99	9,98	17,57	1,36

This is because of the mortar properties the masons wanted

to accomplish. A strong pozzolanic material usually means that the structure is better protected, and can be used as a basis for the next layer. In this case a decorated base of high C.I. mortar in the external layer probably indicates that.

Interesting is the observation of the thickness of the two-layer coatings composition. The internal layer is ranging up to 4cm and the external is much thinner, ranging from 0.30 to 2.2cm. Also, there was a difference in color between the two layers, with the internal one being light red, mainly due to the presence of clay or ceramic material and in some cases whitish, while the external layer usually was white in color (Table III). An indication of the protective role of coatings is the determination of soluble salts [21] carrying out a laboratory test, using liquid chromatography technique (HPLC).

TABLE III: Technological characteristics of two-layer samples

monument/ sample code	stratigraphy	layers' width	color (Munsell scale)
Koursoum mosque/ KK 3	2 layers	internal: 0,7-2cm external: 0,3-1,8cm	internal: light red external: white
Koursoum mosque/ KK 9	2 layers	internal: 3cm external: 1cm	internal: light red external: white
Geni mosque/ EG 4	2 layers	external: 1,8- 2,2cm	internal: light red external: white
Geni mosque/ EG 7	2 layers	external: 1,8- 2,2cm	internal: light red external: white
Orta mosque/ VO 3	2 layers	internal: 0,6- 2,5cm external: 0,1- 0,3cm	internal: white external: white
Orta mosque/ VO 6	2 layers	internal: 0,4- 1,8cm	internal: white
Orta mosque/ VO 7	2 layers	internal: 0,6- 2,5cm	internal: pale yellow
Ahmet Bey mosque/ GA 5	2 layers	internal: 1,6- 2cm external: 0,5- 1,2cm	internal: pink external: white
Iskender Bey mosque/ GI 2	2 layers	internal: 0,6- 0,8cm external: 0,6- 0,7cm	internal: pale red external: white
Moustafa Bey mosque/ SM 3	2 layers	internal: 0,8- 1,9cm external: 0,7- 1cm	internal: pinkish white external: white
Moustafa Bey mosque/ SM 4	2 layers	internal: 1,9- 3,7cm	internal: pinkish white
Moustafa Bey mosque/ SM 9	2 layers	internal: 0,7- 1,9cm external: 0,1- 0,3cm	internal: pinkish white external: white
Mehmet Bey mosque/ SS 2	2 layers	internal: 1,3- 2,7cm external: 0,6- 0,9cm	internal: pinkish white external: white
Echoxi mosque/ DE 2	2 layers	internal: 0,6- 0,8cm external: 0,01- 0,08cm	internal: pale yellow external: white
Echoxi mosque/ DE 5	2 layers	external: 1,8- 2cm	external: white
Arap mosque/ DA 5.3	2 layers	internal: 0,8- 1cm external: 0,3- 0,4cm	internal: white external: white
Arap mosque/ DA 6	2 layers	internal: 0,5- 0,6cm external:	internal: pale yellow external: white

Measurements of the three main soluble salts in the internal layer samples revealed a strong presence in nitrates and sulphates (in relation to measurements in binding mortars), due to the exposition of the coatings to environmental factors (Tab. 4).

The values in chlorides ranged from 0.01%w.w. to 0.35%w.w. (from interior walls). One of the main ways that chlorides salts penetrate is from the ground, where the moisture is increased [22]. Nitrate salts ranged from 0.04%w.w. to 5.56%w.w. (interior walls), probably because of biological factors that often occur in places used as warehouses. Sulphates values ranged from 0.08%w.w. to 2.37%w.w. (from interior walls), penetrating mostly through the underground water or affected by previous interventions.

TABLE IV: Measurements of soluble salts in the internal layer

sample code	Soluble salts Salt content (%w.w.)		
	Cl^-	NO_3^-	SO_4^{-2}
KK 3.1	0,11	0,73	0,70
KK 9α	0,04	0,27	2,37
KK 13	0,75	5,11	0,32
EG 5.2	<0,01	0,04	0,13
VO 7	0,02	0,19	0,08
GA 5	0,01	0,15	0,42
GI 2.1	0,29	2,48	1,39
SM 4	0,03	0,04	1,26
DE 2α	0,31	4,35	0,52
DE 4	0,35	5,56	0,18
DE 5	0,05	0,79	0,15

In the external layer samples, the measurements of water-soluble salts (Tab. 5) were lower in comparison with the measurements in the internal layer samples. The chlorides ranged from <0.01%w.w. to 0.25%w.w., Nitrates ranged from <0.01%w.w. to 1.91%w.w. and Sulphates ranged from 0.03%w.w. to 3.64%w.w. This is an indication of the salt source. The external layers are directly exposed to environment and most of these mosques are in rural environment.

TABLE V: Measurements of soluble salts in the external layer

sample code	Soluble salts Salt content (%w.w.)		
	Cl^-	NO_3^-	SO_4^{-2}
KK 3.2	0,20	0,25	3,64
EG 4	0,13	0,70	0,28
EG 5	0,01	0,04	0,13
VO 8	<0,01	<0,01	0,03
S.S. 2.2	0,27	1,91	2,16

During the laboratory experimental process, macroscopic and microscopic observation of the samples helps to identify some technological characteristics during the fabrication of the coatings.

These are:

1. The presence of special technological figures such as lime lumps and in some cases crushed brick or brick dust (Fig. 1).

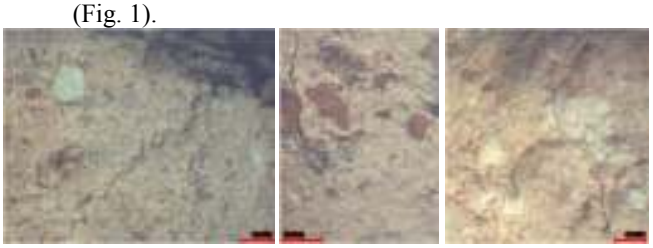


Fig. 1: Microscopic view of the samples DE 4, EG 7, KK 3.1

2. In the majority of the samples there were herbal inclusions [23] from plant fibers such as straws, or twigs (Fig.2), in order to enhance the structure. They reinforce the stability and crack restriction of the material [24]. Herbal fibers can be considered as naturally occurring composites constituted mainly of holocellulose (cellulose, hemicellulose) and lignin with minor contents of sugars, starch, proteins, extractives and ash [25].

Fig. 2: Samples with the presence of herbal fibers in their structure

In one sample the use of two types of herbal origin was recognized (Fig.3).

Fig. 3: Sample from the Orta mosque (VO 3.1)

In another case, the high straw content in the structure of the mortar was estimated at 3%w.w. (Fig.4). The straws found had a variety in length ranging from 2.8 to 4cm and width from 0.2 to 0.35cm, showing that they had undergone some preparation before introduced into the mixture.

Fig. 4: Sample from the Ahmed Bey mosque (GA 5.1)

In two cases, between the two layers of the coating, there was a zone of fibers acting as a grid. In one sample, thin fibers are forming a dividing zone, as a grid, between the two mortar layers (Fig.5).

Fig. 5: The sample from the dome of the Yeni mosque (EG 7)

In another sample, herbal fibers were kneaded within a thin layer of mortar (smaller than 1mm) and this layer was applied between plaster layers. It should have had a strengthening role (Fig.6).

Fig. 6: The sample from the Arap mosque (DA 3)

Only in one fragment animal origin dark-colored fibers, with great hardness were detected. These fibers were not distributed homogeneous in the mortar mass (Fig.7).

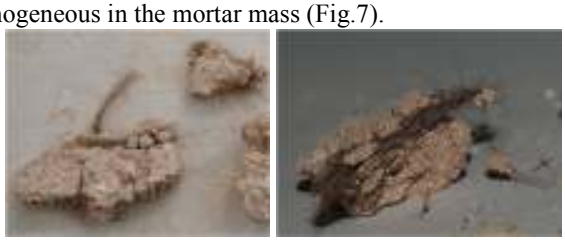


Fig. 7: The sample from the Yeni mosque (EG 7)

3. The determination of the binder in the mortar structure is recorded, by evaluating information from the chemical and microstructural analysis. Measurements in the coating samples showed that they are rich in binder, with the Binder/ Aggregate ratio varying from 1/1 to 4.5/1 (Tab.VI).

TABLE VI The measurements of the Binder/ Aggregate ratio in the samples

sample code	stratigraphy layer	B/A ratio	sample code	stratigraphy layer	B/A ratio
KK 3.1	internal	2,33/1	GA 5.1	external	1,3/1
KK 3.2	external	3/1	GI 3	internal	1,5/1
KK 13	internal	3/1	SM 4	internal	1,4/1
EG 4	external	1/1,45	S.S. 2.1	internal	4,5/1
EG 5.2	internal	1/1	DE 4	internal	1,6/1
VO 7	internal	1,5/1	DE 5	internal	5,2/1
VO 8	external	2,7/1	DA 5.1	internal	1,2/1

Physical properties of the samples were carried out with laboratory tests, concerning the consistency of the materials by determining the open porosity and apparent specific gravity (Tab.VII). Porosity is an important property for the perspiration of the masonry and the removal of moisture, elements that help in the general state of preservation of each construction [26]. Coatings based on lime, can be classified as porous materials, with porosity measurements recorded at fairly high levels from 36.66% to 55.03%.

TABLE VII Physical properties of samples

sample code/ layer	porosity (%)	ap. specific gravity (g/cm^2)
KK 3.1/ internal	38,28%	1,42
KK 9α/ internal	42,87%	1,25
KK 13/ internal	34,92%	1,44
EG 4/ external	29,34%	1,57
EG 5.2/internal	48,61%	1,26
VO 3.1/ internal	43,00%	1,05
VO 7/ internal	37,67%	1,15
VO 8/ external	39,07%	1,28
GA 5.1/ external	55,03%	0,91
GI 2.1/internal	36,66%	1,34

sample code/ layer	porosity (%)	ap. specific gravity (g/cm^2)
GI 2.2/ external	44,03%	1,25
GI 3/ internal	52,69%	1,19
SM 4/internal	41,94%	1,30
SM 7.1/ internal	40,33%	1,20
SM 9/ internal	36,38%	1,44
S.S. 2.1/ internal	44,35%	1,15
DE 4/internal	31,17%	1,42
DA 3/ internal	32,35%	1,16
DA 2/ external	36,25%	1,53
DA 5.1/internal	38,31%	1,42

They are light materials as the specific gravity ranges from 1,05 to 1,44.

The laboratory test on the samples determines the compressive strength of the coatings taking into account the influence of decay over the centuries. Compressive strength was recorded mainly in internal mortar layers due to the sample size. External layers were too thin to measure following this procedure. Thus, the remaining compressive strength, measured, ranged from 0.60 MPa to 3.63 MPa for the samples of the internal layers. In the two samples from the external layer, where the methodology could be applied, low values of 0.58 Mpa were recorded (Tab.VIII). Keeping in mind that compressive strength is not the main property when designing renders/ plasters, it can be concluded that these materials present remarkable mechanical properties.

TABLE VIII: Mechanical properties of samples

sample code/ layer	Compr. Streght (Mpa)	sample code/ layer	Compr. Streght (Mpa)
KK 3.1/ internal	3,63	GI 3/ internal	3,00
KK 13/ internal	3,08	SM 4/internal	2,51
EG 4/ external	0,59	SM 7.1/ internal	0,78
EG 5.2/internal	0,60	S.S. 2.1/ internal	1,13
VO 3.1/ internal	1,69	DE 4/internal	2,38
GA 5.1/ external	0,57		

Finally, the gradation of the aggregates used in the mixture of each coating is calculated from the sieve analysis. This information defines the stability of the mortar structure [27]. In addition, the microscopic observation specifies the type of aggregates, which in all cases concerned the use of natural sand. Differentiation of the gradient range was recorded between the two layers, where in the internal layer the aggregates are usually up to 4mm and in a few cases up to 8mm. In the external layer, the aggregates used were thinner, up to 2mm. Also, there were cases characterized by the lack of aggregates (Fig.8). Fine aggregates (smaller than 1mm) prevail in the structure of many samples.

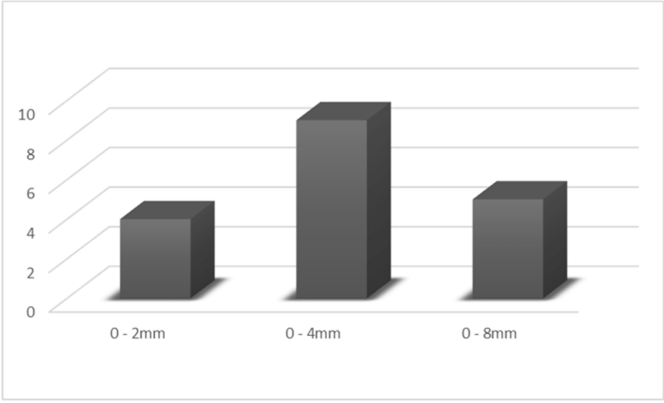


Fig. 8: Range of the aggregate's gradation to the number of samples

III.II Analysis of one-layer coatings

Samples of one-layer were collected from the facades of three mosques (Moustafa mosque SM8, Echoxi mosque DE7, Arap mosque DA2), from the internal walls (Orta mosque VO4, Arap mosque DA3) and from the structure of two minarets (Koursoum mosque KK11, Arap mosque DA8.2).

The samples of this group showed that the mortars are of a lime-pozzolan composition (Tab.IX). The color was white and the thickness ranges from 0.80 to 4cm (Tab.X). Their stability varies from weak, due to disorganization of the structure, up to compact.

TABLE IX: Chemical properties of layer binder

monument name / sample location	sample code	total soluble oxide (%ww)			C.I.
		Ca0	MgO	$SiO_2+Al_2O_3+Fe_2O_3$	
Koursoum mosque/exterior masonry	KK 11	47,6	1,09	27,07	1,48
Orta mosque/ exterior masonry	VO 4	34,97	1,66	20,88	1,54
Arap mosque/ exterior masonry	DA 8.2	29,38	0,27	54,49	5,08

TABLE X: Technological characteristics of one-layer samples

monument/ sample code	stratigraphy	layers' width	color (Munsell scale)
Koursou mosque/ KK 11	1 layer	1,20- 2,50cm	white
Orta mosque/ VO 4	1 layer	0,90- 2,00cm	white
Moustafa Bey mosque/ SM 8	1 layer	0,90- 4,00cm	white
Echoxi mosque/ DE 7	1 layer	2,00- 2,50cm	white
Arap mosque/ DA 2	1 layer	0,80- 1,00cm	white
Arap mosque/ DA 3	1 layer	1,00- 1,60cm	white
Arap mosque/ DA 8.2	1 layer	1,00- 1,50cm	white

In all samples, inclusions such as lime lumps were found. Additionally, charcoal and herbal fibers were included only in one sample (DE 7). Measurements on three coating samples showed that they are materials rich in binder, with the Binder/ Aggregate ratio recorded up to 9/1 (Tab.XI).

TABLE XI: The measurements of the Binder/ Aggregate ratio in the samples

sample code	stratigraphy	B/A ratio
VO 4	1 layer	9/1
DA 8.2	1 layer	9/1
KK 11	1 layer	9/1

Mechanical properties, as the remaining compressing strength measured in the samples (Tab.XII), showed that ranged from 2.14 to 3.08 MPa for the facade samples, while strength was too low for the minaret samples (KK 11, DA 8.2), possibly due to the disorganized structure

TABLE XII: Mechanical properties of samples

sample code/ 1 layer	Compr. Streght (Mpa)
VO 4	1,14
SM 8	2,14
DE 7	3,08
DA 8.2	0,39
KK 11	0,72

Samples of one-layer are characterized as porous materials and allow masonry to breath. Values in open porosity were quite high from 24.39% to 49.92% (Tab.XIII).

TABLE XIII: Physical properties of samples

sample code/ 1 layer	porosity (%)	ap. specific gravity (gr/cm^2)
VO 4	39,87%	1,34
SM 8	32,35%	1,85
DE 7	36,12%	1,37
DA 8.2	24,39%	1,56
KK 11	49,92%	1,09

Finally, the range of aggregate's gradation in the mixture of each coating was up to 8mm. They were natural aggregates (Fig.11). Fine aggregates (smaller than 1mm) prevail in the structure of many samples.

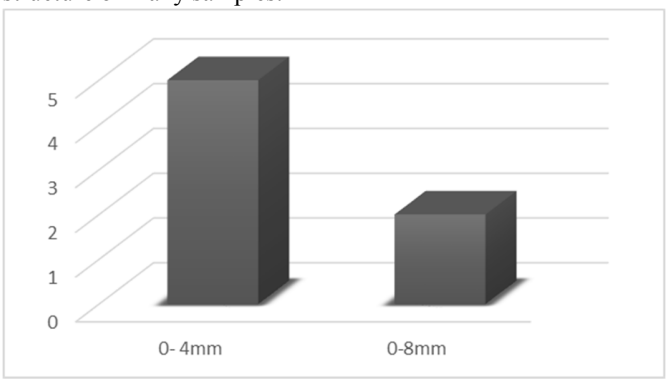


Fig 11: Range of the aggregate's gradation to the number of samples

IV. Conclusions

This study presents on the results of experimental procedures carried out on coatings from nine Ottoman mosques, in order to record their individual and particular technological characteristics. Mosques are preserved in poor conservation condition, mainly due to many changes of use as well as to the lack of systematic maintenance. Moisture and biological factors are the main causes for the decay of materials presented by detachment from the masonry, material loss, and presence of cracks as well as loosening of their structure, salts and various types of deposits and pollutants.

Coatings (renders and plasters) of traditional buildings based on lime (of 1 or 2 layers) had an aesthetic as well a protective role. They are generally considered sacrificial wall coverings, and their main functions are to protect and decorate the walls. In spite of the sacrificial nature, some minimal requirements are imposed to ensure adequate durability of the coatings and prevent their degradation.

The purpose of sampling was to study the characteristics of the mortars and their technical construction as they were carried out during their operation as religious spaces. Thus, they were collected from locations that didn't disrupt the overall integrity but mainly from places that could be clearly identified. The analysis showed that lime-based coatings were used with a variation in the percentage of pozzolanic content. The diachronic practice of using pozzolanic materials has been recorded in the forms as brick dust or natural pozzolans. In addition, in most samples, fibers or animal hair are used to increase the cohesion among mortar's components and reduce shrinkage during setting.

Coatings can be described as porous mortars, mainly consisting of two layers with different characteristics for each layer. The adhesion of the two layers is as strong as with the masonry. This was achieved by selecting proper materials and applying appropriate techniques. The thickness varies from 1-6cm. The internal layer is usually thicker and reddish in color. It is followed by the external one of smaller thickness and white coloring. The application of successive layers functioned very sufficient, by creating a hard and durable skin of the structure as proposal also in literature.

The quality of the constituents presented variation in both layers. Fine sand of siliceous or crushed brick in fine gradation was used in 4,5:1 to 1:1 proportion to binder (by weight). The size of the sand grains was reduced from the internal to the external layer resulting in a smooth finishing layer. The internal layer usually consists of sand up to 8mm. In the external thin layer, the aggregates are usually absent, but there are cases that were measured up to 2 mm.

The aim of this paper is to set out the guidelines for those who work in the field, to provide solutions for specific problems, whether through the use of traditional materials or by deploying modern materials of proven compatibility. In addition, the possibility of studying building materials from a set of monuments, of the same time period and use, aims to develop a methodology for the design of new materials that will be proposed for the restoration of the monument. The main principle governing the mixture of a mortar for replacement purposes (or renovation, as mentioned in EN 998-1), is that the main characteristics of the original material have to be kept as much as possible in order to achieve compatibility and assist to the preservation of the cultural heritage.

Acknowledgment

The fragments of the mortars were taken to the Laboratory of

Building Materials in the School of Civil Engineering of the Aristotle University of Thessaloniki and examined in accordance to the methodology of the laboratory.

REFERENCES

[1]. Final Study- Working paper for the STOA Unit, Technological Requirements for Solutions in the Conservation and Protection of Historic Monuments and Archaeological Remains, M. Cassar (ed.), European Parliament Directorate-General for Research Directorate A,Directorate General for Research, Luxembourg, 2001.

[2]. ICOMOS, Recommendations for the Analysis, Conservation and Structural Restoration of Architectural Heritage, 2003.

[3]. R. Nogueira, A. P. Ferreira Pinto, A. Gomes, Design and behavior of traditional lime-based plasters and renders. Review and critical appraisal of strengths and weaknesses, Cement and Concrete Composites 89, Elsevier Ltd., 2018, pp.192-204

[4]. Y. Govaerts, R. Hayen, M.d.Bouw, A. Verdonk, W. Meulebroeck, S. Mertens, Y. Gregoire, Performance of a lime-based insulating render for heritage buildings, Construction and Building Materials, 159, Elsevier Ltd., 2018, pp. 376-389

[5]. M-d-M Barbero-Barrera, L. Maldonado-Ramos, K. Van Balen, A. Garcia-Santos, F. J. Neila-Gonzalez, Lime render layers: An overwiew of their properties, Journal of Cultural Heritage 15, Elsevier Masson SAS, 2014, pp.326-330

[6]. E. Ipekci, E. Ugurlu Sagin, H. Boke, Interior plastering of Ottoman bath buildings, Case Studies in Construction Materials 11, Elsevier Ltd., 2019, e00295

[7]. J. Eriksson, Jan Erik Lindqvist, Lime render, shrinkage cracks and craftsman ship in building restoration, Journal if Cultural Heritage 37, Elsevier Masson SAS, 2019, pp.73-81

[8]. RILEM TC203 Repair mortars for Historical Masonry, RILEM TECHNICAL COMMITTEE, 2012

[9]. I. Papayianni, M. Stefanidou, Rendering and plasters of Ottoman monuments in Thessaloniki, 2nd International Congress on Studies in Ancient Structures, G. Arun- N. Seckin (eds), Istanbul Turkey 2001, p. 620

[10]. I. Papayianni, M. Stefanidou, Rendering and plasters of Ottoman monuments in Thessaloniki, 2nd International Congress on Studies in Ancient Structures, G. Arun- N. Seckin (eds), Istanbul, 2001, p.620

[11]. R. Nogueira, A.P. Ferreira Pinto, A. Gomes, Design and behavior of traditional lime-based plasters and renders. Review and critical appraisal of strengths and weaknesses, Cement and Concrete Composites, 89, Elsevier Ltd., 2018, pp:192-204

[12]. M. Stefanidou, F. Kesikidou, K. Antoniadis, The Role of Application Techniques for High Performance Traditional Renders, Procedia Environmental Sciences 38, Elsevier B.V., 2017 pp: 242-247

[13]. M. Stefanidou, Methods for measuring porosity of lime-based mortars, Construction and Building Materials, 2010, 24. p. 2572-2578

[14]. ICOMOS, Recommendations for the Analysis, Conservation and Structural Restoration of Architectural Heritage, 2003.

[15]. I. Papayianni, A Holistic Way of Studying Mortars and Bricks of Ancient Masonry for Manufacturing Compatible Repair Materials, Proceedings of the 4th International Symposium on the Conservation of Monuments in the Mediterranean, Rhodes, 1997, pp.265-275.

[16]. M. Stefanidou, Estimation of porosity by using different techniques, Proceedings of the 8th Euroseminar on Microscopy Applied to Buildings Materials, 2001, Athens, pp.619-624.

[17]. I. Papayianni, M. Stefanidou 2016, Comparison of the crushing value of irregular shaped of historic mortars to standard codified compressive strength, 4th International Conference on Historic Mortars, HMC 2016, Santorini, 491-499

[18]. I. Papayianni, The longevity of old mortars, Applied Physics a, Materials Science & Processing, 2006, 83, p.685

[19]. Boynton R.S., 1980, Chemistry and technology of lime and limestone, 2nd edn. Wiley, New York, p.18

[20]. J. Eriksson, Jan Erik Lindqvist, Lime render, shrinkage cracks and craftsman ship in building restoration, Journal if Cultural Heritage 37, Elsevier Masson SAS, 2019, pp.73-81

[21]. I. Papagianni, M. Stefanidou, V. Pachta, S. Konopisi, Content and topography of salts in historic mortars, 3rd Historic Mortars Conference, Glasgow Scotland 2013

[22]. L. Binda, G. Baronio, Mechanisms of masonry decay due to salt crystallization, Durability of Building Materials, 4, p.227-240

[23]. M. Stefanidou, I. Papayianni, V. Pachta, Evaluation of inclusions in mortars of different historical periods from Greek monuments, Archaeometry 54, 4, 2012, pp. 735-751

[24]. A. Izaguirre, J. Lanas, J.I. Alvarez, Effect of a polypropylene fibre on the behavior of aerial lime-based mortars, Construction and Building Materials, 25.2, 2011, pp: 992-1000

[25]. S. Hamza, H. Saad, B. Charrier, Naceur Ayed, Fatima Charrier-El Bouhtoury, Physico-chemical characterization of Tunisian plant fibers and its utilization as reinforcement for plaster-based composites, Industrial Crops and Products, 49, Elsevier B.V., 2013, pp. 357- 365

[26]. M. Stefanidou, V. Pachta, Technological characteristics of coatings applied in monuments of the 4th C BC in Greece, Romanian Journal of Materials 2015, 45 (1). p. 35–42

[27]. M. Stefanidou, I. Papayianni, The role of aggregates on the structure and properties of lime mortars, Cement and Concrete Composites, vol.27, issues 9-10, 2006, pp.914-919

2D Ti3C2@MoO3 Composite as an Efficient Anode Material for High-Performance Supercapacitors

Iffat Ashraf, Saba Ahmad, Muhammad Arslan Raza, Ghulam Ali, Syed Rizwan, Mudassir Iqbal

***Abstract*—** 2D Titanium Carbide (Ti3C2) has attracted significant research interest as an electrode material for energy storage applications. In this paper, we propose a strategy to synthesize a composite electrode comprised of 2D Ti3C2 nanosheets and 1D MoO3 nanobelts by the hydrothermal method of synthesis. Ti3C2@MoO3 composite with unique 2D/1D structure possesses the advantage of excellent electrical conductivity, active sites as well as high charge transfer efficiency. As a result, the Ti3C2@MoO3 composite electrode exhibits a high specific capacitance of 624 Fg-1 at 1 Ag-1 as compared to Ti3C2 MXene (321 Fg-1) and MoO3 nanobelts (258 Fg-1) and excellent cyclic stability with 84 % retention after 5000 cycles. Overall, the novel Ti3C2@MoO3 composite with excellent energy storage features has the potential to be applied commercially as an electrode material for supercapacitors.

***Keywords*—** energy storage, metal oxide, supercapacitor, Ti3C2 MXene.

Iffat Ashraf is with the National University of Science and Technology, Pakistan (e-mail: iffatashraf777@gmail.com).

Hardware Error Analysis and Severity Characterization in Linux-Based Server Systems

N. Georgoulopoulos, A. Hatzopoulos, K. Karamitsios, K. Kotrotsios, A. I. Metsai

***Abstract*—**Current server systems are responsible for critical applications that run in different infrastructures, such as the cloud, physical machines, and virtual machines. A common challenge that these systems face are the various hardware faults that may occur due to the high load, among other reasons, which translates to errors resulting in malfunctions or even server downtime. The most important hardware parts, that are causing most of the errors, are the CPU, RAM, and the hard drive - HDD. In this work, we investigate selected CPU, RAM, and HDD errors, observed or simulated in kernel ring buffer log files from GNU/Linux servers. Moreover, a severity characterization is given for each error type. Understanding these errors is crucial for the efficient analysis of kernel logs that are usually utilized for monitoring servers and diagnosing faults. In addition, to support the previous analysis, we present possible ways of simulating hardware errors in RAM and HDD, aiming to facilitate the testing of methods for detecting and tackling the above issues in a server running on GNU/Linux.

***Keywords*—**Hardware errors, Kernel logs, GNU/Linux servers, RAM, HDD, CPU.

I. Introduction

MODERN servers, including cloud systems and supercomputers, are responsible for critical applications with significant resource requirements. These intense requirements, coupled with the aging of materials, are significant factors that contribute to hardware faults in such servers. These errors can lead to various events, from minor malfunctions to critical errors that cause downtimes. The most frequent errors occur due to faults in the CPU, the RAM and the hard disk drive. For a robust and reliable server function, it is crucial to analyze the above errors and their causes, with the above being the subject of this work.

Even though the advancement of HDDs has been similar to Moore's law, doubling every two years, the demand for storage from large organizations often surpasses their abilities, especially when taking into account factors such as their cost [1]. The above, as well as the importance of the data that may be stored in a server's HDDs [2], constitutes the reliability of these important. A plethora of errors can occur in a hard disk, which should be addressed immediately. Usual examples of HDD errors, either correctable or non-correctable, are the following ones: repeated read/write error, non-accessible files and folders, timeout errors, absence of data from various HDD sectors, disk freeze during system boot etc.

Another important component of servers is the RAM memory. The operating system, the content of program variables, as well as executables and filesystem structures, all make use of the RAM. Overall, all the applications run on a server rely on the above's reliability execution [3]. Taking into account various malfunctions that can occur, data in the random-access memory can be wrongly altered, rending the memory unable to function. Causes for the above malfunctions include ionizing radiation, the aging of materials, etc. [4]-[6]. These malfunctions, caused by the above mechanisms, call for the utilization of DIMM modules in separate RAM chips, with the aim of preserving the error correcting codes (ECCs). The aforementioned correction codes can be used to correct some errors, such as a single-bit error. We note that such an error is called "correctable" (CE) in the relevant bibliography, while the reverse is called an uncorrectable error (UE) [3].

The ECC metadata are exploited for identifying and correcting the errors by a module called the Memory Controller. However, the procedure of correcting the errors can introduce further delays in a server system. Specifically, the procedure for fixing an error requires a hardware exception, the machine check exception (MCE), that he central processing unit is responsible for calling [3]. It should be noted that memory errors can occur, aside from the RAM, in units such as the memory controller itself.

Finally, for the CPU, the dramatic increase of the number of transistors according to Moore's law [7], along with the increase of the cache memory and bus speed, have raised the probability of data corruption, caused by temporary or permanent faults. Combined with the above, the increased adoption of cloud computing systems, of which servers form a crucial component, is increasing the chances for the introduction of CPU errors with great severity, since the services provided can be critical to users and organizations that depend on the cloud. At the same time, a decrease of the time between MCEs can be observed [8], [9].

The subsystem of the CPU responsible for the recording and the detection of CE and UE errors in each of the CPU cores and the northbridge is called the Machine Check Architecture (MCA). This splits the possible machine checks in two types: the silent checks and the MCEs. The former includes errors that are correctable and recorded in model specific registers (MSRs), while the latter are errors that the system cannot correct by itself, and therefore their appearance disrupts the whole system [8]. We can observe a variety of CPU errors, including ECCs, cache-related parity errors and errors of the

Nikolaos Georgoulopoulos and Alkis Hatzopoulos are with the Dept. of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki, Greece (e-mail: ngeorgou@ece.auth.gr, alkis@ece.auth.gr).

Konstantinos Karamitsios, Konstantinos Kotrotsios, and Alexandros I. Metsai are with My Company Projects O.E., Thessaloniki, Greece (e-mail: kk@mycompany.com.gr, kotrotsios@mycompany.com.gr, alexandros.metsai@mycompany.com.gr).

system bus [10].

We describe and analyze selected hard disk (HDD), RAM and CPU errors, observed in kernel ring buffer log files of real-world Linux servers, in Sections II, III and IV, respectively. The following important properties are presented for each error type: the nature of the error, the form of the error message in the system logs, its cause, the error handling and, finally, possible methods to simulate the error. With this work on the hardware errors, we aim to facilitate the analysis of kernel logs, utilized for hardware fault diagnosis and techniques for the detection of these, combined with the system logs [11]-[13]. Moreover, we contribute to the review of possible approaches of the simulation of hardware errors in the aforementioned system components of such a server system [14]-[18].

The family of GNU/Linux operating systems is used by the majority of servers, cloud system, platforms, and data centers, with alternative operating systems holding a much smaller percentage [19]. We utilized two datasets of kernel logs for producing our observations. The first one originates from ten badly maintained servers that host websites, while the second set comes from servers that are properly maintained and belong to the IT Center of a large university (see Section VII). These error handling approaches for each type of error were applied to several of the available servers, with the goal of fixing these issues. Finally, in Section V, we attempt a to provide a prolific discussion for the selected error types and characterization of the severity level is given for each error type. With the above taken into account, we find it important to emphasize that any action for triggering one of the selected hardware errors should not be applied to a production server, but rather to test servers.

II. Analysis of Hard Disk Drive Errors

This section provides an analysis conducted for three separate error types that can occur in the hard disk drives of servers: A) an HDD write error, B) an unrecoverable read error – medium error and, finally, C) a buffer I/O error. The main reasons we selected these specific error types include the following two: these errors are either frequently observed in kernel logs, and therefore their collection and study is made easier, or their occurrence can be simulated without significant effort, with the goal of recording them as actual error messages in the kernel log files.

A. HDD Write Error

1) Description and Causes

One of the most usual error messages that appear in servers running on the GMU/Linux operating system, specifically in the kernel or system log files, is the following HDD write error message: "*writing to 'some_device': No space left on device*". This message's purpose is to inform the user that there is no more free space left in the hard disk, even though it may be obvious to a user or a system administration that the disk is not full. The cause of this type of error is not apparent immediately, and can be attributed to three major causes leading to its generation: a) a previously deleted file is being reserved by another process, b) there are not enough inodes left in the disk to allow for further allocation of space (note that inodes are filesystem metadata that trace file-related information) and c) bad blocks or sectors exist within the filesystem; possible causes may be breakdowns of the filesystem or the deterioration of the disk due to aging, which also causes the disk to malfunction.

2) Error Handling and Simulation

Below we demonstrate the possible ways to tackle an HDD write error, depending on its cause of generation:

a) If there is a deleted file still reserved by another process, the issue can be tackled by restarting this process through the command line, using a GNU/Linux shell such as bash:

sudo systemctl restart <service_name>

b) If the disk has run out of free inodes, the command below can be used to get the available inodes and evaluate if this is actually the cause of the error:

sudo df -i /

In the case where no free inodes exist, the user or system administrator can remove unnecessary files from the system (e.g., from the trash or from temporary file folders) in order to free some inodes and allow the server to continue to function as normal.

c) Finally, for evaluating if the disk contains bad blocks or sectors, the below command can be used:

sudo fsck -vcck <disk_location>

For the simulation of this error, we suggest using the pseudodevice available in the path */dev/full*, which will always return the "*ENOSPACE*" error message when a system user attempts to write upon it. An example of such a command is:

dd if=/dev/zero of=/dev/full

The above will output the error message: "*dd: writing to `/dev/full': No space left on device*".

B. Unrecoverable Read Error – Medium Error

1) Description and Causes

An unrecoverable read error – medium error in a hard disk sector, which is also displayed as an error message in the kernel log, can be interpreted as a read error, for retrieving a sector, that could not be correct even through the system attempted to. This meaning of this is that repeated reads and ECC mechanisms could not retrieve a data block that is not corrupted. This is a strong indication that the disk has deteriorated with the passage of time and could require replacement. We display an instance of this error, recoded in the kernel log, in Fig. 1. As for the possible causes for these errors, they can be attributed to various reasons. The aging of the HDD mentioned above can lead to a lower tolerance to manufacturing malfunctions. Aside from this, the magnetic sectors can naturally weaken. Even more, cosmic radiation can

be identified as a possible cause which can damage the disk. It is obvious that the above reasons can introduce a large degree of randomness in the occurrence of such errors, making the process of simulating these highly useful.

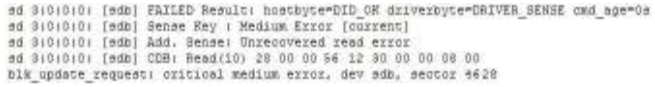


Fig. 1 Unrecoverable read error – medium error message in kernel log

2) Error Handling and Simulation

In the description of the error, we mentioned that frequent appearances of this type indicated that the HDD is approaching the end of its life, though its replacement can be further delayed to reduce costs. A possible way for this is the execution of the "*fsck*" command, that attempts to fix errors. Moreover, the "*smartmontools*" package (S.M.A.R.T. Monitoring Tools) [14] can be utilized for fixing a malfunctioning block with the parallel monitoring of the reports that its two main utility programs provide.

In order to trigger an error of this type, the scsi_debug module [20] can be used. This module is available for all the latest versions of the Linux kernel and supports the production of multiple SCSI (Small Computer Serial Interface) error types along with the usage of various parameters. As an example, the following two commands will trigger an unrecoverable read error – medium error:

sudo modprobe scsi_debug opts=2 every_nth=1
sudo dd if=/dev/sdb of=/dev/null

C. Buffer I/O Error Analysis

1) Description and Causes

This error occurs when a process requests a file stored in the page cache. We display a buffer I/O error in Fig. 2. We notice that the kernel log provides a warning for the user regarding hardware errors in device dm-3 in logic block 0-7. Concerning the case of this type of error, possible reasons include the existence of defective blocks in the HDD or issues with the wiring. In some rare cases, the buffer I/O error can be observed large numbers of HDDs, where multiple errors on these could imply a fault in the disk controller.

2) Error Handling and Simulation

For resolving a buffer I/O error and examining if the error is due to a faulty disk or a faulty wiring, the "*fsck*" command or the smartmontools package [14] are recommended. As for the simulation of a buffer I/O error, this can be achieved with the "*dmsetup tool*", which can create a device mapper with the error target as a parameter, to simulate such errors. An example of such a command is given below (*140* is the number of sectors and */dev/loop0* is the original device):

dmsetup create test --table '0 140 error 1 0 /dev/loop0'

It is important to note that the generation of these errors normally requires the termination of the server's operation. However, this can be tackled by using the below command:

sudo dmsetup remove test

Fig. 2 Buffer I/O error messages in kernel log

III. Analysis of RAM Errors

In this section, we analyze the three most common error types of RAM memories: a) the out-of-memory error, b) the single-bit ECC RAM error and, finally, c) the correctable memory read error. These types of error were selected due to their frequent occurrences in kernel log files, in the same manner as the hard disk errors. Moreover, their appearance can be easily simulated, so as to record them as actual error messages in in kernel logs.

A. Out-Of-Memory Error

1) Description and Causes

An out-of-memory (OOM) error can occur when the system has fallen to low levels of available memory. In any server running on the GNU/Linux operating system, this can lead to "*kernel panic*", a situation that the system will try to avoid [21]. In this manner, the OS will choose to kill a process as a method for protecting the system. The process that is usually killed is the one that has allocated the largest amount of memory. However, these means of protection can cause issues, since the terminated process could be important, either for the user or the system itself. We demonstrate such an error in Fig. 3.

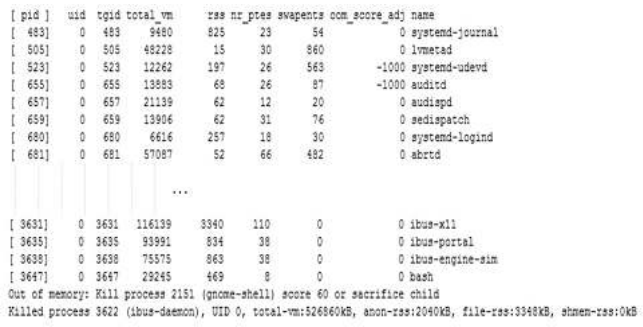


Fig. 3 OOM error message in kernel log

The main causes leading to the introduction of OOM errors are two. The first is attributed to the low capacity in RAM memory for the requirements of the server. As an example, for a server that is meant to handle a high load of services, if its RAM is not sufficient then it is most possible that OOM errors will appear frequently. The second, and harder to detect, causes are possible memory leaks, either from terminated applications and zombie processes, or from processes that are still running, reserving in this way unnecessary resources. These leaks add up and cause the memory allocation to approach its limit.

2) Error Handling and Simulation

Various methods have been proposed for tackling OOM errors. The first and most obvious one is the increase of the total RAM memory of the server, so as to suffice for the actual requirements. Aside from the above, the evaluation of the memory usage of the system can be performed though shell commands like "*ps aux*" and "*free -mt*". As mentioned above, when the system is running out of memory, the process using most of the RAM is identified and considered for termination. Lastly, terminating any stopped processes can be also utilized as a solution, since they are still reserving resources. This can be achieved by running the shell command "*jobs -ps*" to identify the processes IDs of stopped processes/jobs, and the proceeding to kill these using the command "*kill -9 <pid>*".

To simulate this error, one can trigger it manually by running the command:

echo f > /proc/sysrq-trigger

This will call the OOM killer process, which will terminate a process with high memory demands. This will result in the creation of a memory exhaustion event and an OOM error message production in the kernel log. This information will be similar to the bellow message:

localhost kernel: Out of memory: Kill process 3356 (gnome-shell) score 59 or sacrifice child
localhost kernel: Killed process 3356 (ibus-daemon), UID 0, total-vm:458136kB, anon-rss:1024kB, file-rss:648kB, shmem-rss:0kB

B. Single-bit ECC RAM Error

1) Description and Causes

This error message means that a single bit error occurred in a DIMM module of the RAM, and that the ECC mechanism has been triggered to fix the issue. We provide an example of such an error in Fig. 4. We note that if this error occurs frequently in a specific DIMM module, then this could imply that the DIMM has been damaged. In any case, these errors do not affect the system severely, since they can be continuously addressed and corrected. Even more, depending on memory settings, the memory controller module could deactivate the defective DIMM.

Fig. 4 Single-bit ECC RAM error message in kernel log

2) Error Handling and Simulation

As mentioned previously, the frequent occurrence of single-bit ECC errors for a specific DIMM can imply a gradual decline of this DIMM. If this actually the case, various actions can be performed to tackle the problem. First, it can be examined if the DIMM module is correctly connected to its case. Second, the case can be cleared from any traces of dust that can affect its performance. Third, the gold finger connections at the edges of the DIMM can be examined to make sure that they provide proper connectivity. Finally, if none of the above fixes the frequent occurrence of the single-bit ECC errors, it is recommended to replace the DIMM.

As for the possible causes leading to the introduction of this type of error, these can be due to various random factors. To simulate and manually trigger such an error, an old server computer can be used, since its DIMM module of the RAM and its connections with the case will most definitely have degraded, making the occurrences of single-bit ECC RAM errors are highly probable.

C. Correctable Memory Read Error

1) Description and Causes

The occurrence of a correctable memory read error signifies that an error occurred which was corrected from the Error Detection and Correction (EDAC) mechanism of the device [15]. In Fig. 5, we provide an example of this error occurrence, taken from the kernel log. In this case, this message clarifies that a correctable error appeared in DIMM 0 and channel 0 of the memory controller, where EDAC mechanism detected and corrected it.

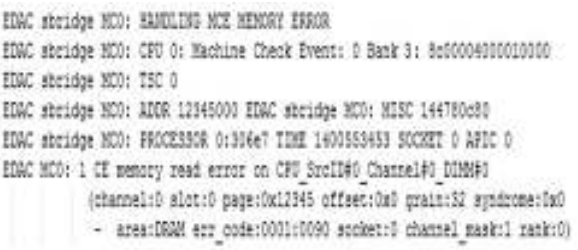


Fig. 5 Correctable memory read error message in kernel log

It is noted by several manufacturers that even a significant number of correctable memory read errors does not damage the RAM [22]. In any case, this is strong indication that the RAM is slowly deteriorating and may cease to operate. As with many similar errors, the main cause for these errors is the aging of the materials, which will lead to failure given enough time.

2) Error Handling and Simulation

Frequent occurrences of correctable memory read errors during small periods of time usually imply the failure of a DIMM module, with the maximum number for normal operation varying between manufacturers. There are two main approaches for addressing these issues. The first one is the detection of the error-occurring point in the memory. This can be achieved through investigating the memory controller file, which can be found at the absolute path "*/sys/devices/system/edac/mc/mc**". This file can inform the user about the row or the DIMM module at which the error appeared, as well as about the number of correctable and

uncorrectable errors that have occurred. For example, using the following command, we can show information about "*ce_count"*, which corresponds to the number of correctable errors:

ls -s /sys/devices/system/edac/mc/mc0

Part of the output is shown as:

0 ce_count 0 csrow1 0 csrow4 0 csrow7 0 reset_counters 0 size_mb

An alternative solution for the appearance of these errors is the preventive maintenance and replacement of a part of DIMMs that encounter these memory read errors. This action can also reduce the possibility of uncorrectable errors and kernel panic events.

For simulating a correctable memory read error, the Error INJection (EINJ) mechanism of the Linux kernel can be used [17]. An example script for triggering the error is given as:

```
cd /sys/kernel/debug/apei/einj
# See which errors can be injected
cat available_error_type
# Set memory address for injection
echo 0x01234560 > param
# Mask 0xfffffffffffff000 - anywhere in this page
echo $((-1 << 12)) > param2
# Choose memory CE
echo 0x8 > error_type
# Perform the injection
echo 1 > error_inject
```

IV. Analysis of CPU Errors

In this section we analyze the main three types of CPU errors: the CPU temperature error, the VB data ECC or parity error and the Lx BTB multi-match error. The CPU constitutes one of the most reliable parts of the system [24], but it is also the most critical part of the hardware. In some cases, the CPU can produce a variety of errors that can affect the performance of server. We selected the following CPU error types since their appearance inside our kernel log dataset was more frequent than other types of such errors. Moreover, they are different regarding their nature and cause. It is important to note that there is no straightforward method for the simulation of these CPU errors.

A. CPU Temperature Error Analysis

1) Description and Causes

This error specifies that the CPU is overheating and activates a mechanism which causes an MCE exception. If this error persists, then so does the overheating of the core, an issue which can lead to permanent damage in the hardware. We provide an example of a CPU temperature error message in Fig. 6, as observed in a kernel log file. Usual causes for overheating issues leading to the appearance of the CPU temperature error include: a problematic CPU heat sink, b) malfunction of the CPU fan, c) a defective thermal pipe, d) an insufficient or aged thermal paste and e) CPU overclocking.

```
CPU0: Core temperature above threshold, cpu clock throttled (total events = 202)
CPU2: Core temperature above threshold, cpu clock throttled (total events = 202)
CPU1: Package temperature above threshold, cpu clock throttled (total events = 202)
CPU3: Package temperature above threshold, cpu clock throttled (total events = 202)
CPU2: Package temperature above threshold, cpu clock throttled (total events = 202)
mce: [Hardware Error]: Machine check events logged
```

Fig. 6 CPU temperature error message in kernel log

2) Error Handling

Methods for addressing CPU errors, with regard to their cause, include a) cleaning the CPU for accumulated dust, b) installing a thermal management tool [23] for effectively monitoring the CPU's temperature, c) updating the BIOS and d) replacement of the thermal paste, fan, heat sink, etc. As mentioned, simulating CPU errors is not a straightforward process.

B. VB Data ECC or Parity Error Analysis

1) Description and Causes

A VB data ECC or parity error specify that the CPU cache diagnosed a parity error and proceeded to correct it automatically. We demonstrate an example of parity error in cache L2 in Fig. 7. Although the parity error in a cache can easily be corrected, frequent occurrences of this error denote that the CPU has begun to malfunction. Various causes can be attributed: a) the CPU may suffer random faults due to radiation strikes, b) the CPU may be underfunctioning due to low power consumption, overheating or overclocking, c) the CPU has begun to break down due to physical decline and aging and, finally, d) dust or dirt inside the CPU fan are causing overheating.

2) Error Handling

We present various methods for tackling a parity error in a cache, depending on its cause. First, inspecting the voltage in the BIOS can be performed, should it not meet the required criteria. In such case, either a replacement of the voltage supply can be conducted, or a motherboard check can be performed for defective capacitances. Second, for cases of high temperature, cleaning the CPU fan and replacing the thermal paste can also be performed. Third, reduction of overclocking or SMALL INCREASES IN THE SUPPLIED POWER CAN ALSO RESOLVE THE issue. Last, if none of the above improves the situation, replacing the CPU is advised.

```
mce: [Hardware Error]: Machine check events logged
[Hardware Error]: Corrected error, no action required.
[Hardware Error]: CPU:0 (15:0:0) MC2_STATUS[-|CE|MiscV|-|-|-|-|CBCC]: 0x8c0040010e01
[Hardware Error]: MC2 Error: VB Data ECC or parity error.
[Hardware Error]: cache level: L2, tx: DATA, mem-tx: RV
```

Fig. 7. VB data ECC or parity error message in kernel log.

C. Lx BTB Multi-Match Error Analysis

1) *Description and Causes*

An Lx BTB multi-match error in Instruction Fetch Unit is

considered harmless and could be omitted, since it can be easily corrected. However, the accumulation of large numbers of such errors could drive the MCA threshold counter to overflow, resulting in thresholding interrupts. In Fig. 8 we provide an example of a multi-match error in the Instruction Fetch Unit, as recorded in the kernel log. For cases where this error is rarely detected by the system, it is probably attributed to sporadic random faults. However, in cases where this error appears frequently, it could signify issues due to the overclocking of the RAM.

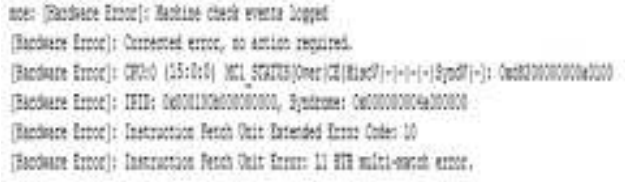


Fig. 8. Example of an L1 BTB multi-match error message in kernel log.

2) *Error Handling*

When this error is generated frequently during a small period, proper clocking of the RAM, in case it has been overclocked, can cause the error to disappear. Alternative solutions include clearing the Counter Present bit of MCA_MISC0 of Instruction Fetch bank, to block the MCA thresholding mechanism from producing several interrupts due to a multi-match error. Aside from that, updating the BIOS firmware to comply with the specifications of the manufacturer can resolve the error.

V. Discussion and Severity Characterization

From the analysis of the selected CPU, RAM and HDD errors in GNU/Linux based server systems we can gather useful insights. First, simulating these hardware errors is only possible for the ones that concern the HDD and RAM, and not for the ones related to the CPU. This highlights the need of mechanisms that facilitate the simulation of such errors, with the introduction of newer error injection utilities. Second, the CPU errors have a larger degree of randomness than the ones that concern the HDD and RAM. Furthermore, with regard to the handling of the errors, we successfully applied all of the proposed approaches to various servers of our computing infrastructure, leading to the disappearance of the error messages in the kernel logs. In addition to this, the correction mechanism of the GNU/Linux operating system managed to catch and fix all the correctable hardware errors that appeared. Finally, this analysis can provide a guide for future fault diagnosis schemes and methods for predicting hardware failures, such as methods based on machine learning and deep learning.

Characterization of the severity level is attempted for the selected hardware errors presented in Sections II, III and IV. Three categories of severity are considered: notice, medium-level and critical. In Table I, the severity level for each selected HW error is displayed. The HDD write error, single-bit ECC RAM error and Lx BTB multi-match error are characterized as notice errors, because there is not a significant negative impact in the system due to their appearance, while minimum attention is needed from the operator to handle them. The buffer I/O error, correctable memory read error and CPU temperature error are classified as medium-level errors, as there is medium impact to a server system and larger error handling effort than notice errors is required. Finally, the unrecoverable read error – medium error, OOM error and VB data ECC or parity error are considered as critical errors, as they can have a big impact in the proper system operation and the support team might need additional hardware or even to replace certain hardware parts due to their deterioration.

TABLE I. Severity Level of the Selected HW Errors

Error Type	***Severity Level***
[HDD] HDD write error [RAM] single-bit ECC RAM error [CPU] Lx BTB multi-match error	Notice
[HDD] Buffer I/O error [RAM] correctable memory read error [CPU] CPU temperature error	Medium Level
[HDD] unrecoverable read error – medium error [RAM] OOM error [CPU] VB data ECC or parity error	Critical

VI. Conclusions

In this work, we presented an analysis of hardware errors that concern the HDD, RAM, and CPU in GNU/Linux based server systems. The causes of each error were described, as well as possible ways to simulate HDD and RAM errors to facilitate their study and the development of methods to overcome these. Furthermore, we provided error handling methods which we successfully tested on our servers. Finally, a severity characterization for each error was introduced depending on the potential impact to a server system and the actions required for fixing an issue. The overall analysis can assist the users for deeper analysis of kernel log files. For future work, hardware failure prediction techniques based on deep learning will be implemented and possible analysis and severity characterization of more HW error types will be considered.

VII. Acknowledgements

This work has been co-funded by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project code: T2EDK-00168). The IT Center mentioned in the introduction belongs to the Aristotle University of Thessaloniki (AUTh) High Performance Computing Infrastructure and Resources.

References

[1] B. David (July 1, 2015). “Prices for Data Storage Equipment and the State of IT Innovation”. The Federal Reserve Board FEDS Notes, 2015

[2] G. Amvrosiadis, A. Oprea and B. Schroeder, “Practical scrubbing: Getting to the bad sector at the right time”, IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2012), pp. 1-12, 2012.

[3] J. Meza, Q. Wu, S. Kumar and O. Mutlu, “Revisiting Memory Errors in Large-Scale Production Data Centers: Analysis and Modeling of New Trends from the Field”, 45th Annual IEEE/IFIP International Conference on Dependable Systems and Networks, pp. 415-426, 2015.

[4] T. C. May and M. H. Woods, “Alpha-Particle-Induced Soft Errors in Dynamic Memories”, IEEE Transactions on Electron Devices, 1979.

[5] C. Constantinescu, “Trends and Challenges in VLSI Circuit Reliability”, IEEE Micro, 2003.

[6] P.-F. Chia, S.-J. Wen and S. Baeg, “New DRAM HCI Qualification Method Emphasizing on Repeated Memory Access”, IRW, 2010.

[7] B. G. Streetman, S. Banerjee, “Solid state electronic devices”, Boston: Pearson. p. 341, 2016.

[8] A. Kleen, “Machine check handling on linux”, SUSE Labs, 2004.

[9] N. Pandit, Z. Kalbarczyk and R. K. Iyer, “Effectiveness of machine checks for error diagnostics”, 2009 IEEE/IFIP International Conference on Dependable Systems & Networks, pp. 578-583, 2009.

[10] Intel Corporation, “Machine Check Architecture”, in Intel® 64 and IA-32 Architectures Software Developer’s Manual Volume 3B: System Programming Guide, Part 2, 2018.

[11] A. Das, F. Mueller, C. Siegel and A. Vishnu, “Desh: Deep Learning for System Health Prediction of Lead Times to Failure in HPC.”, Proceedings of the 27th International Symposium on High-Performance Parallel and Distributed Computing - HPDC ’18, 2018.

[12] I. Giurgiu, J. Szabo, D. Wiesmann and J. Bird, “Predicting DRAM reliability in the field with machine learning”, Proceedings of the 18th ACM/IFIP/USENIX Middleware Conference on Industrial Track - Middleware ’17, 2017.

[13] X. Sun et al., “System-level hardware failure prediction using deep learning”, 56th ACM/IEEE Design Automation Conference (DAC), pp. 1-6, 2019.

[14] E. Nemeth, G. Snyder, T.R. Hein, B. Whaley “Unix and Linux System Administration Handbook”. Pearson Education. p. 366, 2010.

[15] S. M. Hancock, “Tru64 UNIX troubleshooting: diagnosing and correcting system problems”, Digital Press, 2002.

[16] The kernel development community, “Error Detection And Correction (EDAC) Devices” , https://www.kernel.org/doc/html/v4.14/driver-api/edac.html, 2020.

[17] APEI Error INJection, https://www.kernel.org/doc/Documentation/acpi/apei/einj.txt.

[18] Memtest86+, www.memtest.org.

[19] "Usage of operating systems for websites". W3Techs. Technologies, Operating Systems, 7 March 2015.

[20] Scsi_debug adapter driver for Linux, http://sg.danny.cz/sg/sdebug26.html.

[21] G. Kroah-Hartman, “Linux kernel in a nutshell”, O'Reilly Media Inc., p. 59, 2007.

[22] Oracle, “Troubleshooting DIMM Problems”, https://docs.oracle.com/cd/E19121-01/sf.x4250/820-4213-11/dimms.html.

[23] “Linux Thermal Daemon Monitors and Controls Temperature in Tablets, Laptops”, https://www.linux.com/news/linux-thermal-daemon-monitors-and-controls-temperature-tablets-laptops/.

[24] E. B. Nightingale, J. Douceur and V. Orgovan, “Cycles, Cells and Platters: An Empirical Analysis of Hardware Failures on a Million Consumer PCs”, Proceedings of EuroSys 2011, 2011.

Automatic Interpretable ICD-10 Code Assignment of Clinical Narratives Using BERT based model case study of Ethiopia National Classification of Diseases (NcoD)

Prepared by: Dereje Gizaw (PHD candidate)

Computer science and Engineering department, ASTU, 2021

Submitted to: Conference

Abstract

Clinical notes are essential medical documents to record each patient's symptoms. Each record is typically annotated with medical diagnostic codes, which means diagnosis and treatment. ICD coding is valuable and challenging.This paper focuses on predicting diagnostic codes given the descriptive present illness in electronic health records by leveraging alivating the hierarci Al nature of ICD code with recent technology called transformer technology . The challenge this paper addresses are two fold, Firstly, In order to capture the relationship between labels on the hierarchical label structure, investigating how to learn label representation using BERT pre training based models that could leverage the hierarchical relationship among the class labels is the focus of this study. Thus as contribution exploring the greedy search algorithm for the design of meta path using transformer architecture that guide random walk so that generate label sequence . Secondly, beside this , model interpretability for practical automatic (ICD) code assignment in free clinical text National Classification of Disease (NCoD) Ethiopia using MIMIC data set have been studied.

Keyword:ICD coding;BERT; Explainblity;EHR

1. INTRODUCTION

Using AI in developing nations who do not have the resources will diminish the need for outsourcing and can improve patient care. AI can allow for not only diagnosis of patient is areas where healthcare is scarce, but also allow for a good patient experience by resourcing files to find the best treatment for a patient.[1]

The ICD is the international standard diagnostic classification for all general epidemiological, many health management purposes for clinical use. These include the analysis of the general health situation of population groups and monitoring of the incidence and prevalence of diseases and other health problems in relation to other variables such as the characteristics and circumstances of the individuals affected, reimbursement, resource allocation, quality and guidelines. The tenth revision of the ICD came into use in WHO [3]

The (FMoH) Federal ministry of health has long since recognized the benefits of ICT to support and transform the health sector of

Ethiopia [3]. Ethiopia represented national health enterprise architecture based on the interventions in Information Revolution Roadmap document. Under the standards-based registries data dictionary were included to facilitate interoperability and harmonized diagnosis documentation [8] The FMoH had designed two national health data dictionaries (NHDD) i.e. National Classification of Disease (NCoD) and Health Management Information System (HMIS) indicators definitions. The National Classification of Diseases (NCoD) was developed as part of the HMIS reform to organize health information. It is the national standard system of defining, monitoring and reporting health conditions and their outcomes. The preparation of the NCoD started with a countrywide study of samples from service points & administrative units.

The main target of the revised disease classification is morbidity and mortality reporting with focus on priority diseases. The national level report is limited only to diseases that are relevant for epidemiological surveillance, planning and management purposes. The NCoD comprises of three documents that incorporate essential list of diseases corresponding to the diagnostic capacity of the respective level of the health care system. Out of the three components NCoD, the Extended Edition with 2,055 diseases was developed and used as a reference for tertiary level health care providers. The NCoD is a simplified disease classification system; it is organized according to the ICD-10 framework with specification of ICD-codes corresponding to each diagnosis or disease categories. It is also fully mapped to the ICD10.

As preliminary investigation of this study it is one of challenges in building EMR system in Ethiopia healthcare system since it is observed ICD-10 codes assignment in clinical text currently done only manual assignment which resulted in most of hospital in Ethiopia doctors are hesitant due to the complex task and length of medical notes nature of ICD encoding also it leads to error prone , time consuming and needs a lot of trained encoder staff as Ethiopia health context .

Existing clinical literature review indicate clinical narratives free text not addressed are: often, terms in the report diverge strongly from standard reference terminology. Focus automatic ICD-10 assignment to clinical diagnosis and procedures, as Data Representation clinical narratives: out of vocabulary (OOV) words, problematic statements, abbreviations and infrequent words in the normalization of concept makes the ICD code assignment to clinical text specifically , Challenges : Hierarchical dependencies are explicitly introduced into the model is via the use of a hierarchical loss function to penalize hierarchical inconsistencies across the model's prediction layer. [2] The reduction of the number of classes is efficient in recent study.

Due to the advantage of transformer models such as BERT over CNN and RNN is that they can handle long-range dependencies. Hence, if longer text could be fed at once in to BERT encoder it would be possible to find relationship and patter over longer span of text. BERT: Bidirectional bi directional Encoder Representations from Transformers and it is a recent, a deep transformer network model. [6]

Recent models with regards to automatic ICD-10 code assignment for clinical narratives free text state-of–the-art developed are which differentiate this study from previous ones focused on the following models, these are :BERT-ICD by

[5];SnipBERT by [8]; MIMIC-BERT by [9]; Hypercore by [10]; CAMEL and DR-CAMEL by [11].

All of the above models except CAMEL and DR-CAMEL are based BERT : Bidirectional bi directional Encoder Representations from Transformers and it is a recent , a deep transformer network model of Deep neural network for NLP natural language understanding such as automatic ICD code assignment for clinical text in healthcare system. Building a bigger dataset and training on BERT [6] for construction of feature space containing clinical codes from different hierarchical level; Adding more preprocessing to clinical narratives BERT since More examples per class given the number of classes to assign, and better word or concept vector representation of documents in order to accurately assign codes

For instance as challenge in automatic ICD code assignment for clinical text task as multi-label classification problem, the author called [4] stated that in the clinical application of medical code assignment, diagnosis and procedure codes are inferred from lengthy clinical notes such as hospital discharge summaries. However, it is not clear if pertained models are useful for medical code prediction without further architecture engineering.

2.0 Statement of problem

According to [5]comparison of different state of the art models for automatic ICD code assignment as limitation and motivation ,the advantage of transformer models like BERT over CNN and RNN is that they can handle long-range dependencies. Hence, if longer text could be fed at once in to BERT encoder it would be possible to find relationship and patter over longer span of text. Since existing deep learning based models for multi class text classification considered non sequence and long distance semantics ,treat output as independent method but ignoring hierarchical relationship among them which leads to a substantial loss of useful semantics relationship, as nature of ICD code is hierarchal nature. In order to capture the relationship between labels on the hierarchical label structure, exploring to design a meta path using [transformer architecture search as proposed by [7] apply algorithms like greedy search ,to guide random walk and generate label sequence to learn label representation using BERT pre training based models to leverage the hierarchical relationship among the class labels. Beside this , inference interpretability is a crucial challenge and obstacle for practical automatic ICD code assignment in free clinical text, since professional are willing to be convinced by the model insight of vital supporting information or decision-making process will be investigated.

Unlike the general trend in language processing, few transformer model had been reported to reach high performance on this task. [5] The motivation for this study stated that their work sets a solid foundation for further research on ICD coding and suggests that overcoming the exposed limitations of BERT-based models is likely to lead to a new state-of-the-art. Furthermore, the authors believed that the interpretability of ICD coding models is an interesting avenue for future work, which can benefit from a large body of existing research. The exploration in Interpretability which makes how the decision made by the algorithms with justification to convince the clinicians using feature engineering on mapping of clinical notes to ICD-10 code assignment in health care system also investigate in the Logical Inference (sound reasoning) which is a process that implements the entailment

relation between sentences will be considered.

Therefore ,In this study unlike the previous studies, focus on investigation of novel automatic ICD coding assignment based on a hierarchical taxonomy aware fine tuning and BERT pertaining models while decision can be explained as case study and practice oriented on Ethiopian NCoD.

Thus answer the following research questions : to what extent can the capabilities of hierarchical transformer network be transferred and applied to automatic ICD-code assignment Focus in order to capture the categories underlying structure for exploiting the hierarchical of codes and co-occurrence. Beside investigate Explainblity in the design and deployment automatic ICD code assignment of clinical free text in the EMR system considering the discharge summaries clinical notes with MIMIC III data set and also considering Ethiopia context of ICD-10 of NCoD for exploring strategies of greedy search on transformer architecture , BERT based model that leverage the issues of hierarchical nature of ICD code.

3. Conceptual Framework

From the data set MIMIC III unstructured data is preprocessing such that data cleaning and annotation will be conducted then in the state of the art pervious related literature will be analyzed , Reports from ICU department of discharge summaries data base from with the matching from NCoD /ICD-10 codes and a framework will be introduced as model based state of the art [5] as data science research process will be conducted so that an improved or new end to end pipe line model including the health care system as novel model or algorithms for mapping unstructured data in to NCoD or ICD 10 will be designed and implemented as prototype solution with decision can be explained.

4.0 Literature review

The state-of-the-art of automatic ICD coding has been dominated by deep learning models lie CNN , RNN, LSTM but specially in last three years , Since the appearance of the Transformer model [14], transformer-based architectures [12] have become state-ofthe-art in almost every area of Natural Language Processing [13] thanks to their ability to handle long range dependencies.

BERT [6] a bidirectional transformer, is of particular importance since it is the basis of many other language understanding models. Nonetheless, given the specific characteristics of medical text, e.g., specialized vocabulary, models pre trained on generic language, like BERT, do not reach high performance on biomedical language understanding tasks.

Therefore, specialized models, such as BioBERT [14] or ClinicalBERT [14] , pre trained on medical text have been proposed. In particular, the recent PubMed BERT model (Gu et al., 2020) is the state-of-the-art in the BLURB benchmark (Gu et al., 2020), a benchmark for biomedical language understanding which includes , document classification, relation extraction, sentence similarity and evidence-based medical information extraction. Despite its prominence in medical language understanding, automatic ICD coding escapes the set of tasks where BERT-based models excel. To the best of our knowledge, few BERTbased model has been proposed yet that reaches competitive performance on ICD coding on the MIMIC-III dataset. In this study, investigate in detail BERT-based ICD coding and as proposed by [6] identify existing limitations and opportunities.

Considering the following recent models for automatic ICD 10 code assignment for

clinical text , these are :BERT-ICD by [5]; SnipBERT by [8]; MIMIC-BERT by [9]; Hypercore by Pengfei Cao etal,2020]; while also considering the CNN based models like CAMEL and DR-CAMEL by [11]

4.0 Research Methodology

Qualitative approach using case study method which can provide tools for researchers to study complex phenomena with in their contexts. Case study , for the guidelines to assign codes to discharge summaries report in context NCoD will be studied; structured Interview ,literature review and observation for data collection

Beside this, **Experiment Design:** Conduct Experiment based on as Base line: [5] [2] Preprocessing : discharge summaries report from EMR software of ICU department health data set corpus Ethiopia healthcare system consideration for comparison , beside international known ICD 10 code assignment data set called MIMIC-III will be applied and train several BERT models based will be done to answer the research questions .

The performance Evaluating:

The models is evaluated against common baselines beside experimental evaluation of to answer research question

Use multiple methods:case study and experimental. Also will use statistics to assess the outcomes of experiment. Statistically description will be provided on distribution of codes in data raw, test etc.

Experimental Evaluation

Conduct series of experiment on transformer network along with BERT using dataset described as follow: Dataset: MIMIC-III dataset and discharge summaries report beside EMR software of database from Hawassa compressive specialized hospital, SNPR .

Preprocessing and Data set:

Reports from ICU department of discharge summaries data base from with the matching from NCoD /ICD-10 codes of the hospital of Hawassa and MIMIC.Tasks will be defined on dataset explained. Learning experimented with model to check the AOC F1 ,accuracy of it. List of parameters is as follow: Parameter Name : Number of epochs, Batch size, Number of filters, Filter sizes, Embedding dimension, Regularization, Dropout probability

Evaluation criterion as [5] is based on the following metrics: macro ACU and micro ACU beside Accuracy and Error Rate. Since these are primary metrics to evaluate the quality of a classification model. Precision is a measure for which model is correctly predicting any particular category and F1 score is a harmonic mean of p and R .

References

[1] J, Li B , The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries. 2018

[2]. zdravstvena; International Statistical Classification of Diseases and Related health Problems, the 10th Revision,, 1992 .

[3]. FMoH Information Revolution Roadmap, National Health Information Enterprise Architecture, Apr 2016

[4]. Ji etal ,Does the Magic of BERT Apply to Medical Code Assignment? A Quantitative Study,2021

[5]. Pascual, etal , Towards BERT-based Automatic ICD Coding: Limitations and Opportunities,2021

[6]. et al; Bidirectional Encoder Representations from Transformers by 2019

[7]. Isken etal,2017

[8]. Huang etal,Snip BERT,2020

[9].,KbhavaniaSighetal,MIMIC BERT,2021

[10]. Cao etal,hypercore ,2020

[11]. etal,CAMEL and DR-CAMEL, 2018

[12]. Et al., 2020; Lewis et al., 2020; Raffel et al., 2019

[13]. et al., 2018a, 2019

[14].Alsentzer, ClinicalBERT , 2019

Donoho-Stark's and Hardy's Uncertainty Principles for the Short-time Quaternion Offset Linear Canonical Transform

Mohammad Younus Bhat

***Abstract*—** The quaternion offset linear canonical transform (QOLCT) which is time- shifted and frequency-modulated version of the quaternion linear canonical transform (QLCT) provides a more general framework of most existing signal processing tools. For the generalized QOLCT, the classical Heisenberg's and Lieb's uncertainty principles have been studied recently. In this paper, we first define the short-time quaternion offset linear canonical transform (ST-QOLCT) and drive its relationship with the quaternion Fourier transform (QFT). The crux of the paper lies in the generalization of several well known uncertainty principles for the ST-QOLCT, including Donoho-Stark's uncertainty principle, Hardy's uncertainty principle, Beurling's uncertainty principle, and Logarith- mic uncertainty principle.

***Keywords*—** Quaternion Fourier transform,Quaternion offset linear canonical trans- form,Short-time quaternion offset linear canonical transform,Uncertainty principle.

Mohammad Younus Bhat is with the Islamic University of Science and Technology, Awantipora, India (e-mail: gyounusg@gmail.com).

Model Predictive Control of a Vehicle Climate System

K. Poovendran, D Abel

***Abstract*—**Vehicle Cabins are known to be complex non-linear systems, in this work, the thermal behaviour is considered. A cabin model is used within a Model Predictive Control algorithm, the algorithm is altered in order to consider numerous constraints, showing performance benefits over standard Proportional Integral Differentiating Controllers. This is predominantly noticeable in disturbance compensation, steady state tracking, degree of decoupling as well settling time. The control strategy is applied on both a simulation environment and an actual vehicle control unit. The achieved results show potential to use a model predictive control algorithm in future production vehicles.

***Keywords*—**Grey box modelling, Identification for control, Experimental design, Model predictive control

K. Poovendran is with the Department responsible for Thermal Comfort and HVAC Control, Mercedes Benz AG, 1 Benz-Str, Sindelfingen, Germany (phone: +49-176-309-90534; e-mail: karthik.poovendran@daimler.com).

D. Abel is with the Institute of Automatic Control, RWTH Aachen, Campus Boulevard 30 Aachen, Germany (e-mail: d.abel@irt.rwth-aachen.de).

Dimensioning and Fault Tolerance in an H-P2PSIP Architecture for IPv6-Based Mobility Management

Abel DIATTA[1], Mandicou BA[2], Bassirou GUEYE[3] and Ibrahima NIANG[3]
[1]Laboratoire d'Informatique et d'Ingénierie pour l'Innovation (LI3),
[2]Ecole Superieure Polytechnique (ESP) de Dakar, [3]Laboratoire d'Informatique de Dakar
[1] abel.diatta@univ-zig.sn, [2]mandicou.ba@esp.sn, [3]{bassirou.gueye, ibrahima1.niang}@ucad.edu.sn,
[1]Universite Assane Seck de Ziguinchor, [2,3] Universite Cheikh Anta Diop, Dakar, Senegal

***Abstract*—Peer-to-Peer (P2P) networks usage on mobile communication objects is growing more and more. With introduction of telephony in these networks thanks to SIP protocol (P2PSIP), they have become real-time systems. This makes their management more complex. In fact, in addition to having to minimize delays, it is necessary to take into account the loss of links, disconnections, services interruptions, etc., generated not only by users movements but also by nodes arrivals and departures in the system.**

After having set up an architecture to manage nodes mobility in an hierarchical P2PSIP network, in this paper, we first show, by a theoretical dimensioning, that this architecture obeys a scaling up, then we detail the strategy implemented to manage node failures in said architecture. Results obtained from simulations with Omnet++ confirm the architecture scalability but especially the strong robustness of the solution.

Keywords - Dimensioning, Fault tolerance, Mobility management, P2P mobility

I. Introduction

With the advent of Skype, using telephony in peer-to-peer (P2P) networks has become possible thanks to the Session Initiation Protocol (SIP); which gave rise to P2PSIP networks. Since then, P2PSIP networks have become real-time communication systems. However, in such systems, the shorter the session setup time is, the better the service is reliable. To better optimize their use and thus save time, hierarchical P2PSIP (H-P2PSIP) networks have been implemented [1], [2], [3]. They consist of dividing nodes into super nodes and ordinary nodes. Super nodes are great capacities nodes and take care of most complex tasks, whereas ordinary nodes are weak capacities nodes. These latter most often play clients roles.

However, peer-to-peer networks are highly dynamic because nodes arrivals and departures take place at any moment. This frequently causes link losses, resulting in disruption and degradation of the quality of service. In addition, users mobility very often causes disconnections, thus causing interruptions in communications.

By combining the very dynamic nature of P2P networks and the high mobility of users, mobile P2PSIP networks offer all the guarantees to be unstable [4], [5]. It is therefore important to put in place a solution that not only manages this instability and avoids an interruption of services, but also takes into account the increase in the number of mobile objects. It is in this context that we have put in place a hierarchical P2PSIP architecture for real-time communications management based on IPv6 [6], exploiting the advantages offered by IPv6 protocol, especially the multitude of IP addresses, gain time with route optimization, automated handover management [7], etc. In this paper, we present the proposed architecture sizing and the strategy implemented to manage node failures.

The rest of the paper is organized as follows. In Section II, we give an overview on IPv6 mobility. In Section III, we present the architecture. Dimensioning is illustrated in section IV. Section V is devoted to fault tolerance strategies and simulation results. Finally, in Section VI, we give a conclusion and perspectives.

II. IPv6 mobility

Mobility is the set of techniques put in place, in a network composed by several entities or access technologies, to make it possible to establish or maintain a communication when a mobile moves from one zone to another or from one technology to another [8], [9]. With the increase of telephone operators subscribers, the volume of communications has increased significantly. Cells are often saturated due to a lack of free communication channels. To solve this problem, it was necessary either to multiply the number of channels of a cell, or to allow neighboring cells to borrow frequencies from one another according to necessity [10]. These solutions soon came up against limits. For the first, the multiplication of the number of cells consists in dividing cells, which contributes to increase the number of handover and thus, a more difficult management. As for the second, it supposes the enlargement of the frequency band of cells, which could (by dint of widening) create interference.

Telephony over IP is one of the solutions, because it allows to unclog, to a certain extent, cellular networks by offering users the possibility to call themselves via Internet. Users are spared the many expectations, often encountered in GSM networks, for example due to a lack of free channels. Because of all these advantages and also the high prices often imposed by telephony operators, in a concern of benefiting from a quality of service with lower costs, users are moving more and more towards IP communications. As a result, new challenges

are emerging. Since users are often very mobile, they must be given the opportunity to communicate via internet while granting the possibility of displacement during communication. It is in this context that IPv4 and IPv6 mobility solutions have been put in place. With the proliferation of mobile communication devices and the emergence of connected objects, IPv4-based solutions have shown their limitations due to a lack of available addresses. Moreover, in IPv4, with triangular routing, node location often takes longer because of possible encapsulations when it comes to non-routable addresses. By cons, in IPv6, an infinity of addresses is available and the routing obeys the principle of route optimization. This principle consists in avoiding triangular routing of IPv4 by directly sending messages to the destinations. There is no encapsulation required because all addresses are routable. All of these reasons explain IPv6-based solutions popularity.

There are basically two major sets of protocols in IPv6 solutions [7] namely Host-Based Mobility Management Protocols ($HBMMPs$) and Network-Based Mobility Management Protocols ($NBMMPs$). Elements that make up $HBMMPs$ are Home Agents (HAs), Foreign Agents (FAs), Mobility Anchor Points ($MAPs$) and Mobile Nodes (MNs). Three protocols determine their operation [11]: Mobile IPv6 (MIPv6), Hierarchical MIPv6 and Fast handover in MIPv6. For $NBMMPs$, elements are Local Mobility Anchors ($LMAs$), Mobile Access Gateways ($MAGs$) and MNs. $LMAs$ insure HAs's role for MNs and $MAGs$ are movement detectors. In $NBMMPs$, two protocols are implemented [12]: Proxy MIPv6 (PMIPv6) and Fast PMIPv6.

According to [7], $NBMMPs$ offer better performance than $HBMMPs$ in terms of handover latency. This explains the fact that more and more, IPv6 mobility solutions are moving towards $NBMMPs$.

III. Architecture Presentation

In this section we show, on the one hand, the architecture organization, and on the other hand, its functioning.

A. Organization

The architecture is composed of three hierarchical levels as shown in figure 1: $LMAs$ level, $MAGs$ level and simple mobile nodes ($SMNs$) level.

Each SMN has a MAG and a LMA responsible to which it is attached. Likewise, each MAG has a responsible LMA to which it is attached. $LMAs$ and $MAGs$ are respectively organized into rings. Each LMA has a Handover Management Server ($HoMS$) we put in place to manage nodes mobility in the overlay. The $HoMS$ server maintains a Communicating and Locating Node (CLN) table in which are indexed all nodes in communication or in location.CLN table is as follows:

CLN table						
srcSIP@	destSIP@	srcIPv6@	destIPv6@	Call-ID	srcType	destType

- $srcSIP@$: Sip address of $INVITE$ message source;

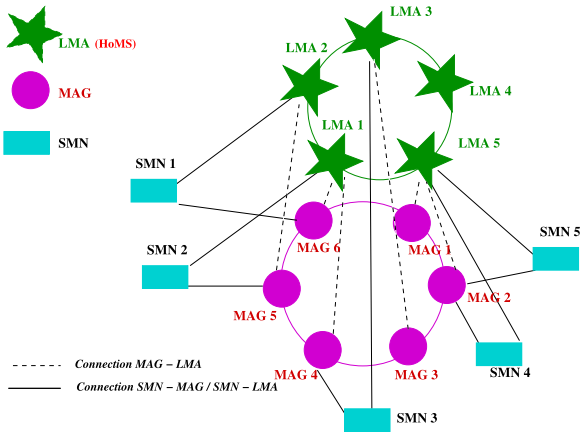


Fig. 1: H-P2PSIP architecture for IPv6-based mobility management

- $destSIP@$: Sip address of $INVITE$ message destination;
- $srcIPv6@$: IPv6 address of $INVITE$ message source;
- $destIPv6@$: IPv6 address of $INVITE$ message destination;
- $Call\text{-}ID$: Call ID generated by SIP protocol. It uniquely identifies the session;
- $srcType$: Type (LMA, MAG or SMN) of source node;
- $destType$: Type of destination node of the call.

$INVITE$ messages are sent to $LMAs$ layer. When an LMA receives an $INVITE$ message, it fills the CLN table with source node information and $Call\text{-}ID$ field. On the other hand, $destIPv6@$, $destID$ and $destType$ fields remain empty until the destination's response. These will allow the $HoMS$ server to detect nodes that are in location phase or in communication. Indeed, when $destIPv6@$, $destID$ and $destType$ are not indicated, it means that source node has launched an $INVITE$ message but has not yet received a response (ie location). When these fields are filled, then source node is in communication.

When an LMA detects a node movement, the $HoMS$ server consults the CLN table to see if moving node is in location or communication phase. Depending on the case, the $HoMS$ server triggers handover management process during location or during communication.

Each node can perform all tasks reserved for lower-level nodes. Upper-level nodes index their successors and lower-level nodes attached to them. Lower-level nodes attached to a higher-level node are called child nodes.

Since the system consists of three layers (LMA, MAG and SMN) and works on an IPv6 physical layer, we define a new addressing system in [6]. Indeed, an IPv6 address is composed of three fields [13]: Home Network Prefix (HNP), Subnet ID and Interface ID.

n bits	m bits	$(128-n-m)$ bits
HNP ID	subnet ID	interface ID

Thus, node IDs in our overlay network are obtained via $sha1$ hash function as follows:

- LMA-ID = $sha1(HNPID)$;
- MAG-ID = $sha1(HNPID + SubnetID)$;
- SMN-ID = $sha1(HNPID + SubnetID + InterfaceID)$.

B. Architecture functioning

An LMA or a MAG has a successor and a predecessor. The successor (respectively predecessor) of an LMA or a MAG is its direct neighbor on its ring in the clockwise direction (respectively counterclockwise).

Considering the logic used in Chord [14] where each node maintains a list of $r = log_2(k)$ successors (with k, the number of nodes in the ring), we define:

- $r_{lma} = log_2(N_{lma})$: size of an LMA successor list;
- $r_{mag} = log_2(N_{mag})$: size of a MAG successor list.

With N_{lma} and N_{mag} the number of $LMAs$ and $MAGs$ in the system.
In $LMAs$ ring, it is possible to have up to 2^n addresses, so at most, the number of $LMAs$ is $N_{lma} = 2^n$. In the same way, the maximum number of $MAGs$ in $MAGs$ ring is $N_{mag} = 2^{(n+m)}$, ie at most ($r_{lma} = n$ and $r_{mag} = n + m$).

When an LMA joins the network, its successor sends to it the list of its successors, to form its own. In the case of a MAG, when it arrives in the system, in addition to the list of successors that its successor MAG sends to it, there is also the one that its LMA sends to it. Similarly, when a SMN integrates the system, its LMA and its MAG send to it their list of successors. So:

- An LMA contains in its index tables, not only IP addresses and identifiers (IDs) of its r_{lma} first successors, but also information of its child nodes.
  Example: Table I shows the index tables of LMA 1.

LMA 1 successors	
ID	IP address
$ID\ LMA\ 2$	$@LMA\ 2$
$ID\ LMA\ 3$	$@LMA\ 3$

LMA 1 child nodes		
Nature	ID	IP address
MAG	$ID\ MAG\ 6$	$@MAG\ 6$
MAG	$ID\ MAG\ 4$	$@MAG\ 4$
SMN	$ID\ SMN\ 2$	$@SMN\ 2$

TABLE I: $LMA1$ index tables

- A MAG contains in its index tables, IP addresses and IDs of its r_{mag} first successors and the r_{lma} first successors of its responsible LMA. It also maintains information of its child $SMNs$. Example: Table II shows the index tables of MAG 4 which responsible LMA is LMA 1.

MAG 4 child nodes	
ID	IP address
$ID\ SMN\ 3$	$@SMN\ 3$

LMA 1 successors	
ID	IP address
$ID\ LMA\ 2$	$@LMA\ 2$
$ID\ LMA\ 3$	$@LMA\ 3$

- A SMN contains in its index tables, IP addresses and IDs of r_{lma} (respectively r_{mag}) first successors of its LMA (respectively its MAG). Example: Table III shows SMN 3 index tables which responsible LMA is LMA 1 and responsible MAG is MAG 4.

MAG 4 successors	
ID	IP address
$ID\ MAG\ 5$	$@MAG\ 5$
$ID\ MAG\ 6$	$@MAG\ 6$
$ID\ MAG\ 1$	$@MAG\ 1$

TABLE II: MAG 4 index tables

LMA 3 successors	
ID	IP address
$ID\ LMA\ 4$	$@LMA\ 4$
$ID\ LMA\ 5$	$@LMA\ 5$

MAG 4 successors	
ID	Adresse IP
$ID\ MAG\ 5$	$@MAG\ 5$
$ID\ MAG\ 6$	$@MAG\ 6$
$ID\ MAG\ 1$	$@MAG\ 1$

TABLE III: SMN 3 index tables

IV. Architectural dimensioning

In this section, we are interested in two types of characteristics: node memory capabilities and node processing capabilities.

A. Dimensioning in terms of memory capacity

To make the sizing, we used some notations which are:

- $\mathcal{N}_X^Y$: number of X-type nodes connected to a Y-type node;
- $\mathcal{N}_{com/loc}$: number of pairs of nodes in communication or location;
- X_c: free memory space of X-type node.

$X, Y \in \{LMA, MAG, SMN\}$

1) For $SMNs$: A simple mobile node holds a table to index the r_{lma} successors of its LMA and another for the r_{mag} successors of its MAG. For each successor of its LMA, SMN maintains information as IP address (128 bits), ID (256 bits), the tunnel ID of the interface via which SMN contacts its LMA successor, in the case its responsible LMA is down (256 bits); which gives a total of 640 bits. To this number, it is necessary to add (1600 + $128 \times s$) bits as it is specified in [15] to store one of its LMA successor; with s, the number of bits of interface ID, ie (s = 128 - n - m) bits, where n and m represent the bit numbers of site ID and subnet ID. So to store all its r_{lma} LMA successors, a SMN needs a memory space of $[(2240 + 128 \times s) \times rlma]$ bits.
In the same way, we show that a SMN needs, for storing the r_{mag} MAG successors, a free memory space of $[(2240 + 128 \times s) \times rmag]$ bits.
So, a SMN finally needs a free memory space of at least:

$$SMN_c = (2240 + 128 \times s) \times (r_{lma} + r_{mag})$$

$$SMN_c = (2240 + 128 \times s) \times (2n + m)$$

By replacing s with its expression, we obtain:

$$SMN_c = [18624 - 128(n + m)] \times (2n + m) \qquad (1)$$

2) For $MAGs$: A MAG has to index its r_{mag} own successors, its r_{lma} LMA successors and the set of $SMNs$ attached to it. In the same way, a MAG needs a free memory space of $(2240 + 128 \times s) \times (r_{lma} + r_{mag})$ bits to index its r_{mag} successors and r_{lma} successors of its LMA.

To index a SMN, in addition to IP address, node ID, tunnel ID of connection interface and (1600 + $128 \times s$); which gives $(2240 + 128 \times s)$ bits, a MAG also needs the sip address (256 bits) of its SMN [6]. So it needs a total free memory space of $(2496 + 128 \times s)$ bits for indexing information from a SMN. To index all $SMNs$ attached to it, the MAG will need a free memory space of $[(2496 + 128 \times s) \times \mathcal{N}_{smn}^{mag}]$ bits.

So, in total, a MAG will need a free memory space of:

$$MAG_c = (2240{+}128s){\times}(r_{lma}{+}r_{mag}){+}(2496{+}128s){\times}\mathcal{N}_{smn}^{mag}$$

$$MAG_c = (2240{+}128s){\times}(2n{+}m){+}(2496{+}128s){\times}\mathcal{N}_{smn}^{mag}$$

However, for $MAGs$, s value is obtained by $s = (128 - n)$ bits. Thus,

$$\begin{aligned} MAG_c &= [2240 + 128 \times (128 - n)] \times (2n + m) + \\ &\quad [2496 + 128 \times (128 - n)] \times \mathcal{N}_{smn}^{mag} \\ &= [2240 + 128 \times (128 - n - m + m)] \times (2n + m) + \\ &\quad [2496 + 128 \times (128 - n)] \times \mathcal{N}_{smn}^{mag} \end{aligned}$$

$$MAG_c = SMN_c{+}128m(2n{+}m){+}(18880{-}128n){\times}\mathcal{N}_{smn}^{mag} \quad (2)$$

3) For $LMAs$: An LMA must index its r_{lma} successors, all $MAGs$ that are attached to it, all $SMNs$ connected to it, and all nodes in communication or in location, ie nodes in CLN table.
By proceeding in the same way, we find that an LMA needs $(2240 + 128 \times s) \times r_{lma}$ bits to index all its successors, of $[(2496 + 128\times) \times \mathcal{N}_{mag}^{lma}]$ bits for all its $MAGs$ and $[(2496 + 128 \times s) \times \mathcal{N}_{smn}^{lma}]$ bits for the set of its $SMNs$. It should be noted, however, that for $LMAs$, $s = 128$ bits.

For nodes in CLN table, an LMA needs source node IP address (128 bits), destination node IP address (128 bits), source node nature (256 bits) and destination node nature (256 bits), session call-id (256 bits) [6]. It is therefore necessary for an LMA free memory space of 1024 bits to index a couple of nodes in communication or one of which is in location phase.

Then, free memory space needed by an LMA to index the set of all its information is:

$$\begin{aligned} LMA_c = (2240 + 128 \times s) \times r_{lma} \ + \\ (2496{+}128{\times}s){\times}(\mathcal{N}_{mag}^{lma}{+}\mathcal{N}_{smn}^{lma}){+}1024{\times}\mathcal{N}_{com/loc} \end{aligned}$$

By replacing s and r_{lma} by their respective expression, we get:

$$LMA_c = 18624n + 18880(\mathcal{N}_{mag}^{lma} + \mathcal{N}_{smn}^{lma}) + 1024\mathcal{N}_{com/loc} \quad (3)$$

4) In summury: Equation (1) makes it possible to determine the minimum values of SMNs. Without detracting generality, we assume that $n = 48$ bits and $m = 16$ bits. Thus, equation (1) shows that a SMN must have a free memory space of at least $142,625$ Ko.

Equation (2) shows that: on the one hand, a MAG must necessarily have a larger free memory space than SMNs, on the other hand, the number of SMNs that can be indexed by a MAG depends on the available MAG memory space. Thus, with equation (2), a MAG which has an available memory space of $1Go$, can index up to 674351 SMNs. Similarly, equation (3) shows that an LMA with an available memory space of $2Go$, can index up to 454928 child nodes $(\mathcal{N}_{mag}^{lma} + \mathcal{N}_{smn}^{lma})$ not being in communication (that is, $\mathcal{N}_{com/loc} = 0$).

The larger the available memory space is, the more a MAG or an LMA can index child nodes.

B. Dimensioning in terms of processing capabilities

In the network, nodes are classified according to their performances. Depending on its capabilities, a node has limits in terms of information to be processed per time unit. According to [15], a $3GHz$ processor can process 2048 bits in $100us$ (microseconds). In our architecture, when a MAG searches, if a given SMN is connected to it, verification is based on SMN ID (256 bits) or on IP address (128 bits); so it is at most 256 bits. Thus a MAG whose processor speed is $3GHz$ can check 80000 SMN in one second on the table of its connected SMNs. For LMAs, location search (to check if a MAG or SMN is connected to it) is also based on node ID or IP address. In the same way, an LMA of $3Ghz$ processor speed will check for 80000 $SMNs$ or $MAGs$ per second. Moreover, during a MAG or a SMN handover, an LMA, via its $HoMS$, checks in the CLN table, if moving node is in communication or location, to trigger the handover management process . Verification is based on source node and destination IP addresses (256 bits). Entire CLN table must be checked because a node can service multiple communications at a time. The higher the number of nodes in communication and/or location is, the longer the LMA takes time to browse the CLN table. This shows that an LMA, to effectively fulfil its role, must have a sufficiently large speed processor.

With multi-core and parallel processing, it is very likely that the number of child nodes processed per time unit $(80000/s)$ will be revised upwards, which increases the system performances and shows the transition to scale of architecture.

V. Fault tolerant strategy and simulation results

A node is considered to be down, if it does not respond, during a given duration, to the messages sent to it. This can be caused either by a physical failure either by the node departure. Depending on broken node nature, disturbances at the network level are different.
At a regular time interval of t seconds, each LMA and each MAG sends a $HELLO$ message to its successor and to lower level nodes connected to it. Each time the $HELLO$

message is sent, LMA or MAG records the moment of the last transmission. Similarly, each time it receives a $HELLO$ response from its successor or one of its lower level nodes, it picks up the instant of last response arrival. Each LMA or MAG that receives a $HELLO$ message from its predecessor, marks the moment it receives the last $HELLO$ from its predecessor. It is the same for $MAGs$ and $SMNs$ which mark the moment of the last $HELLO$ reception from their LMA and/or from their MAG.

A. Fault detection by an LMA

- When an LMA, having sent a $HELLO$ message in advance, remains until next dispatch without receiving a response from a node, it considers the latter to be out. If failed node is a MAG or a SMN, LMA removes it from its child nodes list. If, on the other hand, it is its successor, LMA must look for a new successor. To do this, it sends a $LOOKINGFORNEWSUCCESSOR$ message to all its other r_{lma} successors. This message contains the ID, the IP address and the nature of node that looks for a new successor. Difference between $LOOKINGFORNEWSUCCESSOR$ message and $REGISTER$ message is that any node that sends a $REGISTER$ message to another node of the same nature, seeks to integrate the network. While a node that sends a $LOOKINGFORNEWSUCCESSOR$ message is already in the system. An LMA receiving a $REGISTER$ message from another LMA, insert the latter between it and its successor. Whereas, if it receives a $LOOKINGFORNEWSUCCESSOR$ message from another LMA, it knows that the latter is already in the system, that it already has a predecessor and that it seeks only a successor. Thus, among $LMAs$ that receive the $LOOKINGFORNEWSUCCESSOR$ message, only those with no predecessor will respond because they are the only ones able to satisfy LMA's request for a new successor;
- When it obtains its new successor, LMA updates its successors list;
- It sends a message to all $MAGs$ and $SMNs$ connected to it for an update ;
- The new successor of this LMA sends an $UPDATESUCCESSORLIST$ message with a time-to-live ($TTL = r_{lma}$) to allow predecessors to update their successors list. TTl is decremented after each jump. When it is zero, message is deleted.

B. Fault detection by a MAG

We recall that each MAG maintains a list of r_{lma} successors of its LMA.

- When a MAG, having previously sent a $HELLO$ message, remains until next sending without receiving a response from a node, it considers the latter to be out. If failed node is a SMN, that MAG removes it from its child nodes list. If, on the other hand, it is its successor, this MAG launches the search for a new successor by sending a $LOOKINGFORNEWSUCCESSOR$ message to all its other successors in its r_{mag} successors list;
- When a MAG gets its new successor, it updates its successors list;
- It sends a message to all $SMNs$ connected to it for an update;
- The new successor of this MAG sends an $UPDATESUCCESSORLIST$ message with a $TTL = r_{mag}$ to allow predecessors to update their successors list;
- When a MAG detects that its LMA is out, it sends a $REGISTER$ message to all its successor $LMAs$ for looking for new LMA. An LMA that receives the MAG $REGISTER$ message, sends a response to that MAG. The MAG sends a confirmation to first responding LMA and the two nodes connect to each other.

C. Fault detection by a SMN

SMN does not initiate a $HELLO$ message, it only responds to its LMA and MAG that send it to it.

- When it remains for more than t seconds, without receiving a $HELLO$ message from its LMA, it considers it out of order. In this case, it sends a $REGISTER$ message to all its LMA successors (since it has the list) to search for new LMA. An LMA that receives a $REGISTER$ message from that SMN, sends a response to it which in turn sends a confirmation to the first responding LMA. The two nodes connect to each other and SMN now has a new LMA;
- Likewise, when a SMN detects its MAG failure, it sends a $REGISTER$ message to all r_{mag} successors of its MAG in order to look for a new MAG. Each MAG receiving the $REGISTER$ message from that SMN, sends a response to this SMN. The latter sends a confirmation to the first MAG to respond and the two nodes connect each other. SMN thus has a new MAG.

D. Performance of fault tolerance strategy

In this section, we show the solution performance in terms of generated messages.

1) Analytical performance: When a node is down, consequences are felt at its level and lower layers, because higher-level nodes eventually have lower-level nodes connected to them. Several messages are exchanged to rebuild the network.

Let $\mathcal{N}^{mess}_{X/Faulty}$ be the number of messages exchanged for network reconstitution during an X-type node failure.
$X \in \{LMA, MAG, SMN\}$.

a) LMA failure: An LMA failure is detected, on the one hand by its predecessor (when it does not receive an respond to the $HELLO$ message), on the other hand, by all $MAGs$ and $SMNs$ connected to it (when they stay for a time greater than t seconds without receiving a $HELLO$ message from their LMA). In this case, an LMA which finds that its successor is down, sends to all remaining ($r_{lma}-1$) successors

a $LOOKINGFORNEWSUCCESSOR$ message. Only those without predecessors will respond. At most, there are $(r_{lma}-1)$; so $2\times(r_{lma}-1)$ messages.

An LMA that receives a response from its successors, sends a confirmation (1 message) to the first responding LMA. It sends messages to its ($\mathcal{N}_{smn}^{lma}$) simple mobile nodes and to its ($\mathcal{N}_{mag}^{lma}$) $MAGs$ so that they update their LMA successors list [($\mathcal{N}_{smn}^{lma}+\mathcal{N}_{mag}^{lma}$) messages]. Its new successor sends an $UPDATESUCCESSORLIST$ message with a $TTL=r_{lma}$ to predecessors (r_{lma} messages). In addition, each of the ($\mathcal{N}_{smn}^{lma}$) $SMNs$ and ($\mathcal{N}_{mag}^{lma}$) $MAGs$ sends a $REGISTER$ message to the r_{lma} successors of their LMA that will respond [$2\times(\mathcal{N}_{smn}^{lma}+\mathcal{N}_{mag}^{lma})\times r_{lma}$ messages]. Each of ($\mathcal{N}_{smn}^{lma}$) and ($\mathcal{N}_{mag}^{lma}$) sends a confirmation message to the first responding LMA [($\mathcal{N}_{smn}^{lma}+\mathcal{N}_{mag}^{lma}$) messages]. So, the number of messages generated for network rebuilding, when an LMA is down is

$$\begin{aligned}\mathcal{N}_{LMA/Faulty}^{mess} &= 2(r_{lma}-1)+2(\mathcal{N}_{smn}^{lma}+\mathcal{N}_{mag}^{lma})+\\ &\quad r_{lma}+2(\mathcal{N}_{smn}^{lma}+\mathcal{N}_{mag}^{lma})\times r_{lma}+1\\ &= 2(r_{lma}+1)(\mathcal{N}_{smn}^{lma}+\mathcal{N}_{mag}^{lma})+3r_{lma}-1\\ &= 2(\mathcal{N}_{smn}^{lma}+\mathcal{N}_{mag}^{lma})\times[log_2(\mathcal{N}_{lma})+1]+\\ &\quad 3log_2(\mathcal{N}_{lma})-1\end{aligned}$$

$$\mathcal{N}_{LMA/Faulty}^{mess}\approx \mathbf{O}\big((\mathcal{N}_{smn}^{lma}+\mathcal{N}_{mag}^{lma})\times log_2(\mathcal{N}_{lma})\big)$$

b) MAG failure: A MAG failure is detected on the one hand, by its predecessor and its LMA when they do not receive a response to the $HELLO$ message, and on the other hand, by all $SMNs$ connected to it, when they do not receive a $HELLO$ message from their MAG. When an LMA detects a failure of one of its $MAGs$, it simply deletes it from its connected nodes list. Any message is sent in this case. On the other hand, if a MAG notices its successor failure, it sends to all remaining $(r_{mag}-1)$ successors a $LOOKINGFORNEWSUCCESSOR$ message. Those with no predecessors (at most $(r_{mag}-1)$) can respond $[2\times(r_{mag}-1)]$ messages.

A MAG that receives a response from its successors sends a confirmation (1 message) to the first responding MAG. It sends messages to its ($\mathcal{N}_{smn}$) $SMNs$ so that they update their MAG successors list ($\mathcal{N}_{smn}$ messages). Its new successor sends an $UPDATESUCCCESSORLIST$ message with a $TTL=r_{mag}$ to predecessors (r_{mag} messages). In addition, each of ($\mathcal{N}_{smn}$) $SMNs$ sends a $REGISTER$ message to the r_{mag} successors of their MAG. These will respond ($2\times\mathcal{N}_{smn}\times r_{mag}$ messages). Each of the ($\mathcal{N}_{smn}$) sends a confirmation message to the first responding MAG (N_{smn} messages). So, the number of messages generated for network reconstruction during a MAG breakdown is

$$\begin{aligned}\mathcal{N}_{MAG/Faulty}^{mess} &= 2\times(r_{mag}-1)+\mathcal{N}_{smn}^{mag}+r_{mag}+\\ &\quad 2\times\mathcal{N}_{smn}^{mag}\times r_{mag}+\mathcal{N}_{smn}^{mag}+1\\ &= 2\times\mathcal{N}_{smn}^{mag}\times(r_{mag}+1)+3r_{mag}-1\\ &= 2\mathcal{N}_{smn}^{mag}\times[log_2(\mathcal{N}_{mag})+1]+3log_2(\mathcal{N}_{mag})-1\end{aligned}$$

$$\mathcal{N}_{MAG/Faulty}^{mess}\approx \mathbf{O}\big(\mathcal{N}_{smn}^{mag}\times log_2(\mathcal{N}_{mag})\big)$$

c) SMN failure: When a simple mobile node falls down, there is no message exchange. Its LMA and its MAG, noting that their SMN no longer responds to $HELLO$ messages, remove it from their child nodes list.

Finally, Table IV summarizes the performance study (in terms of generated messages) of fault tolerance strategy.

	Complexity
$LMAfailure$	$\mathbf{O}\big((\mathcal{N}_{smn}^{lma}+\mathcal{N}_{mag}^{lma})\times log_2(\mathcal{N}_{lma})\big)$
$MAGfailure$	$\mathbf{O}\big(\mathcal{N}_{smn}^{mag}\times log_2(\mathcal{N}_{mag})\big)$
$SMNfailure$	0

TABLE IV: Summary of Analytical Performance Study

2) Simulation performance: To measure performances of fault tolerance strategy, we conducted our simulations with OMNeT++[16] which is an object-oriented discrete event simulator. Due to its many use cases, it has become one of the most used simulators. It can be used for modeling wired and wireless communication networks, modeling protocols, queuing networks, multiprocessors and other distributed hardware systems, validating hardware architectures, and so on.
We used R software [17] and igraph [18] library to generate our physical network topologies on which our hierarchical peer-to-peer SIP architecture is based.

Simulations are done in three different scenarios. In the first case, in order to see the solution behavior in terms of robustness, we set the number of nodes in the network and vary the rate of failed nodes from 15% to 90% in steps of 15%. In the second case, to study the solution scaling, we set the rate of failed nodes and vary the number of nodes from 100 to 1000 in steps of 100 nodes. Finally, in the third case, the number of nodes in the network and the rate of faults evolve at the same time. This allows us to study the behavior of our architecture in a context of high disconnections combined with an increase in network size.

In a context, on the one hand of high disconnections as it is in the case of mobile peer-to-peer networks, and on the other hand of real-time communications, we must avoid that a

node stays too long time before it detects its neighbor failure. Thus, $HELLO$ message sending interval should not be wide enough. However, we must also avoid the interval being too small, so as not to overload the network, because this would contribute to delay packets and thus degrade the quality of communications. The interval must be average. So we fixed it at $5ms$.

Different results are illustrated in figures 2, 3 and 4

a) First case: the number of nodes is fixed: for this case, we study the impact of the failed node rate on the costs in terms of messages exchanged. We set the number of nodes in the network to 1000 and vary the rate of failed nodes. Figure 2 shows the results obtained.

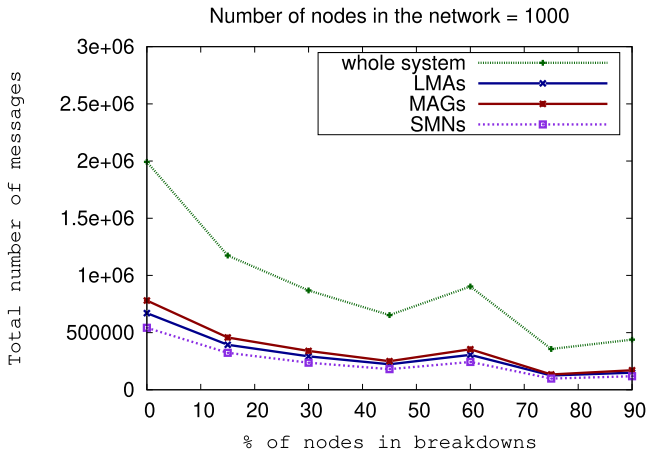


Fig. 2: Maintenance costs with a fixed number of nodes

Figure 2 shows that, the more the nodes are failed, the lower the costs are. Indeed, when a node detects its neighbor failure, it stops sending it the $HELLO$ message. Thus, the more the number of failed nodes increases, nodes to which $HELLO$ message must be sent become fewer. Messages that a node sent at the beginning diminish because possible destinations no longer exist.

In addition, we find that $MAGs$ costs are the highest. Indeed, when a high-level node detects a low-level node failure, there is no message exchange. Low level node is just removed from the list. Thus, costs generated by $LMAs$ exist only in case of other $LMAs$ failure. On the other hand, for $MAGs$, the costs are generated, not only in case of other $MAGs$ failure, but also in case of $LMAs$ failure. In the latter case, all $MAGs$, whose LMA is down, create messages to try to connect to another LMA. Similarly, when $MAGs$ fail, in addition to messages that other $MAGs$ send to $SMNs$ for updates, there are many messages to rebuild the ring. These are often very numerous because predecessors have to make updates.

For $SMNs$, a high level node failure affects them. However, costs generated by $SMNs$ are only messages they send to different successors of their MAG or their LMA.

b) Second case: the rate of failed nodes is fixed: we want to study the architecture scalability but also check its behavior in a breakdowns situation. We are interested in the number of messages generated, when the network grows. So, we change the number of nodes in the system from 100 to 1000 while fixing the rate of nodes in breakdown. Results obtained are shown in figure 3.

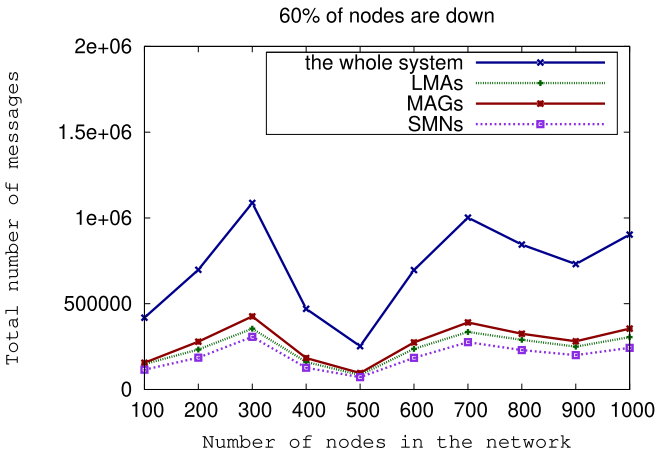


Fig. 3: Maintenance costs with a fixed failure rate

In Figure 3, the same explanations for $LMAs$, $MAGs$ and $SMNs$ are valid. In addition, the figure shows that costs do not have a linear dependence on the number of nodes in the network. For 300 nodes for example, costs are higher than for 900 nodes. Costs depend on failed nodes nature, ie the rate of $LMAs$, $MAGs$ and $SMNs$ down. This proves the architecture scalability, but also and especially the system robustness, because even with a high rate of failures, the system works.

c) Third case: the number of nodes and the failure rate evolve: In this part, we study the system robustness especially when the number of nodes increases. We vary both the number of nodes and the rate of failed nodes and measure the number of messages generated. This allows to see the impact of node failures and the increase in network size, on the number of messages (figure 4).

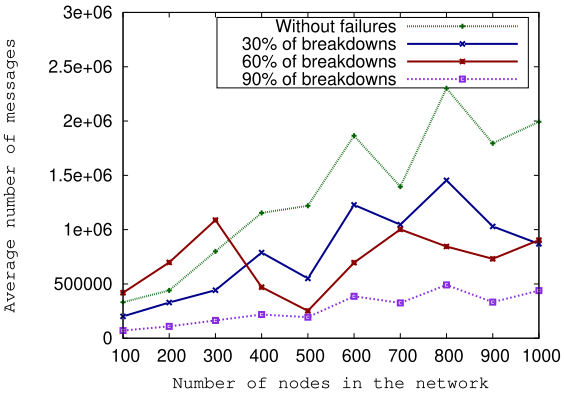


Fig. 4: Maintenance costs with a failure rate and a number of nodes that evolve

With figure 4, we always observe that there is no direct correlation between the number of nodes in the network, the rate of failed nodes and maintenance costs. Same explanations always hold, namely that costs depend on the nature of inoperative nodes and especially their percentage (rate of $LMAs$, $MAGs$ and $SMNs$ down). In addition, we find that costs with no failures remain more important, especially when the network size increases. This is because in this case, the more the number of nodes grows, nodes to which it is necessary to send $HELLO$ messages increase.

Moreover, the results confirm the robustness because even with 90% of failed nodes, the system works.

VI. Conclusion

In this paper, after explaining the roles of different elements of our architecture, we have shown, by theoretical sizing, that higher-level nodes can support a large number of lower-level nodes. We then detailed the procedure implemented to manage node failures in said architecture. Subsequently we showed, thanks to simulations results, the good behavior of the solution in terms of robustness and scalability.

As future work, we are interested in setting up a load balancing strategy but also in experimentally studying the system sizing.

References

[1] W. Liu, J. Song, and J. Yu, "An overlapping structured {P2P} for {REIK} overlay network," *Physics Procedia*, vol. 33, no. 0, pp. 1022 – 1028, 2012, 2012 International Conference on Medical Physics and Biomedical Engineering (ICMPBE2012). [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1875389212014812

[2] O. Bravo, A. Costa, and M. Nicolau, "Design and implementation of a hierarchical sip-based peer-to-peer network," in *Software, Telecommunications and Computer Networks (SoftCOM), 2012 20th International Conference on*, Sept 2012, pp. 1–9.

[3] I. Martinez-Yelmo, C. Guerrero, R. Cuevas, and A. Mauthe, "A hierarchical p2psip architecture to support skype-like services," in *Proceedings of the 17th Euromicro International Conference on Parallel, Distributed and Network-based Processing*, 2009, pp. 316–322.

[4] C.-C. Lai and C.-M. Liu, "A mobility-aware approach for distributed data update on unstructured mobile p2p networks," *Journal of Parallel and Distributed Computing*, vol. 123, pp. 168–179, 2019. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0743731518306956

[5] S. Singh, C. KWirumar, and P. Nath, "Local p2p group (lpg) communication in structured mobile p2p networks," in *J Ambient Intell Human Comput*, no. 11, 2020, p. 3005–3019.

[6] A. Diatta, I. Niang, B. Gueye, and M. Ba, "Pmipv6-based mobility for realtime communications on hierarchical p2psip systems," in *2017 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, June 2017, pp. 962–969.

[7] R. A. Khan and A. H. Mir, "Performance analysis of host based and network based ip mobility management schemes over ipv6 network," in *Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on*, Sept 2014, pp. 1798–1803.

[8] S. Singh, C. Kumar, and P. Nath, "Analysis and modelling the effects of mobility, churn rate, node's life span, intermittent bandwidth and stabilization cost of finger table in structured mobile p2p networks," in *Wireless Netw*, no. 27, 2021, pp. 1049–1062.

[9] T. Lucidarme, "Principes de radiocommunication de troisième génération," *Vuibert, Paris*, 2002.

[10] G. Rodriguez, "Introduction aux réseaux cellulaires - techniques d'accès et de partage de la ressource radio," *Systèmes de Télécommunication Cycle d'harmonisation 2A AST TEL-COM202*, Septembre 2011.

[11] W. S. Hoh, S. Muthut, B.-L. Ong, M. Elshaikh, M. N. M. Warip, and R. B. Ahmad, "A survey of mobility management protocols," in *ARPN Journal of Engineering and Applied Sciences*, vol. 10, no. 19, OCTOBER 2015, pp. 9015 – 9019.

[12] B. H. Song, J. Shin, S. Kim, and J. Jeong, "On pmipv6-based mobility support for hierarchical p2p-sip architecture in intelligent transportation system," in *System Sciences (HICSS), 2015 48th Hawaii International Conference on*, Jan 2015, pp. 5446–5452.

[13] R. Hinden and S. Deering, "Architecture d'adressage du protocole internet version 6 (ipv6)," Avril 2003- Request for Comments 3513.

[14] I. Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, "Chord: A scalable peer-to-peer lookup service for internet applications," *ACM SIGCOMM Computer Communication Review,*, vol. 31, no. 4, pp. 149–160, 2001.

[15] H. Luo, H. Zhang, Y. Qin, and V. C. M. Leung, "An approach for building scalable proxy mobile ipv6 domains," *IEEE Transactions on Network and Service Management*, vol. 8, no. 3, pp. 176–189, September 2011.

[16] "Omnet++ - Objective Modular Network Test-bet in C++," https ://omnetpp.org/doc/omnetpp/SimulationManual.pdf .

[17] https ://www.r-project.org/.

[18] http ://igraph.org/r/.

Smart and Active Package Integrating Printed Electronics

Joana Pimenta, Lorena Coelho, José Silva, Vanessa Miranda, Jorge Laranjeira, Rui Soares

***Abstract*—** In this paper, the results of R&D on an innovative food package for increased shelf-life are presented.

SAP4MA aims at the development of a printed active device that enables smart packaging solutions for food preservation, targeting the extension of the shelf-life of the packed food through the controlled release of active natural antioxidant agents at the onset of the food degradation process. To do so, SAP4MA focuses on the development of active devices such as printed heaters and batteries/supercapacitors in a label format, to be integrated on packaging lids during its injection molding process, promoting the passive release of natural antioxidants after the product is packed, during transportation and in the shelves, and actively when the end-user activates the package, just prior to consuming the product at home. When the active device present on the lid is activated, the release of the natural antioxidants embedded in the inner layer of the packaging lid in direct contact with the headspace atmosphere of the food package starts. This approach is based on the use of active functional coatings composed of nano encapsulated active agents (natural antioxidants species) in the prevention of the oxidation of lipid compounds in food by agents such as oxygen, thus keeping the product quality during the shelf-life, not only when the user opens the packaging and use the product, but also during the period from the food production until it reaches the consumer.

For the development of the smart label, it is important to create a printed heater and battery that can operate in the working conditions associated with this application. The ideal electrical resistance of the heating system was determined to give rise to an optimized design that promotes an even distribution of heat throughout the lid of the food tub. This factor is extremely important because the release of antioxidants needs a homogeneity temperature on the lid. To decrease battery consumption, it was stipulated that the heating system should operate at the lowest possible electrical current.

The printed heating circuit was studied using three different substrates and two different conductive inks. Inks were selected, taking into consideration that the printed circuits will be subjected to high pressures and temperatures during the injection molding process.

As for the battery design, we opted for two different configurations, one interdigitated form and the other in which there are overlapping layers.

This research has been carried out within the Project “Smart and Active Packing for Margarine Product” (SAP4MA) running under the EURIPIDES Program being co-financed by COMPETE 2020 – the Operational Programme for Competitiveness and Internationalization and under Portugal 2020 through the European Regional Development Fund (ERDF).

***Keywords*—** smart package, printed heat circuits, printed batteries, flexible and printed electronic.

J. Pimenta, L. Coelho, J. Silva and V. Miranda are with the CeNTI – Centre for nanotechnology and smart materials, Vila Nova de Famalicão, 4760-034, Portugal (phone: +351 252 104 152; fax: +351 252 327 358; e-mail: jpimenta@centi.pt; lcoelho@centi.pt; jsilva@centi.pt; vmiranda@centi.pt).

J. Larangeira is with the Moldit - Indústria De Moldes, S.A., Oliveira de Azemeis, 3720-903, Portugal (phone: +351 256 690 050; fax: +351 256 690 054; e-mail: jorge.laranjeira@molditindustries.com).

R. Soares is with the CENTIMFE - Centro Tecnológico da Indústria de Moldes, Ferramentas Especiais e Plásticos, Marinha Grande, 2430-028, Portugal (phone: +351 244 545 600; e-mail: rui.soares@centimfe.com).

Development Results of an Energy Supply Platform Based on Various Energy Harvesting Technologies

Matthias Fahland (1), Markus Tuomikoski (2), Bernd Fuchsbichler (3), Jan Hesse (1), Damian Hau (4), Katrina Morgan (5)

***Abstract*—** The widespread introduction of wearable devices is expected to be one of the major trends in the next one or two decades. First applications have already entered the market, like e.g. the smartwatch from Apple or various types of fitness trackers. Wearables will generate completely new opportunities for sensing (e.g. vital parameter monitoring), mobile data storage, wireless communication and internet of things. Apart from legal topics, like e.g. data security, there are also technological bottlenecks. Today design and appearance are significantly limited by given geometries of state-of-the-art components. Therefore, making devices thinner, safer, more flexible and easy to integrate are major goals of current research activities. The energy supply to wearable devices is probably the most serious challenge among these technological bottlenecks. The European funded Smart2Go project tackles exactly this topic. The aim of the Smart2Go project is the creation of an autonomous energy-supply platform. This will be achieved by the combination of a powerful, thin and scalable battery with appropriate energy harvesting technologies. The paper presents the achieved solution. Different energy harvesting approaches exploit ambient energy sources. These can be either radiation energy or thermal energy or mechanical energy, depending on the specific application scenario. The energy is stored in a Li-based thin-film battery. It has a capacity of up to 110 mWh (10 cm² area). The energy flow is managed by an electronic platform as an integral part of the solution. All components are thin and flexible and retain their performance after the bending test. A roll-to-roll manufacturing process for the integration of all the components will ensure suitable production capacity with low manufacturing costs.

***Keywords*—**Energy harvesting, energy storage, organic photovoltaics, piezoelectric sensors, thermoelectric generator, triboelectric generators.

(1) Author is with the Fraunhofer Institute for Organic Electronics, Electron Beam and Plasma Technology, Dresden, Germany, (corresponding author, phone: +49 351 2586 135; e-mail: matthias.fahland@ fep.fraunhofer.de).
(2) Author is with VTT (Finland),
(3) Author is with VARTA Microinnovation (Autria)
(4) Author is with ARMOR ABL (France)
(5) Author is with University of Southampton (UK)

Evaluation of Packaging Elements From a Consumer's Perspective; Empirical Findings from a Study for the Packaged Olive Oil

Konstantoglou A., Fotiadis T. Rotsios, K., Folinas, D.

***Abstract*—** This paper aims to demonstrate the multidisciplinary nature of food packaging. The authors agree with the research results of many studies, which support the need to consider many factors in the design and development of food product packaging, such as marketing, logistics and supply chain management, food technologies, and environmental-friendly practices. Therefore, factors such as product promotion, safety, and ease of handling environmental impact throughout the food life cycle, etc., are examined. The research tries to identify the importance that consumers perceive against the various elements of the packaging of a specific food product, the olive oil. Olive oil has been an integral part of life in Greece from the first stirrings of its civilization. Furthermore, olive oil has always been a significant factor in the Greek economy, and almost every report, study, and piece of research on the country's agricultural sector make a separate reference to it.

Based on the above, the main objective of this study is the identification and categorization of the packaging elements of food products. Primary data were collected through a research initiative via a questionnaire, which was filled by consumers of packaged olive oil in two (2) cities in Northern Greece. A pool of packaging elements was developed, aiming to examine the marketing, logistics and supply chain management, food technologies and environmental friendly (green) practices that have a positive relationship to consumer behavior in buying food products. The main research question of this study is as follows: "What do consumers consider being the key elements of olive oil packaging?" The research findings show that consumers recognize the important role that packaging plays in food safety and quality and in relation to the information that it provides. It was proved that consumers understand and appreciate the multifunctional and multidisciplinary nature of packaging, considering that the informational, operational, physical, and visual elements are of high importance. By identifying the importance that all the above perceive against the various elements of the packaging of retail products, manufacturers can take into consideration the elements that are highly appreciated by all disciplines/views.

***Keywords*—** consumer behavior, food industry, food supply chain, olive oil, packaging, packaging elements.

Dimitris Folinas is with the International Hellenic University, Greece (e-mail: dfolinas@gmail.com).

Data-Driven Decision Making: A Reference Model for Organizational, Educational and Competency-Based Learning Systems

Emanuel Koseos

***Abstract*—**Data-Driven Decision Making (DDDM) refers to making decisions that are based on historical data in order to inform practice, develop strategies and implement policies that benefit organizational settings. In educational technology, DDDM helps schools measure progress towards standards and policies that assess the needs, progress and performance of students and educators. Current research has focused on prediction of student achievement by consideration of Educational Data Mining (EDM) or Competency-based Education (CBE) approaches at a university level. However, we propose a more generic reference model that can be used by these approaches and applies to different organizational levels in education or business institutions. To understand how educators, conceptualize DDDM, data from 45 schools in the Langley, BC, Canada regional school district was mined to elaborate dimensions on how DDDM varies in practice. This study established a viable model that is used to make decisions at different levels of organization. We established key process initiatives (KPIs) used by a machine learning framework implemented in statistical software (SAS, R) to provision and establish best practices for evaluating educational policies and standards by educators. One underlying assumption was that the model will enhance decisions on the allocation of resources and improve teaching and learning. We outline factors that enabled or inhibited various types of DDDM and concluded the model is well suited for applications to educational and business environments.

***Keywords*—**Competency-based learning, data-driven decision making, machine learning, optimization.

E.K. Author is with the Chadron State College, Chadron, NE 69337 USA (phone: 001-778-689-7570; e-mail: emanuel.koseos@eagles.csc.edu).

Technological Applications in Automobile Manufacturing Sector - A Case Study Analysis

Raja Kannusamy

***Abstract*—**The research focuses on the applicable technologies in the automobile industry and their effects on the productivity and annual revenue of the industry. A study has been conducted on 6 major automobile manufacturing industries represented in this research as M1, M2, M3, M4, M5 and M6. The results indicate that M1, which is a pioneer in technological applications, remains the market leader, followed by M5 & M2 taking the second and third positions, respectively. M3, M6 and M4 are the followers and are placed next in positions. It has also been observed that M1 and M2 have entered into an agreement to share the basic structural technologies and they maintain long-term and trusted relationships with their suppliers through the Keiretsu system. With technological giants such as Apple, Microsoft, Uber and Google entering the automobile industry in recent years, an upward trend is expected in the futuristic market with self-driving cars to dominate the automobile sector. To keep up with the market trend, it is essential for automobile manufacturers to understand the importance of developing technological capabilities and skills to be competitive in the marketplace.

***Keywords*—**Automobile manufacturing industries, Competitiveness, performance improvement, technological applications.

I. Introduction

To remain competitive in the global business environment, manufacturing organisations need to consistently improve the quality, reliability, on-time delivery and customer service. Further they need to rapidly develop new products and be flexible in manufacturing their products. The advances in technologies have enabled manufacturing organisations to develop the competitive capabilities required for competing in the market. The customer requirements are continuously changing and so the automobile manufacturers are not only required to achieve the economies of scale but also need to keep up with the changing customer demand. This requires a continuous improvement and flexible approach which could be achieved using the latest technologies in the manufacturing process. This research analyses the technologies used by the major automobile organisations and the effects on the organisation's production and annual revenue.

This research critically evaluates the manufacturing technologies used in the automotive industry and its effects on the performance of the organisation. The technologies used by automobile manufacturing sector have been compared, analysed, and evaluated. This study provides an insight on to the importance of applying technologies in automobile manufacturing sector to remain competitive in the market.

Raja Kannusamy is with the College of Leadership and Business, Brisbane, QLD 4006 AUSTRALIA (corresponding author, phone: 07-31122096; fax: 07-32577195; e-mail: raja@ colab.edu.au).

II. Literature Review

The automobile manufacturing sector remains competitive by continuously implementing new technologies in its production and assembly lines. Technology have not only reduced errors in the production but have greatly increased the efficiency and productivity. The researchers [1] have studied the application of Artificial Intelligence (AI) in manufacturing. The researchers state that the integrated system provides features to digitalization and provide flexible service to the customer. In [2] the researcher has presented that the application of artificial intelligence in manufacturing industry improves the output ratio, quality, and innovation. The researcher also presents that the automation software used in manufacturing organisations very well detects errors in the production line. Reference [3] reports that the automotive processes are used in manufacturing industry for producing better quality with reduced unit cost. The researchers [4] report that the manufacturing process could be monitored using sensors linked to Internet of things (IOT) and big data. The researchers have also presented the details of the Raspberry pi sensor, which could monitor and control the process and could determine the real-time speed using computational models. The researchers further state that an artificial intelligence agent could work with the automobile driver to ensure safety and comfort through integrated voice communication, enabling functions to increase alertness or lower tension, improve lighting, air conditioning, fragrances, and other human-machine interaction functions. Reference [5] presents that in 2017 more than $100 million in Artificial Intelligence have been invested by Toyota Research Institute's subsidiary, Toyota Artificial Intelligence Ventures, and further state that it would spend an additional $100 million to focus on artificial intelligence, automation, and robotics in future. The study reveals that implementing advanced technologies results in fully automated production lines, quick decision-making capabilities, improved production and quality with reduced cost, higher scalability and provides novel opportunities. Reference [6] have highlighted that the next revolution in automation is close and so the manufacturing sector need to be prepared for total integrative systems in which the supply chain, the expert team, the production line, and the quality assurance are all well-organised and coordinated through smart engine that could provide remarkable insights through expertise. The goal of the latest artificial intelligence technology in business is to set a milestone for automation in the mass market [7]. Reference [8] have presented that in Ford, AI is used in

applications such as, process planning of vehicles, ergonomic analysis, customizing and translating the machine language and knowledge management. The researchers [9] have supported this by arguing that the challenges faced in automobile sector could be very well addressed using AI.

III. CASE STUDIES

The automobile manufacturer M1 in its vehicle production process has implemented "lean manufacturing," and "Just in Time (JIT)" approaches [10]. While the philosophy JIDOKA has helped to automate the processes and JIT helped to meet the demand with instantaneous production, M1's manufacturing technology known as "Toyota Production System (TPS)" has helped to develop competitive advantage. JIDOKA denotes that a machine comes to a safe mode any moment when an irregularity arises. The JIDOKA technology is incorporated into the production process and so machines get simpler and more inexpensive, while maintenance becomes less tedious, allowing the construction of simple, slim, flexible lines that can adjust to varying levels of output volume [10]. Thus, M1's JIDOKA idea is based on the consistent growth of both human capabilities and achievements.

M2 is the fifth-largest automobile manufacturer. The Protegé, Mazda3, Mazda6, Miata, RX-8, MPV, and Tribute are among the organization's U.S. brands [11]. The company runs two primary production plants in Japan, as well as 15 other production facilities located across the world, including two Ford Motor Company manufacturing joint ventures located in Michigan and Thailand. M5 invested in M2 in 1979, and as a result it obtained a controlling 33.4% stake in the organization on the verge of bankruptcy in 1996. As a result, to save the company, which was largely dependent on cutting costs, rebuilding, and launching a line of new models, M5 started a turnaround by engaging in cost-cutting, restructuring, and launching new models. M2's unique performance visibility, optimal driving position and heads-up cockpit concept are the core elements of the human centric design philosophy [11]. M2 made use of a line of MZR engines which came in four displacements and engines were built using specialised machines, each of which performed a specific task [12]. Though practiced for a longer time, this was not an effective solution. Skyactiv technology was then introduced and with the use of Skyactiv mills, each engine could be altered to fit the architecture and displacement. CNC machines have helped to produce different engines such as gasoline and diesel mills flexibly at the same time and in line with the changing customer demand. With the use of these technologies, the time to machine an engine has reduced from six hours to one and a half hours.

M4, which has a market capitalization of $42 billion, is India's largest vehicle enterprise and one of the world's top vehicle producers specializing in automobiles, commercial vehicles, passenger cars, transports, trucks, and military vehicles. M4 a wholly owned part of the $100 billion Group, was founded in India in 1945. The company is focusing on designing future mobility with emphasis on quality, innovation, design, and efficiency by investing heavily on technologies and engineering services. The company spends about 8% to 10% of its research and development expenditure towards emerging technologies. Reference [13] presents that the most known brands such as Jaguar, Land Rover and Tata Daewoo belong to the conglomerate, and all are structured differently. The specific areas of technological applications are in machine utilization module, engine volume prediction, maintenance module, smart energy management and predictive quality management. Manufacturing execution system is used to schedule production, improve product quality, and generate real-time reports. Big data is used to analyse production, productivity loss, quality, and efficiency improvement. The integrated manufacturing environment gathers data on real-time losses in the assembly line with the use of Manufacturing Execution System (MES).

M5 founded in 1903, is ranked 11th on the Fortune 500 list of the nation's top corporations. The organization has offered twelve brands under its control including brands that have passed on such as Edsel, Merkur and Mercury. Reference [14] presents that in addition to these, brands such as Mazda Land Rover, Jaguar and Volvo were given off to other businesses. Lincoln and Ford are part of brands under M5's ownership, the group also introduced world's first game sports car under the name Bronco. Additive manufacturing technology is used for quick prototyping, tooling, surface coating and die making. The researcher [15] have analysed the various types of materials used for additive manufacturing and have reported that utilizing laser is the most efficient method for handling metal particles in the automobile sector.

M6 manufactures autos and light business vehicles [16]. The company was established in 1896 and headquartered in Paris, France. The company manufactures motorbikes, bikes, water powered motors, automobile segments and provides after sales services. The business division includes Vehicle Automotive, Vehicle Parts, Transportation and Logistics and Finance Division. The Automotive division performs pre-production planning, factory manufacturing and fleet sales of light commercial vehicles and automobiles branded with the P and C brands. The Automobile Equipment division performs interior frameworks, auto seats, control advancements and components. This division focusses on synchronizing vehicle and cargo travel with vehicle transport. The Finance Division offers retail and discount financing to P and C customers and vendors [16]. The group developed Efficient Modular Platform [EMP2] and Common Modular Platform [CMP] in 2013. EMP2 was developed to manufacture wide range of body styles of compact and family cars and provides a single, streamlined platform for innovative and efficient solutions and it allowed a reduction of 70 kg in vehicle mass, 22% reduction in fuel consumption and elimination of 99% particulates. Common Modular Platform was developed in partnership with Dongfeng Motors China and the platform is electrified to provide efficient solution in terms of scalability, flexibility, and CO2 emissions [16].

IV. Comparative Analysis

Toyota Production System (TPS) comprises of two segments, JIDOKA and Just-in-Time. The objective of TPS is to eliminate the waste termed as 'Muda'. Wastes are associated with overproduction, pausing, overstock, unnecessary movements, and unplanned shipping [10]. The company has teamed up with Robot Wars X so that the mechanical and synergistic robots are used for painting and welding. M1 advances its lean assembly by including both program guided tuggers and program guided trucks. Components are left in the drop-off-zones which are then fed to robots. On completing the procedure, the robots place the components on to the rack which are then moved to the next workstation by automated guided vehicles [17]. Automated Guided Container Transport System Robots help to automate long distance transport thus freeing up yards and improving packaging operations and eliminating the idleness of cranes and holder trailers. The other benefits for the company are excellent quality, increased productivity and on time delivery.

Mono-tsukuri Innovation, the technique used by M2 comprises of bundled product planning, common architectures, and flexible production. The products produced are transferable across different models and standard vehicle architectures. M2 design the vehicles not only based on model-based body forms but also use a component-based construction and sectional shape configuration to fit specific vehicle. This has helped the company to keep pace with the market demand [11]. M2 through its Mono-tsukuri Innovation had the option to create different product offering to keep up the economies of scale inspite of its low production capacity. Managers from product development, manufacturing, purchasing and supplier departments regularly meet to discuss the production plans for the next five to ten years and work closely to accomplish the goal [18]. Thus, M2 could characterize the designs such as a fixed engineering concept, cross sections, machining method and transfer standard among few of the future models. This helped in design standards that are unable to be produced in low-cost regions and to maintain the required productivity level.

The press and weld shops at M3 are automated. The Press Shop pushes sheets of metal into pre-determined forms. Once welding is completed, the sections are connected in the weld shop. Robots enable chipping and to connect in narrow holes to execute spot-welding. The weld shop uses robots to perform three dimensional and three thousand of a millimeter hole directing, which welds various components of a skeleton. This automation enabled M3 to continually increase the usage. This resulted in better weld quality, increased yield, reduced waste, and less variable cost. Moving from labor-intensive to capital-intensive operations require preventive maintenance procedures to be followed for full-scale production to be achieved [19].

M4 outsources engineering services and develops Information Technology products and services. The company is expert in the design space and applies innovative methods in assembly through unified Connected Vehicle Platform (CVP). This enables the company to reduce the processing time and limit the manufacturing cost, thereby reducing the stock.

M5 utilises Additive manufacturing for three-dimensional printing. Prototyping is very expensive and three-dimensional printing replaces the expensive method by enabling to make different multifaceted model plans without the additional tooling and design needed for making the conventional prototype for individual components. This automated design not only helped M5 in achieving the efficiency, but also assisted in keeping up its sales despite market concentration. Additive manufacturing technology is used for making complex shapes along with regular components thereby consolidating production and assembly together. The mechanical properties of a finished component depends on the quality of the alloy material powder used and it is highly unlikely to vary it at the later stage. Specialty interaction, meaning a variation to the added plan is also possible and requires two to three hours to create a shape [20].

M6 adopted a three-step design approach which are initialization, macrosimulation and validation. M6 uses simulation and analytical methods as tools in this process. Simulation is carried out using software such as Arena. Verification and validation of the product functions are carried out using the design simulation. Simulation requires expert skills but is used to analyse the dynamic characteristics which are limited with analytical methods. Initial stages of automobile assembly at M6 involve moderate labor and machinery, robots are used in the intermittent stages for welding and applications, the consecutive stages involve a mix of labor and automation before moving to the paint shop and final assembly. The different technologies used by the manufacturing organisations and the output received have been presented in Table I.

TABLE I
Techniques Implemented and Outputs Achieved

Manufacturing organisations	Techniques Implemented	Outputs achieved
M1	TPS consisting of 1. Jidoka 2. Just-in-time	Improved quality & safety
M2	1. Mono-tsukuri Innovation 2. Skyactiv mills	Optimised manufacturing process & improved collaboration
M3	Automated Press and Weld shops	Improved driving performance & fuel efficiency
M4	1. Machine utilization module 2. Manufacturing Execution System 3. Predictive Quality	Improved efficiency & reduced inventory

	Management	
M5	Additive Manufacturing	Combining manufacturing & assembly in a single step
M6	1. Efficient Modular Platform 2. Simulation techniques	Improved efficiency & better validation of new products

V. Results and Findings

The annual revenue of the different manufacturers and the number of units manufactured by each of them have been compared for the years from 2015 to 2019 and have been presented in Table II. M represents the different manufacturers, and the Production and Revenue details are compared for the years from 2015 to 2019. From the comparative analysis it has been observed that the manufacturer M1 has used the technology very well to improve its productivity and the annual revenue. The results indicate that M2 has moderately used the technology compared to M1 and that there are lots of opportunities for improving the efficiencies further. M4 have managed to transform its production process with the help of technology, however a slight drop in productivity has been observed. This indicates that M4 could further focus on improving its productivity. New technology implementation has helped M4 to manufacture cars at a reasonable price suiting to market needs. M5 managed to maintain the productivity rate steadily with the implementation of new technologies, however a slight mismatch between the organization and the technology implemented have been observed. M6 has experienced a high increase in productivity with very reasonable increase in cost.

TABLE II
Production & Revenue Analysis

M	Production (Millions) Revenue ($Billion)	2015	2016	2017	2018	2019
M1	Production	10.1	10.2	10.5	9	10.7
	Revenue	246	253	253	266	272
M2	Production	15.4	15.9	16.1	16	16.6
	Revenue	26.5	30.2	28.8	31.2	32.1
M3	Production	3.1	2.9	3.3	3.4	3.5
	Revenue	25.8	27.2	26.6	31.2	34.8
M4	Production	1.01	1.08	0.93	1.04	1.03
	Revenue	37.8	39.9	39.4	42.2	43.4
M5	Production	6.4	6.42	6.4	6.4	5.4
	Revenue	158	163	169	176	156
M6	Production	2.9	3.1	3.6	3.9	3.5
	Revenue	54.6	54.1	62.2	74.02	74.7

The analysis indicates that M1 has remained as the market leader throughout the years from 2015 to 2019. M5 has retained its position as the second largest producer of automobiles. M6 has lost its position from being the third largest producer in 2015 to fifth in 2018. M3 improved from being the fourth largest automobile manufacturer in 2016 to the third largest manufacturer in 2018 by implementing latest technologies. M2 has positioned itself as fifth but have improved to fourth in 2018 and M4 has been observed as the sixth largest manufacturer throughout the period. The performance comparison has been presented in Table III. It is clear from the analysis that the adoption of the latest technologies has played a greater role in improving the performance and competitiveness of manufacturing organisations.

TABLE III
Performance comparison of the manufacturers

Manufacturer	2015	2016	2017	2018	2019
M1	1	1	1	1	1
M2	5	5	5	4	4
M3	3	4	4	3	3
M4	6	6	6	6	6
M5	2	2	2	2	2
M6	4	3	3	5	5

VI. Conclusion

Adoption of new technologies have played significant role in improving the performance and the competitiveness of manufacturing organisations. Moving towards a complete autonomous production level is the objective of all manufacturing organisations, however it requires lot of research and developmental efforts and resources. Adapting new technologies as and when the organisations move forward would help the manufacturing organisations to progress towards achieving the complete autonomous production. This study highlights the importance of technological applications and provides an insight on the advantages of adopting new technologies in automobile manufacturing sector.

References

[1] B. Li, B. Hou, X. Lu & C. Yang, "Applications of artificial intelligence in intelligent manufacturing : a review", *Frontiers of Information Technology & Electronic Engineering*, vol. 18, no. 1, pp. 86-96, 2017.

[2] B. Sandeep, "Age of transformation: Emergence of artificial intelligence for manufacturing industries", *Gavesana Journal of Management,* vol. 11, no. 1, pp. 17-26, 2019.

[3] A. Librantz, S. De Araújo, W. Alves, P. Belan, P. R. Mesquita & A. Selvatici, "Artificial intelligence-based system to improve the inspection of plastic mould surfaces", *Journal of Intelligent Manufacturing*, vol. 28, no. 1, pp. 181-190, 2017.

[4] M. Syafrudin, G. Alfian, N. Fitriyani, & J. Rhee, "Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing", *Sensors (Basel),* vol. 18, no. 9, pp. 1-31, 2018.

[5] Toyota Motor Corporation, "Toyota Will Establish New Artificial Intelligence Research and Development Company", *Toyota*, 2015, viewed 5th July 2021 <https://global.toyota/en/detail/10171645>.
[6] Plant Automation Technology, "The Future of Artificial Intelligence in Manufacturing Industries", *Industrial Events*, 2021, viewed 1st July 2021 <https://www.plantautomation-technology.com/articles/the-future-of-artificial-intelligence-in-manufacturing-industries>.
[7] M. Rangaiah, "How Tesla is making use of Artificial Intelligence in its operations?", *Analytic steps*, 2021 viewed 2nd July 2021 <https://www.analyticssteps.com/blogs/how-tesla-making-use-artificial-intelligence-its-operations>.
[8] N. Rychtyckyj, "Intelligent manufacturing applications at Ford Motor Company", IEEE Xplore, *Conference of the North American Fuzzy Information Processing Society*, pp. 298-302, 2005.
[9] A.K. Dogru, & B.B Keskin, "AI in operations management: applications, challenges and opportunities", *Journal of Data, Information and Managemen*t, vol. 2, no. 2, pp. 67-74, 2020.
[10] Toyota 2021, "Understanding the Toyota Production System", *Toyota*, viewed 15th June 2021, <https://www.creativesafetysupply.com/articles/understanding-the-toyota-production-system/>.
[11] Mazda Newsroom, "Technology/Design", *Mazda*, 2021, viewed 9 June 2021 <https://newsroom.mazda.com/en/technology/>.
[12] B. Pope, "Mazda Rethinks Manufacturing Process", *Industry News*, viewed 8 June 2021 <https://www.wardsauto.com/industry/mazda-rethinks-manufacturing-process>.
[13] A. Ghoshal, 2018, "Using new tech to speed up production develop autonomous vehicle", *Tata Motors Petkar in Techcircle*, 2018, viewed 8 June 2021 <https://www.techcircle.in/2018/08/24/using-new-tech-to-speed-up-production-develop-autonomous-vehicle-tata-motors-petkar>.
[14] P. Fretty, 2020, "Manufacturing Technology Leader of the Year: Ford's Gary Johnson", *Industry Week*, viewed 9 June 2021 <https://www.industryweek.com/technology-and-iiot/article/21150030/2020-manufacturing-technology-leader-of-the-year-fords-gary-johnson>.
[15] L. Mortara, J. Hughes, P.S. Ramsundar, F. Livesey & D.R. Probert, "Proposed classification scheme for direct writing technologies", *Rapid prototyping journa*l, vol. 15, no. 4, pp. 299-309, 2009.
[16] PSA Groupe, "Industry 4.0 and future's plant: Innovative industrial efficiency in Groupe PSA", 2017, viewed 1 June 2021 <https://www.groupe-psa.com/en/automotive-group/industrial-performance/>.
[17] RobotWorx, "How Toyota uses automation to improve processes", *Integration*, 2021, viewed 9 June 2021 <https://www.robots.com/articles/how-toyota-uses-automation-to-improve-processes>.
[18] Mazda, "Mazda's manufacturing flexibility", *Mazda Insider*, 2021, viewed 10 June 2021 <https://www.insidemazda.co.uk/2018/01/10/mazdas-manufacturing-flexibility/>.
[19] Suzuki, "Maruti's mega car plants lean on new technology to be future-ready", *Mint*, 2021, viewed 11 June 2021 https://www.livemint.com/Industry/bdnKlSrtBHHLjMeljaJDpI/Marutis-mega-car-plants-lean-on-new-technology-to-be-future.html.
[20] Ford, "Enters Digital Era with Its Advanced Manufacturing Center", *Connected Manufacturing Forum*, 2021, viewed 25 July 2021 <https://connectedmanufacturing.wbresearch.com/blog/ford-motor-company-advanced-manufacturing-center-strategy>.

Theorizing About the Determinants of Sustainable Entrepreneurship Intention and Behavior

Mariella Pinna

***Abstract*—** Sustainable entrepreneurship is an innovative corporate approach to create value combining economic, social and environmental goals overtime. In the last two decades, the interest towards sustainable entrepreneurship has flourished thanks to its potential to answer the current challenges of sustainable development. As a result, scholars are increasingly interested in understanding the determinants of the intentions to become a sustainable entrepreneur and consistent behavior. To date, prior studies provided empirical evidence for the influence of attitudes, perceived feasibility and desirability, values, and personality traits on the decision-making process of becoming a sustainable entrepreneur. Conversely, scant effort has been provided to understand which factors inhibit sustainable entrepreneurial intentions and behaviors, therefore a global understanding of sustainable entrepreneurship decision-making process is missing.

This paper contributes to the debate on sustainable entrepreneurship, by proposing a conceptual model that combines the factors which are predicted to facilitate and hinder the proclivity of individuals to become sustainable entrepreneurs. More in particular, the proposed framework theorizes about the role of the characteristics of the prospective sustainable entrepreneur (e.g., socio-demographic, psychological, cultural), the positive antecedents (e.g., attitude, social feasibility and desirability, among others) and the negative precursors (e.g., neutralization) in influencing sustainable entrepreneurship intentions and subsequent behavior.

The proposed framework is expected to shed further light on the decision-making process of becoming a sustainable entrepreneur, which in turn, is of practical relevance for public policy institutions and the society as a whole to enhance the favorable conditions to create new sustainable ventures.

***Keywords*—** *sustainable entrepreneurship, entrepreneurial intentions, entrepreneurial decision-making, antecedents of entrepreneurial intention and behavior.*

Mariella Pinna is with the University of Sassari, Italy (e-mail: mapinna@uniss.it).

Residual Stress Relaxation for the Assessment of Welded Structures

H. Tryla, L. Barrallier, L. Heraud, P. Bristiel

***Abstract*—**Gas metal arc welding (GMAW) creates strong microstructure gradients generating residual stress fields that influence the mechanical response and service life of welded joints. Under mechanical loadings, the welding residual stresses can be relaxed or redistributed. This work attempts to study the relaxation of welding residual stress generated by GMAW. Three types of lap joints specimens have been tested to determine each fatigue strength for a full tensile uniaxial load. Subsequently, experimental investigations based on interrupted high cycle fatigue tests have been conducted. Residual stresses were analyzed before, during and after cyclic loads by means of X-ray diffraction. The welding residual stress evolution kinetics for high cycle fatigue loading has been identified. Furthermore, metallurgical analysis such as EBSD analysis was carried out on tested specimens to clearly identify the underlying mechanisms of welding residual stress relaxation. In fine, this work aims to propose modeling to consider stress relaxation in numerical fatigue simulations for the assessment of automotive suspension steel parts.

***Keywords*—** gas metal arc welding, relaxation, residual stress, welded structures, X-ray diffraction.

H. Tryla, L. Barrallier and L. Heraud are with the Arts et Métiers Institute of Technology, MSMP, HESAM Université, F 13617 Aix en Provence, France (e-mail: hugo.tryla@ensam.eu, laurent.barrallier@ensam.eu, lorene.heraud@ensam.eu).

H. Tryla and P. Bristiel are with the Groupe PSA, Research and Development Division, Centre d'Expertise de Poissy, 78300 France (e-mail: hugo.tryla@mpsa.com, philippe.bristiel@mpsa.com)

Effect of heat input on the transverse and longitudinal distortion of multilayer butt-welded joints on Submerged Arc Welding Process

M. Bahrami, M. Guillot

***Abstract*—** In any industry, Submerged Arc Welding (SAW) is fully used when high-quality welding joint join plates of higher thickness is needed. There is a number of different weld distortions that Transverse and Longitudinal are more important in heavy industry and cause issues. In this work-study based on experimental results, the effects of welding heat input on longitudinal and transverse distortion for each layer of the weld were investigated. In this paper, high thickness low carbon steel of ASTM A131 grade A equivalent grade A steel approved by shipbuilding classification society (Lloyds and DNV-GL) samples have been welded in V-butt joint preparation with SAW process. Taguchi method is used for the Design of Experiments (DOE). Heat input for each pass of the weld calculated. After 24 hours of the welding operation, both distortions were measured for all specimens. Finally, the effect of the heat input, as well as wire feed speed on both distortions were researched. The ultrasonic test has been used to analyze macro examination and find a welding defect for all samples and a radiography test has been performed for the best samples as well. Finally, the finite element model (FEM) has been simulated and comparison results between FEM and experiments have been shown.

***Keywords*—** Heat Input, Cooling Rate, Distortion, Submerged arc welding, Taguchi method

Milad Bahrami, Ph.D. Candidate, University of Laval (phone: 438-995-2899; e-mail: milad.bahrami.1@ ulaval.ca).

Michel Guillot, Full professor University of Laval, Research Team, Department of Mechanical Engineering, Laval University, Quebec, G1V 0A6 (e-mail: Michel. Guillot @ulaval.ca).

Assessment of Different Additive Manufacturing Routes to Repair Titanium and Aluminium Alloys: Comparison of Power Bed Fusion and Solid-State Material Deposition Processes

L. Toualbi, C. Davoine, M. Thomas, T. Bouilly, D. Miot

***Abstract*—**Repair or restoration of damaged products becomes an industrial issue in order to both reduce the cost and increase the lifetime of critical parts. In the context of the aerospace industry, repair by the additive manufacturing process is a new field of study, requiring the highest attention to safety, i.e., material health. Particular care has to be taken to ensure the reliability of the repaired parts. In order to do that, we have to understand the microstructure change during repair and ensure that the associated mechanical properties are comparable to those of the initial material.
Among the existing additive manufacturing technology, several processes are a potential candidate to be used for repair. Direct energy deposition or power bed fusion processes provide a strong bond between repaired and initial materials but also induce thermal stresses both in the remanufactured part and at the interfacial region, leading to the need of further heat treatment. In the specific case of massive part, this step can be particularly problematic. From this perspective, solid-state material deposition technology appears to be promising for repair.

The aim of this study is to evaluate the suitability of various additive manufacturing processes to repair 10-2-3 titanium and 2219-aluminium alloys. Direct energy deposition, power bed fusion as well as solid-state material deposition routes are considered.

A detailed characterization of the microstructure using a Scanning Electron Microscope permits the comparison of the initial material and the repaired one. The interfacial region is particularly studied using hardness measurements.

The use of Electron Backscatter Diffraction microstructure mapping at the interfacial region gives access to the material deformation history. Microstructural features such as grain size, crystallographic texture and local misorientation, which is related to the strain level, permit understanding microstructural changes during the remanufacturing step.

Mechanical tests are conducted on as-fabricated, i.e., repaired and initial, i.e., healthy samples. The comparison of initial and repaired mechanical properties gives, in a first approach, an evaluation of the performance of the remanufactured samples.

The comparison of the efficiency of several additive manufacturing processes to repair parts made of two different alloy families provide a wide overview of specific challenges in restoration of damaged products. The obtained results permit a better understanding of the microstructure change during repair, which is a key to determine the best suitable remanufacturing process for a given material and use.

***Keywords*—**Repair, remanufacturing, power bed fusion process, solid-state material deposition process, cold spray, titanium alloys, aluminium alloys.

L. Toualbi is with the French Aerospace Lab, 92320 Châtillon (phone: +33 1 46 73 46 55; e-mail: louise.toualbi@onera.fr).

C. Davoine is with the French Aerospace Lab, 92320 Châtillon (e-mail: cecile.davoine@onera.fr).

M. Thomas is with the French Aerospace Lab, 92320 Châtillon (e-mail: marc.thomas@onera.fr).

T. Bouilly is with CNES Launcher Directorate, 75612 Paris (e-mail: thibaut.bouilly@cnes.fr).

D. Miot is with CNES Launcher Directorate, 75612 Paris (e-mail: david.miot@cnes.fr).

Thermo-Mechanical Modelling of the Wire Arc Additive Manufacturing (WAAM) Process

S. Hilal, D. Missoum-Benziane, P. Kerfriden, S. Hendili, M. Mazière

***Abstract*—**Additive manufacturing processes by wire deposition such as the Wire Arc Additive Manufacturing process (WAAM) allows manufacturing large mechanical components by adding successive layers of the molten metallic wire using electrical arc. This process raises an increasing interest and may provide a viable alternative to the conventional manufacturing processes for its many advantages, among which the ability to produce large parts with very high deposition rates in low-cost installations while avoiding the safety and environmental issues raised by the use of metallic powder.

If this process derived from welding is well known, its use at the industrial level requires to better understand the influence of the welding parameters and the deposition strategies on the residual stresses and distortions distributions generated during the manufacturing process. However, the microstructure, the residual strains and stresses, among other quantities of interest, are very difficult and expansive to access experimentally but can be provided by numerical simulations.

Therefore, the proposed research work consists of setting up, calibrating and validating a finite element model to simulate the WAAM process, to then determine the consequences in terms of microstructure, stresses and distortions on industrial parts.

A macroscopic thermo-mechanical model was implemented using the finite element code Code_Aster. Simulations were carried out for different geometries and with different deposition strategies. The implementation of the model requires the determination of input parameters, such as the parameters of the heat source, that can not be directly measured or characterized. In order to calibrate the setup model, instrumented experimental tests are conducted using thermocouples, thermal imaging and 3D scan. To fit the thermal model to the experimental data, a Bayesian calibration of the parameters based on a surrogate model approach was performed. The finite element results are then compared to the experimental data, and they show good agreement.

This work is part of the additive manufacturing platform “Additive Factory Hub” (AFH) involving French manufacturers and academics.

***Keywords*—**additive manufacturing, WAAM process, arc welding, thermo-mechanical simulation, Bayesian calibration.

S. Hilal is with EDF R&D division, 6 quai Watier, 78400 Chatou, FRANCE / and the Centre des Matériaux Mines Paristech, 63-65 rue Henri Auguste Desbrueres, 91300 Evry, FRANCE (e-mail: sami.hilal@edf.fr / sami.hilal@mines-paristech.fr).

D. Missoum-Benziane is with the Centre des Matériaux Mines Paristech, 63-65 rue Henri Auguste Desbrueres, 91300 Evry, FRANCE (e-mail: djamel.missoum-benziane@mines-paristech.fr).

P. Kerfriden is with the Centre des Matériaux Mines Paristech, 63-65 rue Henri Auguste Desbrueres, 91300 Evry, FRANCE (e-mail: pierre.kerfriden@mines-paristech.fr).

S. Hendili is with EDF R&D division, 6 quai Watier, 78400 Chatou, FRANCE (e-mail: sofiane.hendili@edf.fr).

M. Mazière is with the Centre des Matériaux Mines Paristech, 63-65 rue Henri Auguste Desbrueres, 91300 Evry, FRANCE (e-mail: matthieu.maziere@mines-paristech.fr).

Microstructural Evolution of an Interface Region in a Nickel-Based Superalloy Joint Produced by Direct Energy Deposition

M. Ferguson[1,2,a], T. Konkova[1,3], I. Violatos[2]

[1]Design, Manufacturing & Engineering Management (DMEM) Department, University of Strathclyde, 75 Montrose Street, Glasgow, UK

[2]Advanced Forming Research Centre (AFRC), University of Strathclyde, 85 Inchinnan Drive, Inchinnan, Renfrewshire PA4 9LJ, UK

[3]Institute for Metals Superplasticity Problems, Russian Academy of Science, 39 Khalturin Str., Ufa, 450001, Russia

[a]matthew.ferguson@strath.ac.uk

***Abstract*—** Microstructure analysis of additively manufactured (AM) materials is an important step in understanding the interrelationship between mechanical properties and materials performance. Literature on the effect of a laser-based AM process parameters on the microstructure in the substrate-deposit interface is limited. The interface region, the adjoining area of substrate and deposit, is characterized by the presence of the fusion zone (FZ) and heat affected zone (HAZ) experiencing rapid thermal gyrations resulting in thermal induced transformations. Inconel 718 was utilized as a work material for both the substrate and deposit. Three blocks of Inconel 718 material were deposited by Direct Energy Deposition (DED) using three different laser powers, 550W, 750W and 950W, respectively. A coupled thermo-mechanical transient approach was utilized to correlate temperature history to the evolution of microstructure. Thermal history of the deposition process was monitored with the thermocouples installed inside the substrate material. Interface region of the blocks were analyzed with Optical Microscopy (OM) and Scanning Electron Microscopy (SEM) including electron back-scattered diffraction (EBSD) technique. Laser power was found to influence the dissolution of intermetallic precipitated phases in the substrate and grain growth in the interface region. Microstructure and thermal history data were utilized to draw conclusive comparisons between the investigated process parameters.

***Keywords*—** Additive manufacturing, Direct Energy Deposition, Electron Back-Scatter Diffraction, Finite Element Analysis, Inconel 718, microstructure, Optical Microscopy, Scanning Electron Microscopy, Substrate-deposit interface region.

Ferguson, M. is from Design, Manufacturing & Engineering Management (DMEM) Department at the University of Strathclyde, Glasgow, G1 1XJ, (corresponding author, phone: +44141 534 5200; e-mail: matthew.ferguson@strath.ac.uk)

Konkova, T. is from Design, Manufacturing & Engineering Management (DMEM) Department at the University of Strathclyde, Glasgow, G1 1XJ, UK

Violatos, I. is from Advanced Forming Research Centre (AFRC), University of Strathclyde, 85 Inchinnan Drive, Inchinnan, Renfrewshire, Scotland PA4 9LJ, UK

I. INTRODUCTION

INCONEL 718 is a Ni-based superalloy designed to tolerate severe operating conditions that demand mechanical superiority in harsh service environments. Due to its excellent elevated temperature properties (up to 650ºC) that include corrosion and oxidation resistance in combination with high creep and fatigue resistance, the superalloy is commonly used to manufacture aero-engine components and gas turbine disks [1]. The outstanding mechanical performance is due to the microstructure and precipitation kinetics formed during solid solution and subsequent heat treatments. Additive manufacturing (AM) capabilities have significantly evolved over the last few decades by optimizing process parameters for repair and remanufacturing abilities [2]. Direct energy deposition (DED) is an AM technique developed from powder injection laser cladding and rapid prototyping [3], [4]. A localized heat source generates a melt pool on the substrate surface, to which blown powder is directly fed relative to the raster though a nozzle with a carrier gas directly into the focal point of the laser During the build process, feedstock, either powder or wire, is fed directly into the molten pool created by the localized laser heat source which solidifies and successively creates a metallurgical bonded component [5].

The DED process can be used to manufacture components with complex geometries that require intricate manufacturing and to repair worn parts, such as turbine blades, and disks. Repairing gas turbine engine components reduces high manufacturing costs incurred during replacement parts while improving waste reduction.

The final microstructure is determined by the localised thermal history experienced during manufacturing [6]. By correlating process parameters to temperature history and utilizing this data to deduce microstructural features, the mechanical properties can be determined. Numerous studies have examined the microstructure of additively manufactured IN718 in as-deposited and post heat treatment conditions [7]–[13]. Most investigations predominately focus on the deposited

component with the interface region receiving very little attention. The published literature mentions the interface as a band near the substrate/deposit boundary, where the δ-phase had dissolved into the γ matrix [14]–[17]. The δ-phase is incoherent with the γ matrix and forms at grain boundaries, playing a key role in grain size control in wrought IN718[18]. The interface region comprises of the joint layer where the parent material and deposit material are joined during DED. During the build process, this region is subjected to complex thermal gyrations, steep thermal gradients, and rapid cooling rates. When considering the thermal history experienced, the microstructural evolution response is vitally important for an outstanding metallurgical bond. The resultant complex thermal histories influence the microstructure, mechanical properties, and residual stresses in the final part.

To correlate the temperature history and resultant phenomena's, many researchers use the finite element analysis (FEA) to model the DED process and study the effects on the parts. However, existing data on temperature history is still limited [19]–[26].

The objective of this work is to develop a thermal model of the DED process and validate the predicted thermal response with in-situ temperature measurements. The experimentally validated thermal model will be utilized to deduce the evolution of microstructure in the interface of DED IN718, based on the simulated temperature history experienced.

II. Methodology

A. Experimental Procedure

A 2.5kW continuous wave YAG fibre laser deposition system from nLIGHT® Alta™ which reproduced a 2 mm laser spot diameter at the substrate surface was used in the experiments. The configuration consists of a coaxial Hybrid™ AMBIT™ series mill head which is fitted to a Nikken rotary/tilt unit which allows 5-axis of deposition. Process parameters used in the experiments were chosen from literature review [8], [11], [14], [16], [27]–[30] and these can be seen in Table 1. The coaxial nozzle with powder convergence point focused on deposition substrate surface was utilized. Argon was used as the carrier and shielding gas with a flow rate of 5 L min^{-1} and 8 L min^{-1}, respectively. The surface of the wrought IN718 substrate was polished by abrasive paper to remove oxide layer and cleaned with acetone before the DED process. Each sample consisted of a single clad track and an offset ten-layered solid block where each layer consisted of ten adjacent tracks deposited unidirectionally. A 50s dwell time was considered for the block to reach room temperature between single and block deposition. After the DED, samples were left at room temperature to cool.

TABLE I Process Parameters Used For Experimental Deposition

Laser Power (W)	Scanning Velocity (mm/s)	Powder feed rate (g/min)	Increment of Z axis (mm)	Overlap (%)
950	10	6	0.5	20
750	10	6	0.5	20
550	10	6	0.5	20

In this research, temperature measurements were performed using thermocouples of type-N with a 1.0 mm diameter. Nicotherm D™ Sheath have a measuring range from -270ºC to 1250ºC and were calibrated in-house using a Martel Electronics Precision Temp Calibrator. A total uncertainty of ± 2.4 ºC was found for temperature measurements. A total of twelve (*TC1-12*) 1.1 mm diameter holes were drilled via w-EDM to a distance of 23± 0.2 mm, leaving approximately 0.7 mm between deposited surface and thermocouple tip. The temperatures were recorded at the start, middle and end of the single track and repeated for solid blocks as cooling rates would differ depending on thermocouple location. Each thermocouple and corresponding hole were captured. A National Instrument CompactDAQ chassis was connected to a PC and recorded the temperatures in high resolution and high speed in Labview® at a rate of 50Hz. Temperature measurements were recorded for the single clad track and solid block in each power case. The 950W experiment was repeated twice for validity. The thermocouples were inserted as functionally possible without damaging the flexible sheath and the depth of each thermocouple was measured post deposition.

An inverse heat transfer coefficient (HTC) experimental trial was carried out by placing the thermocouple equipment in a substrate, placed in a Carbolite® furnace, heated to set temperature followed by cooling via natural convection in air. The temperature history was collected.

After DED, samples were sectioned using w-EDM. The microstructure of both parallel section (X-Y plane) and transverse section (X-Z plane) were observed by Optical Microscope (OM) using a Leica™ DM12000 and Scanning electron microscopy (SEM) equipped with electron back-scattered detection using an FEI Quanta™ 250 fitted with Oxford Instrument EBSD detector. Samples were prepared by standard technique grinding and polishing process which is finished by 0.02µm colloidal silica suspension. For the OM, samples were etched using Kallings No.2 reagent with water.

B. Numerical Simulation

An experimentally validated numerical transient thermal finite-element (FE) simulation was conducted in the commercial welding MSC software Simufact Welding 8.0. The module incorporates a "quiet-element technique" to model powder addition. In general, a finer mesh will improve the accuracy of the calculation although at the penalty of computational time. A mesh sensitivity study was performed to optimize the convergence of accuracy and computational time. Considering the interface region is subjected to the largest temperature gradients and most concentrated stresses, a refinement was used in this region with coarser mesh away from this region. All meshes were defined in ABAQUS® finite element code imported to Simufact. In this simulation, the mesh was divided by linear brick elements (DC3D8), comprising of 179,039 mesh elements and 190518 mesh nodes. Temperature-dependent material properties based on [31], [32] were used to eliminate simulation error induced by mismatched physical properties at different temperatures.

A moving 3D double ellipsoidal heat source model developed by *Goldak* [33] simulated the heat input from laser deposition. Based on microstructural observations, the geometry of the melt

pool including length, depth and width could be experimentally validated by the single track.

A power absorption efficiency of 0.37 [34] was selected, however needs determined using calibration experiments. Convective and radiative boundary conditions were applied to all free surfaces, including those of newly deposited material.

At the clamped boundaries, the substrate is in contact with metallic surroundings and the convective heat transfer coefficient was replaced with a contact heat transfer coefficient $\propto$ with a value of 1200W/(m^2.K) [24] to simulate strong heat loss from metallic clamp contact.

III. Calibration of Thermal Model

Fig. 1 shows the thermal history captured for different thermocouples in the substrate, which were compared to *'particle placements'* temperature history in Simufact welding 8.0. Simulations were calibrated by measuring the heat source geometry using OM, performing an inverse simulation for the process efficiency, η, and calculating temperature-dependent convection coefficient values using inverse heat transfer coefficients. The simulated temperature history for all twelve thermocouples is in close agreement with the experimental thermocouple measurements which capture similar peak temperatures and trends.

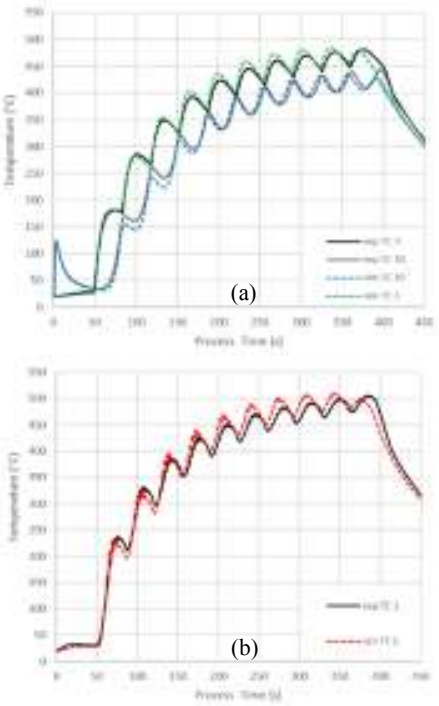


Fig. 1 Comparison between experimental data and simulated data for (a) TC10 (Single clad track) & TC3 (first TC at track 1 of solid block) and (b) TC 1 (final TC at track 10 of deposited block)

IV. Results and Discussion

A. Interface Investigation

The interface region between the as-deposited clad and substrate material revealed a white band as previously cited [14]–[16], [35] that, for 950W, ranged from 330µm at the melt pool midpoint (maximum temperature) to 220µm towards the edge of the melt pool due to the gaussian distribution. The bright band decreased in depth with decreasing laser power, the 750W had a maximum depth of 230µm with 135µm at the edge and the 550W had a maximum of 210µm with 190µm at the edge.

Fig. 2 shows a comparison between an optical micrograph capturing the cross-sectional (x-z) interface region (Track 6, P=950W) and simulated thermal gradient for track 6. The interface region can be described as the clad zone (CZ), the partially melted zone (PMZ), the heat affected zone (HAZ) and the DA IN718 parent material (PM), which experience slightly different temperature histories.

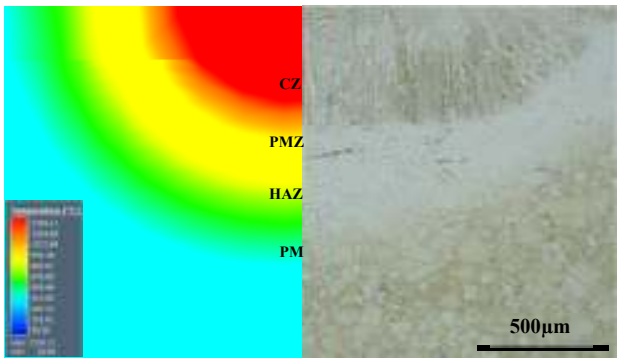


Fig. 2 (a) Thermal gradient for melt pool, P=950W (b) Optical microscope of interface region as-deposited Inconel 718 superalloy

Analyzing the bright band near the substrate/deposit interface, the equiaxed grains evident apparent in the PM were less obvious, (Fig. *2*(b)). In the PMZ region, there is strong evidence to support that the δ-phase were dissolved into the γ matrix. This dissolution of δ-phase (around 996-1015°C [36]) would effectively 'unpin' the microstructure, which could result in grain growth and produce a higher Nb content in this region. If the as-deposited samples were subjected to an aging heat treatment, this could promote a higher density of strengthening phases dispersed across the interface region and thus improved strength [15]. Fig. 3 reveals the presence of solid solution carbides which were randomly dispersed throughout the substrate and remained across the PMZ due to their relatively high solvus temperature (around 1277°C [37]). The microstructural features coupled with the data predicted by the thermal model (Fig. 4) collate this PMZ region experiences a

temperature history between 996°C and 1277°C during the first layer of deposition. Both experimental and simulation data captured agree with previously published literature [14], [15].

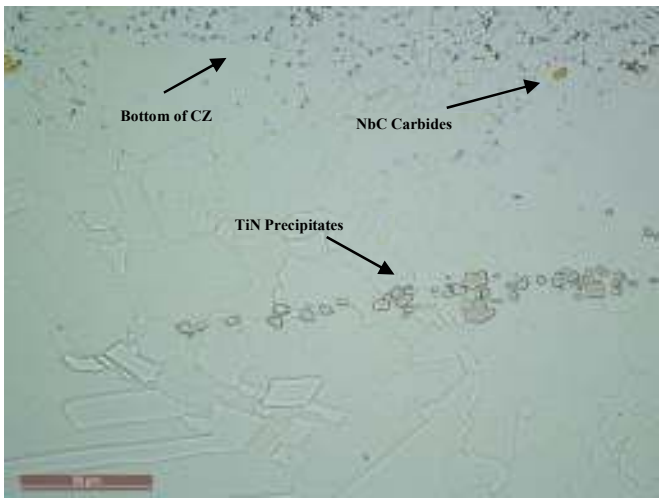


Fig. 3 Optical micrograph of Interface for P=950W, detailing the presence of both NbC and TiN solid solution precipitates

The temperature history experienced at the PMZ was retrieved from the thermal models as shown in Fig. 4. Fig. 4This information will be vital to correspond evolution of microstructure and microstructural features. Due to the sluggishness of δ, Alloy 718 is thermo-mechanically stable up to around 650°C, whereas key thermal induced phase changes occur above this temperature [38]. The 650°C application limit is due to the conversion of γ'' to δ which is accelerated when thermal exposure is over 650°C. At ~900°, δ takes preferential precipitation, so no precipitation of strengthening phases γ'' can occur. This preferential precipitation is up to ~1010°C [39].
The role of Nb segregation on phase precipitation is largely influenced by the solidification sequence, but a heat treatment effect from laser attenuation will play a significant role in microstructural evolution.

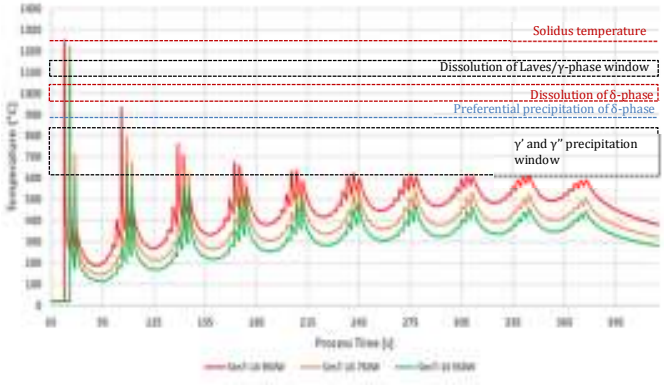


Fig. 4 HAZ temperature history retrieved from thermal model experienced for all three process parameter conditions

Although the presence of δ-phase will inhibit grain growth, the dissolution of δ phases back into the γ-matrix will produce a higher Nb density in that region and subsequent layer addition could replicate or initiate DA aging treatment, thus improving strength at the interface [15]. At temperatures 985°C, the δ-phase was seen to dissolve by almost 50pct, and was completely dissolved by 1010°C [14], [40]. In the higher temperature regions, the dissolving reaction of Laves/γ eutectic initiates around 1080-1140°C [41] and sufficient Laves phase is dissolved at 1150°C [42] (homogenization temperature) which could release more Nb content that could redistribute homogenously in the austenitic matrix. In addition to the alloying elements being dissolved back into the matrix, the Laves phase is well documented as detrimental to mechanical properties [15], [42], [43]. In Sui et al's study, it was determined that the failure mechanism of LMD IN718 was microscopic holes coalescence ductile fracture and that Laves phase were the main nucleases for microscopic hole formation [14].
Although standard homogenization of IN718 normally occurs for 1h and the DA requires a 2-step process for 8 hours each step, the temperature history for all three power cases in the PMZ region is sufficient for partial dissolution of δ-phase. Moreover, the subsequent layer deposition resulted in a heat treatment effect and a temperature retained close to the γ'/γ'' precipitation. Therefore, the microstructural response to temperature requires a more in-depth investigation.

V. Microstructural Evolution

In the cladding zone, it could be seen that columnar dendrites grew epitaxially along the deposition direction from the substrate zone (Fig. 2). This characteristic of directional solidification occurs when the heat loss direction during the solidification of the melt pool is about perpendicular to the surface of the substrate resulting in the epitaxial directional growth [14], [44]. The dendrite formation along the substrate interface is an outcome of the localized, instantaneous cooling rates occurred during solidification. Fig. 5 represents the cross-sectional microstructure of LMDed IN718, consisting of mainly austenitic matrix and continuous networks of Laves phase that occurs parallel to the build direction. As successive layers are added, the isotherm penetrates the surface of the previously deposited layer and remelts a region, the dendrites continue to grow along previous direction and form the combination of continuous and course Laves with strong directional selectivity [42]. The ultra-rapid cooling rate 10^3-10^6 °C/s of laser metal deposition limits the network to mainly consist of primary dendrite arms located very close to one another, hindering the growth of secondary/tertiary arms. The epitaxial grown dendrites curve towards the heat source to ensure the maximum thermal gradients present at solid-liquid interface are maintained with proceeding growth [8].
The primary dendrite arm spacing of columnar dendrites at the interface were measured for Layer 1, Track 6 in all three power cases, using ImageJ analysis. These measurements are to quantitively compare microstructures and to understand the cooling rates experienced from solidification, these results are shown in Table II.

Table II Primary Dendrite Arm Spacing Measurements

Power	*950W*	*750W*	*550W*
Average PDAS (μm)	4.2	3.5	3.1
Standard deviation (μm)	0.87	0.72	0.64

The cooling rate is expected to increase with decreasing the laser power considering that coarser dendrites were given sufficient time to evolve during solidification. The highest cooling rate, based on the smallest PDAS will produce a very fine dendritic structure. A reason for this phenomenon occurring is due to the melt pool size. The 950W inputs more energy to the melt pool region, therefore creating a slightly larger melt pool than the other two process parameters. A slightly larger melt pool size will in effect take incrementally longer for all the material to reach solidification temperature given the ability to retain more thermal energy [43].

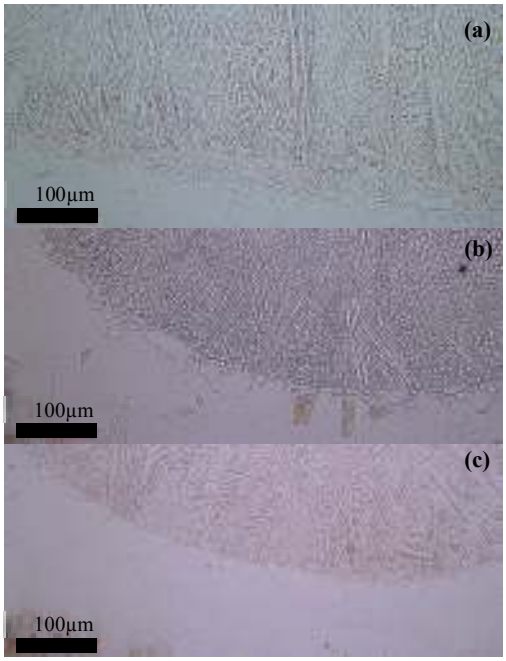


Fig. 5 Optical micrograph of cross section (x-z) for (a) P=950W, (b) P=750 and (c) P=550W

Fig. 6 shows an SEM image of the cross-sectional morphology for the as-deposited clad layer in the P=950W sample, specifically right at the interface region. The solidification sequence for as-deposited IN718 is well understood to be L→L + γ →L + Nb/γ → L + γ → Laves/γ [45]. And due to the low content of C present, formation of Nb carbides is negligible. Therefore, the dark grey phase in Fig. 6 can be identified as the austenite matrix γ phase and the dispersive white bright phase as interdendritic Laves phase, which have been identified using Energy Dispersive Spectrometer (EDS) and TEM in literature [9], [15]. The relatively thick and continuous networks of Laves phase were found in all three power cases. Due the high thermal gradient and instantaneous cooling rates experienced at the FZ, the fine columnar dendrite morphology is observed with interdendritic laves forming irregular shaped but continuous networks from the interface region. Brittle laves phase is detrimental to mechanical properties [46] with a wide-range-melting and precipitation temperature of 650°→1100°C which the upper limit melting temperature can vary with constituents. Homogenization is often employed to completely irradicate the brittle interdendritic phase as it depletes the γ matrix of Nb, a key element in the strengthening γ'' phase [7].

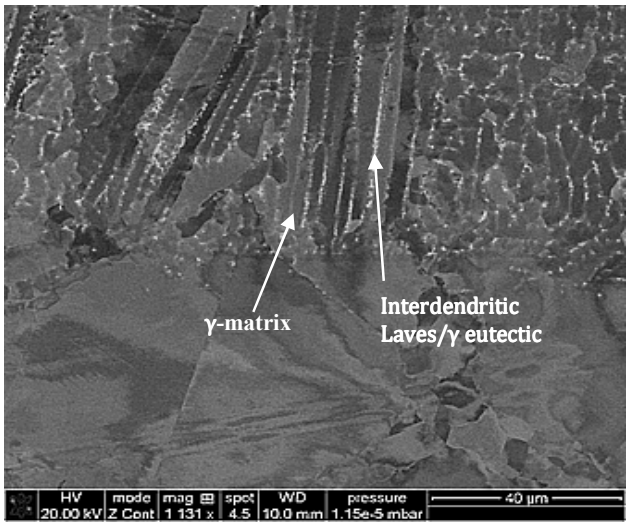


Fig. 6 Interface region P=950W, showing Austenitic γ-matrix with interdendritic Laves/γ eutectic

During the DED process, most of the heat is dissipated through the substrate and previously deposited layers which both act as heat sinks. Heat flow direction can be explained using Fig. 7, the black arrows indicate directional solidification. During the solidification of the melt pool, heat flow direction is approximately perpendicular to the surface of pre-deposited layers and substrate. In this direction, the maximum thermal gradient is achieved, leading to the directional growth of dendrites grown from the bottom up. The directional solidification of the dendrites is explained with respect to maximum heat flux vector, Q [14]. Q_h represents the horizonal heat flux through previously deposited material due to deposition direction and Q_v is the heat flux through the wrought substrate. Heat loss is much lower in the horizontal direction than in vertical direction, therefore the resultant vector Q, is approximately parallel to the build direction and grows epitaxially from the bottom up, aligned with the greatest thermal gradient, towards the scanning direction resulting from competitive growth of dendrites [9], [47]. This directional solidification would be in the tailing edge of the melt pool, the last section to solidify.

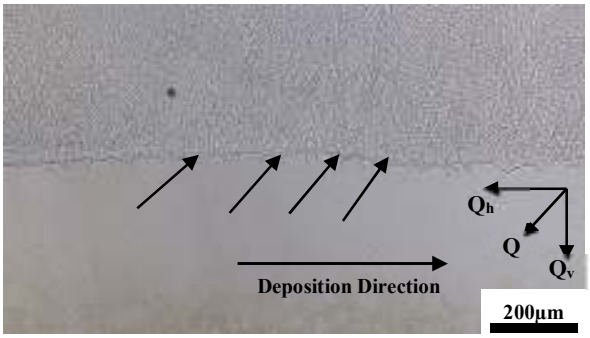


Fig. 7 Directional solidification, P=750W

A. Grain Selection and Texture

Fig. 8 shows the IPF map generated from the EBSD analysis at the interface region of the 950W sample, with build direction along the growth direction (Y). Most grains grow in the direction of {100} family of grain planes with a preferential growth <001> for face-centred alloy IN718 [48]. Along the interface, depending on solidification conditions, nucleation or growth may be predominant. At the melt pool center where thermal gradient is the steepest and one of the <100> directions of the substrate is parallel to the heat flow direction, solidification occurred by epitaxial growth of the PM crystal. As the laser metal deposition was carried out on a polycrystalline substrate, epitaxial grain growth was attained when the substrate grains had a <100> crystallographic direction parallel to the local heat flow direction.

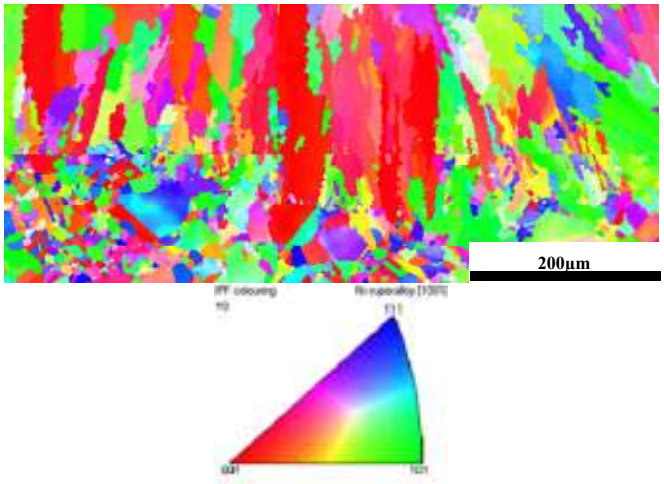


Fig. 8 The as-deposited Inconel 718 superalloy interface region using 950W (Orientated with build direction along the growth direction (Y))

The interface investigation revealed, nucleation of new grains in the CZ which appeared to occur where, in the parent material, no grains have a <100> crystallographic orientation parallel to the heat flow direction. The grain morphology is mostly columnar with a common <100> crystallographic orientation which have grown rapidly, impeding the growth of other grains, this agrees strongly with Dinda et al's investigation [49]. Given the small overlap between adjacent tracks, most grains grew curved towards the heat source to attain maximum thermal gradient with proceeding growth like the dendritic morphology.

VI. Conclusion

A thermal model of the DED process for three process parameters of laser power, P=550W, P=750W and P=950W was developed. An experimentally validated thermal model was implemented to predict the temperature history experienced during DED of IN718.

The interface region was diluted into four key aspects; clad zone (CZ), partial-melted zone (PMZ), heat affected zone (HAZ) and parent material (PM) with varying thermal histories, therefore varied microstructural response. The microstructure of the CZ was found to be non-uniform, consisting of columnar dendrites growing epitaxially along the build direction with interdendritic Laves phase. However, the PM microstructure was equiaxed with a uniform distribution of δ phase. The gaussian shape of the melt pool was indicative of dendrite growth towards maximum temperature. The dendrites were found to grow with respect to maximum thermal gradient, epitaxially from the substrate material. The PDAS decreased with decreasing laser power, therefore higher localized cooling rates were expected, given alterations to mechanical performance. During solidification of the LMD IN718, grains with the <100> crystallographic orientation grew epitaxially from the substrate and continued to grow into the deposit.

The thermal history captured by the simulation will be imperative to explain the microstructural response for the as-deposited interface region.

Acknowledgements

The authors would like to acknowledge the technical staff at the Advanced Forming Research Centre (AFRC) for their assistance with the laser metal deposition and an extended thank you to the microscopy laboratory technicians of the AFRC for their support with the microstructure analyses.

References

[1] E. Hosseini and V. A. Popovich, "A review of mechanical properties of additively manufactured Inconel 718," *Addit. Manuf.*, vol. 30, no. September, p. 100877, 2019.

[2] L. J. Kumar and C. G. K. Nair, "Laser metal deposition repair applications for Inconel 718 alloy," in *Materials Today: Proceedings*, 2017, pp. 11068–11077.

[3] B. Graf, S. Ammer, A. Gumenyuk, and M. Rethmeier, "Design of experiments for laser metal deposition in maintenance, repair and overhaul applications," *Procedia CIRP*, vol. 11, pp. 245–248, 2013.

[4] S. Sreekanth, E. Ghassemali, K. Hurtig, and S. Joshi, "Effect of Direct Energy Deposition Process Parameters on Single-Track Deposits of Alloy 718," *Metals (Basel).*, vol. 10, no. 1, p. 96, 2020.

[5] P. Alvarez, M. Montealegre, J. Pulido-Jiménez, and J. Arrizubieta, "Analysis of the Process Parameter Influence in Laser Cladding of 316L Stainless Steel," *J. Manuf. Mater. Process.*, vol. 2, no. 3, p. 55, 2018.

[6] L. Costa, R. Vilar, T. Reti, and A. M. Deus, "Rapid tooling by laser powder deposition: Process simulation using finite element analysis," *Acta Mater.*, vol. 53, no. 14, pp. 3987–3999, Aug. 2005.

[7] H. Qi, M. Azer, and A. Ritter, "Studies of standard heat treatment effects on microstructure and mechanical properties of laser net shape manufactured INCONEL 718," *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.*, vol. 40, no. 10, pp. 2410–2422, 2009.

[8] L. L. Parimi, R. G. A., D. Clark, and M. M. Attallah, "Microstructural and texture development in direct laser fabricated IN718," *Mater. Charact.*, vol. 89, pp. 102–111, Mar. 2014.

[9] E. L. Stevens, J. Toman, A. C. To, and M. Chmielus, "Variation of hardness, microstructure, and Laves phase distribution in direct laser deposited alloy 718 cuboids," *Mater. Des.*, vol. 119, pp. 188–198, Apr. 2017.

[10] X. Zhao, J. Chen, X. Lin, and W. Huang, "Study on microstructure and mechanical properties of laser rapid forming Inconel 718," *Mater. Sci. Eng. A*, vol. 478, no. 1–2, pp. 119–124, Apr. 2008.

[11] M. Ma, Z. Wang, and X. Zeng, "Effect of energy input on microstructural evolution of direct laser fabricated IN718 alloy," *Mater. Charact.*, vol. 106, pp. 420–427, Aug. 2015.

[12] Q. Zhang, J. Yao, and J. Mazumder, "Laser Direct Metal Deposition Technology and Microstructure and Composition Segregation of Inconel 718 Superalloy," *J. Iron Steel Res. Int.*, vol. 18, no. 4, pp. 73–78, Apr. 2011.

[13] A. S. Johnson, S. Shao, N. Shamsaei, S. M. Thompson, and L. Bian, "Microstructure, Fatigue Behavior, and Failure Mechanisms of Direct Laser-Deposited Inconel 718," *Jom*, vol. 69, no. 3, pp. 597–603, 2017.

[14] S. Sui, J. Chen, R. Zhang, X. Ming, F. Liu, and X. Lin, "The tensile deformation behavior of laser repaired Inconel 718 with a non-uniform microstructure," *Mater. Sci. Eng. A*, vol. 688, pp. 480–487, Mar. 2017.

[15] R. G. Ding, Z. W. Huang, H. Y. Li, I. Mitchell, G. Baxter, and P. Bowen, "Electron microscopy study of direct laser deposited IN718," *Mater. Charact.*, vol. 106, pp. 324–337, 2015.

[16] L. J. Kumar and C. G. K. Nair, "Laser metal deposition repair applications for Inconel 718 alloy," *Mater. Today Proc.*, vol. 4, no. 10, pp. 11068–11077, Jan. 2017.

[17] E. Lertora, C. Mandolfino, and C. Gambaro, "Mechanical Behaviour of Inconel 718 Thin-Walled Laser Welded Components for Aircraft Engines," *Int. J. Aerosp. Eng.*, vol. 2014, pp. 1–9, 2014.

[18] W. J. Sames, K. A. Unocic, R. R. Dehoff, T. Lolla, and S. S. Babu, "Thermal effects on microstructural heterogeneity of Inconel 718 materials fabricated by electron beam melting," *J. Mater. Res.*, vol. 29, no. 17, pp. 1920–1930, 2014.

[19] E. R. Denlinger, J. C. Heigel, P. Michaleris, and T. A. Palmer, "Effect of inter-layer dwell time on distortion and residual stress in additive manufacturing of titanium and nickel alloys," *J. Mater. Process. Technol.*, vol. 215, pp. 123–131, 2015.

[20] J. C. Heigel, P. Michaleris, and E. W. Reutzel, "Thermo-mechanical model development and validation of directed energy deposition additive manufacturing of Ti–6Al–4V," *Addit. Manuf.*, vol. 5, pp. 9–19, Jan. 2015.

[21] P. Michaleris, "Modeling metal deposition in heat transfer analyses of additive manufacturing processes," *Finite Elem. Anal. Des.*, vol. 86, pp. 51–60, 2014.

[22] E. R. Denlinger and P. Michaleris, "Effect of stress relaxation on distortion in additive manufacturing process modeling," *Addit. Manuf.*, vol. 12, pp. 51–59, 2016.

[23] J. Ding *et al.*, "Thermo-mechanical analysis of Wire and Arc Additive Layer Manufacturing process on large multi-layer parts," *Comput. Mater. Sci.*, vol. 50, no. 12, pp. 3315–3322, 2011.

[24] T. R. Walker, C. J. Bennett, T. L. Lee, and A. T. Clare, "A validated analytical-numerical modelling strategy to predict residual stresses in single-track laser deposited IN718," *Int. J. Mech. Sci.*, vol. 151, pp. 609–621, Feb. 2019.

[25] S. M. Kelly, S. S. Babu, S. A. David, T. Zacharia, and S. L. Kampe, "A thermal and microstructure model for laser deposition of Ti-6Al-4V," in *Cost - Affordable Titanium*, 2004, pp. 45–52.

[26] M. Söderberg, A. Lundbaćk, and L. E. Lindgren, "Modeling of metal deposition," *Finite Elem. Anal. Des.*, vol. 47, pp. 1169–1177, 2011.

[27] R. G. Ding, Z. W. Huang, H. Y. Li, I. Mitchell, G. Baxter, and P. Bowen, "Electron microscopy study of direct laser deposited IN718," *Mater. Charact.*, vol. 106, pp. 324–337, 2015.

[28] D. Liu, J. C. Lippold, J. Li, S. R. Rohklin, J. Vollbrecht, and R. Grylls, "Laser engineered net shape (LENS) technology for the repair of Ni-base superalloy turbine components," *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.*, vol. 45, no. 10, pp. 4454–4469, 2014.

[29] Z. Liu, H. Kim, W. Liu, W. Cong, Q. Jiang, and H. Zhang, "Influence of energy density on macro/micro structures and mechanical properties of as-deposited Inconel 718 parts fabricated by laser engineered net shaping," *J. Manuf. Process.*, vol. 42, pp. 96–105, Jun. 2019.

[30] S. M. Thompson, L. Bian, N. Shamsaei, and A. Yadollahi, "An overview of Direct Laser Deposition for additive manufacturing; Part I: Transport phenomena, modeling and diagnostics," *Addit. Manuf.*, vol. 8, pp. 36–62, Oct. 2015.

[31] E. R. Denlinger, V. Jagdale, G. V. Srinivasan, T. El-Wardany, and P. Michaleris, "Thermal modeling of Inconel 718 processed with powder bed fusion and experimental validation using in situ measurements," *Addit. Manuf.*, vol. 11, pp. 7–15, 2016.

[32] D. Dye, B. A. Roder, S. Tin, M. A. Rist, J. A. James, and M. R. Daymond, "Modeling and measurement of residual stresses in a forged IN718 superalloy disc," *Proc. Int. Symp. Superalloys*, pp. 315–322, 2004.

[33] A. P. Chakravarti, J. A. Goldak, and A. S. Rao, *Thermal Analysis of Welds.*, vol. 2. 1985.

[34] R. R. Unocic and J. N. DuPont, "Process Efficiency Measurements in the Laser Engineered Net Shaping Process," *Metall. Mater. Trans. B Process Metall. Mater. Process. Sci.*, vol. 35, no. 1, pp. 143–152, 2004.

[35] Y. Ono, T. Yuri, N. Nagashima, H. Sumiyoshi, T. Ogata, and N. Nagao, "High-cycle fatigue properties of Alloy718 base metal and electron beam welded joint," *Phys. Procedia*, vol. 67, pp. 1028–1035, 2015.

[36] W. Di Cao, "Solidification and solid state phase transformation of Allvac® 718Plus™ alloy," *Proc. Int. Symp. Superalloys Var. Deriv.*, pp. 165–177, 2005.

[37] Y. Zhang, Z. Li, P. Nie, and Y. Wu, "Effect of heat treatment on niobium segregation of laser-cladded IN718 alloy coating," *Metall. Mater. Trans. A Phys. Metall. Mater. Sci.*, vol. 44, no. 2, pp. 708–716, 2013.

[38] S. Azadian, L.-Y. Wei, and R. Warren, "Delta phase precipitation in Inconel 718," *Mater. Charact.*, vol. 53, no. 1, pp. 7–16, Sep. 2004.

[39] J. F. Radavich, "The Physical Metallurgy of Cast and Wrought Alloy 718," in *Superalloys 718 Metallurgy and Applications*, Vol. 1989., TMS, 1989, pp. 229–240.

[40] Y. Desvallees, M. Bouzidi, F. Bois, and N. Beaude, "Delta Phase in INCONEL 718: Mechanical Properties and Forging Process Requirements," pp. 281–291, 2012.

[41] Y. Zhang, L. Yang, T. Chen, W. Zhang, X. Huang, and J. Dai, "Investigation on the optimized heat treatment procedure for laser fabricated IN718 alloy," *Opt. Laser Technol.*, vol. 97, pp. 172–179, 2017.

[42] Y. Zhang, L. Yang, W. Lu, D. Wei, T. Meng, and S. Gao, "Microstructure and elevated temperature mechanical properties of IN718 alloy fabricated by laser metal deposition," *Mater. Sci. Eng. A*, vol. 771, no. October 2019, p. 138580, Jan. 2020.

[43] A. Segerstark, J. Andersson, L.-E. Svensson, and O. Ojo, "Microstructural characterization of laser metal powder deposited Alloy 718," *Mater. Charact.*, vol. 142, pp. 550–559, Aug. 2018.

[44] F. Liu *et al.*, "Microstructural changes in a laser solid forming Inconel 718 superalloy thin wall in the deposition direction," *Opt. Laser Technol.*, vol. 45, pp. 330–335, Feb. 2013.

[45] Y. Chen *et al.*, "Study on the element segregation and Laves phase formation in the laser metal deposited IN718 superalloy by flat top laser and gaussian distribution laser," *Mater. Sci. Eng. A*, vol. 754, pp. 339–347, Apr. 2019.

[46] S. Sui, J. Chen, X. Ming, S. Zhang, X. Lin, and W. Huang, "The failure mechanism of 50% laser additive manufactured Inconel 718 and the deformation behavior of Laves phases during a tensile process," *Int. J. Adv. Manuf. Technol.*, vol. 91, no. 5–8, pp. 2733–2740, Jul. 2017.

[47] F. Liu *et al.*, "The effect of laser scanning path on microstructures and mechanical properties of laser solid formed nickel-base superalloy Inconel 718," *J. Alloys Compd.*, vol. 509, no. 13, pp. 4505–4509, Mar. 2011.

[48] Y. Zhang, L. Yang, J. Dai, Z. Huang, and T. Meng, "Grain growth of Ni-based superalloy IN718 coating fabricated by pulsed laser deposition," *Opt. Laser Technol.*, vol. 80, pp. 220–226, 2016.

[49] G. P. Dinda, A. K. Dasgupta, and J. Mazumder, "Texture control during laser deposition of nickel-based superalloy," *Scr. Mater.*, vol. 67, no. 5, pp. 503–506, Sep. 2012.

Streaming Communication Component for Multi-Robots

George Oliveira, Luana D. Fronza, Luiza Medeiros, Patricia D. M. Plentz

***Abstract*—**The research presented in this article is part of a wide project that proposes a scheduling system for multi-robots in intelligent warehouses employing Multi-Robot Path-Planning (MPP) and Multi-Robot Task Allocation (MRTA) to reconcile multiple restrictions (task delivery time, task priorities, charging capacity and robots battery capacity). We present the software component capable of interconnecting an open streaming processing architecture and ROS (Robot Operating System), ensuring communication and message exchange between robots and the environment in which they are inserted. Simulation results show good performance of our proposed technique for connecting ROS and streaming platforms.

***Keywords*—**Complex distributed systems, mobile robots, smart warehouses, streaming platforms.

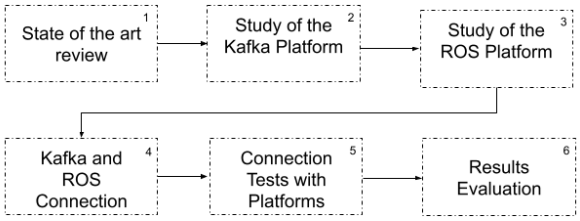
I. Introduction

MULTI-ROBOT systems are groups of robots designed to perform complex tasks, which are unattainable or have a high cost, if they are performed by a single robot. This type of system needs to respect restrictions imposed upon it by the environment and/or by the infrastructure limitations (both physical and in terms of software). In this research area, two themes have fundamental importance: path planning (MPP) and task allocation (MRTA). MPP consists of finding the best path between two positions, while MRTA allocates tasks to a set of robots while taking restrictions (like task priority or maximum load) into account, and reaching one or more goals (time optimization or energy consumption, for example) [1].

MPP and MRTA are used in various scenarios, one of said scenarios being smart warehouses. Hundreds of thousands of products are stored in smart warehouses, and need to be transferred from one location to another. These environments have automated control of product entry and exit, being completely operated by robot teams [2]. The purchase orders arrive at these warehouses through computational platforms that are capable of processing large volumes of data. Data streaming platforms allow for continuous processing of orders and are thoroughly used in smart warehouses of large shopping companies, like Amazon and Alibaba. Besides managing orders, these systems can also manage the environment that is receiving said orders, and its temperature, lighting, motion tracking, and others. These companies have their own proprietary systems that join the concepts of MPP, MRTA, and data stream processing. Since these systems are proprietary, the scientific community does not have access to them, rendering research and development of advanced solutions impossible.

The literature shows that most of the works implement MPP and MRTA separately in the solutions for task allocation [1], [3], [4]. However, smart warehouses require the use of MPP and MRTA simultaneously, to meet the many restrictions involved with these environments. Algorithms of MPP and MRTA are implemented in robotic computational systems, like the ROS [5]. ROS is an open source platform that gives support to the creation of robotic systems, with emphasis on the development of algorithms for locomotion of mobile robots in an environment, and the execution of tasks [6]. The platform offers packs that abstract sensors, actuators, and message exchange mechanisms between robots. However, these systems do not offer communication with data streaming platforms, necessary in smart warehouses for the management of orders and of the warehouse itself.

The research presented in this article is part of a wider project that proposes a scheduler system for the multi-robot in smart warehouses, using MPP and MRTA to join multiple restrictions (task deadlines, task priority, cargo capacity, and robot battery). In this article, we present a software component capable of joining an open streaming processing architecture and ROS, guaranteeing the communication and message exchange between robots and the environment, in which they are. To reach the general goal, the following specific goals were set: (1) investigate the use of an open data streaming platform to receive the data from the environment and the robot team and processing of this data; (2) proposition and development of a technique to connect this platform to the ROS robotic system; (3) simulation of the proposed component, comparing it with existing solutions. To fulfill the research goal, the following investigation and development steps were carried out as shown in Fig. 1.



George Oliveira and Patricia D. M. Plentz are with the Graduate Program in Computer Science, Federal University of Santa Catarina, Florianopolis, Brazil (e-mails: george.oliveira@posgrad.ufsc.br and patricia.plentz@ufsc.br).

Luana D. Fronza and Luiza Medeiros are with the Informatics and Statistics Department, Federal University of Santa Catarina, Florianopolis, Brazil (e-mails: fronza.lu, luizadmeds@gmail.com).

This study was financed in part by the Coordination for the Improvement of Higher Education Personnel - Brazil (CAPES) - Finance Code 001 and CAPES/PRINT - Call no. 41/2017 Senior Visiting Professor at Biomimetics and Intelligent Systems Group - BISG, University of Oulu, Finland

Fig. 1 Search and Development Steps

First, a state-of-the-art review was initiated to locate the project and clarify its contribution to the scientific community. Next, a study was carried out on Kafka [7], an Apache project for distributed processing of data streams that large distributed applications, such as Uber, Spotify, Netflix, Twitter, Linkedin and PayPal use to process requests. With the investigation of how the platform functions, we validated the idea that this would be an interesting computational tool for the processing of large volumes of data, as required in the management of mobile robot teams.

The next step was a study about the ROS platform, and which of its libraries and tools would be necessary for the following stage. After all these studies had been carried out, we investigated a way of connecting these two tools, performing a reading of their documentation. The Kafka platform has a connection API; that is, a software structure that allows reusable or standard integrations for flow events inside and outside external systems, such as databases. The ROS platform consists of several software packages, which can be created to perform numerous tasks, and among them, a connection between robotic systems and other agents. Thus, the plan of creating a ROS package (or enhancing an existing one) that worked in conjunction with the Kafka Connection API was defined, so that the tests of exchanging messages between platforms could then be carried out. In the next sections we show the enhancing ROS package developed in this research.

This paper is organized in the following way: in this section, we presented a general context view of this research and the general and specific goals. In Section II, we bring the works related to the theme of this research. We briefly explain how the proposed MPP-MRTA architecture of mobile robots is organized in Section III. Our contribution is described in Section IV, where the proposed algorithms for connecting the platforms are explained. In Section V, we present the tests that were carried out to prove the efficiency of the connection algorithm. Finally, in Section VI is our conclusion about the work, in addition to an overview of future works.

II. Background

A. ROS – Robot Operating System

ROS is an open source software architecture to develop robotic applications. It has a set of libraries and it is composed by conventions and tools that simplify the creation of complex and robust robot behavior through a wide variety of robotic platforms [5]. ROS core components are arranged in communications infrastructure, robot-specific features and tools. The first one, communications infrastructure, is responsible to deliver some facilities as publish/subscribe anonymous message passing, recording and playback of messages and request/response remote procedure calls. Moreover, ROS enables configuration information sharing through global-key value store. Robot-specific features embrace libraries and tools which provide standard robot messages, robot geometry library, robot description language and others capabilities. Concerning tools, command-line and rviz (three-dimensional visualization of many sensor data types) receive highlights besides the rqt, a Qt-based framework for developing graphical interfaces for robots.

B. Kafka

Kafka is a computational platform that captures streams of events in real-time from sources like mobile devices, sensors, as well as cloud and databases services [7]. Also like ROS, Kafka is a publisher/subscriber system that stores and processes event streams as they occur or retrospectively.

Kafka is a distributed system highly scalable and fault-tolerant, composed of servers and clients that communicate via a high-performance TCP network protocol. It could be integrated with others systems through Kafka Connect API. Usually, these APIs evolve a series of complex operations to integrate large distributed systems. That is why bridge codes are developed to reduce these complexities.

III. Related Works

Bridge codes for integrating different and complex distributed systems are necessary to reduce the effort and time spent on this kind of task. Few published works approach the problems related to this task. In fact, the major platforms use proprietary solutions to integrate distributed robotic systems platforms.

In [8], the authors propose a robot software bridge for interconnecting OPRoS [9] with ROS without modification of these source codes. They implemented the proposed bridge and evaluated it on the remote robot control application based on those systems. To demonstrate the feasibility of the robot software bridge, the bridge prototype was developed in C++ and constructed as a testbed. The authors do not comment on difficulties related to this proposed solution developed through a complex programming language like C++, which raises challenges for the solution's scalability.

ROS2 based communication architecture is presented in [10] together with an industrial use-case of a collaborative and intelligent automation system. The authors describe the communication layer in ROS and ROS2 but not emphasize the bridge code for connecting those systems with a streaming data system, like Kafka. They fail to address how to connect robotic platforms with complex distributed systems that deal with large amounts of external data.

The main contribution of [11] is designing an architecture that embraces edge, fog and cloud concepts in the context of IoT applications. It focuses on delivering services for vehicle driving while gathering and processing data. A fog computing server model that generates a high definition (HD) map node of the ROS environment is used to connect ROS and Kafka. Although the paper describes a highly important topic, the bridge code to connect ROS and Kafka is not explained.

To the present date, we have not found articles that propose a software component to join MPP, MRTA and data streaming algorithms coming in from a monitored environment, and computational systems for managing large volumes of orders. Considering this kind of research highly relevant for the

improvement of the warehouse's performance, this paper presents a proposal to contribute to the scientific mobile robotic community. In the next section, we present the proposed architecture that integrates MPP and MRTA. After that, we describe a software component to interconnect MPP, MRTA, and data streaming for smart warehouses.

IV. MPP-MRTA Architecture for Mobile Robots

Software architecture for mobile robots in smart warehouses should integrate a lot of information from the environment, robots and orders that arrive in the system. A way to do this integration is to join MRTA and MPP strategies with data streaming systems to reach better results in terms of increasing the number of fulfilled tasks.

The distributed robotic software used in this paper is composed of six modules, as shown in Fig. 2. The main execution cycle of this architecture begins in the Orders module where the products lists are received from the data streaming system. These lists are sent to the Task Organizer module, which will break it in tasks. The Cost Estimator module receives data from three different sources: Task Organizer module, Monitor and Execution Profiles. With these data, it estimates the cost for each robot-task pair. The Task Allocator module to define a set of actions that each robot must perform uses this value. This module also receives information about robots and tasks. The first information comes from the Monitor and the second one comes from the Task Organizer. The Path Planner module receives this set of actions, computes routes for each robot and delivers a route to each robot.

Each robot receives its route and tries to execute. If the route is executed with success, the robot finishes its task. On the other hand, if some problem occurs, then the robot triggers a message to the Monitor, which will check the Robot Failure Profile (RFP) to detect the problem and define some solution based on the Execution Profiles.

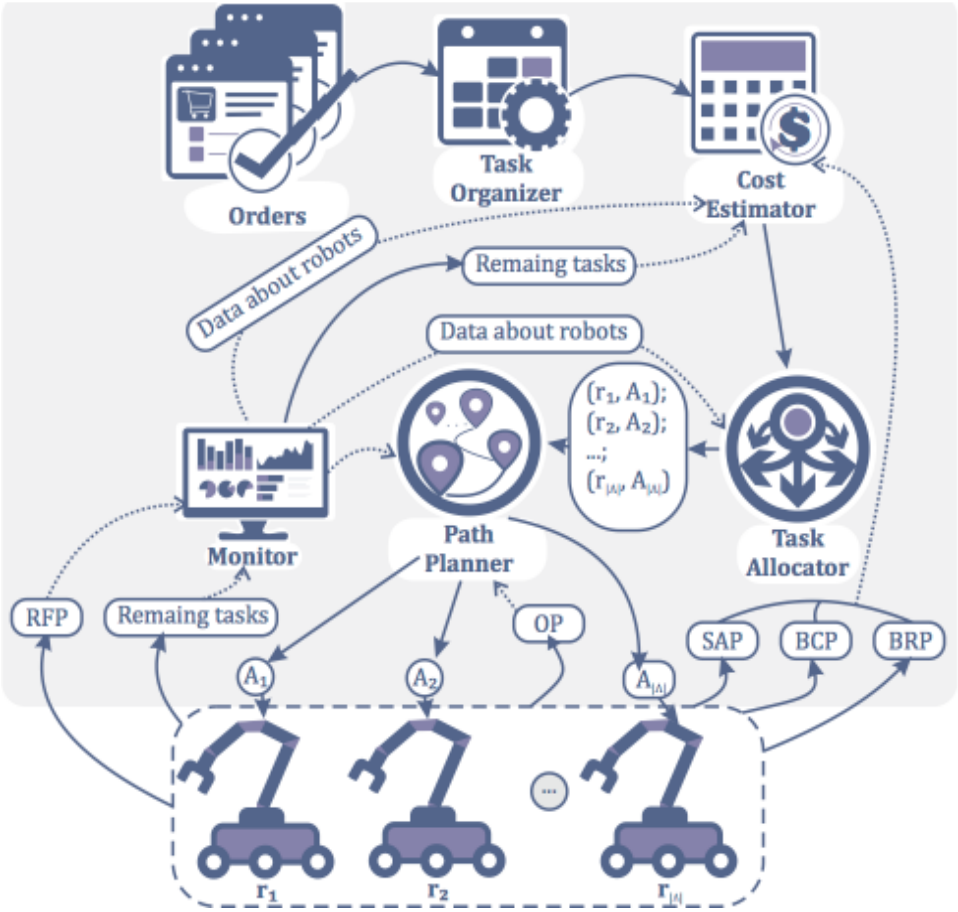


Fig. 1 Proposed architecture for integrating MPP and MRTA

The Orders module refers to a list of products itself. We assume that the proposed architecture is integrated with the warehouse order processor. As soon as a purchase is finished, the corresponding order enters into a queue to be processed by the Scheduler System. Here, we highlight the proposed software component capable of interconnecting an open streaming processing architecture (a warehouse order processor) and ROS, ensuring communication and message exchange between robots and the environment in which they are inserted. This component will be described in the next section.

The Task Organizer receives several tasks at a time. It creates a task for each processed product. Each task is associated with a single product, having information about the location, height

on the shelf, and the total weight to be carried. Recurrent products in different orders are mapped into the same task.

V. Development of the Bridge Connector

The proposed software component for interconnecting a warehouse order processor and ROS to provide communication and message exchange between mobile robots in this kind of environment was developed using available API at Kafka and ROS platforms. It was used a basis code of an existing algorithm that performed the task of bridging the two platforms [12]. We improved this code and developed other algorithms that simulate the production and processing of orders to start performing the initial tests. This source code (also called a bridge) became the main connection technique between ROS and Kafka needed for the project. Algorithm 1 presented in Fig. 3 shows the beginning of the function that makes the connection between the two platforms. On line 3 to line 6, the parameters passed during algorithm execution are recognized. On line 7, a *KafkaConsumer* is created regarding a specific topic, which will start the consumption of Kafka messages linked to this topic. After this connection has been made, the import of the type of message occurs, which will be sent throughout the platforms (line 8). This code initially supports messages of the *std_msgs/String, geometry_msgs/Twist* and *sensor_msgs/Image* type, but there is also a possibility of inserting new types. Line 9 shows ROS and Kafka are connected, creating a *ROS Publisher*, which receives the message, consumed by the *KafkaConsumer* and publishes it on the related ROS topic.

Algorithm 1 ros_publish

```
1: rospy.init_node("ros_publish")
2: rospy.on_shutdown("self.shutdown")
3: rospy.get_param("localhost:9092")
4: rospy.get_param("ros_topic")
5: rospy.get_param("kafka_topic")
6: rospy.get_param("msg_type")
7: Consumer = KafkaConsumer(params)
8: msg_func = import_msg_type(msg_type.param)
9: Publisher = rospy.Publisher(topics params)
```

Fig. 3 Parameter definitions and creation of a *KafkaConsumer*

An important part of the connection algorithm is the task of converting the received and sent messages to String format, and calling the function that publishes the message to the ROS topic, passing the parameter as the message already in its correct format, as shown by Algorithm 2 presented in Fig. 4.

Line 2 to line 7 of Algorithm 2 show that for each message that arrives in the streaming system, a string JSON (JavaScript Object Notation[1]) is created and a JSON message converter is applied in this string along with the type of message. The converted message is published in the corresponding topic allowing the other modules of the system to use this message.

Algorithm 2 ros_message

```
1: while not rospy.is_shutdown():
2:   for msg in self.consumer:
3:     json_str = json.dumps(msg.value)
4:     ros_msg =
       json_message_converter.convert_json_to_ros_
       message(msg_type, json_str)
5:     self.publisher.publish(ros_msg)
6:   end for
7: end while
```

Fig. 4 Function that publishes the message to the ROS topic

VI. Simulation Results

To test the aforementioned code, a python simulation algorithm was written, which creates messages using Kafka modules, receives these messages in real time, and processes them using the ROS modules. The processing code compiles tasks between the available robots in the warehouse, basing itself on how many items need to be retrieved and taken to the unloading area. It also organizes the order in which these tasks must be executed, taking into account the priority of messages that are received on the ROS topic.

First, an algorithm was created which simulates the production of orders in real time. This algorithm produces a message to each topic, and delivers it to the *kafka_topic*. Turtlesim simulator [13] was used, with the goal of defining which parameters should be taken into account, and what information a message regarding the order should contain. The following items are defined as essential information: number, status, order priority, items contained in the order, and, for each item, a number, status, and its coordinates inside the warehouse. The coordinates of the order-unloading place, and of the robot's initial location also were taken into account.

When all of the simulation code is executed, the result is what is registered in Fig. 5. First, the messages are created, published to a Kafka topic, and copied to the ROS topic through the code that joins the platforms. After that, these orders are organized into priority lines, in which the most urgent orders are placed at the beginning, and the normal orders are placed towards the end, so that the orders can be sent, one by one, to the robots in the warehouse. Finally, after task division and allocation between available robots, each one executes its task, simultaneously to the system.

To integrate two complex distributed systems like ROS and Kafka is not a trivial task. Because they can be applied to a lot of different applications, each system has several modules and configurations that difficult the development of new solutions for integrated them. There are implementations that deliver a more direct access to the real-time publish-subscribe (RTPS) wire transfer protocol, but this kind of solution increases the complexity of the entire system. This paper proposes a good

[1] JSON is a lightweight format for storing and transporting data.

solution for integrate data streaming with a distributed robotic architecture in a simple and clear way. Others tests will be developed mainly considering real mobile robots.

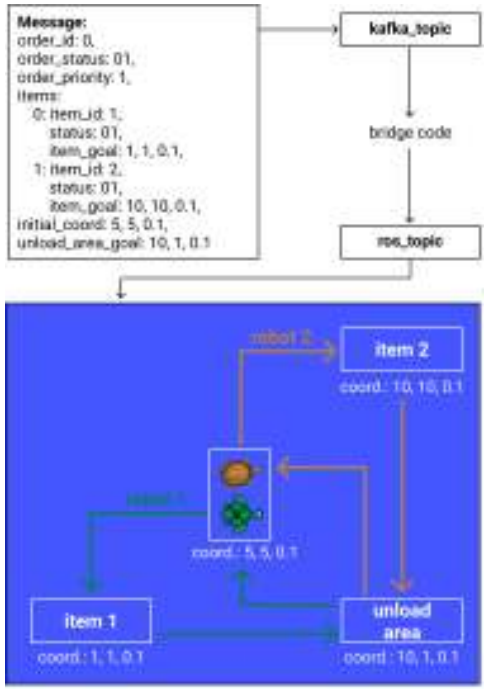


Fig. 5 Simulation done with Turtlesim simulator

VII. Conclusions and Future Works

Using Kafka as a data streaming platform and ROS platform to employ MPP and MRTA in smart warehouses is an important field of research in the distributed systems context. However, it is a challenge to use ROS in such systems because of communications difficulties [10]. A few ROS based frameworks try to handle these communication problems. In this paper, we present a simple bridge code, which reaches adequate and satisfactory behavior. Indeed, besides giving us a way to automate the everyday tasks of a warehouse and its ongoing activities, the algorithm presented also allows us to divide tasks according to priority and robot availability. This implementation considered two order priorities (urgent and normal orders). The next step is to design a comprehensive set of priorities taking into account the large variety of products and their specificities. Moreover, it is important to incorporate environmental factors into the algorithm to monitor things like temperature and humidity inside the warehouse and have a more powerful MPP system guided by sensors to measure the distance between the robots and their surroundings. These directly affect the way the products are organized to be delivered to the robots.

References

[1] B. B. Choudhury and B.B. Biswal (2008). An optimized multi robot task allocation. In International Conference on Emerging Trends in Engineering and Technology, pages 320–325, 2008.

[2] Zhi Li, Ali Vatankhah Barenji, Jiazhi Jiang, Ray Zhong, and Gangyan Xu (2018). A mechanism for scheduling multi robot intelligent warehouse system face with dynamic demand. Journal of Intelligent Manufacturing, pages 1–12, 2018.

[3] H. Mitiche, D. Boughaci, and M. Gini (2015). Efficient heuristics for a time-extended multi-robot task allocation problem. In International Conference on New Technologies of Information and Communication, 2015.

[4] Sengang Ye, Haiping Ma, Sheng Xu, Wenqiang Yang, and Minrui Fei (2015). An effective fireworks algorithm for warehouse-scheduling problem. Transactions of the Institute of Measurement and Control, 39:75–85, 08 2015.

[5] ROS - Robot Operational System (2021). Available at www.ros.org.

[6] Dongning Rao, Guodong Hu, and Zhihua Jiang (2020). Probplan: A framework of integrating probabilistic planning into ROS. IEEE Access, 8:1–3, 06 2020.

[7] Kafka – A Distributed Streaming Platform (2021). Available at https://kafka.apache.org

[8] Kang, Jeong & Yu, Dong & Park, Hong. (2012). A robot software bridge for interconnecting OPRoS with ROS. 296-297. 10.1109/URAI.2012.6462998

[9] OPRoS (2021). Available: http://www.opros.or.kr

[10] Endre Eros, Martin Dahl, Kristofer Bengtsson, Atieh Hanna, Petter Falkman, (2019). A ROS2 based communication architecture for control in collaborative and intelligent automation systems, Procedia Manufacturing, Volume 38, Pages 349-357, ISSN 2351-9789, https://doi.org/10.1016/j.promfg.2020.01.045.

[11] Lee, Junwon; Lee, Kieun; Yoo, Aelee; Moon, Changjoo. 2020. "Design and Implementation of Edge-Fog-Cloud System through HD Map Generation from LiDAR Data of Autonomous Vehicles" *Electronics* 9, no. 12: 2084. https://doi.org/10.3390/electronics9122084

[12] ROS Package: ros_kafka_connector (2021). Available at https://github.com/isakdiaz/ros-kafka-connector.

[13] Turtlesim Simulator (2021). Available at http://wiki.ros.org/turtlesim

On quasi-LPV System Identification with (deep) Recurrent Neural Networks

Alexander Rehmer[1] and Andreas Kroll[2]

Abstract—This paper presents an advancement in the identification of affine quasi-LPV systems using Recurrent Neural Networks (RNNs). The proposed approach is based on a structured RNN developed by Lachhab et al. [1]. Through the extension of the original model by so-called *gates*, structures which are found in Gated Units such as the Long short-term memory (LSTM) network [2] or the Gated Recurrent Unit (GRU) [3], the models approximation capability is increased, while still preserving the affine quasi-LPV structure, which allows the model to be directly employed in polytopic LPV controller synthesis. The performance of the proposed approach is demonstrated by comparison with two other RNN-based approaches for affine quasi-LPV identification on two nonlinear system identification benchmark problems.

Keywords—Linear Parameter Varying Systems, Machine Learning, Nonlinear System Identification, Recurrent Neural Networks.

I. Introduction

THE linear parameter varying (LPV) framework has proven to be a suitable tool for the modeling and control of high complex nonlinear systems [4, 1]. Among all synthesis approaches the polytopic approach is the most popular [5]. However, the polytopic synthesis approach is limited in that it requires an affine/polytopic LPV represenation of the plant and the implementation complexity of polytopic LPV controllers grows exponentially with the dimension of the time-varying affine parameter vector, also called scheduling-vector [5]. How to arrive at an aLPV description of a nonlinear plant is in itself non-trivial. If a physical model is available, procedures for automatically deriving an affine quasi-LPV representation exist [6, 7], but often result in a large number of scheduling variables. This in turn necessitates the application of scheduling dimension reduction (SDR) techniques [8], which may fail to preserve the dynamics of the original model while reducing the number of scheduling parameters. Another approach is to identify an affine quasi-LPV model directly from input-output data, with the objective of keeping the dimension of the time-varying parameter vector as little as possible. If the time-varying parameters of the LPV system are assumed to be unknown and depend on the input and/or the state, the task of parameter estimation amounts to solving a nonlinear optimization problem. If no knowledge regarding the nonlinear behavior is present, a black-box approach such as an artificial neural network (ANN) may be employed. As is the case with all model structures, ANNs can be differentiated into the external and the internal dynamics approach [9]. In the external dynamics approach a static approximator, such as a feedforward ANN, is provided with past inputs and past outputs. Methods for extracting an aLPV representation from a feedforward ANN were e.g. shown in [10] and the references therein. The external dynamics approach can however be susceptible to noisy inputs and outputs and the combinatorial problem of choosing the appropriate number of lags for input and output, especially for multiple-input and multiple-output (MIMO) systems, can be substantial. To circumvent these shortcomings the internal dynamics approach may be employed, which is usually realized via Recurrent Neural Networks (RNNs). To the authors best knowledge, there are mainly two approaches for identifying affine quasi-LPV models from data using RNNs: RNNs which nonlinearly interpolate between multiple local linear models, and a specially structured RNN (S-RNN) developed in [11] for which a systematic method for deriving an affine quasi-LPV representation was presented in [1]. Based on the latter approach, a deep RNN architecture for affine quasi-LPV system identification, will be developed via using the concept of *gates*. The proposed deep RNN is potentially capable to approximate nonlinear systems with fewer time-varying parameters compared to the two other approaches, which will be demonstrated on two case studies.

[1]Alexander Rehmer is with the Department of Measurement and Control, University of Kassel, Kassel, Germany alexander.rehmer@mrt.uni-kassel.de

[2]Andreas Kroll is the Head of the Department of Measurement and Control, University of Kassel, Kassel, Germany andreas.kroll@mrt.uni-kassel.de

II. Preliminaries

A linear parameter-varying (LPV) model $\mathcal{P}(\boldsymbol{\rho}_k)$ is a state space model whose matrices depend on a time-varying parameter vector $\boldsymbol{\rho}_k \in \mathbb{R}^{n_\rho}$. A general LPV plant can be written as

$$\begin{bmatrix} \boldsymbol{x}_{k+1} \\ \boldsymbol{y}_k \end{bmatrix} = \begin{bmatrix} \mathcal{A}(\boldsymbol{\rho}_k) & \mathcal{B}(\boldsymbol{\rho}_k) \\ \mathcal{C}(\boldsymbol{\rho}_k) & \mathcal{D}(\boldsymbol{\rho}_k) \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{u}_k \end{bmatrix} \tag{1}$$

with the state $\boldsymbol{x}_k \in \mathbb{R}^{n_x}$, input $\boldsymbol{u}_k \in \mathbb{R}^{n_u}$ output $\boldsymbol{y}_k \in \mathbb{R}^{n_y}$ and matrix-valued functions $\mathcal{A}(\boldsymbol{\rho}_k) : \mathbb{R}^{n_\rho} \to \mathbb{R}^{n_x \times n_x}$, $\mathcal{B}(\boldsymbol{\rho}_k) : \mathbb{R}^{n_\rho} \to \mathbb{R}^{n_x \times n_u}$, $\mathcal{C}(\boldsymbol{\rho}_k) : \mathbb{R}^{n_\rho} \to \mathbb{R}^{n_y \times n_x}$, $\mathcal{D}(\boldsymbol{\rho}_k) : \mathbb{R}^{n_\rho} \to \mathbb{R}^{n_y \times n_u}$, which encode the dependence of the system matrices on the time-varying parameter vector $\boldsymbol{\rho}_k$. $\boldsymbol{\rho}_k$ is assumed to be an unknown function of $[\boldsymbol{u}_k, \boldsymbol{x}_k]$ and therefore encodes the nonlinear behavior of the plant. In the general LPV representation (1) the system matrices can be arbitrary nonlinear functions of $\boldsymbol{\rho}_k$, which makes system identification and controller synthesis very difficult. Hence, often LPV models with *affine* parameter dependence are considered as a special case of (1):

$$\begin{bmatrix} \boldsymbol{x}_{k+1} \\ \boldsymbol{y}_k \end{bmatrix} = \left(\begin{bmatrix} \boldsymbol{A}_0 & \boldsymbol{B}_0 \\ \boldsymbol{C}_0 & \boldsymbol{D}_0 \end{bmatrix} + \sum_{i=1}^{n_\theta} [\boldsymbol{\theta}_k]_i \begin{bmatrix} \boldsymbol{A}_i & \boldsymbol{B}_i \\ \boldsymbol{C}_i & \boldsymbol{D}_i \end{bmatrix} \right) \begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{u}_k \end{bmatrix} \tag{2}$$

$\boldsymbol{\theta}_k = \boldsymbol{\theta}_k(\boldsymbol{x}_k, \boldsymbol{u}_k)$ is the *affine* time-varying parameter vector. From the affine LPV representation a polytopic representation

can be constructed. In a polytopic representation the affine parameter vector $\boldsymbol{\theta}_k$ varies within a convex polytope with vertices $\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_N$, such that the affine parameters result from the convex combination of these N vertices:

$$\boldsymbol{\theta}_k \in Co\left\{\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_N\right\} \coloneqq \alpha_1 \boldsymbol{\omega}_1 + \cdots + \alpha_N \boldsymbol{\omega}_N \tag{3}$$

$\alpha_i \in \mathbb{R}$ denote the barycentric (or polytopic) coordinates, which are non negative and normalized to sum up to 1:

$$\begin{aligned} \sum_{i=1}^{N} \alpha_i(\boldsymbol{\theta}_k) = 1, \quad \alpha_i(\boldsymbol{\theta}_k) \geq 0 \quad \forall i = 1, \ldots, N \\ \forall \boldsymbol{\theta}_k \in Co\left\{\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_N\right\} \end{aligned} \tag{4}$$

The affine system (2) can then be written as a convex combination of the *frozen* linear time invariant (LTI) systems at the vertices of the polytope:

$$\begin{bmatrix} \boldsymbol{x}_{k+1} \\ \boldsymbol{y}_k \end{bmatrix} = \sum_{i=1}^{N} \alpha_i(\boldsymbol{\theta}_k) \begin{bmatrix} \mathcal{A}\left(\boldsymbol{\theta}_k\right) & \mathcal{B}\left(\boldsymbol{\theta}_k\right) \\ \mathcal{C}\left(\boldsymbol{\theta}_k\right) & \mathcal{D}\left(\boldsymbol{\theta}_k\right) \end{bmatrix} \Bigg|_{\boldsymbol{\theta}_k = \boldsymbol{\omega}_i} \begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{u}_k \end{bmatrix} \tag{5}$$

On how to derive the barycentric coordinates from the affine parameter $\boldsymbol{\theta}_k$ see for example [5, S. 324]. The purpose of this paper is to provide an identification approach for affine/polytopic LPV models, which can directly be utilized for polytopic controller synthesis.

III. System identification of affine quasi LPV Systems

In this section, the two main approaches for the identification of affine quasi-LPV models are shortly revisited and their mayor advantages and shortcomings are pointed out. Based on this analysis a model which allows for the incorporation of deep neural network structures in affine quasi-LPV system identification is proposed.

A. Radial Basis Function Recurrent Neural Network (RBF-RNN)

The approach of approximating nonlinear systems via a weighted combination of local linear models has been successfully applied in countless engineering applications, see [12] and the references therein. A local linear state-space model can be written as

$$\begin{bmatrix} \boldsymbol{x}_{k+1} \\ \boldsymbol{y}_k \end{bmatrix} = \sum_{i=1}^{n_p} p_i(\boldsymbol{\phi}) \begin{bmatrix} \boldsymbol{A}_i & \boldsymbol{B}_i \\ \boldsymbol{C}_i & \boldsymbol{D}_i \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_k \\ \boldsymbol{u}_k \end{bmatrix} \tag{6}$$

with $\boldsymbol{A}_i \in \mathbb{R}^{n_x \times n_x}$, $\boldsymbol{B}_i \in \mathbb{R}^{n_x \times n_u}$, $\boldsymbol{C}_i \in \mathbb{R}^{n_y \times n_x}$ and $\boldsymbol{D}_i \in \mathbb{R}^{n_y \times n_u}$. $\boldsymbol{A}_i$, $\boldsymbol{B}_i$, $\boldsymbol{C}_i$ and $\boldsymbol{D}_i$ can be interpreted as local linearizations of the nonlinear plant and $p_i : \mathbb{R}^{n_\phi} \to \mathbb{R}$ are the weighting functions of the local models. $\boldsymbol{\phi}_k \in \mathbb{R}^{n_\phi}$ is assumed to be unknown and usually comprises $\boldsymbol{x}_k$ and $\boldsymbol{u}_k$. The model in (6) is equivalent to the widely used Takagi-Sugeno (TS) fuzzy model, where the weighting functions are interpreted as fuzzy membership functions. It is common to define p_i as normalized radial basis functions (RBF) with centers $\boldsymbol{c_i}$ and widths w_i

$$p_i = \frac{e^{-(\boldsymbol{\phi}_k - \boldsymbol{c}_i)^T \operatorname{diag}(\boldsymbol{w}_i)(\boldsymbol{\phi}_k - \boldsymbol{c}_i)}}{\sum_{i=1}^{n_\theta} e^{-(\boldsymbol{\phi}_k - \boldsymbol{c}_i)^T \operatorname{diag}(\boldsymbol{w}_i)(\boldsymbol{\phi}_k - \boldsymbol{c}_i)}}. \tag{7}$$

Then (6) can also be considered a special structured Recurrent Neural Network with radial basis activation functions. Defining the weighting functions p_i as the affine time varying parameters θ_i one arrives at an affine quasi-LPV representation. The main advantages of the local linear approach are, that is has been shown to be universal approximators and that they can be sensibly initialized: By uniformly distributing the RBFs over the scheduling space and initializing all local models as a globally identified linear model (e.g. subspace) the initial nonlinear model equals the linear model [12]. Major drawbacks are, that many local models might be necessary for modeling highly nonlinear systems, which corresponds to a highly dimensional parameter vector $\boldsymbol{\theta}_k$. Also the number of model parameters scales badly with $\dim(\boldsymbol{\theta}_k)$, since increasing $\dim(\boldsymbol{\theta}_k)$ by one corresponds to adding a fully parameterized linear state-space model.

B. Structured RNN (S-RNN) for affine quasi-LPV system identification

The model equations of the structured RNN for quasi-LPV identification developed by Lachhab et al. [1] are given by

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{A}_1 \boldsymbol{f}_1\left(\boldsymbol{E}_1 \boldsymbol{x}_k\right) + \boldsymbol{B}_1 \boldsymbol{f}_2\left(\boldsymbol{E}_2 \boldsymbol{u}_k\right) \\ \boldsymbol{y}_k &= \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{C}_1 \boldsymbol{f}_3\left(\boldsymbol{E}_3 \boldsymbol{x}_k\right) \end{aligned} \tag{8}$$

with $\boldsymbol{A} \in \mathbb{R}^{n_x \times n_x}$, $\boldsymbol{B} \in \mathbb{R}^{n_x \times n_u}$, $\boldsymbol{A}_1 \in \mathbb{R}^{n_x \times n_{\theta_1}}$, $\boldsymbol{B}_1 \in \mathbb{R}^{n_x \times n_{\theta_2}}$, $\boldsymbol{E}_1 \in \mathbb{R}^{n_{\theta_1} \times n_x}$, $\boldsymbol{E}_2 \in \mathbb{R}^{n_{\theta_2} \times n_u}$, $\boldsymbol{C} \in \mathbb{R}^{n_y \times n_x}$, $\boldsymbol{C}_1 \in \mathbb{R}^{n_x \times n_{\theta_3}}$, $\boldsymbol{E}_3 \in \mathbb{R}^{n_{\theta_3} \times n_x}$, $\boldsymbol{f}_1 : \mathbb{R}^{n_{\theta_1}} \to \mathbb{R}^{n_{\theta_1}}$, $\boldsymbol{f}_2 : \mathbb{R}^{n_{\theta_2}} \to \mathbb{R}^{n_{\theta_2}}$ and $\boldsymbol{f}_3 : \mathbb{R}^{n_{\theta_3}} \to \mathbb{R}^{n_{\theta_3}}$. $\boldsymbol{f}_1$, $\boldsymbol{f}_2$ and $\boldsymbol{f}_3$ are nonlinear activation functions, but can not be chosen arbitrarily, as will be shown in the following. In order to rewrite the model in (8) in an affine quasi-LPV representation, the time-varying parameters $\boldsymbol{\theta}_k$ have to be defined as

$$\begin{aligned} \left[\boldsymbol{\theta}_1\right]_i &= \frac{f_1\left(\boldsymbol{e}_{1,i}\boldsymbol{x}\right)}{\boldsymbol{e}_{1,i}\boldsymbol{x}}, \qquad i = 1, \ldots, n_{\theta_1} \\ \left[\boldsymbol{\theta}_2\right]_i &= \frac{f_2\left(\boldsymbol{e}_{2,i}\boldsymbol{u}\right)}{\boldsymbol{e}_{2,i}\boldsymbol{u}}, \qquad i = 1, \ldots, n_{\theta_2} \\ \left[\boldsymbol{\theta}_3\right]_i &= \frac{f_3\left(\boldsymbol{e}_{3,i}\boldsymbol{x}\right)}{\boldsymbol{e}_{3,i}\boldsymbol{x}}, \qquad i = 1, \ldots, n_{\theta_3} \end{aligned} \tag{9}$$

$\boldsymbol{e}_{j,i}$ denotes the i-th row of the matrix $\boldsymbol{E}_j$. With these definitions, (8) can be rewritten in affine quasi-LPV form:

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \left(\boldsymbol{A} + \sum_{i=1}^{n_{\theta_1}} \boldsymbol{A}_1 \boldsymbol{e}_{1,i} \left[\boldsymbol{\theta}_{1,k}\right]_i\right) \boldsymbol{x}_k + \\ &\qquad \left(\boldsymbol{B} + \sum_{i=1}^{n_{\theta_2}} \boldsymbol{B}_1 \boldsymbol{e}_{2,i} \left[\boldsymbol{\theta}_{2,k}\right]_i\right) \boldsymbol{u}_k \\ \boldsymbol{y}_k &= \left(\boldsymbol{C} + \sum_{i=1}^{n_{\theta_3}} \boldsymbol{C}_1 \boldsymbol{e}_{3,i} \left[\boldsymbol{\theta}_{3,k}\right]_i\right) \boldsymbol{x}_k \end{aligned} \tag{10}$$

Mayor advantages of the model in (8) compared to the local linear RBF-RNN approach are, that it distinguishes between three sets of affine parameters $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$, which encode the nonlinearity in the system, input and output matrix respectively. It also scales better with increasing dimension of the affine parameter vector in terms of the number of model

parameters: Increasing $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$ or $\boldsymbol{\theta}_3$ amounts to adding $2n_x$, $n_x + n_u$ or $n_y + n_x$ parameters to the model respectively. Also stability conditions can be derived for this model, which to the authors best knowledge is not possible for the general local linear approach. The most significant disadvantages of this approach are

- only $\tanh(\cdot)$ can be used as an activation function and the usage of a bias is not permitted
- $\boldsymbol{f}_j$ must be shallow neural networks, i.e. a single layer of nonlinear neurons
- $\mathcal{B}(\boldsymbol{\theta})$ cannot depend on $\boldsymbol{x}_k$, only on $\boldsymbol{u}_k$, the opposite is true for $\mathcal{A}(\boldsymbol{\theta})$ and $\mathcal{C}(\boldsymbol{\theta})$

These shortcomings are caused by the definition of the time-varying parameters (9), which are essentially created by factoring out the argument in the nonlinear activation functions $\boldsymbol{f}_j$. This creates a discontinuity at $\boldsymbol{x}_k, \boldsymbol{u}_k \to 0$, which is only removable if $\boldsymbol{f}_j$ itself tends to zero, hence a $\tanh$-activation or any activation function crossing the origin has to be used. This can easily be verified va L'Hospital's Rule. For the same reason a bias cannot be used. Factoring out is also not possible, if $\boldsymbol{f}_j$ comprises more than one layer, i.e. deep network structures cannot be used. The absence of a bias in $\boldsymbol{f}_j$ causes all $\boldsymbol{\theta}_j$ to be bell-shaped axially symmetric functions whose width is determined by $\boldsymbol{e}_{j,i}$. This is a very restrictive assumption since $\boldsymbol{\theta}_j$ is supposed to encode the nonlinear behavior of the plant. The rough equivalent of this would be a local linear model, where all local models are centered at the origin of the scheduling space and only the widths of the RBFs can be adjusted.

IV. Gated-RNN (G-RNN) for affine quasi-LPV system identification

The goal is to modify the S-RNN (8) in a way that alleviates its shortcomings while keeping its benefits, i.e. the equivalence to an affine quasi-LPV representation, the good scaling behavior and the stability results. To this end, so-called *gates*, which are found in Gated Units such as the LSTM and the GRU, are integrated into the S-RNN. Gates are neural network structures consisting of two single-layered neural networks, where the output of one layer is weighted by the activation of the other layer. Typically the logistic function $\sigma(\cdot)$ and $\tanh(\cdot)$ are used as activation functions. The equation for the exemplary gate depicted in Fig. 1 is

$$\boldsymbol{z} = \boldsymbol{\sigma}\left(\boldsymbol{W}_{\sigma}\boldsymbol{o} + \boldsymbol{b}_{\sigma}\right) \circ \mathbf{tanh}\left(\boldsymbol{W}_{\tanh}\boldsymbol{o} + \boldsymbol{b}_{\tanh}\right)$$

with $\boldsymbol{z}_k \in \mathbb{R}^{n_o}$, $\boldsymbol{z}_k \in \mathbb{R}^{n_z}$, $\boldsymbol{\sigma} : \mathbb{R}^{n_o} \to \mathbb{R}^{n_z}$, $\mathbf{tanh} : \mathbb{R}^{n_o} \to \mathbb{R}^{n_z}$ and the operator $\circ$ denotes the hadamard product. While gates foremost serve the purpose of alleviating the

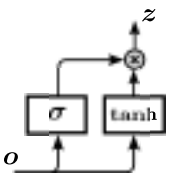


Fig. 1: General structure of a gate.

vanishing gradient problem in Gated Units [2], they have also shown to increase the capability to approximate quite complex nonlinear functions with comparable few parameters due to the superposition of different activation functions [13]. They have also shown to have a positive impact on the initialization dependence of Recurrent Neural Networks [14]. In this application, they will be employed to increase the S-RNNs (8) approximation capabilities in order to approximate highly nonlinear systems while keeping $\dim(\boldsymbol{\theta})$ as little as possible.

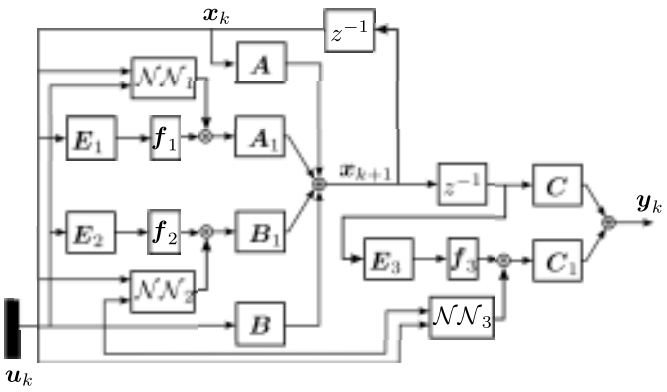


Fig. 2: Structured RNN with *gates* for affine quasi-LPV system identification.

The proposed gated RNN (G-RNN) is depicted in Fig. 2 and follows the equation

$$\begin{aligned}\boldsymbol{x}_{k+1} =& \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{A}_1\mathcal{NN}_1(\boldsymbol{x}_k, \boldsymbol{u}_k) \circ \boldsymbol{f}_1\left(\boldsymbol{E}_1\boldsymbol{x}_k\right) \\ &+ \boldsymbol{B}_1\mathcal{NN}_2(\boldsymbol{x}_k, \boldsymbol{u}_k) \circ \boldsymbol{f}_2\left(\boldsymbol{E}_2\boldsymbol{u}_k\right) \\ \boldsymbol{y}_k =& \boldsymbol{C}\boldsymbol{x}_k + \boldsymbol{C}_1\mathcal{NN}_3(\boldsymbol{x}_k, \boldsymbol{u}_k) \circ \boldsymbol{f}_3\left(\boldsymbol{E}_3\boldsymbol{x}_k\right)\end{aligned} \tag{11}$$

The output of each nonlinear layer is now multiplied element-wise with the output of a Neural Network $\mathcal{NN}_j$. The time-varying parameters $\boldsymbol{\theta}_j$ are again defined by factoring out the argument of $\boldsymbol{f}_j$

$$\begin{aligned}\left[\boldsymbol{\theta}_1\right]_i &= \left[\mathcal{NN}_1(\boldsymbol{x}, \boldsymbol{u})\right]_i \frac{f_1\left(\boldsymbol{e}_{1,i}\boldsymbol{x}\right)}{\boldsymbol{e}_{1,i}\boldsymbol{x}}, \qquad i = 1, \ldots, n_{\theta_1} \\ \left[\boldsymbol{\theta}_2\right]_i &= \left[\mathcal{NN}_2(\boldsymbol{x}, \boldsymbol{u})\right]_i \frac{f_2\left(\boldsymbol{e}_{2,i}\boldsymbol{u}\right)}{\boldsymbol{e}_{2,i}\boldsymbol{u}}, \qquad i = 1, \ldots, n_{\theta_2} \\ \left[\boldsymbol{\theta}_3\right]_i &= \left[\mathcal{NN}_3(\boldsymbol{x}, \boldsymbol{u})\right]_i \frac{f_3\left(\boldsymbol{e}_{3,i}\boldsymbol{x}\right)}{\boldsymbol{e}_{3,i}\boldsymbol{x}}, \qquad i = 1, \ldots, n_{\theta_3}\end{aligned} \tag{12}$$

With these definitions the model (11) can be written as in (10). Again, no bias can be included in the shallow $\boldsymbol{f}_j$-layers in order to preserve the equivalence to an affine quasi-LPV representation and $\tanh$ has to be chosen as an activation function to remove the discontinuity at $\boldsymbol{x}_k, \boldsymbol{u}_k \to 0$. However, there are very little to no prerequisites regarding the Neural Networks $\mathcal{NN}_j$: They can have both $\boldsymbol{x}_k$ and $\boldsymbol{u}_k$ as inputs, can possess an arbitrary number of layers and the user is free to specify any activation function as well as biases. However, if a proof of stability following the one in [1] is desired, the activation functions should be bounded, since the proof of stability makes use of the mean value theorem and the concept of Lipschitz continuity. Also, in order for $\boldsymbol{\theta}_j$ to be continuous, the limit at the possible discontinuity must exist. Application

of L'Hospital's Rule yields

$$\lim_{\boldsymbol{x},\boldsymbol{u}\to\mathbf{0}}[\boldsymbol{\theta}_j]_i = \mathcal{NN}_j(\boldsymbol{x}=\mathbf{0},\boldsymbol{u}=\mathbf{0}) \quad j=1,2,3.$$

Hence, if the output of $\mathcal{NN}_j$ at the origin exists, which is usually the case for Feed Forward Neural Networks with bounded activation functions, $\boldsymbol{\theta}_j$ is a continuous function.

Through the incorporation of an additional (possibly deep) neural network into the structured RNN, it is now possible for the time varying parameters $\boldsymbol{\theta}_j$ to represent potentially arbitrary functions of $\boldsymbol{x}$ and $\boldsymbol{u}$. This is especially important, since as mentioned above, the complexity of quasi LPV controller implementation grows exponentially with $\dim(\boldsymbol{\theta})$. E.g. if a highly nonlinear system is to be approximated via the local linear RBF-RNN approach (6) or the S-RNN approach (8), a better model fit can almost exclusively be achieved via increasing the dimension of $\boldsymbol{\theta}$, which is undesirable. The proposed G-RNN (11) offers the possibility to increase the approximation capability of the model by adding neurons and/or layers to the $\mathcal{NN}_j$, hence increasing only the number of model parameters instead of increasing the dimension $\boldsymbol{\theta}$.

V. Case Studies

The proposed model structure for quasi LPV system identification will be evaluated on two case studies by comparison with the two other approaches described in section III.

A. Silver Box

The Silver Box [15] is an electrical circuit mimicking the behavior of a nonlinear mechanical resonating system with moving mass m, viscous damping d and nonlinear spring $k(y)$:

$$m\ddot{y} + d\dot{y} + k(y)y = u.$$

The nonlinear position-dependent spring follows the equation

$$k(y) = a + by^2. \tag{13}$$

The Silver Box was chosen, because it is widely used as benchmark for nonlinear system identification [16, 17] and can be transformed into an affine quasi-LPV representation. Therefore it is possible to gain some notion regarding the parameter variation, that has to be learned by the Recurrent Neural Networks. Rewriting the Silver Box as a quasi-LPV system yields

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \begin{bmatrix}0 & 1\\ -\frac{a+bx_1^2}{m} & -\frac{d}{m}\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}0\\ 1\end{bmatrix}u \tag{14}$$

One possible affine representation is

$$\begin{bmatrix}\dot{x}_1\\ \dot{x}_2\end{bmatrix} = \left(\begin{bmatrix}0 & 1\\ -\frac{a}{m} & -\frac{d}{m}\end{bmatrix} + \theta_k\begin{bmatrix}0 & 0\\ -\frac{b}{m} & 0\end{bmatrix}\right)\begin{bmatrix}x_1\\ x_2\end{bmatrix} + \begin{bmatrix}0\\ 1\end{bmatrix}u \tag{15}$$

with

$$\theta_k = -x_1^2. \tag{16}$$

The available data were preprocessed by eliminating the offset and subsequently divided into training-, validation- and test-data in the following manner: One period of a multisine with 1342 excited odd harmonics, maximal frequency $f_{\max} \approx 200\text{Hz}$ and phases uniformly distributed in $[0, 2\pi)$ was used for parameter estimation. The validation data set is comprised of a white gaussian noise with linearly increasing amplitude and the corresponding system response. As test data set a multisine with Schröder phases and the corresponding system response, which is partly shown in Fig. 3, was chosen. The measured data can be considered virtually noise-free.

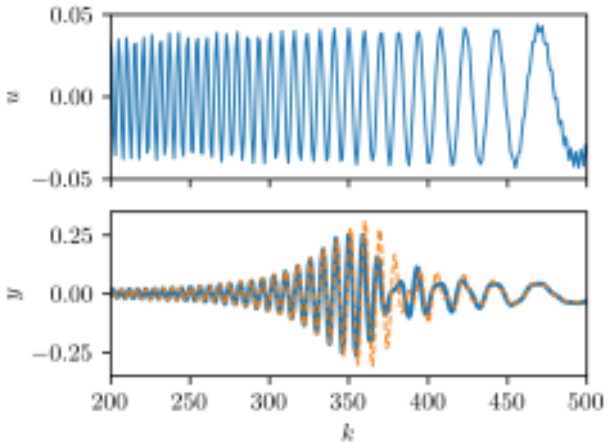


Fig. 3: Multisine signal with Schroeder phases with corresponding system response (—) used for model testing and simulated response of the identified linear subspace model (—) with 46 % BFR.

B. Bioreactor

This simplified bioreactor (17) was proposed by [18] and since used as a benchmark for nonlinear system identification and control numerous times, see e.g. [12, 19, 20].

$$\begin{aligned}\dot{x}_1 &= -x_1u + x_1(1-x_2)e^{x_2/\gamma}\\ \dot{x}_2 &= -x_2u + x_1(1-x_2)e^{x_2/\gamma}\frac{1+\beta}{1+\beta-x_2}\\ y &= x_1\end{aligned} \tag{17}$$

The main challenge of this benchmark, beside the severe nonlinearity is, that the system is nonlinear in both of the states, of which only one is observed. In accordance with common practice the parameters were chosen to be $\beta = 0.02$ and $\gamma = 0.48$, the system was simulated using the forward Euler method with a step size of $h = 0.01$ time units and resulting data sets were downsampled by a factor of 50. Respectively one data set for training, validation and testing was generated. The system was excited with an Amplitude Modulated Pseudo Binary Random (APRBS) signal with range determined from a uniform random distribtion over [0,0.7] and holding times drawn from a uniform random distribution over [1250,2500] time units. The input was limited to this range to prevent pushing the reactor outside of the stable region around the equilibrium $\boldsymbol{x}_e = [0.107, 1]^T$, which was also chosen as the initial state for all conducted experiments. Fig. 4 shows the first half of the test data set. No noise was added to the simulated system response, so theoretically a perfect model fit is achievable.

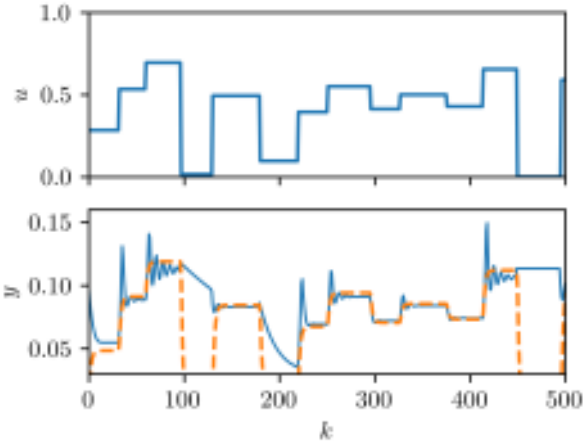


Fig. 4: APRBS signal with corresponding system response (—) used for model testing and simulated response of the identified linear subspace model (—) with 0 % BFR.

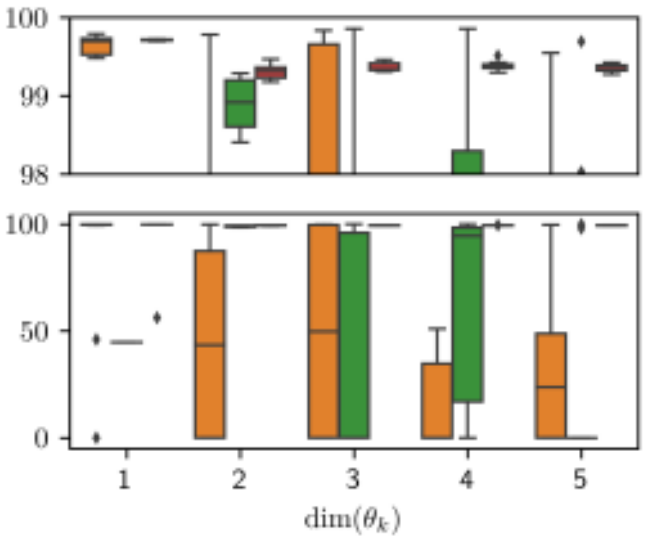


Fig. 5: BFR in % of the RBF-RNN (■), the S-RNN (■), and proposed G-RNN (■) depending on $\dim(\boldsymbol{\theta}_k)$ on the Silver Box benchmark.

VI. Identification Procedure and Results

In each case a linear subspace model was identified using the Matlab® System Identification Toolbox™ implementation of the N4SID-Algorithm.

The parameters of the linear model were used to initialize the appropriate parts of the nonlinear models. The remaining parameters were initialized by drawing from a normal random distribution, except for the matrices $\boldsymbol{A}_1$, $\boldsymbol{B}_1$ and $\boldsymbol{C}_1$ in the S-RNN (8) and G-RNN (11), which were initialized to have a small gain in order for the nonlinear disturbance to be as little as possible in the early stages of training and $\boldsymbol{c}_i$ in (6) were determined from a uniform random distribution over the range of the scheduling variables.

Implementation and gradient-based optimization of all models was performed using Casadi [21] and Ipopt [22]. Each model configuration was initialized 10 times. The optimization procedure was stopped, if either the error did not further decrease or a maximum of 1000 iterations was reached. The performance of each model configuration is measured in terms of the Best Fit Rate (BFR) on the test data set. The BFR corresponds to the coefficient of determination, but can only take on positive values:

$$\mathrm{BFR} = 100\% \cdot \max\left(1 - \frac{\|\boldsymbol{y}_k - \hat{\boldsymbol{y}}_k\|_2}{\|\boldsymbol{y}_k - \bar{\boldsymbol{y}}\|_2}, 0\right)$$

The BFR of the linear subspace model of the Silver Box was 99% on the training data set, but only and 46% on the test data set, see Fig. 3. Likewise the BFR for the linear model of the Bioreactor was 64 % on the training data set, but 0 %, i.e. worse than the mean value, on the test data set, see Fig. 4.

For the Silver Box Benchmark the dimension of the time varying parameter vector $\boldsymbol{\theta}$ is varied from 1 to 5. Since the input and output matrix are not time varying, $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$ are set to zero in (8) and (11), s.t. $\boldsymbol{\theta} = \boldsymbol{\theta}_1$. $\mathcal{NN}_1$ was chosen to be only one layer of neurons with logistic activation functions. The number of neurons was chosen equal to $\dim(\boldsymbol{\theta}_1)$. The results for scheduling with the input and state, i.e. $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{u}, \boldsymbol{x})$ are summarized as a box plot in Fig. 5. It can be observed, that both the S-RNN and the local linear RBF-RNN achieve very good performances on the test data set, if the model structure fits the true structure of the plant fairly well. For the S-RNN (8), the optimal choice is $\dim(\boldsymbol{\theta}_1) = 1$, since the nonlinearity assumption of the model fits the actual nonlinearity of the plant (16) almost perfectly. The nonlinearity assumption of the RBF-RNN also fits the nonlinearity of the plant well, but since all matrices are assumed to be time varying, which is not the case with the Silver Box, more than one time varying parameter is necessary to compensate for that mismatch. Both the S-RNN and the local linear RBF-RNN exhibit considerable variation in model performance with increasing $\dim(\boldsymbol{\theta})$. The proposed G-RNN on the other hand exhibits exceptional convergence behavior with almost no fluctuations regarding the fit on the test data set: Out of a total of 50 identified G-RNNs, every one had a BFR of more than 99 %. The simulation results of the best model of each approach are shown in Fig. 6. All three models approximate the plant almost perfectly. Only at very high amplitudes the simulation error of the G-RNN is slightly higher, a sign that the progressive stiffness of the spring has not been approximated as well. For the Bioreactor Benchmark (17) $\dim(\boldsymbol{\theta}_3)$ was chosen to be zero, s.t. $\boldsymbol{\theta} = [\boldsymbol{\theta}_1, \boldsymbol{\theta}_2]$ for the S-RNN (8) and G-RNN (11). In order to enable a fair comparison to the local linear RBF-RNN and present the results in a compact manner, $\dim(\boldsymbol{\theta}_1)$ was chosen to be equal to $\dim(\boldsymbol{\theta}_2)$. $\mathcal{NN}_1$ was chosen to be a two-layered neural network with 5 neurons with $\tanh$-activation in the first and $\dim(\boldsymbol{\theta}_1)$ neurons with logistic-activation in the second layer. $\mathcal{NN}_2$ is a single layer of $\dim(\boldsymbol{\theta}_2)$ neurons with logistic activation. The results for scheduling with the input and state, i.e. $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{u}, \boldsymbol{x})$ are summarized as a box plot in Fig. 7. On the Bioreactor Benchmark the structured RNN is clearly outperformed by the local linear approach and the gated structured RNN, most likely due to the restrictive assumptions discussed in section III-B. Both the local linear approach and the gated structured

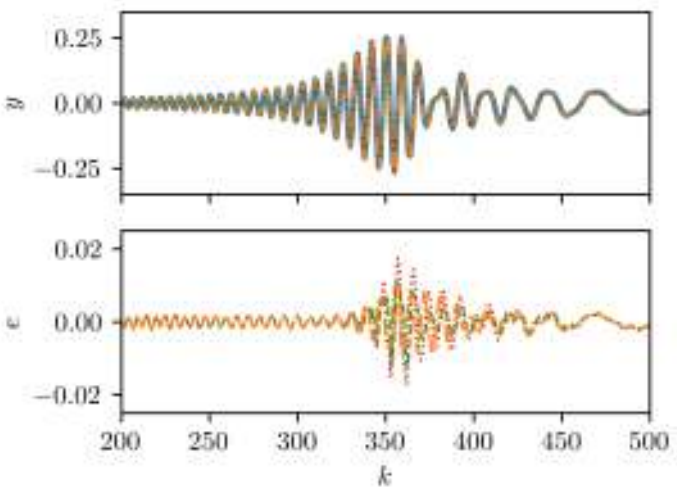


Fig. 6: Top: True system response (—) and simulated responses of the RBF-RNN (—) with $\dim(\boldsymbol{\theta}) = 3$, S-RNN with $\dim(\boldsymbol{\theta}_1) = 3$ (—) and G-RNN (—) with $\dim(\boldsymbol{\theta}_1) = 1$. Bottom: Corresponding simulation error.

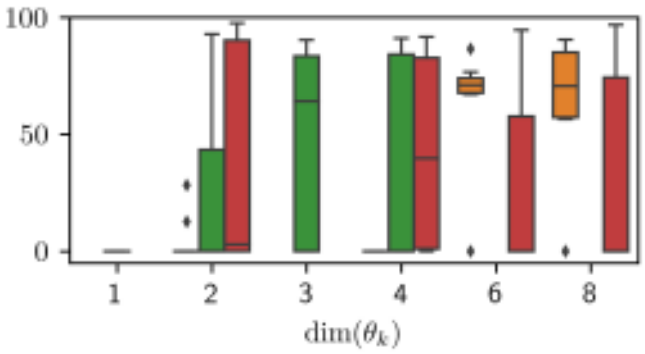


Fig. 7: BFR in % of the RBF-RNN (■), the S-RNN (■), and proposed G-RNN (■) depending on $\dim(\boldsymbol{\theta}_k)$ on the Bioreactor benchmark.

RNN show high fluctuations in performance. However, the proposed G-RNN consistently produces the model with the best fit over all choices of $\dim((\theta))$. The best identified model is a G-RNN with only two time varying parameters and a BFR of 97 %. The simulated response of the respective best identified model of each approach is shown in Fig. 8. Clearly all models struggle to some extent with the two large downwards steps at $k \approx 100$ and $k \approx 450$, as did the linear model. However, while the S-RNNs and the RBF-RNNs simulated responses are considerably off, the proposed G-RNN produces a perfect response for the first downwards step and is also reasonable close to the true response for the second downwards step.

VII. Conclusion & Outlook

A (deep) Recurrent Neural Network for affine quasi-LPV system identification that achieves a better model fit than current approaches with fewer affine parameters was proposed.

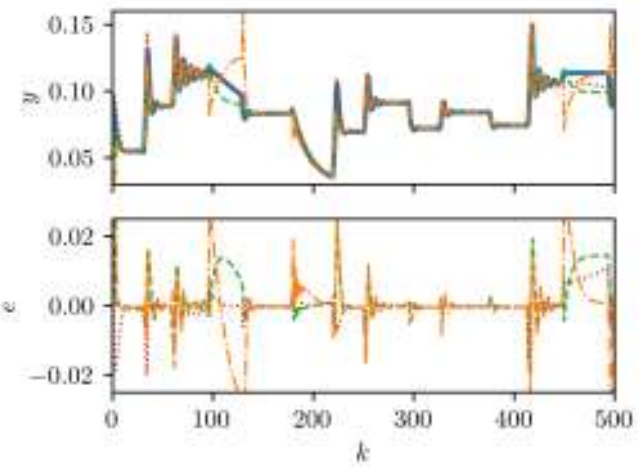


Fig. 8: Top: True system response (—) and simulated responses of the RBF-RNN (—) with $\dim(\boldsymbol{\theta}) = 2$, S-RNN (—) with $\dim(\boldsymbol{\theta}) = 8$ and G-RNN (—) with $\dim(\boldsymbol{\theta}) = 2$. Bottom: Corresponding simulation error.

The proposed model structure was compared to existing approaches on two nonlinear system identification benchmarks with very promising results. For future research, a systematic way for initializing the nonlinear part of the model should be derived and compared to existing initialization procedures for (Deep) Neural Networks, e.g. [23]. Also at least an empirical analysis regarding the choice of the structure of the introduced Neural Networks $\mathcal{NN}_j$, i.e. number of layers and neurons as well as the choice of the activation functions should be conducted. Finally, since the model was developed for polytopic LPV controller synthesis, the control of a controller based on the proposed model needs to be evaluated.

Acknowledgment

This work was supported by the European Union (EU), the European Fund for Regional Development (EFRE) and the state of Hesse.

References

[1] N. Lachhab, H. Abbas, and H. Werner. “A Neural-Network Based Technique for Modelling and LPV Control of an Arm-Driven Inverted Pendulum”. In: *Proceedings of the 47th IEEE Conference on Decision and Control*. 2008, pp. 3860–3865.

[2] F.A. Gers, J. Schmidhuber, and F. Cummins. “Learning to forget: continual prediction with LSTM”. In: *9th International Conference on Artificial Neural Networks ICANN 99*. 1999. DOI: 10.1049/cp:19991218.

[3] K. Cho et al. “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”. In: *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 8. 2014, 1724–1734. DOI: 10.3115/v1/D14-1179. URL: http://www.aclweb.org/anthology/D14-1179.

[4] C. Hoffmann and H. Werner. “A Survey of Linear Parameter-Varying Control Applications Validated by Experiments or High-Fidelity Simulations”. In: *IEEE Transactions on Control Systems Technology* 23 (2015), pp. 416–434.

[5] C. Hoffmann. “Linear Parameter-Varying Control of Systems of High Complexity”. PhD thesis. Technische Universitat Hamburg-Harburg, 2016.

[6] A. Kwiatkowski, M.-T. Boll, and H. Werner. “Automated generation and assessment of affine LPV models”. In: *Proceedings of the 45th IEEE Conference on Decision and Control*. IEEE. 2006, pp. 6690–6695.

[7] F. Casella, F. Donida, and M. Lovera. “Automatic generation of LFTs from object-oriented non-linear models with uncertain parameters”. In: *6th Vienna International Conference on Mathematical Modeling*. 2009, pp. 11–13.

[8] P.J.W. Koelewijn and R. Tóth. “Scheduling Dimension Reduction of LPV Models-A Deep Neural Network Approach”. In: *2020 American Control Conference (ACC)*. IEEE. 2020, pp. 1111–1117.

[9] O. Nelles. *Nonlinear System Identification: From Classical Approaches to Neural Networks and Fuzzy Models*. Berlin Heidelberg, Germany: Springer, 2001. DOI: https://doi.org/10.1007/978-3-662-04323-3.

[10] H. Abbas and H. Werner. “Polytopic quasi-lpv models based on neural state-space models and application to air charge control of a SI engine”. In: *IFAC Proceedings Volumes* 41.2 (2008), pp. 6466–6471.

[11] P. Gil et al. “On state-space neural networks for systems identification: Stability and complexity”. In: *2006 IEEE Conference on Cybernetics and Intelligent Systems*. IEEE. 2006, pp. 1–5.

[12] V. Verdult. “Nonlinear System Identification: A State-Space Approach”. PhD thesis. University of Twente, 2002.

[13] A. Rehmer and A. Kroll. “On the vanishing and exploding gradient problem in Gated Recurrent Units”. In: *21st IFAC World Congress*. Berlin, Germany, 2020.

[14] A. Rehmer and A. Kroll. “On Using Gated Recurrent Units for Nonlinear System Identification”. In: *18th European Control Conference, ECC 2019, Naples, Italy, June 25-28, 2019*. IEEE, 2019, pp. 2504–2509. DOI: 10.23919/ECC.2019.8795631. URL: https://doi.org/10.23919/ECC.2019.8795631.

[15] T. Wigren and J. Schoukens. “Three free data sets for development and benchmarking in nonlinear system identification”. In: *Proceedings of the Xth European Control Conference (ECC)*. Zurich, Switzerland. 2013, pp. 2933–2938. DOI: 10.23919/ECC.2013.6669201.

[16] V. Verdult. “Identification of Local Linear State-Space Models: The Silver-Box Case Study”. In: *IFAC Proceedings Volumes* 37.13 (2004). 6th IFAC Symposium on Nonlinear Control Systems 2004 (NOLCOS 2004), Stuttgart, Germany, 1-3 September, 2004, pp. 393–398. ISSN: 1474-6670. DOI: https://doi.org/10.1016/S1474-6670(17)31255-7. URL: https://www.sciencedirect.com/science/article/pii/S1474667017312557.

[17] L. Ljung et al. “Deep learning and system identification”. In: *Proc IFAC Congress, Berlin*. 2020.

[18] L. H. Ungar. “A Bioreactor benchmark for adaptive network based process control”. In: *Neural networks for control*. Ed. by W. T. III Miller, R. S. Sutton, and P. J. Werbos. MIT Press, 1990, pp. 387–402.

[19] D. C. Dracopoulos and R. Piccoli. “Bioreactor control by genetic programming”. In: *International Conference on Parallel Problem Solving from Nature*. Springer. 2010, pp. 181–188.

[20] M. Ö. Efe. “A comparison of networked approximators in parallel mode identification of a bioreactor”. In: *Advances in Engineering Software* 41.9 (2010), pp. 1132–1147.

[21] J. A. E. Andersson et al. “CasADi – A software framework for nonlinear optimization and optimal control”. In: *Mathematical Programming Computation* (In Press, 2018).

[22] A. Wächter and L. T. Biegler. “On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming”. In: *Mathematical programming* 106.1 (2006), pp. 25–57.

[23] X. Glorot and Y. Bengio. “Understanding the difficulty of training deep feedforward neural networks”. In: *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics (AISTATS)*. 2010, pp. 249–256.

Performance Comparison of Deep Convolutional Neural Networks for Binary Classification of Fine-Grained Leaf Images

Kamal KC, Zhendong Yin, Dasen Li and Zhilu Wu

Abstract—Intra-plant disease classification based on leaf images is a challenging computer vision task, due to similarities in texture, color and shape of leaves with a slight variation of leaf spot; and external environmental changes such as lighting and background noises. Deep Convolutional Neural Network (DCNN) have proven to be an effective tool for binary classification. In this paper, two methods for binary classification of diseased plant leaves using DCNN are presented; model created from scratch and transfer learning. Our main contribution is a thorough evaluation of 4 networks created from scratch and transfer learning of 5 pre-trained models. Training and testing of these models were performed on a plant leaf images dataset belonging to 16 distinct classes, containing a total of 22,265 images from 8 different plants, consisting a pair of healthy and diseased leaves. We introduce a deep CNN model, Optimized MobileNet. This model with depthwise separable CNN as building block attained an average test accuracy of 99.77%. We also present a fine-tuning method by introducing the concept of convolutional block which is a collection on different deep neural layers. Fine-tuned models proved to be efficient in terms of accuracy and computational cost. Fine-tuned MobileNet achieved an average test accuracy of 99.89% on 8 pairs of [healthy, diseased] leaf ImageSet.

Keywords—Deep Convolution Neural Network, Depthwise Separable Convolution,Fine-grained classification, MobileNet, Plant disease, Transfer Learning.

I. Introduction

GLOBAL food demand is estimated to double from 2005 to 2050 [1] due to a growing world population and increasing per capita food consumption. To meet this demand, a 2.4% per year rate of yield gains (non-compounding) would be required [2]. Crop yield can be increased by increasing arable land. However, in recent years, there has been a drastic decrease in the quantity of arable lands due to soil erosion and urbanization. Each year about 10 million hectares of cropland are lost due to soil erosion, thus reducing the cropland available for world food production [3]. Shrinking of natural resources such as water, farmlands, arable soil, biodiversity etc. is one of the reasons for yield loss. Moreover, there is a substantial difference between potentially achievable crop yields and actual yields, named as yield gaps, which is caused by crop loss through pest, diseases and climate changes. Roughly, direct yield losses caused by pathogens, animals, and weeds are altogether responsible for losses ranging between 20 and 40% of global agricultural productivity and can be prevented, or reduced, by plant or crop protection measures [4], [5]. Closing the yield gap is a potential solution to increase worldwide food production. Many researches in the field of agro-industry, such as modern irrigation, use of fertilizers, use of genetically engineered [6] or hybrid seeds that are naturally inclined to grow faster, stronger, precision agriculture [7] or combination of these techniques have been per-formed to increase crop yield. Yield loss due to plant diseases is inevitable and hence requires monitoring and implementation of correction measures. Disease manifestation is primarily detected by visual examination. Implementation of computer vision algorithms for automatic detection and identification of diseases based on plant images is a challenging task but is proven to be effective and efficient.

Intra-species image classification is a challenging task due to small inter-class variance. Intra-species visual variability and view conditions variability make the identification more realistic but more complex. Intra-species visual variability due to leaf's color, shape, margin appearance and number and relative positions of leafets makes the classification challenging [8]. Recognition of images belonging to same species but different breeds or images of same categories but different sub-categories gives rise to a sub-field of image recognition, fine-grained categorization or classification which aims to distinguish subordinate categories belonging to the same basic level category. This categorization is considered a difficult task as the variance between intra-class is large while the inter-class variance is low [9]. This generally refers to classifying different breeds of dogs, different models of cars or even different species of birds and flowers. Various datasets such as Fine-Grained Visual Classification of Aircraft (FGVC-Aircraft), Stanford dogs, NABirds, Oxford 102 Flowers, Comprehensive Cars (CompCars) etc. have been created for fine-grained image classification [10], [11], [12], [13], [14]. Breeds are dened as intra-species groups thus referring to Stanford dogs challenge as intra-species classification. The similarity of visual traits among various breeds may not be as complex as the task at hand; identifying plant disease within the same plant. The distinguishing factor between the diseased and non-diseased plants can be a minor spot invisible to naked eyes, as it could be early stage in manifestation of disease. This fine-grained, intra-species classification is a crucial step to increase crop yield. The similarity of visual traits among various breeds may not be as complex as the task at hand; identifying plant disease within the same plant. The distinguishing factor between the diseased and non-diseased plants can be a minor spot invisible to naked eyes, as it could be early stage in manifestation

Kamal KC, Z. Yin, D. Li, and Z. Wu are with the School of Electronics and Information Engineering, Harbin Institute of Technology. (email: kamalkc@hit.edu.cn, yinzhendong@hit.edu.cn, sen09024@hit.edu.cn, wuzhilu@hit.edu.cn). Z. Yin is the corresponding author.

of disease. This fine-grained, intra-species classification is a crucial step to increase crop yield.

Deep learning models have been outperforming machine learning models with the ability to use raw data directly without feature engineering and intervention of human experts. A convolutional neural network (CNN) is a deep neural network (DNN) inspired by the animal visual cortex, primarily employed for computer vision tasks. Various CNN architectures were proposed to be used for object recognition. Among them LeNet [15] and AlexNet [16] have been considered as first of their kinds to revolutionize CNN for computer vision tasks. LeNet was used to classify tiny images of handwritten digits (28×28 pixels), MNIST dataset [17], while the AlexNet which is a deeper network, is considered as a breakthrough in object recognition which set a new benchmark for ILSVRC, ImageNet Large Scale Visual Recognition Challenge [18]. Transfer learning is a machine learning method

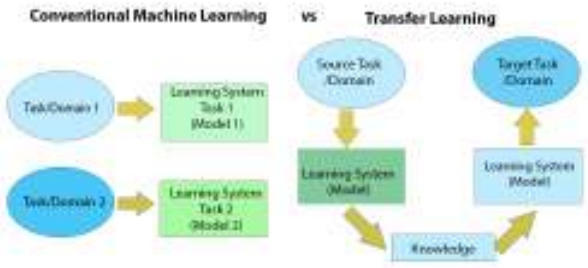


Fig. 1. Conventional Machine learning model vs Transfer Learning

that focuses on storing knowledge gained while solving one problem and reuse as the starting point for a different but related problem. While traditional machine learning method works on creating models per task, transfer learning reuses the learned knowledge for similar task as can be seen in Fig 1. Transfer learning is categorized into three sub-settings; inductive transfer learning, transductive transfer learning, and unsupervised transfer learning, based on different situations between the source and target domains and tasks [19]. In inductive transfer learning, the source and target tasks are different irrespective of source and target domains. In this case, some labelled data in the target domain are required to induce an objective predictive model in the target domain to achieve high performance in the target task by transferring knowledge from the source task. In the transductive transfer learning, the source and target tasks are same, while the source and target domains are different. In unsupervised transfer learning, source and target tasks are different but slightly related and focus is on solving unsupervised learning tasks in the target domain, such as clustering, dimensionality reduction and density estimation since there are no labelled data available in both source and target do-mains while training. Transfer learning helps to reduce the computational cost and overcomes the complication of requiring plethora of data that a model needs to learn.

II. Related Works

It is of great importance to diagnose the plant diseases at early stages in order to implement appropriate and timely preventive measures to avoid future losses. Numerous image processing techniques have been used in the course of time to detect and classify images. Due to the advancement in machine learning, various ML techniques have been introduced for classification of plant types and diseases. A lot of research has been conducted in achieving the utmost efficiency for identification and classification of images. Machine Learning models such as Support Vector Machine (SVM), Artificial Neural Network (ANN), K-Nearest Neighbor (KNN) and deep learning model like Convolutional Neural Network (CNN) has been explored for plant disease classification based on images. A survey on image analysis using machine learning algorithms such as Gaussian models, linear regression, logistic regression, support vector machines, Gaussian mixture model, latent linear models, sparse linear models, Gaussian mixture models, ensemble learning, directed graphical models, undirected graphical models, clustering, Gaussian processes, Dirichlet processes, and deep learning is done to demonstrate its outstanding predictive power [20].

A. Machine Learning in Agriculture

Early detection and classification of plant diseases like fungi, bacteria, virus, nematodes and deficiency diseases affecting different agriculture/horticulture crops was done for 900 plant parts images belonging to 6 classes. Features were selected based on color and textures were classified using ANN and SVM with average classification accuracy of 87.48% and 92.17% [21]. Fuzzy C-means clustering algorithm was applied for identification of disease in wheat leaves based on seven features with an accuracy of 88% [22]. A total of 137 grape leaf images belonging to two diseases; downy mildew and powdery mildew, were classified using SVM after segmenting the image via k-means clustering techniques. The system reached 88.89% average accuracy when classified based on texture and color features [23]. An algorithm for binary classification of healthy and diseased soybean leaves was implemented in four major steps. Pre-processing was done by converting RGB image to HSV (Hue Saturation Value) color space, ROI (Region of Interest) was extracted using multi-thresholding, image was segmented based on color and cluster based methods along with SIFT (Scale Invariant Feature Transform) technique and finally, SVM classifier was used for automatic classification which achieved an accuracy of 93.79% [24]. A mobile clientserver architecture for leaf disease detection and diagnosis was proposed where the mobile client captures and pre-processes the leaf image, segments diseased patches in it and transmits to the Pathology Server, reducing transmission cost. The Server performs the computational tasks: combination of Gabor wavelet transform (GWT) and gray level co-occurrence matrix (GLCM) for feature extraction and k -Nearest Neighbor classification. This automated mobile version system was experimented on 297 images from five different plant leaf diseases and proved to be 93% accurate [25]. 300 potato leaf images belonging to 3 classes were classified after 60%-40% train-test split using multi-class SVM with linear kernel. This algorithm attained an accuracy of 95% [26].

Groundnut leaf disease extraction and classification using color imagery by employing color imaginary transform and color co-occurence matrix in junction with Backpropagation neural network (BPNN)was done with 97.41% accuracy for four different diseases [27]. Using a dataset of 13,689 images of diseased apple leaves belonging to four diffetent diseases, a DCNN model based on AlexNet with addition of GoogleNet Inception module was implemented. This model automatically discovered the discriminative features of leaf images successfully and enabled an end-to-end learning pipeline with 97.62% accuracy [28]. Fine-tuned GoogleNet achieved an accuracy of 99.18% while classifying a dataset containing 14,828 tomato leaf images infected with nine diseases which dramatically outperformed shallow models [29]. Using a public dataset of 54,306 images of diseased and healthy plant leaves, a deep convolutional neural network was trained to identify 14 crop species and 26 diseases (or absence thereof). Transfer learned GoogleNet achieved an accuracy of 99.35% on 80-20 train-test data split [30]. 12,673 samples containing leaf images of four classes, including the healthy leaf images, were obtained from the PlantVillage database taken under uncontrolled environment. The implemented CNN model achieved 99.32% classification accuracy. The dataset was divided 70% for the training, 10% for validation and 20% for testing [31]. Several CNN model architectures were trained, with the best performance reaching a 99.53% success rate in identifying the corresponding [plant, disease] combination (or healthy plant) from an open database of 87,848 images, containing 25 different plants in a set of 58 distinct classes of [plant, disease] combinations, including healthy plants [32].

B. Fine-grained Image Analysis

Fine-grained visual classification (FGVC) datasets contain large visual diversity within a class and visual similarity, often confusing samples among classes. The large variation in data size per class adds tediousness to the classification problem. This makes classification tedious for such datasets. To achieve better performance on fine-grained images, a two-stream model combining vision and language (CVL) for learning latent semantic representations was proposed. The vision stream learned deep representations from the original visual information via DCNN while the language stream utilized the natural language descriptions, which pointed out the discriminative parts or characteristics for each image, and provided a flexible and compact way of encoding the salient visual aspects for distinguishing sub-categories. Since the two streams are complementary, combining the two streams achieved better classification accuracy [33]. Fine-tuned baseline models using pairwise classification (PC) optimization procedure substantially improved accuracy on various FGVC datasets. PC-DenseNet-161 obtained 86.87% accuracy on Caltech-USD Birds (CUB-200-2011) data which is 2.66% improvement in comaprision to baseline Densenet-161. Similarly, PC-ResNet on Standford cars and PC-BilinearCNN on Flower-102 attained 93.43% and 93.65% accuracy respectively [34].

A method for fine-grained plant classification for the leaf image in which complementary global and patch-based features were combined at each hierarchical level (genus and species) was presented by use of pre-trained and fine-tuned CNNs.The experimental results based on the ImageCLEF 2015 plant recognition dataset proved that the proposed coarse-to-fine classification strategy was better than flat classification approaches considering each individual category (family, genus, and species) [35]. The best performance on the fine-grained CiradRiceSeed dataset reached up to 88.67% good classification rate and top-5 classification accuracy of 99.82%, which was attained by CNN pre-trained on ImageNet along with histogram equalization [36]. Identification of plant species from pictures of bark and leaves using textural recognition with the proposed Ffirst method lead to state-of-the-art results. The experiment further illustrated that classification accuracy with CNN reached $99.9-100\%$ accuracy on the MEW dataset with 153 plant species [37].

III. Materials and methods

A. Proposed Approach

Convolutional layers are the basic building blocks of a deep CNN model. Accuracy of model is dependent on the depth of the model as evident by the performance of LeNet [15], AlexNet [16], VGG-16, and VGG-19 [38] on ImageNet dataset, as well as the core building blocks within the network. Various state-of-the-art deep learning models such as Inception [39], ResNet [40], Xception [41], etc., which performed well in ILSVRC [18] make use of CNN and it's variant. We propose two methods for image classification viz: model creation from scratch and fine-tuning a transfer learned model.

1) Model from Scratch: Standard DCNN comprises of different types of hidden layers between input and output layers. A typical ConvNet (Convolutional Network) consists of a series of convolutional layers, activation layers, and is subsequently followed by pooling, normalization and fully connected layers. We make use of two types of convolutional building blocks: conventional CNN and depthwise separable CNN while designing models from scratch. The network with conventional CNN is a 19 layered network inspired by VGG19 and the networks with depthwise separable CNN are inspired by MobileNet [42]. The convolution output is given as

$$y_{i^{l+1},j^{l+1},d} = \sum_{i=0}^{H}\sum_{j=0}^{W}\sum_{d=0}^{D} f_{i,j,d} \times x^{l}_{i^{l+1}+i,j^{l+1}+j,d}. \quad (1)$$

where, $x^{l}_{i,j,d}$ is the input of $l-th$ layer, $y_{i,j,d}$ is the output , and $f_{i,j,d}$ is the filter with D filters of spatial size span $H \times W$.

The depthwise separable convolution splits feature learning, operated by standard convolutions, over two simpler steps: a spatial feature learning step and a channel combination step. The depthwise separable convolution output is obtained as

$$y_{i^{l+1},j^{l+1},d} = \sum_{d=0}^{D} f_{d} \times \sum_{i=0}^{H}\sum_{j=0}^{W} f_{i,j} \times x^{l}_{i^{l+1}+i,j^{l+1}+j} \quad (2)$$

In equation (2), f_d is a 1×1 convolution filter.

2) Transfer Learning and fine-tuning: Multitask learning improves generalization by leveraging the domain-specific information contained in training data of related tasks by learning source and target task simultaneously [43]. Domain Adaptation is a process that attempts to alter the source domain in a way to bring the distribution of source close to that of the target domain. Transfer learning focuses on improving the learning of predictive function in target domain using knowledge in both source and target domains when domains or tasks between source and target differ. In transfer learning, layers with reusable features of a pre-trained model trained on a large readily available dataset and on a completely different task are determined and the output of that layer is used as input features to train a much smaller network that requires smaller number of parameters. This smaller network only needs to learn the relations for the specific problem having already learned about patterns in the data from the pre-trained model. Transfer learning virtually creates a shallow network within a deep network by utilizing previously learned knowledge. Transfer learning is done either by using the model as feature extractor or by fine-tuning the model for better performance.

Transfer learning is employed when training dataset has smaller amount of data and is similar to pre-trained dataset. Transfer learning is implemented either by creating a suitable model from scratch, training the model on dataset with plethora of akin data, and finally fine-tuning the model on available dataset or by reusing state-of-the-art model pre-trained on standard dataset with surfeit and analogous data, and fine-tuning in accordance to available data. The latter is preferred as this overcomes the hassle of creating a model and saves time for training on different set of data.

Given a source domain $\mathcal{D}_S = \{\mathcal{X}_S, P(X_S)\}$, where $\mathcal{X}_S$ is a feature space and $P(X_S)$ is a marginal probability distribution in which $X_S = \{x_{S_1}, ..., x_{S_n}\} \in \mathcal{X}_S$, source task $\mathcal{T}_S = \{\mathcal{Y}_S, f_S(\cdot)\}$, where $\mathcal{Y}_S$ is label space and $f_S(\cdot)$ an objective predictive function, target domain $\mathcal{D}_T = \{\mathcal{X}_T, P(X_T)\}$, where $\mathcal{X}_T$ is a feature space and $P(X_T)$ is a marginal probability distribution in which $X_T = \{x_{T_1}, ..., x_{T_n}\} \in \mathcal{X}_T$ and target task $\mathcal{T}_T = \{\mathcal{Y}_T, f_T(\cdot)\}$, where $\mathcal{Y}_T$ is label space and $f_T(\cdot)$ an objective predictive function, such that $0 \leq n_T \ll n_S$. Predictive function $f_S(\cdot)$ is learned from source training data, which consists of pairs $\{x_{S_i}, y_{S_i}\}$ and $f_T(\cdot)$ is learned from target training data, which consists of pairs $\{x_{T_i}, y_{T_i}\}$ in junction with $f_S(\cdot)$.

While classifying diseased plant leaf images based on ImageNet dataset [44], the source task $\mathcal{T}_S$ and the target task $\mathcal{T}_T$ are different (i.e. $\mathcal{T}_S \neq \mathcal{T}_T$). The label spaces between these two tasks are different (i.e.$\mathcal{Y}_S \neq \mathcal{Y}_T$). Inductive transfer learning is the best solution to solve such problem. In inductive transfer learning, the common features can be learned by solving an optimization problem [19], given as

$$\begin{aligned} \arg\min_{A,U} \sum_{t\in\{T,S\}} \sum_{i=1}^{n_t} L(y_{t_i}, \langle a_t, U^T x_{t_i}\rangle) + \gamma\|A\|_{2,1}^2 \\ s.t. \quad U \in \mathbf{O}^d. \end{aligned} \tag{3}$$

In this equation, S and T denote the tasks in the source domain and target domain, respectively. $A = [a_S, a_T] \in R^{d\times 2}$ is a matrix of parameters. U is a $d \times d$ orthogonal matrix (mapping function) for mapping the original high-dimensional data to low-dimensional representations. The (r, p)-norm of A is defined as $\|A\|_{r,p} := (\sum_{i=1}^{d} \|a^i\|_r^p)^{\frac{1}{p}}$. The optimization problem (3) estimates the low-dimensional representations $U^T X_T$, $U^T X_S$ and the parameters, A, of the model at the same time.

B. Methodology

The choice of using existing deep architectures was motivated by their spectacular results in computer vision tasks. We used four models from scratch. VGG19 and MobileNet were recreated based on works by Simonyan & Zisserman [38], and Google Inc. [42] using conventional convolution, and depthwise separable convolution respectively. Reduced MobileNet is a layer pruned version of MobileNet [45]. Table I shows the architecture of optimized mobilenet. All the input images were resized to 224×224. Base Mobilenet was used for the experiments. Hyper-parameters such as α and ρ, that are used for thinner models in Mobilenet were set to 1.

TABLE I
OPTIMIZED MOBILENET ARCHITECTURE

Layer Size/Stride	Output Shape	Parameters
3,3 Conv,32/S2	112,112,32	992
3,3 DepthwiseConv, 32/S1 1,1 PointwiseConv, 64/S1	56,56,64	2720
3,3 DepthwiseConv, 64/S2 1,1 PointwiseConv, 128/S1	28,28,128	9536
3,3 DepthwiseConv, 128/S1 1,1 PointwiseConv, 128/S1	28,28,128	18524
3,3 DepthwiseConv, 128/S2 1,1 PointwiseConv, 256/S1	14,14,256	35456
3,3 DepthwiseConv, 256/S1 1,1 PointwiseConv, 256/S1	14,14,256	69888
3,3 DepthwiseConv, 256/S2 1,1 PointwiseConv, 512/S1	7,7,512	136448
3,3 DepthwiseConv, 512/S1 1,1 PointwiseConv, 512/S1	7,7,512	270848
3,3 DepthwiseConv, 512/S2 1,1 PointwiseConv, 1024/S1	4,4,1024	535040
Global Average Pooling	1,1,512	0
Dense	1,1,2	2050
Softmax	1,1,2	0

Transfer learning makes use of previously learned knowledge on new tasks. Reusable features extracted by models pre-trained on ImageNet was applied to re-train the model on PlantLeaf dataset. Transfer learning of a model is generally done in two ways: model used as feature extractor and fine-tuning the model. Transfer learning works on the concept of layer freeze. Freezing a layer is referred to as not updating layer weights during training in order to avoid making changes on previously extracted better features by filters in the earlier layers. Parameters corresponding to the frozen layers are referred as non-trainable parameters, whereas the network trains with remaining trainable parameters. In contrast to the backpropagating and updating the weights of all the layers in the network, fine-tuning drastically reduces the computational cost. There is an inverse relationship between number of frozen layers and number of trainable parameters. Feature

extractor is implemented by replacing the final output layer with the suitable classifier and freezing the weight for entire network except that of the final fully connected layer, whose neurons have full connections to all activations in the previous layer. The rest of the network is treated as a fixed feature extractor while the reusable features are entirely extracted from ImageNet. In fine-tuning, not only the classifier is replaced but the weights of the pre-trained network is also fine-tuned by continuing the backpropagation. The number of layers required to fine tune depends upon the data used and the type of network. It is possible to fine-tune all the layers of the model, or it is possible to keep some of the earlier layers fixed to avoid overfitting and only fine-tune some higher-level portion of the network. We opted for fine-tuning instead of feature extractor for higher accuracy in expense of slightly higher computational cost [46].

C. Convolutional Block

Convolutional block is referred to as collection of layers in a model comprising of a convolutional layer and all the layers before the succeeding convolutional layer or a group of convolutional layers together with other layers depending upon the model architecture. Convolutional blocks are frozen and unfrozen instead of individual layers which helps to reach the desired accuracy faster. Trainable blocks are the convolutional blocks that are not being frozen. Fig 2 shows the convolutional blocks and fine-tuning process for MobileNet model. Trainable parameters are the total number of parameters that gets re-trained on the new data. The PlantLeaf dataset was split

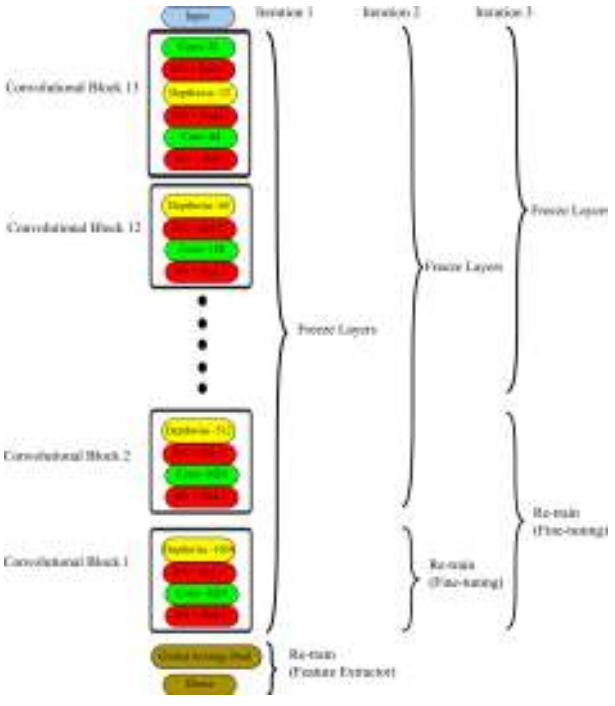


Fig. 2. MobileNet retrained as feature extractor and fine-tuned for better accuracy by retraining convolutional blocks.

into three sets: training, validation, and testing. Validation data was used to tune the network parameters and hyper-parameters to prevent overfitting. Unseen test data was used to get a generalized measure of classification accuracy. The train-validation-test data split percentage of 70:20:10 was used to avoid overfitting [30]. After the model was completely trained, the model was used to evaluate the accuracy for 10% of the data. All the experiments were run for 500 epochs.

D. DataSet

The dataset used in our experiments was obtained from PlantVillage dataset. All the dataset were divided into training, validating and testing set with 70:20:10 split respectively. This PlantLeaf dataset contains plant leaf images collected from 8 different plants as seen in fig 3. A total of 22,265 images spread over 16 classes. Each class represents a combination of [plant type, disease type/healthy]. Total number of images per class ranges from 64 to 4048. This imageset contains images with clean background and leaf images insitu with lots of background noise. A detailed information about the images is shown in Table II. Both the healthy and diseased leaf images of cherry are clean images, while that of banana and cabbage are cluttered images. Others are combination of both type of images. In regards to images from squash leaves, all the healthy leaves are taken in noisy condition while the diseased leaves are entirely with clean background. In addition to the tabulated data, corn leaves infected with leaf blight is added making three sets of healthy/commonrust, healthy/leafblight and commonrust/leafblight to check the robustness of the model. All the images of corn leaves infested with leafblight disease was taken in clean background. From a total of 985 images, 688 were used for training, 199 for validation and the remaining 98 images were used for testing.

Fig. 3. Few samples from the dataset which includes an image from each class.

IV. Result

VGG19, MobileNet, Reduced MobileNet and Optimized MobileNet achieved an average test accuracy of 99.65%,

TABLE II
PLANTLEAF DATASET

Plant	Disease	Clean Images			Clutter Images			Images in the class
		Train	validation	Test	Train	Validaion	Test	Total
Apple	BlackRot	444	115	62	55	27	9	712
Apple	Healthy	1083	307	154	151	44	21	1760
Banana	Healthy	0	0	0	1151	328	164	1643
Banana	Speckle	0	0	0	2300	656	328	3284
Cabbage	Blackrot	0	0	0	46	12	6	64
Cabbage	Healthy	0	0	0	294	84	42	420
Cherry	Healthy	599	169	86	0	0	0	854
Cherry	Powdery Mildew	737	210	105	0	0	0	1052
Corn	Common Rust	826	244	122	309	78	35	1614
Corn	Healthy	563	183	78	2269	628	327	4048
Grape	Healthy	303	77	43	127	45	18	613
Grape	Leaf Blight	752	216	108	0	0	0	1076
Pepper	Bacterial Spot	698	199	100	0	0	0	997
Pepper	Healthy	1031	303	143	390	102	60	2029
Squash	Healthy	0	0	0	184	53	27	264
Squash	Powdery Mildew	1286	366	183	0	0	0	1835

99.73%, 99.72% and 99.77% respectively, for 8 pairs of [healthy, diseased] plant leaf images. VGG19 achieved the highest accuracy of 99.79% for apple leaves, MobileNet achieved the highest accuracy of 100% for cabbage, cherry, grape and squash leaf images, Reduced MobileNet achieved highest accuracy of 100% for cabbage, cherry and squash leaves, and Optimized MobileNet achieved highest accuracy of 100% for 5 plant leaf images. Table III shows the individual classification accuracy attained by four models created from scratch on 8 pairs of healthy and diseased leaf images.

Transfer Learning was performed on six different state-of-the-art models pre-trained on ImageNet. Table IV shows the accuracies obtained while employing fine-tuned models on PlantLeaf dataset. Fine-tuned MobileNet achieved an average accuracy of 99.89% on 8 pairs of plant leaves, which outperformed models created from scratch. This MobileNet obtained 100% testing accuracy on five pairs of healthy and diseased dataset. Fine-tuned InceptionV3 and Nasnetmobile performed well with an average testing accuracy of 99.84% and 99.82% respectively. Fine-tuned VGG19 and Densenet121 achieved an average accuracy 99.8% and 99.77% respectively. All of the fine-tuned models achieved 100% accuracy on 5 sets of leaf images. InceptionV3 attained 100% on 6 out of 8 pairs of datasets.

Three sets of corn leaf images were classified using various forms of MobileNet model. The accuracies are shown in figure 4. Fine-tuned MobileNet achieved higher accuracy on all three sets of corn leaf images. It attained a test accuracy of 99.82%, 100% and 99.6% on three corn leaf images binary sets healthy,commonrust, healthy, leafblight and commonrust,leafblight respectively.

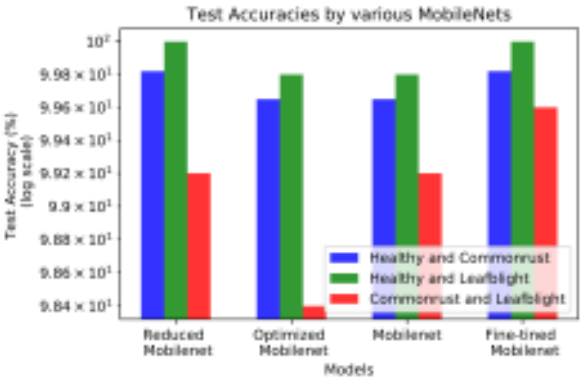


Fig. 4. Accuracy attained by various versions of Mobilenets on three sets of corn leaf images

V. CONCLUSION

In this work, various deep learning models were created from scratch and compared with different fine-tuned state-of-art deep learning models, pre-trained on ImageNet, for efficient classification of plant diseases based on RGB leaf images of healthy and diseased plants. Layer pruned version of MobileNet, Reduced MobileNet, proved to be at par with MobileNet but slightly lesser accurate compared to fine tuned models, while Optimized MobileNet outperformed all the models created form scratch. Depthwise separable CNNs performed exceptionally well with drastic decrements in number of parameters and thus, computational cost.

ImageNet proved to be a reliable source domain for transferring learned features on intra-plant leaf disease classification. Features learned from ImageNet proved to be effective for fine-grain classification. Fine-tuned InceptionV3, achieved a success rate of 100% on testing data for 6 out of 8 sets of plants, while fine-tuned Mobilenet achieved highest average test accuracy. The experiment on the squash leaf images, which had higher inter class correlation compared to intra class correlation demonstrates that alienation in inter class data causes high accuracy. Similarly, the experiment on cabbage leaf images, which had scarce data show that inadequacy of data for this dataset is not an issue for successfully training the models. All the models achieved 100% accuracy on these subset of data. The binary classification of 16 different classes from 8 various plant species using models from scratch and fine tuned standard models is compared. Fine-tuned models proved to be effective irrespective of imbalance of data between classes or inadequacy of data for training. Fine-tuned MobileNet outperformed all other versions of MobileNets

TABLE III
COMPARING ACCURACIES ON PLANTLEAF DATASET FOR MODELS CREATED FROM SCRATCH

Plant (Healthy/Disease)	VGG19	MobileNet	Reduced-Mobilenet	Optimized-MobileNet
Apple (Heathy/Blackrot)	99.756	**99.53**	**99.53**	99.1
Banana (Healthy/Speckle)	98.95	98.95	**99.44**	99.39
Cabbage (Healthy/Blackrot)	**100**	**100**	**100**	**100**
Cherry (Healthy/Powderymildew)	**100**	**100**	**100**	**100**
Corn (Healthy/Commonrust)	99.28	99.65	**99.82**	99.65
Grape (Healthy/Leafblight)	**100**	**100**	99.29	**100**
Pepper (Healthy/Bacterialspot)	99.34	99.67	99.65	**100**
Squash (Healthy/Powderymildew)	**100**	**100**	**100**	**100**

TABLE IV
COMPARING ACCURACIES OF FINE-TUNED MODELS ON PLANTLEAF DATASET

Plant (Healthy/Disease)	VGG19	InceptionV3	MobileNet	NasNetMobile	DenseNet121
Apple (Heathy/Blackrot)	**99.56**	99.1	99.53	99.1	99.17
Banana (Healthy/Speckle)	99.37	99.6	**99.77**	99.6	99.17
Cabbage (Healthy/Blackrot)	**100**	**100**	**100**	**100**	**100**
Cherry (Healthy/Powderymildew)	**100**	**100**	**100**	**100**	**100**
Corn (Healthy/Commonrust)	99.46	**100**	99.82	99.82	99.82
Grape (Healthy/Leafblight)	**100**	**100**	**100**	**100**	**100**
Pepper (Healthy/Bacterialspot)	**100**	**100**	**100**	**100**	**100**
Squash (Healthy/Powderymildew)	**100**	**100**	**100**	**100**	**100**

while classifying three sets of corn leaf images. The higher classification accuracies of healthy and diseased leaf images compared to two diseased leaf images indicate that the intra-class variance is higher than inter-class variance on the latter imageset. The proposed deep learning approach showed higher efficacy on available dataset and its potential depends on quality and quantity of available data. This paper explored the potential of efficient network architecture and various network models that can be easily matched to the design requirements for mobile and embedded vision applications.

To overcome the disadvantage of conventional CNNs which demand high hardware cost for training the models and for real-time implementation, massive parallelization on Graphical Processing Units (GPUs) or other deep-learning-specialized hardware units, such as the recently introduced Tensor Processor Units (TPUs) are introduced. From the network design point of view, the processing speed might be increased by quantization or pruning or special deep learning architectures, such as MobileNets. This study explored various deep learning models and transfer learning techniques by creating and comparing their efficacy on plant leaf images, while keeping computational demands small making it suitable for real-time application.

FUNDING

The research presented in this article was supported by the National Natural Science Foundation of China (Grant Nos. 61471142, 61601145, 61571167 and 61871157).

COMPETING INTERESTS

The authors declare no competing financial interests.

REFERENCES

[1] D. Tilman, C. Balzer, J. Hill, and B. L. Befort, "Global food demand and the sustainable intensification of agriculture," *Proceedings of the National Academy of Sciences*, vol. 108, no. 50, pp. 20 260–20 264, 2011. [Online]. Available: https://www.pnas.org/content/108/50/20260

[2] D. Ray, N. Mueller, P. West, and J. Foley, "Yield trends are insufficient to double global crop production by 2050," *PloS one*, vol. 8, p. e66428, 06 2013.

[3] D. Pimentel and M. Burgess, "Soil erosion threatens food production," *Agriculture*, vol. 3, no. 3, pp. 443–463, 2013.

[4] E.-C. Oerke, "Crop losses to pests," *The Journal of Agricultural Science*, vol. 144, no. 1, pp. 31–43, 2006.

[5] S. Savary, A. Ficke, J.-N. Aubertot, and C. Hollier, "Crop losses due to diseases and their implications for global food production losses and food security," *Food Security*, vol. 4, 12 2012.

[6] P. Ronald, "Plant genetics, sustainable agriculture and global food security," *Genetics*, vol. 188, pp. 11–20, 2011.

[7] F. Yandun Narvaez, G. Reina, M. Torres-Torriti, G. Kantor, and F. A. Cheein, "A survey of ranging and imaging techniques for precision agriculture phenotyping," *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 6, pp. 2428–2439, 2017.

[8] H. Goëau, P. Bonnet, A. Joly, N. Boujemaa, D. Barthélémy, J.-F. Molino, P. Birnbaum, E. Mouysset, and M. Picard, "The imageclef 2011 plant images classi cation task," in *ImageCLEF 2011*, 2011, pp. 0–0.

[9] X. He and Y. Peng, "Fine-grained image classification via combining vision and language," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017, pp. 7332–7340.

[10] S. Maji, J. Kannala, E. Rahtu, M. Blaschko, and A. Vedaldi, "Fine-grained visual classification of aircraft," Tech. Rep., 2013.

[11] A. Khosla, N. Jayadevaprakash, B. Yao, and L. Fei-Fei, "Novel dataset for fine-grained image categorization," in *First Workshop on Fine-Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition*, Colorado Springs, CO, June 2011.

[12] G. Van Horn, S. Branson, R. Farrell, S. Haber, J. Barry, P. Ipeirotis, P. Perona, and S. Belongie, "Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 595–604.

[13] M.-E. Nilsback and A. Zisserman, "Automated flower classification over a large number of classes," in *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*. IEEE, 2008, pp. 722–729.

[14] L. Yang, P. Luo, C. Change Loy, and X. Tang, "A large-scale car dataset for fine-grained categorization and verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3973–3981.

[15] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 11 1998.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012,

pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

[17] Y. LeCun, C. Cortes, and C. Burges, "Mnist handwritten digit database," *ATT Labs [Online]. Available: http://yann. lecun. com/exdb/mnist*, vol. 2, 2010.

[18] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. S. Bernstein, A. C. Berg, and F. Li, "Imagenet large scale visual recognition challenge," *CoRR*, vol. abs/1409.0575, 2014. [Online]. Available: http://arxiv.org/abs/1409.0575

[19] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, Oct 2010.

[20] U. B. Gewali, S. T. Monteiro, and E. Saber, "Machine learning based hyperspectral image analysis: a survey," *arXiv preprint arXiv:1802.08701*, 2018.

[21] D. Pujari, R. Yakkundimath, and A. S. Byadgi, "Svm and ann based classification of plant diseases using feature reduction technique," *IJIMAI*, vol. 3, no. 7, pp. 6–14, 2016.

[22] D. Majumdar, A. Ghosh, D. K. Kole, A. Chakraborty, and D. D. Majumder, "Application of fuzzy c-means clustering method to classify wheat leaf images based on the presence of rust disease," in *Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications (FICTA) 2014*. Springer, 2015, pp. 277–284.

[23] P. B. Padol and A. A. Yadav, "Svm classifier based grape leaf disease detection," in *2016 Conference on advances in signal processing (CASP)*. IEEE, 2016, pp. 175–179.

[24] Y. Dandawate and R. Kokare, "An automated approach for classification of plant diseases towards development of futuristic decision support system in indian perspective," in *2015 International conference on advances in computing, communications and informatics (ICACCI)*. IEEE, 2015, pp. 794–799.

[25] S. Prasad, S. K. Peddoju, and D. Ghosh, "Multi-resolution mobile vision system for plant leaf disease diagnosis," *Signal, Image and Video Processing*, vol. 10, no. 2, pp. 379–388, Feb 2016. [Online]. Available: https://doi.org/10.1007/s11760-015-0751-y

[26] M. Islam, A. Dinh, K. Wahid, and P. Bhowmik, "Detection of potato diseases using image segmentation and multiclass support vector machine," in *2017 IEEE 30th canadian conference on electrical and computer engineering (CCECE)*. IEEE, 2017, pp. 1–4.

[27] M. Ramakrishnan *et al.*, "Groundnut leaf disease detection and classification by using back probagation algorithm," in *2015 International Conference on Communications and Signal Processing (ICCSP)*. IEEE, 2015, pp. 0964–0968.

[28] B. Liu, Y. Zhang, D. He, and Y. Li, "Identification of apple leaf diseases based on deep convolutional neural networks," *Symmetry*, vol. 10, no. 1, p. 11, 2018.

[29] M. Brahimi, K. Boukhalfa, and A. Moussaoui, "Deep learning for tomato diseases: classification and symptoms visualization," *Applied Artificial Intelligence*, vol. 31, no. 4, pp. 299–315, 2017.

[30] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," *Frontiers in plant science*, vol. 7, p. 1419, 2016.

[31] S. Wallelign, M. Polceanu, and C. Buche, "Soybean plant disease identification using convolutional neural network," in *The Thirty-First International Flairs Conference*, 2018.

[32] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.

[33] X. He and Y. Peng, "Fine-grained image classification via combining vision and language," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5994–6002.

[34] A. Dubey, O. Gupta, P. Guo, R. Raskar, R. Farrell, and N. Naik, "Pairwise confusion for fine-grained visual classification," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 70–86.

[35] V. M. Araújo, A. S. Britto, A. L. Brun, A. L. Koerich, and L. E. Oliveira, "Fine-grained hierarchical classification of plant leaf images using fusion of deep models," in *2018 IEEE 30th International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2018, pp. 1–5.

[36] J. Champ, T. Lorieul, P. Bonnet, N. Maghnaoui, C. Sereno, T. Dessup, J.-M. Boursiquot, L. Audeguin, T. Lacombe, and A. Joly, "Categorizing plant images at the variety level: Did you say fine-grained?" *Pattern Recognition Letters*, vol. 81, pp. 71–79, 2016.

[37] M. Šulc and J. Matas, "Fine-grained recognition of plants from images," *Plant methods*, vol. 13, no. 1, p. 115, 2017.

[38] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[39] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *CoRR*, vol. abs/1409.4842, 2014. [Online]. Available: http://arxiv.org/abs/1409.4842

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016, pp. 770–778.

[41] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," *CoRR*, vol. abs/1610.02357, 2016. [Online]. Available: http://arxiv.org/abs/1610.02357

[42] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *CoRR*, vol. abs/1704.04861, 2017. [Online]. Available: http://arxiv.org/abs/1704.04861

[43] R. Caruana, "Multitask learning," *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, Jul. 1997. [Online]. Available: https://doi.org/10.1023/A:1007379606734

[44] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A Large-Scale Hierarchical Image Database," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.

[45] K. Kamal, Z. Yin, M. Wu, and Z. Wu, "Depthwise separable convolution architectures for plant disease classification," *Computers and Electronics in Agriculture*, vol. 165, p. 104948, 2019.

[46] K. Kamal, Z. Yin, B. Li, B. Ma, and M. Wu, "Transfer learning for fine-grained crop disease classification based on leaf images," in *2019 10th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS)*. IEEE, 2019, pp. 1–5.

Automatic Atlas Generation: Development of a Plugin for QGIS

Bensaid, A., Slaoui E., Nedjai, R., Al-Juhaishi M.

Abstract**—**the proposed plugin allows the generating of an automatic atlas of spatially referenced data layers, by free or supervised crossing of several thematic parameters in the form of a graphical interface. It is developed in Python language to be integrated directly into the environment of the Qgis 3 geographic information system. In the end, this tool offers the opportunity for different users of map production at the national level (Water Agency, planning agency, local authority, design offices) to generate an atlas of maps in an automatic way. The Atlas offers a panorama of several thematic maps: (demography, employment, housing, disability, land use, etc.). In the long run, the result will be the creation of a dynamic atlas adapted to the needs of the researcher as well as those of the actors in charge of the management and planning of the territory. One of the advantages of this plugin is that it is quite possible to integrate geographic data into a Webmapping interface. It should also be noted that this tool allows you to connect to a PostGres database remotely for the making of the maps.

Keywords**—** Atlas, Card, GIS, Plugin, Python

A. Bensaid is with the Department of Geography, University of Orleans France (phone: 336 69 25 53 91; e-mail: abdelkrim.bensaid@univ-orleans.fr).

E. Slaoui student at the Department of Geography, University of Orleans France (phone: 332 38 49 24 90;e-mail: el-mehdi.slaoui@etu.univ-orleans.fr).

R. Nedjai is with the Department of Geography, University of Orleans France (phone: 332 38 49 24 90; e-mail: rachid.nedjai@univ-orleans.fr).

M. Al-Juhaishi is with, Computer Center and Department of Civil Engineering, University of Baghdad, Iraq (mrdh74@yahoo.com).

Applying the Crystal Model Approach on Light Nuclei for calculating radii and density distribution

A. Amar

***Abstract*—** New model namely crystal model has been modified to calculate radius and density distribution of light nuclei up to ^{8}Be. The crystal model has been modified according to solid state physics which uses the analogy between nucleon distribution and atoms distribution in the crystal. The model has analytical analysis to calculate the radius where the density distribution of light nuclei has obtained from analogy of crystal lattice. The distribution of nucleons over crystal has been discussed in general form. The equation has been used to calculate binding energy was taken from solid state model of repulsive and attractive force. The numbers of the protons were taken to control repulsive force where the atomic number was responsible for the attractive force. The parameter has been calculated from crystal model was found to be proportional to the radius of nucleus. The density distribution of light nuclei was taken a summation of two clusters distribution as in ^{6}Li=alpha+deuteron configuration. A test has been done on the data obtained for radius and density distribution using double folding for d+^{6,7}Li with M3Y nucleon-nucleon interaction. Good agreement has been obtained for both radius and density distribution of light nuclei. The model failed to calculate the radius of ^{9}Be, so modifications should be done to overcome discrepancy.

***Keywords*—** nuclear physics, Nuclear Lattice, study nucleus as crystal, light nuclei till to ^{8}Be.

I. Introduction

Here new philosophy of the scientific outcome should be changed by the time. Quantum mechanics still one suitable and sophisticated tool to give us a scope on the structure of the nucleus. As mentioned before, the light nuclei tend to form clusters as for ^{6,7}Li and ^{7,8,9}Be nuclei [1,2]. In an article review [3] the author has discussed the anthropic considerations and the role of Hoyle-state and its closeness of the triple-alpha process to the energy of ^{4}He+^{8}Be. Many subjects have been discussed in previous review which discussed the nuclear forces and chiral symmetry. Monte Carlo and ab-initio solutions have been used to study nuclear systems up to four nucleons. The quark mass dependence on nuclear forces is fulfilled to about 95%. For a long time, it was a goal of nuclear physics and astrophysics to find the mechanism of nuclear reactions and the structure of the nucleus for light, intermediate and heavy nuclei [4]. A lot of attempts have been done on nuclear systems to obtain and discuss the ambiguities of nuclear parameters like coupled-cluster approach [5,6] ab-initio calculations [7,8] and Coupled channel scattering [9-12] which have been studied by using lattice calculations at crystal lattice range in Fermi scale. Variation Monte Carlo method also, has been used in the same track [13,14]. Guatam and Dean Lee paved the way for general method for all nuclear reactions calculations on the crystal lattice. They have been studied the neutron radiative reaction (n,γ) on the lattice, they also included halo nuclei like ^{14}C(n, γ)^{15}C as example of applicability of their theory by calculating photo-nuclear reaction rates lattice [4]. Michelle Pine et al. [15] have demonstrated and tested the adiabatic projection method on neutron-deuteron scattering. They calculated s-wave phase shift and inelastic scattering reactions on the lattice for neutron-deuteron scattering.

The cluster structure was a helpful phenomenon that scientists used to simplify their studies. Ulf Meißner has discussed clustering using ab-initio method on the lattice, he focused on alpha clustering because its properties as high binding energy and spin (iso-spin) saturated. Nuclei like ^{12}C, ^{16}O and ^{24}Mg have been included in the study [16].The authors have asked a question; how large a nucleus can be calculated from first principles using the framework of chiral nuclear EFT and what are the remaining challenges [17]; where they answered their question by calculating ground states of α-type nuclei from ^{4}He to ^{28}Si. The Nuclear Lattice (EFT) or Nuclear Lattice Simulations method has been used by authors of the previous article succeeded in explanation of clustering in lattice without any constrains which can be used as a benchmark for ab-initio calculations of larger nuclei using chiral nuclear EFT. Matter nuclear distribution for ^{6}Li, ^{6}He and ^{9}Be has been discussed theoretically using the three body models in the [18].

Binding energy and nuclear density distribution have been calculated for light nuclei depending on alpha as a core of a cluster. They choose ^{6}Li, ^{6}He as (α+n+n) and ^{9}Be as α+α+n where in case of ^{6}Li it was an approximation to take the neutron instead of the proton in the configuration. The authors have summarized the form of nuclear density distribution of light nuclei in the form of summation of two Gaussian in the [19]. The authors in [1] have concluded the possibility of using dynamic cluster in the future to study ^{8}Be and ^{12}C.The research [20] has been performed to study nucleus as crystal and cluster is dominated for light till to ^{8}Be. Our work is devoted to study the form of expected nuclear density distribution for light nuclei.

II. Crystal Model Approach

The potential between nucleons inside the nucleus will be calculated using the equations, which have been used to calculate a parameter proportional to the radius of the nucleus [21]:

$$U = (4\varepsilon)\left(\frac{a}{r^{12}} - \frac{b}{r^{7}}\right) \quad (1)$$

A. Amar is with the Physics Department, Faculty of Science, Tanta University, Tanta, Egypt (corresponding author, phone: 002-01022168487; e-mail: amar.physics@yahoo.com).

where a is taken to be the number of the protons and b is the mass number of the nucleus and ε is taken an arbitrary constant. The first term of equation (1) has been used as the repulsion force where second term is taken to be the attractive force. The crystal model has been developed to calculate a parameter that proportional to the radii of light nuclei and to expect the form of matter density distribution of light nuclei. The cluster model and crystal model are close to each other where the distribution of nucleons inside nuclei in our point of view as the distribution of atoms on crystal. ^{6}Li as example contains alpha as a base and two nucleons on the upper corners of crystal one proton on the corner and on the other corner neutron is located (Deuteron is located on the upper face of the crystal, and alpha is on the lower face.) as shown in Fig. 2, and 3. The density distribution of ^{6}Li could be taken as a summation of two densities of alpha and deuteron of the general form:

$$\rho(r) = A\{B\exp(-C) + D\exp(-E) - F\} \quad (2)$$

where A, B, D and F are constants where C and E are functions of r . Let $A = \rho_0$, B=D=F=1 and hence equation (2) could be modified to take the form:

$$\rho(r) = A\{[1 + \exp(-B)] + [1 + \exp(-B)] - 1\} \quad (3)$$

where $A = \rho_0$, and B is a function of r. For single parameterization, the nucleon density distribution in the nucleus is approximated by:

$$\rho(r) = \left(\frac{3}{2\pi R_m^2}\right)^{3/2}\exp\left(\frac{-3r^2}{2R_m^2}\right) \quad (4)$$

where R_m is the rms of matter radius of the nucleus. The standard Fermi of the nuclear matter distribution is written in the form:

$$\rho(r) = \rho_0\left(1 + e^{\frac{(r-R)}{a}}\right)^{-1} \quad (5)$$

which is suitable for description of intermediate and heavy nuclei. For light nuclei, it is suitable to modify eq. (3) to be symmetrized in Fermi-density [17]:

$$\rho(r) = \rho_0\{(1 + e^{\frac{(r-R)}{a}})^{-1} + (1 + e^{\frac{(r-R)}{a}})^{-1} - 1\} \quad (6)$$

$$\rho(r) = \rho_0 \frac{\mathrm{Sinh}\left(\frac{R}{a}\right)}{\mathrm{Cosh}\left(\frac{R}{a}\right) + \mathrm{Cosh}\left(\frac{r}{a}\right)} \quad (7)$$

With normalization

$$\rho_0 = \frac{3}{4\pi R^3}\frac{1}{1+\left(\frac{\pi a}{R}\right)^2} \quad (8)$$

where R is the half-density radius.

If we compare between equation (6) and equation (3) from crystal model approach we can find that first term of eq. 9 $[\rho_0(1 + e^{\frac{(r-R)}{a}})^{-1}]$ represents the density of alpha in our model as simple Gaussian distribution where second and third term $\{\rho_0[\left(1 + e^{\frac{(r-R)}{a}}\right)^{-1} - 1]\}$ is represent the density distribution of deuteron. Where B is taken to be $\frac{(r-R)}{a}$ from the analogy with our expectation and those parameters from [17]. R_m (rms of nuclear matter radius) is related to R and a by the relation:

$$\rho_0 = \left(\frac{3}{2}\right)^{1/2} R\left[1 + \left(\frac{7}{3}\right)\left(\frac{\pi a}{R}\right)^2\right]^{\frac{1}{2}} \quad (9)$$

In another form, from the principles of nuclear structure which states that overlapping between two constituents of nucleus is forbidden [18]. The density distribution of many body systems may be written:

$$\rho(R) = \sum_{\mathrm{cluster=i,j,k}} \rho_{\mathrm{cluster}}(R) \quad (10)$$

Applying equation (10) on ^{7}Li for example, it will be written as:

$$\rho(R) = \rho_\alpha(R) + \rho_t(R) \quad (11)$$

Also, the calculations of alpha nucleus have been done here using the relation:

$$\rho(r) = 0.17\{(1 + e^{\frac{(r-1.97)}{0.65}})^{-1}\} \quad (12)$$

where R=1.97fm, a=0.65fm, and ρ_0=0.17(1/fm3).

III.. Results and Discussion

A. Calculating Binding Energy for Light Nuclei

By imagining the nucleus as a crystal, where the nucleons are arranged on the corners of the crystal with analogy of atoms in the solid-state physics. The model has been developed by us to study light nuclei [20] succeeded to explain up to alpha particles. For nuclei heavier than alpha, the model should be modified to explain the distribution of the nucleons inside the nucleus. The distribution of the nucleons in case ^{8}Be as hexagonal crystal is discussed. The two alphas will be taken on the up and down faces of hexagonal crystal. The situation of ^{8}Be is very interesting as the life-time is about 10^{-16} sec which reflects instability of such nuclei. Dynamic cluster model should be applied in case of instability from mechanical force. The distribution of nucleons inside the nucleus starts with cubic (on the corners of the crystal as one proton beside one neutron). Tin has special case looks like ^{8}Be nucleus where it experiences allotropic change. There is a change in color from white tin into gray by changing the temperature which introduces another crystal structure. The same situation takes place in case of ^{8}Be nucleus where the structure in the moment of formation is cubic and during about 10^{-16} sec, the structure changes into hexagonal close packed in a process known as dynamic cluster of the nucleus. The change in structure inside the nucleus of ^{8}Be increases the volume of it, decreases the nuclear density of ^{8}Be nucleus. As in tin, the change of its structure rumble the buttons of soldier's clothes, also, ^{8}Be will disintegrate into two alphas. The shape of cubic will change from the energy loss which produces stretching. The shape of the nucleus in case of ^{8}Be will change from cubic into hexagonal closed packed (HCP) as shown in Fig.1 where two alphas will occupy the two faces of HCP. The radius of ^{8}Be have been calculated from equation (1) to be 0.8697992638 where U was taken to be Q-value as shown in Table 1 compared with calculated radii from literature [22]. A new parameter α (R_1/R_2) has been calculated by dividing the radius of the nucleus taken from literature and those calculated from the crystal model in the present work. Monte Carlo transfer matrix algorithm has been achieved to give a probe that atomic number A≤8 is completely different from higher atomic number [23]. The value of transformation parameter α (R1/R2) has been calculated from the crystal model of nucleus (CMN) is ~2.2-3.5 which has low value at

lower binding energy as in case of deuteron 2.59. As seen in Table 1, the calculated value of transformation parameter for ^{7}Li is smaller than that value calculated for ^{6}Li, which agrees with the values of them in the literature [2, 24, 25] (see Table 1).

The aim now is to calculate the parameters of the hexagonal close packed (HCP).

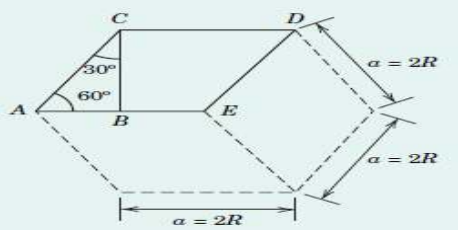


Fig.1. Determination the parameters of hexagonal close packed

The volume of unit cell could be obtained by the product of the base area times the cell height, c. The lattice parameter (a) has a relation with the nucleon radius r as:

$$a=2r \quad (13)$$

$$V_c = 6r^2c\sqrt{3} \quad (14)$$

The volume of ^{8}Be nucleus has been calculated before and after separation into two alphas. r is taken to be 1 to 1.25fm and c is about two alphas radius to be 2×1.6755fm= 3.351fm. So, the minimum volume of ^{8}Be is:

$$V_c = 6r^2c\sqrt{3} = 6 * 1.25 * 1.25 * 3.351 * \sqrt{3} = 54.41\ fm^3 \quad (15)$$

Dynamical torque is required to break down ^{8}Be into two alphas; the dynamical torque is responsible for separation of two alphas from each other. The secrete on such configuration arises from the distance c between two alphas arranged on the faces of hexagonal close packed. The large distance between the two alphas gives an opportunity for two alphas to separate from each other. It has been mentioned that decreasing on the temperature lead to changing the color of tin. The same situation arises at ^{8}Be because the changing on the allotropic transformation produces stretching and dynamic torque leads to separate ^{8}Be into two alphas. Another approach could be applied if the transformation parameter (α) is used for ^{8}Be shown in Table 1 used to calculate the radius of ^{8}Be on the separation moment; which was found in the range 3.05fm. The radius of ^{8}Be on the moment of separation into two alphas is nearly equal the radius of ^{20}Ne which contains more than the twice number of nucleons of ^{8}Be nucleus and equal ^{24}Mg which three times of nucleon number of ^{8}Be. The solution of equation one succeeded to give us the transformation parameter (α) which proportional to the radius of nuclei from deuteron up to ^{8}Be as shown in Table 1. The binding energy of 8Be has been calculated from equation (1). There is a condition which should be applied on ^{8}Be, which is r≤3.5fm. The binding energy of ^{8}Be nucleus that calculated from crystal model approach: $U = (4 \times 10^3)\left(\frac{4}{(R*\alpha)^{12}} - \frac{8}{(R*\alpha)^{7}}\right)$ which applied on the nuclei under consideration with R and α taken from Table 2. The binding energy of 8Be has been calculated:

$$U = (4 \times 10^3)\left(\frac{4}{(3.35)^{12}} - \frac{8}{(3.35)^{7}}\right) = 52\text{MeV} \quad (16)$$

Where U is binding energy of ^{8}Be and ε is extracted from experimental data from literature and has the value 103 MeV.

B. Matter Density Distribution for Light Nuclei

An approximation has been applied on the nuclei using crystal model approach that the density distribution ^{6}Li, ^{7}Li, and ^{7}Be nuclei are taken as a summation of two Gaussian distributions as one for the first cluster is alpha and the other part of nucleus is the residual cluster of nuclei. For example, ^{6}Li is taken as d+alpha or t+^{3}He (two configurations with the same probability) and the density distribution of ^{6}Li now is a summation of two Gaussian one for deuteron and the other for alpha. That is could be described for ^{7}Li≡α+t configuration as in cluster structure and hence alpha will be the core of the nucleus and triton as valence (See Fig. 2 and 3).

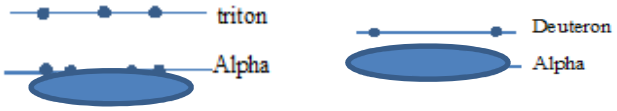


Fig. 2. Cluster model distribution for ^{7}Li and ^{6}Li

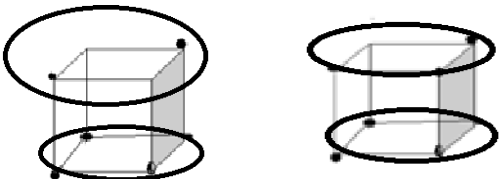


Fig. 3. Distribution of the nucleons on crystal model for ^{6,7}Li where a) the distribution of the nucleons of ^{6}Li as alpha on the base and deuteron on the upper face, b) ^{7}Li case with alpha on the base and triton on the upper face

The imagination of ^{6}Li in the present work agrees with the phenomenological charge and matter distributions and the shell-model distribution have the form [26]:

$$A\rho(r) = a\alpha^3 e^{-\alpha^2 r^2} + (b + c\beta^2 r^2)\beta^3 e^{-\beta^2 r^2} \quad (17)$$

The constants of ^{6}Li were taken from Table 3.

Also, density distribution of ^{7}Li nucleus was taken in the form which is two summations of Gaussian [27]:

$$\rho(R) = \rho_0\left(1 + \alpha(\frac{R}{a})^2\right)\exp\left(-(\frac{R}{a})^2\right), \quad (18)$$

Where for harmonic oscillator a=1.77fm and α=0.327fm. ^{6}Li is taken in crystal model approach to be alpha on the base of the lattice and deuteron on the upper face of crystal which is shown in Fig. 4.

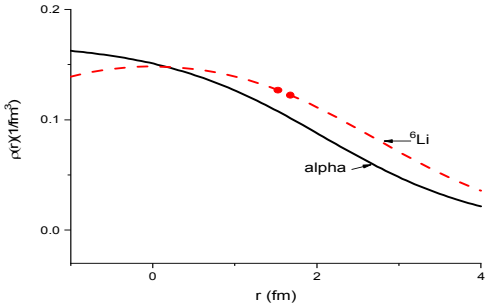


Fig. 4. Matter density distribution ^{6}Li nucleus from crystal model approach

The calculations have been done for ^{6}Li (^{6}Li≡α+d):

$$\rho(r) = 0.17\{(1+e^{\frac{(r-2.5)}{1.2}})^{-1} + (1+e^{\frac{(-r-2.5)}{1.2}})^{-1} - 1\} \quad (5')$$

where R=2.5fm, a=1.2fm (as very smooth nucleus) and ρ_0=0.17(1/fm^3). Also, the calculations of alpha nucleus have been done here using the relation:

$$\rho(r) = 0.17\{(1+e^{\frac{(r-1.97)}{0.65}})^{-1}\} \quad (6')$$

where R=1.97fm, a=0.65fm, and ρ_0=0.17(1/fm^3).

Double folding calculations have been applied on deuteron elastically scattered by ^{6,7}Li using density distribution from the present analysis as shown in Fig. 5 and 6.

Double folding model has been used to calculate microscopic potential in the present work. The real part of the calculated potential is given in the following form [28]:

$$V_F(r) = \int dr_p \int dr_t \rho_p(r_p)\rho_t(r_t)v(r_{pt}) \quad (19)$$

Where ρp is the matter density distribution of the projectile, ρt is the matter density distribution of the target and rpt = r + rt - rp. M3Y form has been used for nucleon-nucleon interaction at the present calculation:

$$v(r) = 7999\frac{\exp(-4r)}{4r} - 2134\frac{\exp(-2.5r)}{2.5r} - 276(1-\frac{0.005E}{A_p})\delta(R) \quad (20)$$

The form of nuclear central optical potential during double folding model calculations was taken in the shape: U(r)=VF(r)+iW(r), where VF(r) is calculated from (19) and W(r) is phenomenological Woods-Saxon potential. The imaginary part was included into the Fresco program [29] to reproduce differential cross section.

Double folding calculations for ^{6}Li+d with alpha transfer has been done using density distribution of ^{6}Li from crystal model approach at deuterons energy 25MeV and 14.7MeV was found to be in a satisfied manner with experimental data as shown in Fig. 5 and 6. The importance of using double folding potential appears at backward angles and the effect of spectroscopic amplitudes is very obvious for ^{7}Li elastically scattered by deuteron.

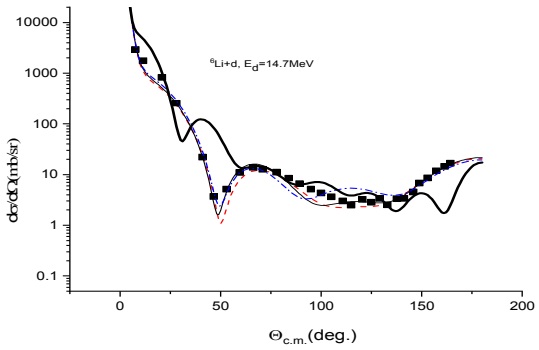


Fig. 5. Represents double folding calculations for ^{6}Li+d for elastic scattering with alpha transfer using density distribution of ^{6}Li from crystal model approach at deuterons energy 14.7MeV where solid line represents CM, dash line is optical model and dash-dot line represents coupled reaction channel. Experimental data were taken from [30].

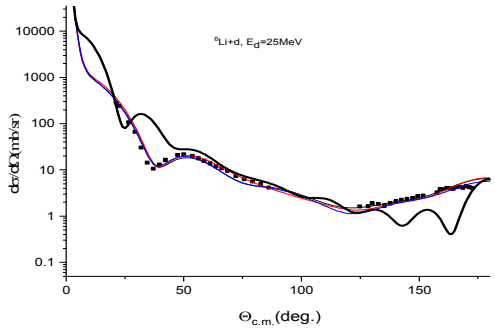


Fig. 6. Represents double folding calculations for ^{6}Li+d for elastic scattering with alpha transfer using density distribution of ^{6}Li from crystal model approach at deuterons energy 25MeV where solid line represents CM, dash line is optical model and dash-dot line represents coupled reaction channel. Experimental data were taken from [31].

The density distribution of ^{7}Li nucleus is simpler than ^{6}Li as triton density distribution is not complicated as deuteron and will be taken in the form:

$$\rho(r) = \rho_0\{Aexp(-B) + Aexp(-C) - D\} \quad (21)$$

where ρ_0, A, and D are constants but B and C are functions of r. The same situation may be applied on ^{7}Li where the base is alpha and triton on the upper face of crystal as shown in Fig. 2 and 3. The density distribution of ^{7}Li is taken:

$$\rho(r) = 0.17\{(1+e^{\frac{(r-2.4)}{1.1}})^{-1} + (1+e^{\frac{(r-2.4)}{1.1}})^{-1} - 1\} \quad (22)$$

where R=2.4fm, a=1.1fm, and ρ^0=0.17(1/fm^3).

^{8}Be was assumed as alpha+alpha configuration that is because it decays into alpha+alpha during 10^{-16}s. The distribution of the nucleons on ^{8}Be is shown in Fig. 9 and 10. Dynamic cluster is needed to explain ^{8}Be decay into two alphas where crystal model could achieve such mission with analogy of solid state physics. The density distribution of ^{8}Be could be written in the form:

$$\rho(r) = \rho_0 A\{Bexp(-C) + Bexp(-C) - 1\} \quad (23)$$

The analysis of equation (8) gives an impression that the density distribution of ^{8}Be is two identical Gaussians with the form Bexp(-C).

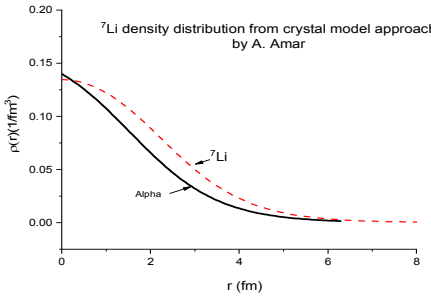


Fig. 7. Nuclear density distribution of ^{7}Li from the crystal model approach

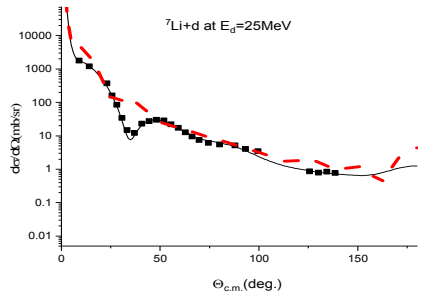


Fig. 8. Double folding calculations for deuteron elastically scattered by ^{7}Li with alpha transfer using density distribution of ^{7}Li from crystal model approach at deuterons energy 25MeV. Solid line represents CM and dash line is coupled reaction channel where experimental data were taken from [32].

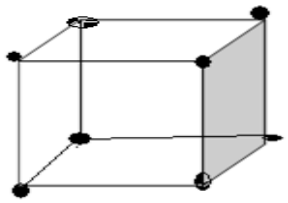


Fig. 9. The distribution of nucleons for ^{8}Be nucleus

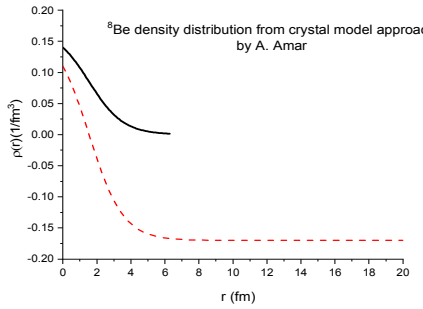


Fig. 10. Nuclear matter distribution of ^{8}Be nucleus using parameters from crystal model approach

$$\rho(r) = 0.17\{(1+e^{\frac{(r-1.97)}{0.65}})^{-1} + (1+e^{\frac{(r-1.97)}{0.65}})^{-1} - 1\}$$

General form for light nuclei

$$\rho(r) = 0.17\{(1+e^{\frac{(r-1.97)}{0.65}})^{-1} + (1+e^{\frac{(r-1.97)}{0.65}})^{-1}\} \quad (24)$$

^{8}Be density distribution is just summation of two alphas. Where R=1.97fm, a=0.65fm, and ρ_0=0.17(1/fm^3). So, from the matter density distribution for ^{8}Be we can expect the continuum state of ^{8}Be which reflects the impossibility of staying two alphas together. ^{7}Be is stable nucleus as it contains four protons and three neutrons with the configuration ^{7}Be≡^{4}He+^{3}He. The density distribution of alpha is just a simple Gaussian distribution {Aexp(-B)} where A, B and C are functions on r in the form:

$$\rho(r) = A\{exp\,(-B) + exp(-C) - 1\} \quad (25)$$

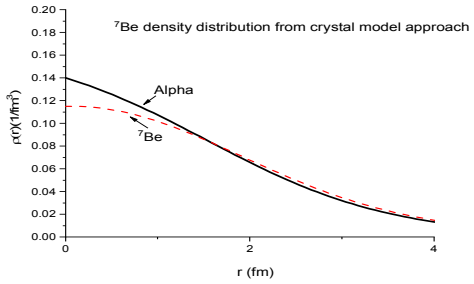


Fig. 11. ^{7}Be density matter distribution from crystal model approach

The same calculations will be applied on ^{7}Be and the results are presented as:

$$\rho(r) = 0.17\{(1+e^{\frac{(r-1.97)}{0.65}})^{-1} + (1+e^{\frac{(r-1.97)}{0.65}})^{-1} - 1\} \quad (26)$$

Hence the density distribution of light nuclei could be studied using crystal model approach. In the near future further study will be done for ^{12}C, ^{16}O and ^{20}Ne. As discussed in [20], the matter density is a summation of two clusters forming the light nucleus.

IV. Conclusion

The crystal model approach produces an imagination for nuclear structure and arrangement of nucleons inside light nuclei. Transformation parameters which proportional to the radii of nuclei have been calculated using crystal model approach. The crystal lattice space has been calculated for different nuclei. The binding energies for light nuclei have been calculated. The distribution of nucleons on the space of crystal has been discussed in satisfied manner which reflects the success of the model to interpret phenomena in nuclear systems such as ^{8}Be decay. Density distribution also has been predicted according to our model, crystal model, for such nuclei under investigation. The model failed to calculate the radius of 9Be, so modifications should be done on equation one to overcome discrepancy. Double folding calculations have been done for deuterons elastically scattered by ^{6,7}Li using matter densities distribution from crystal model approach.

TABLE I

CALCULATED PARAMETERS FROM CRYSTAL MODEL OF NUCLEUS APPROACH (CMN)

Z	A	R1 from Literature [22]	R2 parameter calculated from CMN)	(a) lattice spacing	U or (Q-vale) MeV ref.[24,25]
1	2 (deuteron)	2.1421 fm	0.8270	2.52fm	2.21
2	4 (Alpha)	1.6755 fm	0.7515	1.970fm	28.4
3	6 (Li)	2.589 fm	0.7690	2.1fm	32.1
3	7 (Li)	2.444 fm	0.7530	2.24fm	39.2
4	7 (Be)	2.646 fm	0.7718	2.24fm	46.55
4	8 (Be)	---	0.8690		53.2
4	9 (Be)	2.519 fm	0.7501	2.141fm	58.2

TABLE II
TABLE 2. EXTRACTED BINDING ENERGY FROM CM

Z	A	Extracted value of α from CMN (present work)	Binding energy From Present work in (MeV)
1	2 (deuteron)	3.35	*2.193*
2	4 (Alpha)	2.5	*27.79*
3	6 (Li)	2.6	*30.33*
3	7 (Li)	2.6	*35.02*
4	7 (Be)	2.4	*51.01*
4	8 (Be)	2.5	*52*

TABLE III
PARAMETERS DEFINING GROUND STATE DISTRIBUTIONS OF ^{6}LI

Distribution	a	b	c	α (fm-1)	β (fm-1)
Aρph	1.07	-0.19	0.130	0.5388	0.384
Aρm	1.07	-0.20	0.138	0.575	0.398
AρSM	0.71	0	0.239	0.613	0.505

ACKNOWLEDGMENT

We acknowledge helpful discussions with Prof. Dr. Osama Hemada, Prof. Dr. Reda Elshater. We acknowledge Dr. Ahmed Elbanna and Prof. Dr. Mohammed Shehab as Computing process has been provided by them.

REFERENCES

[1] R. Guardiola , I. Moliner , M.A. Nagarajan , "Alpha-cluster model for ^{8}Be and ^{12}C with correlated alpha particles," Nuclear Physics A 679, 2001, 393–409.

[2] A. Amar, "Spectroscopic information of ^{6}Li from d, ^{3}He and alpha elastically scattered by ^{6}Li," International Journal of Modern Physics E Vol. 23, No. 08, 1450041, 2014.

[3] Ulf-G. Meißner, Sci. Bull. (2015) 60(1):43–54, Physics & Astronomy, Dean Lee, Ulf-G. Meißner , Keith A. Olive, Mikhail Shifman , and Thomas Vonk,"θ-dependence of light nuclei and nucleosynthesis," Physical Review Research 2, 033392, 2020.

[4] Dean Lee, Gautam Rupak, "Radiative capture reactions in lattice effective field theory," Physical Review Letters, PRL111, 032502, 2013.

[5] O. Jensen, G. Hagen, T. Papenbrock, D. Dean, J. Vaagen, "*Computation of spectroscopic factors with the coupled-cluster method*," Phys. Rev. C 82, 014310, 2010.

[6] G. Hagen, N. Michel, " Elastic proton scattering of medium mass nuclei from coupled-cluster theory," Phys. Rev. C 86, 021602, 2012.

[7] Petr Navrátil, R. Roth, S. Quaglioni, "Ab initio many-body calculation of the ^{7}Be(p,g)^{8}B radiative capture," Phys. Lett. B 704, 379, 2011.

[8] Petr Navrátil, S. Quaglioni, "*Ab Initio* Many-Body Calculations of the H3(d,n)He4 and He3(d,p)He4 Fusion Reactions," Phys. Rev. Lett. 108, 042503, 2012.

[9] M. Lage, U.-G. Meißner, and A. Rusetsky," A method to measure the antikaon–nucleon scattering length in lattice QCD," Phys. Lett. B 681, 439, 2009.

[10] V. Bernard, M. Lage, U.-G. Meißner, and A. Rusetsky, "Scalar mesons in a finite volume, J. High Energy Phys. 01, 2011, 019.

[11] H. B. Meyer, "Photodisintegration of a Bound State on the Torus," arXiv:1202.6675.

[12] R.A. Briceno and Z. Davoudi, "Moving Multi-Channel Systems in a Finite Volume with Application to Proton-Proton Fusion," arXiv:1204. 1110.

[13] K.M. Nollett, R. Wiringa, "Six-body calculation of the α-deuteron radiative capture cross section," Phys. Rev. C 84, 024319 83, 041001, 2011.

[14] I. Brida, S.C. Pieper, R. Wiringa, Phys. Rev. C 2011.

[15] Michelle Pine et al., Eur.Phys. J. A, 2013, pp. 49-151.

[16] Ulf-G. Meißner, The 8th International Workshop on Chiral Dynamics, CD2015 29 June 2015 - 03 July 2015 Pisa,Italy.

[17] T.A. Lähde, E. Epelbaum, H. Krebs, D. Lee, U.-G. Meißner and G. Rupak, "Lattice Effective Field Theory for Medium-Mass Nuclei," Phys. Lett. B 732 (2014) 110 [arXiv:1311.0477 [nucl-th].

[18] B.A. Urazbekov, A.S. Denikin, Eurasian Journal of Physics and Functional Materials, 2018, 2, pp. 17-22.

[19] Yu. N. Eldyshev et al, Sov. Jour. Nucl. Phys. 16, 1973, pp. 282.

[20] A. Amar, O. Hemada, Study an approach of crystal model, World Academy of Science, Engineering and Technology, International Journal of Nuclear and Quantum Engineering,Vol:15, No:2, 2021.

[21] James F. Shackelford, Introduction to Materials Science for engineer, Pearson; 8 edition (April 12, 2014).

[22] I. Angelia, K.P. Marinova, Atomic Data and Nuclear Data Tables 99 2013, pp. 69–95.

[23] Bu Borasoy, Evgeny Epelbaum, Hermann Krebs, Dean Lee, Ulf-G. Meißner, R. Bijker, F. Iachello, Phys. Rev. Lett. 112, 2104, 15, 152501.

[24] Ghahramany et al. Journal of Theoretical and Applied Physics 2012, 6:3, http://www.jtaphys.com/content/6/1/3, New approach to nuclear binding energy in integrated nuclear model.

[25] D.R. Tilley , J.H. Kelley, J.L. Godwina,, D.J. Millener, J.E. Purcell, C.G. Sheua,, H.R. Weller Nuclear Physics A 745, 2004, pp. 155–362.

[26] K.H. Bray T, Mahavir Jain Tt, K.S. Jayaraman, G. Lobianco, G.A. Moss W.T.H. Van Oers And D.O. Wells F. Petrovich, Nuclear Physics A l89, 1972, pp. 35-64.

[27] B. A. Watson, P.P. Singh, and R.E. Segal, Phys. Rev. Vol. 182, No. 4 1969, pp. 977-989.

[28] J. Cook, Nuclear Physics A382, 1982, pp. 61-70; P. E. Hodgson, The Deuteron-Nucleus Optical Potential, Nuclear Physics Laboratory, Oxford, 1966; A. Amar, Int. J. Modern. Phys. E, Vol. 23, No. 8, 2014, 1450041, 3: P. E. Hodgson, Rep. Prog. Phys. 34, 1971, pp. 765.

[29] I.J. Thompson, Fresco 2.0, Department of physics, University of Surrey, Guildford GU2 7XH, England, 2006.

[30] S. Matsuki, S. Yamashita, K. Fukunaga, D.C. Nguyen, N. Fujiwara, T. Yanabu, J, Jpj, 26, (6), 1344, 1969, 06.

[31] J. Arvieux, R,CEA-R-3392 ,1 , 196803.

[32] N. Burtebayeva et al. Acta Physica Polonica B, No. 5, Vol. 46, 2015.

A Real-Time Monitoring System of the Supply Chain Conditions, Products and Means of Transport

Dimitrios E. Kontaxis, George Litainas, Dimitrios P. Ptochos, Vaggelis P. Ptochos
Sotirios P. Ptochos, Dimitrios Beletsis, Konstantinos Kritikakis, Milan Sunaric

Abstract — Real-time monitoring of the supply chain conditions and procedures is a critical element for the optimal coordination and safety of the deliveries, as well as for the minimization of the delivery time and cost. Real time monitoring requires IoT data streams, which are related to the conditions of the products and the means of transport (e.g., location, temperature/humidity conditions, kinematic state, ambient light conditions, etc.). These streams are generated by battery-based IoT tracking devices, equipped with appropriate sensors, and are transmitted to a cloud-based back-end system. Proper handling and processing of the IoT data streams, using predictive and artificial intelligence algorithms, can provide significant and useful results, which can be exploited by the supply chain stakeholders in order to enhance their financial benefits, as well as the efficiency, security, transparency, coordination and sustainability of the supply chain procedures. The technology, the features and the characteristics of a complete, proprietary system, including hardware, firmware and software tools - developed in the context of a co-funded R&D program - are addressed and presented in this paper.

Keywords—IoT embedded electronics, real-time monitoring, tracking device, sensor platform.

I. Introduction

ONE of the largest and emerging markets worldwide is related to the services and the automation of the supply chain, aiming at the safe, cost-effective and seamless transport of goods anywhere in the world. RETIMO which stands for 'Real Time Monitoring' is a research program/project [1] which started in 2019 and was proposed, managed and developed by the Greek company 'PEOPLE Private Company' (PPC). The program is co-funded by the ESPA Partnership for Growth Framework 2014-2020, titled 'Research-Create-Innovate'. The target of the RETIMO project is the design and development of a complete system, which comprises innovative IoT hardware, firmware and software tools, to support supply chain stakeholders throughout the transportation and handling procedures of goods, towards the optimal coordination and safety of their delivery, as well as the minimization of the delivery time and cost. Over the last five years, PEOPLE has focused its research efforts on the global supply chain market and is the patent holder of the lightest metal Shipping Container in the world (the 'Container 2.0'), which contributes to a considerable reduction of the transportation costs and the environmental footprint of the supply chain.

Managing the transportation of huge volumes of goods presents challenges which can be tackled effectively only with the adoption of modern IoT technologies, based on smart tracking devices and information systems (IS), able to manage and process huge volumes of data (procedures known as 'big data analysis'), in order to provide significant and exploitable results, within a required timeframe. In the context of the RETIMO project, PEOPLE aimed at developing an 'eco-system' of proprietary tracking devices and an IS, which would support the vision for enhanced efficiency, security, transparency, coordination and sustainability of the supply chain. Therefore, the system consists of:

- smart IoT tracking devices, which can be installed on different entities of the supply chain, like containers, trains, trucks and pallets,
- a back-end system, which carries out data processing, analysis and presentation, and offers interconnection with existing solutions, like automated Warehouse Management Systems (ERP, WMS) and electronic monitoring of cargo and transportation procedures (e-freight, e-documents, etc.),

and provides the following features:

- grouping of the supply chain entities (containers, trains, trucks, pallets),
- periodic transmission of IoT data streams, which source form the IoT tracking devices,
- real-time alarms and notifications, sourcing from the IoT tracking devices and the back-end system, respectively,
- use of blockchain technology, in order to enhance the transparency and 'accountability' of the collected IoT data streams,
- data analytics and decision support, based on the collected IoT data streams, using predictive algorithms and artificial intelligence (AI, machine learning).

The paper is organized as follows: Section II provides an overview of the structure of the system developed in the context of the RETIMO project. Section III describes the technical characteristics and functionality of the IoT tracking devices. In Section IV, the technology of the back-end system, as well as the methodology followed for its development, is addressed. Section V concludes the paper.

II. System Description

The back-end system collects IoT data streams from the IoT tracking devices through the cellular network, exploiting the LTE Cat M1 or GPRS radio access technologies (RAT) and embedded SIMs (eSIM). The IoT data steams include device location and status. Data are analyzed and processed by the back-end system and results are presented to the end user, via an appropriate frond-end application. The user can interact with the frond-end in order to view the status of all tracking

Dimitris Kontaxis is with the PEOPLE Private Company, Greece (e-mail: dkontaxis@people-t.com).

devices, historic data and reports, as well as to remotely configure various parameters of the devices, like:

- Periodicity of transmissions.
- Activation/deactivation of individual alarms.
- Alarm levels/thresholds of all embedded sensors.

The structural parts of PEOPLE's 'eco-system' are depicted in Fig. 1 and are briefly described subsequently:

1) The back-end system: a cloud application software, able to communicate and interact with the IoT tracking devices.
2) The IoT tracking devices:

a) 'DoubleL': a battery-powered tracking device, which can be installed in containers or mounted on trains and trucks.
b) 'Flex': a battery-powered tracking device, which can be mounted on small trucks, vehicles, pallets and boxes.
c) 'Coin': a battery-powered temperature, humidity & light level data logger, which is installed on pallets and in product transport boxes.

3) The Mobile Application: an application for mobile phones with Android or IOS operating systems. It 'interacts' with the back-end system via the mobile's Internet connection and with the IoT tracking devices via the mobile's Bluetooth radio. The communication with the devices is encrypted and is used for remote firmware updates, configuration and data access.

III. RETIMO IoT Tracking Devices

The IoT tracking devices were designed following the lines of the '1220-2005 IEEE Standard for Application and Management of the Systems Engineering Process' [2] and comply with the provisions and the relevant EN standards of EU Directives and Regulations.

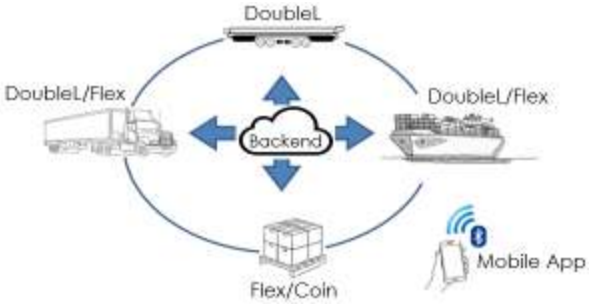


Fig. 1 IoT real-time monitoring system of the supply chain

Tests related to the RE Directive [3], the LV Directive [4], the EMC Directive [5], as well as various performance tests (EN 12830, IK, IP, vibration and shock) were carried out with device prototypes by labs certified with ISO 17025. Moreover, appropriate materials and production procedures were defined in order to ensure the compliance of the devices with the RoHS and WEEE Directives [6], [7]. The IoT tracking devices are depicted in Fig. 2.

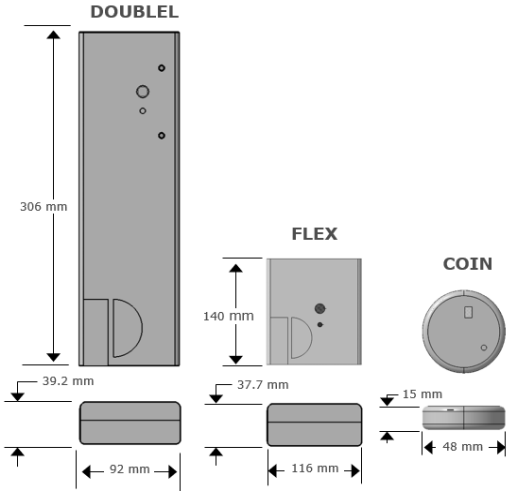


Fig. 2 IoT tracking devices

A. Device Functionality

1) DoubleL and Flex IoT Tracking Devices

These devices were designed to measure and report to the back-end system, via periodic transmissions over the cellular network, the following parameters:

- Location data (coordinates) and time (unix timestamp), via assisted GNSS.
- Location data via the cellular network.
- Number of satellites detected by the GNSS receiver
- Velocity in m/s.
- Ambient light level in Lux.
- Door status: open/closed (via light level and magnetic field detection).
- Temperature level with typical accuracy ± 0.1 ºC and resolution 0.01 ºC.
- Humidity level with typical accuracy ± 1.5% Relative Humidity (RH) and resolution 0.01% RH.
- Battery voltage and state of charge (SoC).
- Bluetooth MAC addresses of all neighboring devices (DoubleL, Flex, Coin).
- Message ID, hash key, BLE Network ID and configuration parameters (like the firmware version and the alarm thresholds).

Except from the periodic transmissions, the DoubleL and Flex IoT tracking devices will transmit various alarms to the back-end system, in real-time, whenever these alarms occur:

- Door open/close alarm.
- Motion start/stop alarm.
- Shock alarm.
- Tilt alarm.
- High/low temperature alarm.
- High/low humidity alarm.
- Low battery level alarm.
- Alarm of loss of Bluetooth connectivity with neighboring devices (DoubleL, Flex, Coin).

IoT data streams are transmitted from the DoubleL and Flex IoT tracking devices with the HTTPS protocol with TLS 1.2. Whenever a message is transmitted from a tracking device to

the back-end, the back-end acknowledges (ack) data reception and provides to the device, along with the ack, various configuration parameters or information for a remote firmware update. During the next transaction, namely in the context of a scheduled/periodic transmission or an alarm, the back-end system will verify the correct application of the configuration parameters or the successful update of the firmware version of the device. If the tracking device does not receive an ack from the back-end or if data transmission fails for any reason (e.g., due to connectivity issues), the device will store the message and transmit it to the back-end along with the HTTPS message of the next transaction. This ensures data integrity and seamless data flow from the device to the back-end. The aforementioned mechanism is depicted in Fig. 3. Regarding the format of the message payload, this is a proprietary JSON structure which employs a blockchain mechanism, with the use of a hash key in every HTTPS message, enhancing security against rogue actions.

2) Coin IoT Tracking Devices

The Coin devices are wireless multi-sensors able to measure temperature, humidity and light level. They are not equipped with a cellular transceiver or a GNSS receiver, achieving in this manner low cost, small volume and weight (which makes them suitable for use even in very small boxes). The Coin devices transmit stored periodic measurements to the back-end, either via the Mobile Application (see Section II) or via neighboring DoubleL/Flex devices, sharing the same Bluetooth Network ID. Moreover, their firmware and configuration parameters (i.e., clock, measurement frequency, alarm thresholds) are remotely programmable from the Mobile Application or neighboring DoubleL/Flex devices.

B. IoT Tracking Device Hardware

The electronics of the DoubleL and Flex IoT tracking devices comprise high quality integrated circuits (ICs), passive components and battery, which were selected in order to support the SRS (System Specification Requirements) of the project.

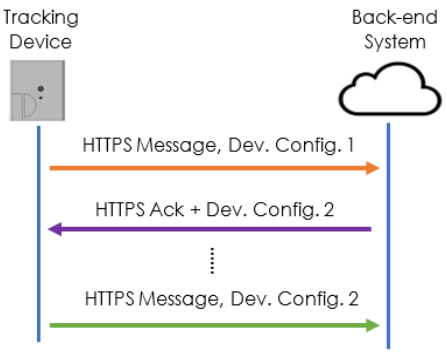


Fig. 3 Back-end and IoT tracking device transaction mechanism

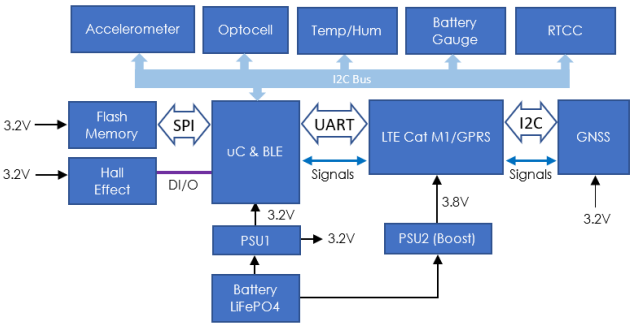


Fig. 4 Hardware architecture of the IoT tracking devices

The most important electronic components, along with information regarding their communication interface and manufacturer, are provided in Table I. Fig. 4 depicts the hardware architecture of the IoT tracking devices. It must be noted that the only difference between the DoubleL and Flex devices is the battery type: DoubleL is equipped with a rechargeable LiFePO4 prismatic battery, with capacity 25 Ah (3.2 V), while Flex carries a much smaller rechargeable LiFePO4 32700 cell, with capacity 6 Ah.

TABLE I
DOUBLEL/FLEX HARDWARE - MAIN ELECTRONIC COMPONENTS

Electronic Component	Comm. protocol	Mfn
Microcontroller & BLE	UART/I2C/SPI	Laird
Real Time Clock & Calendar	I2C	ST
GNSS Receiver	I2C	Ublox
LTE Cat M1/ GPRS Transceiver	UART/I2C	Ublox
Flash Memory	SPI	Winbond
Accelerometer & Gyro sensor	I2C	ST
Temperature & Humidity sensor	I2C	Sensirion
Battery Gauge sensor	I2C	Texas Inst.
Optocell sensor	I2C	Texas Inst.
Hall Effect sensor	Digital I/O	Texas Inst.

The hardware of the Coin devices is a subset of the electronics used for the DoubleL and Flex devices, including only a Microcontroller-BLE unit, a Real Time Clock & Calendar (RTCC), a temperature - humidity sensor and an optocell. The battery used for the Coin devices is a primary Li cell 2477, 1000 mAh (3.3 V).

The BoM (Bill of Materials) of the hardware was selected for high quality and credibility, wide operating temperature range (at least -20 ºC to 60 ºC) and low current consumption, which ensures long lifetime, reaching:

- up to 2 years for the DoubleL devices,
- up to 8 months for the Flex devices and
- up to 1 year for the Coin devices,

depending on the use profile (i.e., number of transmissions per hour) and the 'operational environment' (i.e., cellular and satellite connectivity).

IV. RETIMO BACK-END SYSTEM

As it is mentioned in the previous sections, the back-end system interacts with the IoT tracking devices: it collects, analyzes and processes IoT data streams sourcing from the

devices, configures their parameters and updates their firmware.

The development of the back-end system was based on the following phases:

1) Analysis of system requirements. In this phase, the requirements of the system were analyzed – defined, based on intensive 'interaction' with potential customers and users of the system, mainly stakeholders of the supply chain 'industry'.
2) Definition of software technologies and tools. Based on the requirements defined in the first phase and comparing different existing solutions, the .NET Framework with C# and the SQL Server (RDBMS) were qualified as the most appropriate software tools for the development of the back-end system. Furthermore, the Identity Server 4, which implements the OpenID Connect (OIDC) standard, was selected for the authentication procedures.
3) Design and modeling. In this phase, tools like the UML (Unified Modeling Language) and Microsoft Visio were employed in order to 'visualize' and simply the software development phase.
4) Software development. In this phase, the software applications and the database of system were developed, using the tools and following the lines defined in the two previous phases.
5) Software testing. During this phase, various tests were performed in order to detect and eliminate possible bugs and evaluate the quality of the software and its compliance with the initial requirements. Software testing included the following discrete levels:

a) Unit tests. Tests which checked the correct execution of routines/functions, using test data.
b) Integration tests. These tests verified the successful collaboration between different software components.
c) System tests. With these tests, the correct operation of specific use-cases was evaluated.
d) Acceptance tests. They were performed with the assistance of potential users of the system (see also Section V), in order to verify that the system meets the needs of their business/organization. In the context of this test level, a considerable number of tracking devices was distributed to collaborating companies presenting intensive product transportation activity (pharmaceutical and food).
e) Durability tests. The operation of the software was tested for 'worst-case' scenarios, namely in adverse conditions (e.g., unexpected shutdowns of the server), which is very important for the development of recovery plans.
f) Scalability tests. The software was tested to evaluate its stable operation in case of increased numbers of users and IoT tracking devices.

6) System deployment and maintenance. Once the necessary tests had been performed, the product was launched on the Microsoft Azure Cloud platform and is maintained on a daily basis.

V. Conclusions

In this work, the technology, the features and the basic characteristics of a complete, proprietary real-time monitoring system, which aims at the supply chain 'industry' and comprises hardware, firmware and software tools, were presented. Extensive laboratory and field tests, as well as actual trials carried out by potential users, proved the suitability of the system for the supply chain market and product handing procedures.

Acknowledgments

The work presented in this paper is co-funded by the ESPA Partnership for Growth Framework 2014-2020, Research-Create-Innovate, under the code T2EΔK-04894. The authors would like to thank in particular the company Creative Systems Engineering (CSE), partner in the RETIMO project.

References

[1] ESPA Partnership for Growth Framework 2014-2020, Research-Create-Innovate, code T2EΔK-04894.
[2] IEEE 1220-2005 - Standard for Application and Management of the Systems Engineering Process.
[3] Directive 2014/53/EU of the European Parliament and of the Council of 16 April 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of radio equipment and repealing Directive 1999/5/EC Text with EEA relevance.
[4] Directive 2014/35/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of electrical equipment designed for use within certain voltage limits Text with EEA relevance.
[5] Directive 2014/30/EU of the European Parliament and of the Council of 26 February 2014 on the harmonisation of the laws of the Member States relating to electromagnetic compatibility (recast) Text with EEA relevance.
[6] Directive 2011/65/EU of the European Parliament and of the Council of 8 June 2011 on the restriction of the use of certain hazardous substances in electrical and electronic equipment Text with EEA relevance.
[7] Directive 2012/19/EU of the European Parliament and of the Council of 4 July 2012 on waste electrical and electronic equipment (WEEE) Text with EEA relevance.

Blue Whale Body Condition from Photographs Taken over a 14-Year Period in the North East Pacific: Annual Variations and Connection to Measures of Ocean Productivity

Rachel Wachtendonk, John Calambokidis, Kiirsten Flynn

***Abstract*—** Large marine mammals can serve as an indicator of the overall state of the environment due to their long lifespan and apex position in marine food webs. Reductions in prey, driven by changes in environmental conditions can have resounding impacts on the trophic system as a whole; this can manifest in reduced fat stores that are visible on large whales. Poor health can lead to reduced survivorship and fitness, both of which can be detrimental to a recovering population. A non-invasive technique was used for monitoring blue whale health and for seeing if it changes with ocean conditions. Digital photographs of blue whales taken in the NE Pacific by Cascadia Research and collaborators from 2005-2018 (n=3,545) were scored for overall body condition based on visible vertebrae and body shape on a scale of 0-3 where a score of 0 indicated best body condition and a score of 3 indicated poorest. The data was analyzed to determine if there were patterns in the health of whales across years and whether overall poor health was related to oceanographic conditions and predictors of prey abundance on the California coast. The year was a highly significant factor in body condition (Chi-Square, $p<0.001$). The proportion of whales showing poor body condition (scores 2 & 3) overall was 33% but by year varied widely from a low of 18% (2008) to a high of 55% (2015). The only two years where >50% of animals had poor body condition were 2015 and 2017 (no other year was above 45%). The 2015 maximum proportion of whales in poor body condition coincide with the marine heat wave that affected the NE Pacific 2014-16 and impacted other whale populations. This indicates that the scoring method was an effective way to evaluate blue whale health and how they respond to a changing ocean.

***Keywords*—** blue whale, body condition, environmental variability, photo-identification.

Rachel Wachtendonk is with the University of Algarve, Portugal (e-mail: a67278@ualg.pt).

Fully Coupled Porous Media Model

Nia Mair Fry*, Matthew Profit† and Chenfeng Li†

* Energy Safety Research Institute (ESRI)
Swansea University
Fabian Way, Crymlyn Burrows, Skewen, Swansea SA1 8EN
e-mail: 485783@swansea.ac.uk, web page: https://www.swansea.ac.uk/

† Rockfield Global
Rockfield Ethos, Kings Road, Prince of Wales Dock, Swansea Waterfront
e-mail: matthew.profit@rockfieldglobal.com, web page: https://www.rockfieldglobal.com/

† Energy Safety Research Institute (ESRI)
Swansea University
Fabian Way, Crymlyn Burrows, Skewen, Swansea SA1 8EN
e-mail: c.f.li@swansea.ac.uk, web page: https://www.swansea.ac.uk/

ABSTRACT

This work focuses on the development and implementation of a fully implicit-implicit, coupled mechanical deformation and porous flow, finite element software tool. The fully implicit software accurately predicts classical fundamental analytical solutions such as the Terzaghi consolidation problem. Furthermore, it can capture other analytical solutions less well known in the literature, such as Gibson's sedimentation rate problem and Coussy's problems investigating wellbore stability for poroelastic rocks. The mechanical volume strains are transferred to the porous flow governing equation in an implicit framework. This will overcome some of the many current industrial issues, which use explicit solvers for the mechanical governing equations and only implicit solvers on the porous flow side (i.e., the so-called staggered coupling schemes). This can potentially lead to instability and non-convergence issues in the coupled system, plus giving results with an accountable degree of error. The specification of a fully monolithic implicit-implicit coupled porous media code sees the solution of both seepage-mechanical equations in one matrix system, under a unified time-stepping scheme, which makes the problem definition much easier. When using an explicit solver, additional input such as the damping coefficient and mass scaling factor is required, which are circumvented with a fully implicit solution. Further, improved accuracy is achieved as the solution is not dependent on predictor-corrector methods for the pore fluid pressure solution, but this comes at the potential cost of reduced stability. In testing of this fully monolithic porous media code, there is the comparison of the fully implicit coupled scheme against an existing staggered explicit-implicit coupled scheme solution across a range of geotechnical problems with important discussions on accuracy, stability and run-times.

REFERENCES

[1] Numerical Modelling in Geomechanics, M. Pastor, C. Tamagnini, Kogan Science, 2003.

[2] The Finite Element Method: Its Basis and Fundamentals, O.C. Zienkiewicz, R.L. Taylor, J.Z. Zhu, Elsevier Butterworth-Heinemann, 2005.

[3] Poromechanics, O. Coussy, Wiley, 2004.

[4] The Progress of Consolidation in a Clay Layer Increasing in Thickness with Time, R.E. Gibson, Géotechnique, Vol. 8 Issue 4, 1958.

Review of Carbon Materials - Application in Alternative Energy Sources and Catalysis

Marita Pigłowska[1], Beata Kurc[2], Maciej Galiński[2]

[1]*Poznań University of Technology, Faculty of Chemical Technology, student*
[2] *Institute of Chemistry and Electrochemistry, Faculty of Chemical Technology, Poznan University of Technology, Berdychowo 4, PL-60965 Poznan, Poland*

***Abstract*—**The application of carbon materials in the branches of the electrochemical industry shows an increasing tendency each year due to the many interesting properties they possess. These are, among others, a well-developed specific surface, porosity, high sorption capacity, good adsorption properties, low bulk density, electrical conductivity and chemical resistance. All these properties allow for their effective use, among others in supercapacitors, which can store electric charges of the order of 100 F due to carbon electrodes constituting the capacitor plates. Coals (including expanded graphite, carbon black, graphite carbon fibers, activated carbon) are commonly used in electrochemical methods of removing oil derivatives from water after tanker disasters, e.g. phenols and their derivatives by their electrochemical anodic oxidation. Phenol can occupy practically the entire surface of carbon material and leave the water clean of hydrophobic impurities. Regeneration of such electrodes is also not complicated, it is carried out by electrochemical methods consisting in unblocking the pores and reducing resistances, and thus their reactivation for subsequent adsorption processes. Graphite is commonly used as an anode material in lithium-ion cells, while due to the limited capacity it offers (372 mAh g^{-1}), new solutions are sought that meet both capacitive, efficiency and economic criteria. Increasingly, biodegradable materials, green materials, biomass, waste (including agricultural waste) are used in order to reuse them and reduce greenhouse effects and, above all, to meet the biodegradability criterion necessary for the production of lithium-ion cells as chemical power sources. The most common of these materials are cellulose, starch, wheat, rice, and corn waste, e.g. from agricultural, paper and pharmaceutical production. Such products are subjected to appropriate treatments depending on the desired application (including chemical, thermal, electrochemical). Starch is a biodegradable polysaccharide that consists of polymeric units such as amylose and amylopectin that build an ordered (linear) and amorphous (branched) structure of the polymer. Carbon is also used as a catalyst. Elemental carbon has become available in many nano-structured forms representing the hybridization combinations found in the primary carbon allotropes, and the materials can be enriched with a large number of surface functional groups. There are many examples of catalytic applications of coal in the literature, but the development of this field has been hampered by the lack of a conceptual approach combining structure and function and a lack of understanding of material synthesis. In the context of catalytic applications, the integrity of carbon environmental management properties and parameters such as metal conductivity range and bond sequence management should be characterized. Such data, along with surface and textured information, can form the basis for the provision of network support services.

Keywords— Carbon materials, catalysis, BET, capacitors, lithium ion cell

1. Introduction

Thanks to the widespread use of hard coal in Poland, many people deal with it on a daily basis. Many kilograms of fuel are bought before the heating season. It's not everything. The properties of hard coal make it possible to obtain (thanks to gasification, hydrogenation or smelting) coke, various types of engine fuels and fuel gas or benzene. The metallurgical industry cannot do without it, it is used in the production of cosmetics, medicines, and even in the jewelry industry.

Activated carbon is also an ingredient in whitening toothpastes. Due to the content of tannins, such a paste helps to remove discoloration resulting, for example, from frequent drinking of coffee or tea. In addition, activated charcoal in toothpaste helps to eliminate

unpleasant smell from the mouth and soothes irritated gums. It is worth adding that the mentioned properties of activated carbon can also be used by crushing the tablet into a brush, but remember not to use it too often, as it may damage the enamel.

Fig.1. Application of carbon materials

Carbon is also used in the production of protective mask inserts and in the production of water and air filters. It is also used for the clarification of protein hydrolysates, musts, wines and juices, for the production of spirit, pure sugar, for the separation of gas mixtures, for the purification of vapors and gases. On the other hand, very black varieties of activated carbon are used for coloring.

2. Carbon in supercapacitors - overwiev

Recently, electrochemical capacitors have attracted increasing attention due to the rapidly growing need for clean energy sources, as it can provide high power density, fast charge/discharge rate and long cycle lifetime. Electrochemical double layer capacitors (EDLCs) using carbonaceous materials as electrodes are extensively developed due to their low cost and high operation stability. Among them, carbon with large specific surface area is widely employed. However, its poor conductivity arising from non-graphitized structure has limited its charge/discharge performance at high charge/discharge current densities.

The value of specific surface area of carbon materials is limited by the thickness and specific weight of each layer of carbon. The maximum specific surface area of carbon materials is about 2600 m^2 g^{-1} [1]. However, they are found in the literature reports of carbon materials having a specific surface area of the order of 2900 m^2 g^{-1} [2], of 3000 m^2 g^{-1} [3-5], and even in the range 3200-3487 m^2 g^{-1} [6-9], which is caused method of determining (BET adsorption isotherm N_2). There are many reports in the literature regarding EDLC capacitors which show the possibility to use for the construction of different carbonaceous materials such as activated carbons in powder form [10, 11], carbon materials obtained from carbides [12, 13], woven carbon fiber [14, 15], carbon nanofibers [16, 17], carbon onions [18], carbon nanotubes [18-20], the carbon nanospheres, etc. It should be noted that the capacity of the carbon/ electrolyte is usually given terms of carbon, with the simultaneous assertion that the coals "have the capacity". This statement is wrong, the more that highly porous carbon electrodes have a lower mass than the electrolyte contained in them [1].

The popularity of carbon as an electrode material is due to its chemical and physical properties, namely:

- high temperature stability,
- controlled pore structure,
- processability and compatibility in composite materials,
- relatively low cost,
- good conductivity

Carbon materials are often used electrode materials in the construction of electrochemical capacitors. As is known, the ability to conduct electric current is due to the presence of graphene planes in which the carbon atoms C have sp2 hybridization (as opposed to non-conductive carbon varieties sp3). The graphite resistance is high and depends on the structure and porosity. Carbon electrodes are available with a surface area of up to 2500 m^2/g. Carbon material is used in the form of powder, fabrics, felt or fibers. Electricity storage on carbon electrodes is capacitive in an electrochemical double layer. These are the so-called electrochemical ECDL capacitors. Progress in the field of nanotechnology allows us to predict that in the coming years the use of carbon nanomaterials heads in the form of single and multi-wall nanotubes or nanoparticles will allow for better results in relation to the specific capacity of the electrode materials.

3. Graphite – important materials

The technological industry, especially the electrochemical industry, enables the use and dissemination of various allotropic types of carbon. Electrodes used in electrochemical energy sources require a large amount of conductive carbon - in order to maintain proper conductivity [21, 22].

Graphite ceases to fulfill its role as an active material in lithium-ion cells due to intensified intercalation and other solutions are sought [21, 22]. Graphite is often used to dope the dry mass during the formation of the electrode paste to increase the stability of the material. The greatest development of carbonaceous materials in lithium-ion cells concerns their use as an anode active material or as a cathode conducting solvent [23]. Carbon anode materials include: HOPG (highly oriented pyrolytic graphite), natural graphite, artificial (synthetic) graphite, hard and soft carbons, carbon nanotubes, materials with low mesophase temperature and MCMB (mesocarbon microspheres) [21, 23]. Hard carbons show the greatest possible capacity. Outstanding performance, flat discharge profile, high operating voltage, significant efficiency (also at low temperatures) and electrical conductivity are observed for synthetic graphite. It is also characterized by a low cost [23]. Mesophasic carbon fibers (MCFs), e.g. Melbon 3100, exhibit a radial crystal orientation. This allows for the trouble-free insertion of lithium ions on either side of the material surface. The sidewall is covered with a thin, porous, graphene layer. It protects the fibers from changing shape during swelling and volume reduction. Fibers reaching a thickness of 7 μm are used in anodic coating systems. Toschiba uses fibers from 1 to 7 mm thick and the batteries thus created show excellent low temperature performance compared to conventional lithium-ion cells. Acetylene or carbon black also finds applicability, while it is characterized by a lower but satisfactory effectiveness and an initial irreversible capacity [23]. Carbon mesophase microspheres (MCMB), manufactured by Osaka Gas Co., are highly effective and can be easily applied to copper anode collectors. The price and the production process still require optimization due to the huge volume of solvent needed to wash and recover the beads from the division matrix [23].

4. Hard carbons

Hard carbons are formed when thermosetting resins are heated from materials such as sugars, proteins, cellulose. The process takes place under vacuum or inert gas. As a result of

thermal decomposition, water and gases are released (syneresis). Then, the appropriate structure of carbon is obtained, the particles of which aggregate, creating a complex structure unable to transform into the form of graphite. The product shows high hardness, graphite structure and voids. Their significant advantage is the inability of these forms to decompose PC (propylene carbonate) used as a solvent. Its disadvantage is decomposition with the release of a large amount of heat in the presence of lithium graphite. A favorable phenomenon is the slight change in volume during the loading and unloading processes. It should be noted that due to slow diffusion during reversible intercalation, hard carbons show little efficiency. Sony uses hard carbons in the anodes due to the increased lifetime of the cells. Replacing metallic lithium with carbon in Li-ion is to significantly increase the safety of battery use. The main distinguishing feature of such a solution is the possibility to control the formation of the passive SEI layer on the anode (Solid Electrolyte Interphase) on carbon or graphite at a specified time. It is impossible with metallic lithium. The SEI coating is usually formed during the first charging cycle. The revision of its formation and destruction is crucial for best battery performance and safety [23].

In recent years, attempts have been made to dope carbon with boron. The unique properties of boron, such as the introduction of additional electronic acceptor levels by its atoms, increase the specific capacities [23]. CNTs carbon nanotubes are innovative. They are made of rolled graphene layers, significantly resembling building bricks. Configuratively, they create the form of a chair, rectal and chiral. The use of CNTs allows to increase the specific surface area of the adsorbent, co significantly improves electrochemical applicability. Depending on the degree of graphitization, the SWCNT variant (single-walled) and carbon nanotubes with higher capacity than MWCNT (multi-walled) graphite are used. SWCNTs have a lithium reversible capacity of 460 mAh g^{-1}, while MWCNTs have up to 980 mAh g^{-1}.

5. Activated carbons

Due to high surface area activated carbons in electrochemical capacitors can be stored a large amount of cargo. Activated carbons are characterized by good electrical conductivity and electrochemical stability. Electrical double layer can be formed rapidly, hence the electric double layer capacitors, characterized by a rapid increase in power compared to conventional batteries, for which may be a limiting factor in the faraday reaction kinetics. Also, the chemical properties of the electrodes of the capacitor, such as the distribution of functional

groups on the surface of carbon electrodes, affects the performance of the capacitors, which share a certain faraday reaction [24-27].

M. Nakamura et al [27] have studied the effects of physical properties of activated carbon on the properties of the electrical double layer of capacitor. The authors showed that the rest potential of the activated carbon is proportional to the logarithm of the oxygen content or the concentration of the acid functional groups on the surface of activated carbon.

According to the IUPAC classification of porous structure of activated carbons can distinguish three types of pores: micropores (pore diameter 1 or 2 nm), mesopores (with a diameter of 2 - 50 nm) and macropores (diameter above 50 nm). For most activated carbons total macropore part of the surface is negligible and is less than 2 $m^2 g^{-1}$. A significant portion of the surface of activated carbons are micropores that due to the small diameter may not be available for the electrolyte ions.

Theoretically, the larger the surface area of the activated carbon, the greater the obtained specific capacitance of the capacitor.

6. Carbon nanotubes

Of carbon nanotubes (CNTs) can be distinguished: single-walled carbon nanotubes SWCNT [28-32], double-walled carbon nanotubes DWCNT [33, 34], multi-walled carbon nanotubes MWCNT [35-40], fluoridated FCNT carbon nanotubes [41], as well as activated carbon nanotubes ACNT [42, 43].

Carbon nanotubes have high mesoporous, easy accessibility for the electrolyte ions and excellent electrical conductivity.

Recently, electric double layer capacitor (EDLC) has attracted extensive interest because of its advantages such as fast charge-discharge rate, high power-density in discharging and recharging, excellent cycle lives on the order of 105 to 106, many times those of batteries, and environmental friendliness [44-46]. Surface active carbon materials are recognized as an essential component for EDLC. EDLC properties have recognized to depend strongly upon physicochemical properties such as surface area and pore size of carbon materials [47, 48], precursor of carbon materials for carbon electrode [49, 50].

Various carbon materials could be used as supercapacitors electrodes, typical examples include carbon nanotubes (CNTs) [51, 52] and ordered mesoporous carbon (OMC) [53, 54]. CNTs are used as electrodes to increase the power density as a result of their unique tubular structures and superior electrical properties, which favor fast ions and electrons transportation.

However, the high production cost and entanglement problem limit their wide spread applications [55, 56]. OMC possesses excellent electrochemical performance due to its high surface area, narrow pore size distribution and uniform pore connection. But the conventional hard-template nanocasting method to prepare OMC is obviously high-cost Modifications of carbon materials are very important due to the use of a given material and due to its electrochemical performance. Hence the advantages and disadvantages very often.

7. Carbon aerogels, xerogels and crogels

Organic aerogels CA (specific surface area > 400 m^2 g^{-1}) consisting of interconnected colloidal particles, plates or fibers. are obtained resin: resorcinol-formaldehyde, cresol-formaldehyde, phenol-resorcinol-formaldehyde, phenol-furfural, melamine-formaldehyde, polyurethanes and polyureas. The structure of the carbon airgel consists of a three-dimensional network of interconnected nano-particles having a diameter of 5-10 nm [57, 58].

Organic aerogels pyrolyzed under an inert atmosphere gives the carbon aerogels monolithic form, a powder, microspheres or in the form of composite thin films. The monolithic carbon airgel structure, relatively high specific surface area and electrical conductivity make carbon aerogels may be used in electrical double layer capacitors. The structure and properties of the carbon airgel depend on the synthesis conditions, the pyrolysis temperature, the presence of dopant activation.

Literature review of electrochemical capacitors in which the electrode material as aerogels have been used (CA), cryogels (CGC) and carbon xerogels (XGC) are summarized in Table 1.

Table 1. Literature review of the electric double layer capacitor as an electrode material containing aerogels, cryogels and xerogels carbon [59-68]

Electrolyte	Electrode	Counter-electrode
4M KOH	CA	CA
1M H_2SO_4	CA+RuO_2	G
5M KOH, 0,5M Et_4NBF_4 in PC	CA	CA
30% H_2SO_4	CA	CA
4M KOH	ACA	ACA
1M H_2SO_4	CA	CA
1M H_2SO_4	CA	Pt
1M H_2SO_4	PAN-CA	Pt
1M H_2SO_4	CA	Pt
aquous solution NaF	CA	CA2
6M H_2SO_4	CA	CA
0,8M Et_4NBF_4 in PC	AC or CA	AC or CA
0,8M Et_4NBF_4 in PC	CA	CA
30 wt.-% KOH	CA	CA
1,5M Et_3MeNBF_4 in PC	CA (surfactant)	CA (surfactant)

6M KOH, 1M Na_2SO_4, 2M $(NH_4)_2SO_4$	CA	CA
PVA-KOH lub 6M KOH	CA	CA
1M Et_4NBF_4 in PC	ACA (surfactant) or CA	ACA (surfactant) or CA
30% KOH	CA	CA
0,8M Et_4NBF_4 in PC	AC (surfactant) CA (surfactant)	AC (surfactant) CA (surfactant)
$BMPyrNTf_2$, $EMImNTf_2$	XGC, CGC	pMeT
$BMPyrNTf_2$, $EMImNTf_2$	XGC, CGC	Ag
1M H_2SO_4	CA/SnO_2	CA/SnO_2
1M Et_4NBF_4 w PC	CA/AC	CA/AC
1,5M Et_3MeNBF_4 in PC	CA	CA
1,5M Et_3MeNBF_4 in PC	CA	CA
30% KOH	CA	CA
1M H_2SO_4	NCA	Pt
$EMImNTf_2$, $BMPyrNTf_2$	XGC	XGC
4M KOH	CA	G
30% KOH	ACA	ACA

CA - carbon aerogel, G - graphite, PC - propylene carbonate, PAS - poly (acrylonitrile), AC - activated carbon, ACA - activated carbon airgel, pMet - poly (3-methylthiophene), XGC - carbon xerogel, CGC - carbon cryogel NCA - carbon airgel enriched with nitrogen

8. Catalysis

The rapid development of modern society is based on massive energy consumption. However, the over-extraction and use of conventional fossil fuels leads to increasing environmental problems and problems related to the energy crisis. It is essential to search for sustainable and environmentally friendly sources of energy. Due to its high heat of combustion and clean combustion product, hydrogen is considered a very promising energy carrier. As a clean and carbon neutral strategy, the Electrocatalytic Hydrogen Release (HER) reaction plays an important role in generating high purity hydrogen energy and potential in combination with renewable resources such as solar, wind and hydro power. The production of renewable energy, the extraction of chemical raw materials and the production of functional material from biomass in bulk are a promising direction of research. It is one of the important industrial methods of converting plant biomass to produce commercial activated carbon with adsorptive properties and supports for heterogeneous active ingredients of the catalyst. It is worth noting that catalytically active material derived from biomass will be beneficial in creating a perfect, consistent cycle system. Plant biomass components, especially structural uniqueness, provide rich development space for the design of efficient active catalysts for the conversion and storage of renewable energy. Nanocarbon catalysts with structures regulated at the atomic-molecular level, excellent electronic conductivity,

customizable physicochemical properties and solid mechanical-chemical stability show a promising potential for building an efficient catalytic system.

To achieve high efficiency of HER catalysts, three key factors must be taken into account: the number of active sites, the intrinsic activity of each active site, and electrical conductivity. Due to the abundance of earth and excellent electrical conductivity, carbon materials have attracted a lot of attention, especially the emergence of novel advanced carbon materials such as graphene, carbon nanotubes, etc. These properties, developed surface, electrochemical stability and nanostructures are features that contribute to their use as electrocatalysts [69-72].

Conclusion

One of the greatest achievements of modern nanotechnology is invention materials that were used to build supercapacitors, also called ultracapacitors or double layer capacitors. The prefix "super" were used in the name of these capacitors due to their large capacity. It decides about the effective surface of the electrodes and a very small distance between them.

The electrodes are made of electrically conductive activated carbon, the structure of which resembles a sponge. Such large surfaces with little material is obtained thanks to the use of carbon nanotubes, which are the main compound building electrodes in a supercapacitor.

Carbon nanotubes are cylindrical forms in the hollow tubes of nanometer size made of coiled carbon graphene (monatomic graphite layer). They are placed on one plane supermolecular structures. Nanotubes have a large surface area and high electrical conductivity, so they can be used in construction nanoelectrodes in a supercapacitor. They are excellent heat conductors. Bindings the bonds that make up the nanotubes are stronger than the bonds that make up the diamond, which is why they show enormous tear resistance.

Electrochemical capacitors using the so-called pseudocapacity redox is the second type of condensation electrochemical tracks widely used in practice. These are the so-called redox electrochemical capacitors or redox supercapacitors. These devices, apart from using the electric charge of the double layer, are a source of current from charge transfer reaction, concerning substances adsorbed on the surface and undergoing transformation in its Faradayan nature. The difference between the usual redox reaction and the discussed phenomenon is the availability of the substrate. In capacitors, the charge transfer is strictly limited to the electrode surface. The process takes place in many electrons and in a wide range of potentials. In this type of devices it uses there are materials capable of undergoing redox surface reactions (e.g. ruthenium oxides, electroactive polymers).

Graphite and its derivatives are characterized by good biocompatibility. Therefore, much attention is paid to research aimed at the use of nanotubes and graphene in sensors for determining the concentration of active substances in drugs, e.g. paracetamol, and biomarkers in biological matrices. As shown, the detection of compounds of physiological importance, e.g. dopamine, ascorbic acid and uric acid, can take place without the participation of enzymes. Parallel to this research, prototypes of enzymatic biomarker sensors are being constructed, in which the proteins responsible for catalyzing the redox reactions of analytes (osidase and dehydrogenase) are deposited on nanotubes or graphene. Carbon nanotubes and graphene provide a good platform to directly study the mechanisms of redox reactions involving proteins such as myoglobin, hemoglobin and glucose oxidase. It has been observed that in the case of a graphene electrode or a carbon nanotube, it is possible to transfer an electron between the electrode and the protein without intermediating substances. It is very important from the point of view of research on the kinetics of reactions responsible for the occurrence of life processes.

Acknowledgments

This work is supported by the Polish Ministry of Science and Education, project "The best of the best! 4.0", number 0911/MNSW/4968 – M.P. and grant 0911/SBAD/2102—B.K., M.G..

References

[1] A. Lewandowski, M. Galinski, J. Power Sources 173 (2007) 822.
[2] G. Chun-yu, W. Cheng-yang, Compos. Sci. Technol. 67 (2007) 1747.
[3] L. Wang, M. Fujita, M. Inagaki, Electrochim. Acta 51 (2006) 4096.
[4] D.-W. Kim, J.M. Ko, W.J. Kim, J.H. Kim, J. Power Sources 163 (2006) 300.
[5] W.-J. Kim, D.-W. Kim, Electrochim. Acta 53 (2008) 4331.
[6] H. Zhang, W. Zhang, J. Cheng, G. Cao, Y. Yang, Solid State Ionics 179 (2008) 1946.
[7] M.S. Balathanigaimani, W.-G. Shim, M.-J. Lee, Ch. Kim, J.-W. Lee, H. Moon, Electrochem. Commun. 10 (2008) 868.
[8] B. Xu, F. Wu, R. Chen, G. Cao, S. Chen, Z. Zhou, Y. Yang, Electrochem. Commun. 10 (2008) 795.
[9] V. Khomenko, E. Raymundo-Piñero, F. Béguin, J. Power Sources 177 (2008) 643.
[10] W. Quiao, S.-H. Yoon, I. Mochida, Energ. Fuels 20 (2006) 1680.
[11] K. Yuyama, G. Masuda, H. Yoshida, T. Sato, J. Power Sources 162 (2006) 1401.
[12] J. Chmiola, G. Yushin, Y. Gogotsi, C. Portet, P. Simon, P.L. Taberna, Science 313 (2006) 1760.
[13] J.A. Fernández, M. Arulepp, J. Leis, F. Stoeckli, T.A. Centeno, Electrochim. Acta 53 (2008) 7111.
[14] A. Lewandowski, A. Olejniczak, J. Power Sources 172 (2007) 487.
[15]I.V. Sheveleva, L.A. Zemskova, S.V. Zheleznov, A. Voit, N.N. Barinov, S.V. Sukhoverstov, V.I. Sergienko, Russ. J. Appl. Chem. 80 (2007) 924.
[16] C. Merino, P. Soto, E. Vilaplana-Ortego, J.M. Gomez de Salazar, F. Pico, J.M. Rojo, Carbon 43 (2005) 551.
[17] S.-H. Yoon, S. Lim, Y. Song, Y. Ota, W. Qiao, A. Tanaka, I. Mochida, Carbon 42 (2004) 1723.
[18] C. Portet, G. Yushin, Y. Gogotsi, Carbon 45 (2007) 2511;
[19] B.-J. Yoon, S.-H. Jeong, K.-H. Lee, H.S. Kim, Ch.G. Park, J.H. Han, Chem. Phys. Lett., 388 (2004) 170.
[20] C. Du, J. Yeh, N. Pan, Nanotechnology 16 (2005) 350.
[21] L. S. Roselin, R. Shin Juang, C.-T. Hsieh , S. Sagadevan , A. Umar, R. Selvin, H. Hegazy, Materials 12 (2019) 1229

[22] M. Salamia, H. Louis, S. Zafar , O. O. Amusan, Asian Journal of Nanoscience and Materials 2(1) (2018) 1-26
[23] H. J. Santner, K. C. Möller, W. Kohs, C. Veit *Springer* (2006) 161-167, 157
[24] M. Ishikawa, A. Sakamoto, M. Morita, Y. Matsuda, K. Ishida, J. Power Sources 60 (1996) 233.
[25] A. Yoshida, S. Nonaka, I. Aoki, A. Nishino, J. Power Sources 60 (1996) 213.
[26] H. Shi, Electrochim. Acta 41 (1996) 1633.
[27] M. Nakamura, M. Nakanishi, K. Yamamoto, J. Power Sources 60 (1996) 225.
[28] T. Liu, T.V. Sreekumar, S. Kumar, R.H. Hauge, R.E. Smalley, Carbon 41 (2003) 2427.
[29] C. Zhou, S. Kumar, C.D. Doyle, J.M. Tour, Chem. Mater. 17 (2005) 1997.
[30] C. Zhou, T. Liu, T. Wang, S. Kumar, Polymer 47 (2006) 5831.
[31] J. Oh, M.E. Kozlov, B.G. Kim, H.-K. Kim, R.H. Baughman, Y.H. Hwang, Synthetic Met. 158 (2008) 638.
[32] P.W. Ruch, R. Kötz, A. Wokaun, Electrochim. Acta 54 (2009) 4451.
[33] C. Portet, P.L. Taberna, P. Simon, E. Flahaut, J. Power Sources 139 (2005) 371.
[34] Y. Honda, M. Takeshige, H. Shiozaki, T. Kitamura, K. Yoshikawa, S. Chakrabarti, O. Suekane, L. Pan, Y. Nakayama, M. Yamagata, M. Ishikawa, J. Power Sources 183 (2008) 1580.
[35] T. Bordjiba, M. Mohamedi, L.H. Dao, J. Power Sources 172 (2007) 991.
[36] Ch.-Ch. Hu, J.-H. Su, T.-Ch. Wen, J. Phys. Chem. Solids 68 (2007) 2353.
[37] Ch.-W. Huang, Y.-T. Wu, Ch.-Ch. Hu, Y.-Y. Li, J. Power Sources 172 (2007) 460.
[38] Ch.-T. Hsieh, Y.-W. Chou, W.-Y. Chen, J. Solid State Electrochem. 12 (2008) 663.
[39] W. Xing, S. Zhuo, H. Cui, W. Si, X. Gao, Z. Yan, J. Porous Mat. 15 (2008) 647.
[40] J.P. Zheng, C.M. Pettit, P.C. Goonetilleke, G.M. Zenger, D. Roy, Talanta 78 (2009) 1056.
[41] J.Y. Lee, K.H. An, J.K. Heo, Y.H. Lee, J. Phys. Chem. B 107 (2003) 8812.
[42] Ch. Li, D. Wang, T. Liang, X. Wang, L. Ji, Mater. Lett. 59 (2004) 3774.
[43] Q. Jiang, Q. Zhang, B. Du, R.L. Zou, Y.H. Liu, Y. Zhao, J. Mater. Sci.: Mater. Electron 19 (2008) 241.
[44] H. Jiang, P.S. Lee, C.Z. Li, Energy Environ. Sci. 6 (2013) 41.
[45] J.K. Miller, P. Simon, Science 321 (2008) 651.
[46] G.P. Wang, L. Zhang, J.J. Zhang, Chem. Soc. Rev. 41 (2011) 797.
[47] C.C. Hu, K.H. Chang, M.C. Lin, Y.T. Wu, Nano Lett. 6 (2006) 2690.
[48] H. Jiang, T. Zhao, C.Z. Li, J. Ma, J. Mater. Chem. 21 (2011) 3818.
[49] R.K. Paul, M. Ghazinejad, M. Penchev, J. Lin, M. Ozkan, S. Ozkan, Small 6 (2010) 2309.
[50] Y.F. Yan, Q.L. Cheng, G.C. Wang, C.Z. Li, J. Power Sources 196 (2011) 7835.
[51] J.P. Randin, E. Yeager, *J. Electroanal. Chem.*, 36,257 (1972).
[52] J.P. Randin, E. Yeager, *Electroanalytical Chemistry and Interfacial Electrochemistry*, 58,313(1975).
[53] S. Iijima, Nature, 354,56 (1991).
[54] N.M. Rodriquez, Mater. Res.,8, 3233 (1993).
[55] B.E. Conway (Edior), Electrochemical Suoercapacitors "Scientific fundamental and technological applications", p.12, Kluwer Acedamic/Plenum Publishers (1999).
[56] C. Dekker, Phys. Today, 5, 22 (1999).
[57] S-H. Yoon, A. Tanaka, S. Lim, Y. Korai, I. Mochida, B. Ahn, K. Yokogawa, C-W.Park, In Processing of Carbon 2003. International Conference on Carbon, Oviedo, Spain, p76 (2003).
[58] S.-W. Hwang, S.-H. Hyun, J. Non-Cryst. Solids 347 (2004) 238.
[59] L. Zhang, H. Liu, M. Wang, W. Liu, Rare Metals 25 (2006) 51.
[60] H. Pröbstle, C. Schmitt, J. Fricke, J. Power Sources 105 (2002) 189.
[61] B. Fang, Y.-Z. Wei, K. Suzuki, M. Kumagai, Electrochim.Acta 50 (2005) 3616.
[62] B. Fang, Y.-Z. Wei, K. Maruyama, M. Kumagi, J. Appl. Electrochem. 35 (2005) 229.
[63] J.M. Miller, B. Dunn, T.D. Tran, R.W. Pekala, J. Electrochem. Soc. 144 (1997) L309.
[64] C. Schmitt, H. Pröbstle, J. Fricke, J. Non-Cryst. Solids 285 (2001) 277.
[65] N. Yoshizawa, H. Hatori, Y. Soneda, Y. Hanzawa, K. Kaneko, M.S. Dresselhaus, J. Non-Cryst. Solids 330 (2003) 99.
[66] H. Talbi, P.-E. Just, L.H. Dao, J. Appl. Electrochem. 33 (2003) 465.
[67] H. Pröbstle, M. Wiener, J. Fricke, J. Porous Mat. 10 (2003) 213.
[68] D. Kalpana, N.G. Renganathan, S. Pitchumani, J. Power Sources 157 (2006) 621.
[69] C. Arbizzani, S. Beninati, M. Lazzari, F. Soavi, M. Mastragostino, J. Power Sources 174 (2007) 648.
[70] H. Ren, Z.-H. Huang, Z. Yang, S. Tang, F. Kang, R. Lv, *J. Energy Chem.* 2017, 26, 1217– 1222
[71]X. Wang, X. Gan, T. Hu, K. Fujisawa, Y. Lei, Z. Lin, B. Xu, Z. H. Huang, F. Kang, M. Terrones, R. Lv, *Adv. Mater.* 2017, 29, 1603617

[72]Y. Lei, S. Pakhira, K. Fujisawa, X. Wang, O. O. Iyiola, N. Perea Lopez, A. Laura Elias, L. P. Rajukumar, C. Zhou, B. Kabius, N. Alem, M. Endo, R. Lv, J. L. Mendoza-Cortes, M. Terrones, *ACS Nano* 2017, 11, 5103–5112

Application of Az Compound as a Bio-Desalination anda Bio-Treatment of Saline Wastewater

Eman Marei

***Abstract*—** This study aims to investigate the ability of AZ compound as a biological compound to treat saline wastewater after primary treatment as a new technique for bio-desalination, bio-treatment and bio-removal of different heavy metals of saline wastewater. Saline wastewater samples were collected from Ras-sudr research station, Cairo, Egypt after primary treatment. This treatment by AZ compound contained mixture free cells of different bacterial cultures. The AZ compound was prepared under lab condition.The obtained results indicated that the percentages of removal rate were found to be 81.10% BOD, 83.22 % COD, 45.71% ammonia nitrogen and 95.77% phosphate due to using AZ compound. As a result of using AZ compound high bio-removal of different heavy metal (58.10 to 80.70 %) was achieved after 24 hrs. Moreover, higher reduction in percentage of TDS (45.00%) was recorded after 24 hrs. due to application of AZ compound. On the basis of the obtained results, it can be concluded that the bio-desalination and bio-treatment of saline wastewater samples with AZ compound could be achieved after 24 hrs. The removal and reduction percentage were found to be 81.10% for organic matter and 45.0 % for TDS.

***Keywords*—** bio-desalination, saline water, wastewater, bio-treatment, bio-sorption heavy metals.

Eman Marie is with the Ain Shams University, Egypt (e-mail: wel_memo@yahoo.com).

Comparative Transcriptome Profiling of Low Light Tolerant and Sensitive Rice Varieties Induced by Low Light Stress at Active Tillering Stage

Darshan Panda, Lambodar Behera, M. J. Baig, Sudhanshu Sekhar

***Abstract*—** Low light intensity is a significant limitation for grain yield and quality in rice. However, yield is not significantly reduced in low-light tolerant rice varieties. The work, therefore, planned for comparative transcriptome profiling under low light stress to decipher the genes involved and molecular mechanism of low light tolerance in rice. At the active tillering stage, 50% low light exposure for one day, three days, and five days were given to Swarnaprabha (low light tolerant) and IR8 (low light sensitive) rice varieties. Illumina (HiSeq) platform was used for transcriptome sequencing. A total of 6,652 and 12,042 genes were differentially expressed due to low light intensity in Swarnaprabha and IR8, respectively, as compared to control. CAB, LRP, SBPase, MT15, TF PCL1, and Photosystem I & II complex related gene expressions were mostly increased in Swarnaprabha upon the longer duration of low light exposure, which was not found in IR8 as compared to control. Their expressions were validated by qRT-PCR. The overall study suggested that the maintenance of grain yield in the tolerant variety under low light might be the result of accelerated expression of the genes, which enable the plant to keep the photosynthetic processes moving at the same pace even under low light.

***Keywords*—** rice, low light, photosynthesis, yield.

Darshan Panda is with the Ravenshaw University, India (e-mail: darshan.panda216@gmail.com).

Pregnancy and Calving Rates Following Ai with Buffalo Sexed Semen Through Modified Swim Up Method

Asma-ul-Husna, Shamim Akhter

***Abstract*—** Sperm Pre-selection is of significance for production of desired sex of the offspring in buffalo farming. Sex separation of spermatozoa is being practiced in developed countries using flow cytometry, which is highly expensive, damaging to sperm and produce low sperm number per dose. Present study was planned to evaluate modified swim up method for production of X and Y bearing Nili Ravi buffalo sperm and success rate was confirmed through *in-vivo* fertility trials. To assess this, five mature Nili Ravi buffalo bulls were used to collect semen kept at SPU, Qadirabad, District Sahiwal, Pakistan, with artificial vaginą (40 ˚C), 2 consecutive ejaculates per week were collected for three weeks period. Semen ejaculates met the qualifying criteria of (volume > 1mL, sperm concentration > 1.5 billion/mL, motility > 60%, abnormalities <20%) were used for further processing. Qualifying ejaculates were pooled and divided into two aliquots. The first aliquot was processed by routine procedure (control), while the second aliquot was processed by modified swim-up technique. After separation, semen was diluted in *tris*-citric acid extender and cryopreserved using standard techniques. Sexed semen doses were evaluated for field fertility trials during peak breeding season to check the accuracy of method. 200 inseminations were performed with cryopreserved sexed and unsexed (control) semen in Tehsil Pattoki District Kasur. The artificially inseminated animals were examined for pregnancy rate through rectal palpation at least 3 month after insemination under field conditions. Calving ratio of female and male calves were recoded after parturition. The data on conception rate and calving rate of the semen sexed by modified swim up was analyzed using chi-square statistics. Results showed significantly higher conception rate and higher production of female buffalo calf (calving rate i.e. ~78%) through modified swim up method as compared to control. A modified swim up method proved to be an effective alternative method for Nili-Ravi buffalo sperm sexing as validated by *in vivo* fertility trial for production of female calf.

***Keywords*—** Cryopreservation, Modified swim up, Nili Ravi buffalo, Sperm Sexing.

Asma Ul Husna is with the University of Sialkot, Pakistan (e-mail: asmaulhusna_88@yahoo.com).

Effect of Heparin and Calcium Ionophore a23187 on Dromedary Camel Epididymal Sperm Functions and In -Vitro Fertilization

Hagar Elsharnoby, Nehal Abu-Elnaga, Omaima Kandil

***Abstract*—** The objectives of this study were to evaluate the effect of the use of heparin (Hep) and heparin + calcium ionophore A23187 (Hep+Ca) and incubation time on dromedary camel epididymal sperm functions (Individual motility, sperm morphology, membrane integrity and acrosome reaction (AR)) and their effect on in vitro fertilization (IVF) and subsequent embryo development of dromedary camel oocytes. In Experiment 1, epididymal sperms were collected using the retrograde flushing technique from the mature dromedary camel using 1 mL of S-TALP medium supplemented with either hep. (10mg/mL) or Hep+Ca (0.1 mM/mL). Individual motility, sperm morphology, membrane integrity, and AR were recorded at 0, 1, 2 and 3 h post-incubation for all treatments. Experiment 2, the effect of adding either hep. (10mg/mL) or Hep+Ca (0.1 mM/mL) in vitro fertilization media (FM) (FM+Hep or FM+Hep+Ca) on in vitro developmental competence of dromedary camel embryos was calculated . Our results show that, the use of Hep+Ca media significantly increase sperm motility, normal morphology, membrane integrity, and AR when compared with hep. group. The fertilization, cleavage, and the morula rate were significant increase (P value<0.01) in FM+Hep+Ca media when compared with FM+Hep media. In conclusion Hep+Ca media enhance epididymal camel sperm function and improve IVF more than hep. group.

***Keywords*—** dromedary camel, camel epididymal sperm capacitation, acrosome reaction, in vitro embryo production.

Elsharnoby, H. A., Abu-Elnaga, N. A. are with the Embryology, Zoology and Entomology Department, Faculty of Science, Al-Azhar University (Girls).

Kandil O. M. is with the Embryo Biotechnology, Department of Animal Reproduction & Artificial Insemination, Veterinary Research Division, National Research Centre (e-mail: Omaima_mk@yahoo.com).

Overcome Effect of Heat Stress on Early Reproduction Success in Buffalo

Omaima Kandil, Ragab Dohreig , Ahmed Abdoon, Esraa Ismael, Hosam Tawfik, Ahmed Shaban, Pascal Mermillod

***Abstract*—**Heat stress (HS) is a major problem for animal breeding. Aims of work are study: 1) Effect of season in oocyte quality and developmental competence of buffalo oocytes 2) Effect of heat stress & IGF in oocyte maturation rate (MR) and embryo development 3) Transfer of in vitro produced embryo (IVPE) during hot season to Egyptian buffalo. Ovaries were collected during summer and winter seasons from abattoir 1) Excellent (Ex) and good (G) buffalo oocytes were matured (IVM) in TCM, fertilized and cultured in SOFM with or without BOEC for 8 days. 2) Ex & G oocytes, IVM in TCM+ IGF (41◦C /1hr- then 38.5◦C vs. 38.5◦C), then fertilized and cultured as described. Blastocyst were Fixed for cell counting. 3) Two fresh IVPE were transferred to each buffaloes (n=5). There were significant ($P<0.01$) increase in the recovery rate of Ex and G oocytes in winter than summer season. BOEC vs. without BOEC significantly increase the blastocyst rate and their cell number ($P<0.1$) in winter than summer. 41◦C for one hour during IVM significantly ($P<0.01$) decreased the MR, cleavage and blastocyst than in 38.5◦C. Three calves were born through transfer of IVPE (3/5).

In conclusion: Heat stress either during hot season or experimental 41◦C during in-vitro maturation of buffalo oocytes significantly decrease in-vitro oocyte developmental competence. Supplementation of IGF on IVM medium and co-culture (BOEC) decreased effect of HS, and improved viability of IVPE and success of calving of two calves (Emy& Medo & Amer) through transfer of IVP buffalo embryos.

***Keywords*—** Buffalo, In vitro Embryo Production, Buffalo Oviduct Epithelial Cells, Heat stress, Embryo Transfer, Insulin Growth Factor.

Kandil O.M., Abdoon A.S.S., Esmail E.,Tawfik H. are with the Dep. of animal reproduction and artificial insemination, Veterinary research division, National research center, Cairo, Egypt (e-mail: omaima_mk@yahoo.com).

Dohreig R.M.A. is with the Unit of Ultrasound, Haram Institute of Reproduction, Giza, Egypt.

Shaban A. is with the Water Pollution Dept, Environmental Science Division (NRC).

Corbin E., Mermillod P. are with the Physiologie de la Reproduction et des Comportements, UMR 6175 INRA, CNRS, Université de Tours, Haras Nationaux, 37380 Nouzilly, France.

Extracellular Vesicles: An Insight into Endometritis in Arabian Mares

Sally Ibrahim, Mohamed Hedia, Mohamed Taqi, Mohamed Derbala, Karima Mahmoud, Youssef

***Abstract*—** However, the role of EVs in uterine lavage and serum during endometritis in mares is not investigated yet. Hence, our aim was (I) to identify and characterize EVs from low volume uterine lavage (LVL) and serum in mares suffering from endometritis, compared to healthy (control) ones. This study was conducted on 50 Arabian mares (4-8 years old), and mares were divided into; 30 sub-fertile mares suspected of endometritis (diseased group) and 20 fertile mares not suspected of endometritis (control group), at a number of stud farms. Consecutively, samples (blood, uterine swabs and LVL) were collected. The EVs was isolated from LVL through differential centrifugation followed by ultracentrifugation, and from serum via ExoQuick® Ultra Kit. Afterwards, EVs were characterized and identified using a transmission electron microscope and specific markers (CD63, CD81). Moreover, the particle size of EVs was determined using Dynamic light scattering (DLS) instrument (NICOMP 380 ZLS). The EVs were successfully isolated and characterized from LVL and serum with a good yield as well as purity. It was also noticed that the size of EVS in both LVL and serum was larger (224.2±2.58 nm and 200±8.22 nm, respectively) in mares with endometritis, compared to control healthy ones (112.2±8.65 nm and 101.8±9.77 nm, respectively). So, extracellular vesicles might be used as potential novel biomarkers for the early diagnosis of endometritis in Arabian mares.

***Keywords*—** Arabian mares, Endometritis, Extracellular vesicles, Low volume uterine lavage, Serum.

Sally Ibrahim is with the National Research Center Inc, Egypt (e-mail: sally_rashad2004@yahoo.com).

Traces of the Past: The Croatian Nationalist Behavior

Elena Fagogeni

***Abstract*—** To date, the anger and aggressiveness of the Croats against the Serbs are clearly still there. Anti-Serb slogans like "Srbe na vrbe" (Hang Serbs on willows) can be easily found in Croatia, while politicians' tactics to blame Serbs during election campaigns are notable. But what makes Croats so provocative many years after the end of the Yugoslav wars? The purpose of this paper is to present the most common anti-Serb acts that have been observed recently by both civilians and politicians and then to analyze the roots of this phenomenon that began long before the Yugoslav wars during World War II. the rise of the ultra-nationalist organization Ustaša (Croatian Revolutionary Movement) and the establishment of the Jasenovac concentration camp. Having a clear view of the past and present, the next question to be answered is the main reason that far-right perceptions in Croatia are still widely accepted. Last but not least, there will be possible scenarios about Croatia's future based on the country's background.

***Keywords*—** Balkans, genocide, nationalism, World War II.

Elena Fagogeni is with the Ionian University, Greece (e-mail: elenafag@yahoo.com).

Reviewing the Government Policy Response to Child Labour in Bangladesh: A Political Economy Analysis

Md Mahmudul Hoque

***Abstract*—**Child Labour is a widespread issue in Bangladesh. Bangladesh's National Child Labour Survey 2013 finds that some 1.2 million children are still trapped in its worst forms. A 2010 UNICEF report estimated that one out of every six children is working which results in 7.4 million children being engaged in labour across the country. The country ratified the International Labour Organization's Convention on the Worst Forms of Child Labour 1999, and national legislation on hazardous child labour was subsequently revised. The Government enacted a set of policies including the National Child Labour Elimination Policy 2010 to get rid of the worst form of child labour by 2021. Despite a legal and administrative framework in place, the issue of child labour remains a big threat to sustainable development because of several socially diverse reasons. This paper examines government policy responses to the issue of child labour. Taking a policy level political economy analytical framework, the paper provides an explanation of how and why certain policies came into effect and practice in the country. After reviewing pertinent literature in the field and key policy documents of the country, several key informant interviews were conducted to explore the interests and incentives of key policy level actors, to examine the politics of policymaking and implementation in a continued socio-political context, and to illustrate the issues and dynamics in this regard. The paper concludes with the argument that government policies to tackle child labour mainly serves the interests of the political elites and fail to focus on the wellbeing of working children. The country needs to refresh its conceptual and pragmatic understanding of child labour and create a
more resilient social protection system for working children fitting in the post-pandemic context.

***Keywords* —** Bangladesh, Child Labour, Political Economy, Working Children

The Author is a doctoral researcher at the Institute of Development Studies in the University of Sussex; Brighton, UK, BN19RE, Phone: +447448388416; Email: m.hoque1@ids.ac.uk

Women in Nazi Germany: Were They Moulded to be Instruments of Destruction or They Willingly Did So?

Archi Biswas

***Abstract*—** During the Third Reich, women of Nazi Germany had to cope up with the constantly changing wills placed upon them due to various political, economic and social pressures as well the events of the World War II. The goal of the dissertation is to navigate the roles of women in Nazi Germany and to critically analyze as to whether they were shaped to be the instruments of destruction or they took part in it voluntarily. The German women had played a crucial role in the growth of the Nazi party and have contributed to the war efforts while some of them were also a part of the resistance. The Jewish women experience during Holocaust were not of much different from those of men but it would be entirely misguiding to put them into the same category. Therefore, a part of this dissertation will also be dedicated to the other minority women such as Roma, Blacks and Lesbians.

***Keywords*—** Holocaust, Hitler, Nazi Germany, women.

Archi Biswas is with the Presidency University, India (e-mail: mona42bis@gmail.com).

Resistance Movements during the Holocaust

Urmimala Dev

***Abstract*—** My dissertation will focus on the Resistance Movements during the Holocaust. This paper will also examine the different repressive policies employed to put them down and up to how much was these movements successful in achieving their goals. The Holocaust was one of the gravest tragedies in our collective history. Amidst the horror of the merciless massacres perpetrated against anyone considered "undesirable", there were people brave enough to stand up against the Nazi oppression. This paper aims to talk about these brave souls who resisted Nazi plans for selective extermination, putting their own lives in danger. I will use newspaper cuttings, books written by survivors serving as a first-hand account of the events, transcripts from interviews, pictures from several archives clicked during that time, and several others as my primary sources.

***Keywords*—** holocaust, Nazi, resistance, second world war.

Urmimala Dev is with the Presidency University, Kolkata, India (e-mail: urmimaladev98@gmail.com).

History of Holy Water and the Discovery of the Mount Zion Fountain: A Unifying Spring for People of All Faiths

Brandon WinsGaris

Abstract— *The History of Holy Water* is a cross-cultural, non-proselytizing nonfiction historical review of Holy Water, one that illustrates how water has been used through the ages for spiritual purposes in three well-known religions (Judaism, Christianity, and Islam); and it documents the landmark discovery of the Mount Zion Fountain and Jerusalem Aqueduct System, which extends from the Solomon's Pools near Bethlehem to the fountain and beyond.

The story of Holy Water and the fountain are the story of humanity, a shared ancestry. Followers of various faiths are too often divided, but the singular, most common characteristic shared amongst world religions is the ritual use of Holy Water and its association to the divine. In that sense, this book will serve as a bridge across the often-raging river of disagreement between those who value both spirituality and unity.

While other titles have been published about the physical and spiritual importance of water, as well as Biblical archeology topics, *The History of Holy Water* is unique and contains a great deal of unique content. It also has a unifying interfaith foundation that will bring people together adding broader appeal. *The History of Holy Water* advocates for unity and the restoration of a functional fountain on Mount Zion in Old Jerusalem as an interfaith and unification symbol representing and promoting peace.

Ritual purification is a feature of many religions not only to include Judaism, Islam and Christianity, but also Hinduism, Buddhism, Sikhism, the Bahá'í Faith, Shinto, Taoism, and the Rastafari movement. The aim of these rituals is to remove specifically defined uncleanliness prior to a particular type of activity, and especially prior to the worship of a deity. This ritual uncleanliness is not identical with ordinary physical impurity, such as dirt stains; nevertheless, body fluids are generally considered ritually unclean. The purpose of ritual purification is to cleanse the inner soul, from the sinful material world, to have a purer divine connection to the almighty allowing the purest spirit of light to shine from your soul creating a higher state of wellbeing of the divine self to be amongst the presence of the Lord God Almighty! This purified, more divine self, then, serves the greater truth in life by creating and surviving in a better and more divine Earthly world as it is in heaven.

Through the comparative study of the major monotheistic Abrahamic religions, it will yield a deeper understanding of the fundamental philosophical concerns of ritual purification through use of water as the medium to wash away the sins of the world in the following religions: Judaism, Islam and Christianity.

B. W. Author is an advocate for the Mount Zion Fountain, Corpus Christi, TX 78418 USA corresponding author, phone: 361-500-1814 (e-mail: brandonwgaris@gmail.com, website: www.BrandonWinsGaris.com).

You shall be holy, for I the Lord your God am holy.

In a divided world, water can unite us.

Keywords—Holy Water, Interfaith, Mount Zion Fountain, Spiritual Purification.

I. Ritual Purification Overview

As a sacred element of life, water is considered a purifier in most religions, however in general people are unaware of the rituals of Holy Water in differing religions and cultures from one another. For example, Judaism uses the *Mikveh,* a holy bath of water, which offers the individual, the community, and the nation of Israel the remarkable gift of purity and holiness. The role of immersion in the *Mikveh* has offered a gateway to purity ever since the creation of man. In Islam, the five daily prayers can be done in most cases (*Tayammum*) after completing washing certain parts of the body using clean water (*wudu*). One of the most prominent features of the Islamic faith is the practice of Taharah, meaning purification, through purity, cleanliness of body, clothes, place, and all aspects of Muslim's life. Taharah (Purification) is a prerequisite for all Islamic 'Ibadah (Worship). In, Christianity, the baptism in water of Jesus Christ is an important moment in Christian theology, and is the third most important feast of the Church called Theophany or Epiphany, following Easter and Pentecost. For roles of consecration, Christian's use Holy Water for the sacrament of baptism to cleanse away the original sin, blessings of persons, places, and objects, or as a means of repelling evil.

Through God's design the King of King's has created this sacrament of divine grace through Holy Water that can cleanse away the profane, creating purification and spiritual rebirth. Be holy as God is holy, be cleansed by the washing of water through the Word!

2 Peter 3:5

[5] But they deliberately forget that long ago by God's word the heavens came into being and the earth was formed out of water and by water.

Taking the bride to the bath house (Mikveh), Shalom Koboshvili, 1939.

Ritual purification is a feature of many religions, not only to include Judaism, Islam and Christianity, but also Hinduism, Buddhism, Sikhism, the Bahá'í Faith, Shinto, Taoism, and the Rastafari movement. The aim of these rituals is to remove specifically defined uncleanliness prior to a particular type of activity, and especially prior to the worship of a deity. This ritual uncleanliness is not identical with ordinary physical impurity, such as dirt stains; nevertheless, body fluids are generally considered ritually unclean. The purpose of ritual purification is to cleanse the inner soul, from the sinful material world, to have a purer divine connection to the almighty allowing the purest spirit of light to shine from your soul creating a higher state of wellbeing of the divine self to be amongst the presence of the Lord God Almighty! This purified more divine self then serves the greater truth in life by creating and surviving in a better and more divine Earthly world as it is in heaven.

Through the comparative study of the major monotheistic Abrahamic religions, this book will help yield a deeper understanding of the fundamental philosophical concerns of ritual purification through use of water as the medium to wash away the sins of the world in the following religions: Judaism, Islam and Christianity.

Proverbs 18:4

The words of the mouth are deep waters, but the fountain of wisdom is a rushing stream.

Judaism

Pool of a medieval mikvah in Speyer, dating back to 1128.

The Hebrew Bible has many rituals of purification relating to menstruation, childbirth, sexual relations, nocturnal emission, unusual bodily fluids, skin disease, death, and animal sacrifices. Modern mainstream Judaism is based on a combination of the Hebrew Bible and Jewish oral law, which includes the Mishnah and Gemarrah (together comprising the Talmud) in addition to other rabbinic commentaries; this oral law further specifies regulations for ritual purity, including obligations relating to excretory functions, meals, and waking. The regulations of biblical and oral law generally prescribe a form of water-based ritual washing in Judaism for removal of any ritual impurity, sometimes requiring just washing of the hands, and at other times
requiring full immersion; the oral law requires the use of *un-drawn water* for any ritual full
immersion - either a natural river/stream/spring, or a special bath (a *Mikvah*), which contains rain-water.

These regulations were variously observed by the ancient Israelites; contemporary Orthodox Jews and (with some modifications and additional leniencies) some Conservative Jews continue to observe the regulations, except for those tied to sacrifice in the Temple in Jerusalem, as the Temple no longer fully exists. These groups continue to observe many of the hand washing rituals. Of those connected with full ritual immersion, perhaps the quintessential immersion rituals still carried out are those related to nidda, according to which a menstruating woman must avoid contact with her husband, especially avoiding sexual contact, and may only resume contact after she has first immersed herself fully in a mikvah of *living water* seven days after her menstruation has ceased.

In December 2006, the Committee on Jewish Law and Standards of Conservative Judaism re-affirmed the traditional requirement that Conservative women ritually immerse following menstruation. In doing so, it adopted multiple opinions regarding details, including an opinion reaffirming traditional (Orthodox) practices and concepts, an opinion adapting certain leniencies, including counting seven days from start of menstruation rather than its end, and an opinion reformulating the theological basis of the practice, basing it on concepts other than ritual purity. Classical ritual immersion and associated requirements are generally not observed by Reform Judaism or Reconstructionist Judaism, with the exception that both generally include immersion as part of the ritual for Conversion to Judaism, although Reform Judaism does not require it.

Ezekiel 36:25

I will sprinkle clean water on you, and you will be clean

Islam

The *al-Kas* ablution fountain, washing station in the Al-Aqsa Mosque on the Temple Mount, Jerusalem, Israel.

Islamic ritual purification is particularly centered on the preparation for ritual prayer; theoretically ritual purification would remain valid throughout the day, but is treated as invalid on the occurrence of certain acts: flatulence, sleep, contact with the opposite gender (depending on which school of thought), unconsciousness, and the emission of blood, semen, or vomit. Some schools of thought mandate that ritual purity is necessary for holding the Qur'an.

Ritual purification takes the form of ablution, in a lesser form (*wudu*), and greater form (*ghusl*), depending on the circumstance; the greater form is obligatory by a woman after she ceases menstruation, on a corpse that didn't die during battle and after sexual activity, and is optionally used on other occasions, for example just prior to Friday prayers or entering ihram.

An alternative "dry ablution" (*tayammum*), involving clean sand or earth, is used if clean water is not available or if suffering from an illness which would be worsened by the use of water; this form is invalidated in the same circumstances as the other forms, and also whenever water becomes available and safe to use. And is also necessary to be repeated (renewed) before every obligatory prayer.

The obligatory activities of the lesser form include beginning with the intention to purify oneself, washing of the face, arms, head, and feet. While some optional acts also exist such as recitation of the Bismillah, Oral hygiene / the brushing of the teeth, the washing of the mouth, nose at the beginning, washing of arms to the elbows and washing of the ears at the end; additionally, recitation of the Shahadah. The greater form (*ghusl*) is completed by first performing *wudu* and then ensuring the entire body is washed. Some minor details of Islamic ritual purification may vary between different schools of thought as well as scholarly opinions and between the different sects of Islam.

The Noble Quran, Surah Al-Anfal [8:11]

[*Remember*] *when He overwhelmed you with drowsiness* [*giving*] *security from Him and sent down upon you from the sky, rain by which to purify you and remove from you the evil* [*suggestions*] *of Satan and to make steadfast your hearts and plant firmly thereby your feet.*

Christianity

Giotto di Bondone (1267-1337), Cappella Scrovegni a Padova, Life of Christ, Washing of Feet.

Baptism, as a form of ritual purification, occurs in several religions related to Judaism, and most prominently in Christianity; Christianity also has other forms of ritual purification. In older churches, and modern Roman Catholic churches, there are a number of lavers around the building for the laity to use as ritual symbolism of cleansing themselves, usually by dipping the fingertips in the holy water, and then making the sign of the cross. In traditional liturgical churches a laver, often embedded in the wall, exists for the priest and deacon to wash their hands before celebrating the Eucharist.

Many ancient churches were built with a large fountain in the courtyard. It was the tradition for Christians to wash before entering the church for worship. This usage is also legislated in the *Rule of St. Benedict*, as a result of which, many medieval monasteries were built with communal lavers for the monks or nuns to wash up before the Daily Office.

Traditionally, Christianity adhered to the biblical regulation requiring the purification of women after childbirth; this practice was adapted into a special ritual known as the churching of women, for which there exists liturgy in the Church of England's Book of Common Prayer, but its use is now rare in Western Christianity. The churching of women is still performed in a number of Eastern Orthodox, Oriental Orthodox and Eastern Catholic churches.

Roman Catholics, Eastern Orthodox and High church Anglicans are also traditionally required to regularly attend confession, as a form of ritual purification from sin, especially as preparation before receiving the Eucharist. For Catholics, this is required at least once a year and required for those who are guilty of unconfessed mortal sins.

In Reformed Christianity, ritual purity is achieved through the Confession of Sins, and Assurance of Forgiveness, and Sanctification. Through the power of the Holy Spirit, believers offer their whole being and labor as a 'living sacrifice'; and cleanliness becomes a way of life;

Romans 12:1-2

A Living Sacrifice

12 *Therefore, I urge you, brothers and sisters, in view of God's mercy, to offer your bodies as a living sacrifice, holy and pleasing to God—this is your true and proper worship.* [2] *Do not conform to the pattern of this world, but be transformed by the renewing of your mind. Then you will be able to test and approve what God's will is—his good, pleasing and perfect will.*

John 13:1-17

Jesus Washes His Disciples' Feet

13 *It was just before the Passover Festival. Jesus knew that the hour had come for him to leave this world and go to the Father. Having loved his own who were in the world, he loved them to the end.*

[2] *The evening meal was in progress, and the devil had already prompted Judas, the son of Simon Iscariot, to betray Jesus.* [3] *Jesus knew that the Father had put all things under his power, and that he had come from God and was returning to God;* [4] *so he got up from the meal, took off his outer clothing, and wrapped a towel around his waist.* [5] *After that, he poured water into a basin and began to wash his disciples' feet, drying them with the towel that was wrapped around him.*

[6] *He came to Simon Peter, who said to him, "Lord, are you going to wash my feet?"*

[7] *Jesus replied, "You do not realize now what I am doing, but later you will understand."*

[8] *"No," said Peter, "you shall never wash my feet."*

Jesus answered, "Unless I wash you, you have no part with me."

[9] *"Then, Lord," Simon Peter replied, "not just my feet but my hands and my head as well!"*

[10] *Jesus answered, "Those who have had a bath need only to wash their feet; their whole body is clean. And you are clean, though not*

every one of you." [11] For he knew who was going to betray him and that was why he said not every one was clean.

[12] When he had finished washing their feet, he put on his clothes and returned to his place. "Do you understand what I have done for you?" he asked them. [13] "You call me 'Teacher' and 'Lord,' and rightly so, for that is what I am. [14] Now that I, your Lord and Teacher, have washed your feet, you also should wash one another's feet. [15] I have set you an example that you should do as I have done for you. [16] Very truly I tell you, no servant is greater than his master, nor is a messenger greater than the one who sent him. [17] Now that you know these things, you will be blessed if you do them.

Matthew 28:18:20

The Great Commission

[18]Then Jesus came to them and said, "All authority in heaven and on earth has been given to me. [19]Therfore go and make disciples of all nations, baptizing them in the name of the Father and of the Son and of the Holy Spirit, [20]and teaching them to obey everything I have commanded you. And surely I am with you always, to the very end of the age."

II. Fountain of Life

The *Fountain of Life***,** or in its earlier form the *Fountain of Living Waters*, is a Christian iconography symbol associated with baptism, first appearing in the 5th century in illuminated manuscripts and later in other art forms such as panel paintings.

The symbol is usually shown as a fountain enclosed in a hexagonal structure capped by a rounded dome and supported by eight columns. The fountain of living waters, *fons vivus* is a baptismal font (a water fountain in which one is baptized, and thus reborn with Christ), and is often surrounded by animals associated with Baptism such as deer. The font probably represents the octagonal Lateran Baptistery in Rome, consecrated by Pope Sixtus III (432-440), which was iconographically associated with the fountain of the water of life mentioned in;

Revelation 21:6

[6] He said to me: "It is done. I am the Alpha and the Omega, the Beginning and the End. To the thirsty I will give water without cost from the spring of the water of life.

Revelation 7:17

[17]For the Lamb at the center of the throne
will be their shepherd;
'he will lead them to springs of living water.'
'And God will wipe away every tear from their eyes.'"

The best examples date from the Carolingian period: the Godescalc Evangelistary made to commemorate the Baptism of the son of Charlemagne in 781, and in the Soissons Gospels.

Godescalc Evangelistary, commemorating the Baptism of Charlemagne's son in Rome in 781 with an image of the Fountain of Life.

Proverbs 10:11

The mouth of the righteous is a Fountain of Life!

Proverbs 13:14

The teaching of the wise is a Fountain of Life, turning a person from the snares of death!

Proverbs 14:27

The fear of the Lord is a Fountain of Life, turning a person from the snares of death!

In the Ghent Altarpiece: The Adoration of the Lamb by Jan Van Eyck (1438), the Lamb of God stands upon an altar dressed as for the Mass of the Precious Blood, with a blood-red frontal: the Lamb's blood is caught in a chalice, and its Eucharistic intention is signaled by the dove of the Holy Spirit above. In the foreground, offering the other means of grace is the Fountain of the Living Water surrounded by the faithful.

The central "Adoration of the Mystic Lamb" panel. The groupings of figures are, from top left anti-clockwise: the male martyrs, the pagan writers and Jewish prophets, the Male saints, and the female martyrs.

In Flanders at the close of the Middle Ages an intense devotion to the Precious Blood of Christ gave rise to an iconographic tradition of the 15th and 16th centuries, which rendered the theological concept of Grace, expressing Roman Catholic dogma allegorically as a fountain of blood. This transformation was first addressed in Evelyn Underhill in 1910, taking her point of departure an Assembly of Saints and the Fountain of Life of 1596 in Ghent, in which blood from the five Holy Wounds of Christ into the upper basin of a "Fountain of Life" and streams out through openings in the lower "Fountain of Mercy". Saints and martyrs, patriarchs and prophets hold golden chalices of blood, which some empty into the fountain and the faithful hold out their hearts to receive droplets of blood.

The Mother of God of the *Life-giving Spring* or *Life-giving Font* is an epithet of the Holy Theotokos that originated with her revelation of a sacred spring in Valoukli, Constantinople, to a soldier named Leo Marcellus, who later became Byzantine Emperor Leo I (457-474). Leo built the historic Church of St. Mary of the Spring over this site, which witnessed numerous miraculous healings over the centuries, through her intercessions, becoming one of the most important pilgrimage sites in Greek Orthodoxy. Thus, the term "*Life-giving Font*" became an epithet of the Holy Theotokos and she was represented as such in iconography.

The feast day of the Life-giving Spring is celebrated on Bright Friday in the Orthodox Church, and in those Eastern Catholic Churches following the Byzantine Rite. Additionally, the icon of the Theotokos the "Life-giving Spring" is commemorated on April 4/17 in Slavic Orthodox churches.

Isaiah 41:18
[18]I will make rivers flow on barren heights,
And springs within the valleys.
I will turn the desert into pools of water,
And the parched ground into springs.

Isaiah 49:10
[10]They will neither hunger nor thirst,
nor will the desert heat or the sun beat down on them.
He who has compassion on them will guide them and lead them beside springs of water.

Numbers 21:16-17
"Gather the people together and I will give them water."
[17]Then Israel sang this song:
"Spring up, O well! Sing about it"

Job 29:23
They waited for me as for showers
And drank in my words as the Spring rain.

Psalms 114:8
Who turned the rock into a pool, The hard rock into springs of water.

III. ZAMZAM WELL

Pilgrims visiting Zamzam Well in Mecca, Saudi Arabia.

The *Well of Zamzam* (or the *Zamzam Well*, or just *Zamzam*; Arabic: زمزم) is a well located within the Masjid al-Haram in Mecca, Saudi Arabia, 20 m (66 ft) east of the Kaaba, the holiest place in Islam. According to Islamic belief, it is a miraculously-generated source of water from God, which began thousands of years ago when Abraham's (Ibrāhīm) infant son Ishmael ('Ismā'īl) was thirsty and kept crying for water. Millions of pilgrims visit the well each year while performing the Hajj or Umrah pilgrimages, in order to drink its water.

Muslims believe that the Zamzam well is a contemporary miracle, never having gone dry despite the millions of liters of water attributed to the well consumed every year. It had been deepened several times in history during periods of severe droughts.

Traditional Origin of the Zamzam Well:

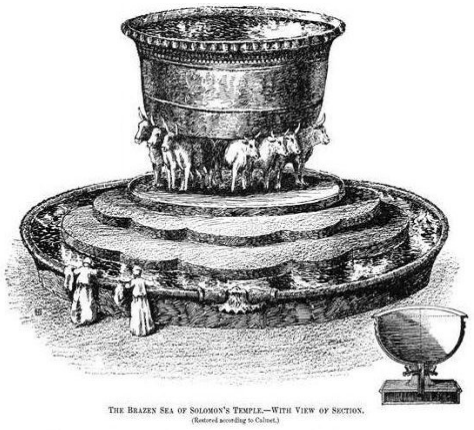
Islamic history states that the Zamzam Well was revealed to Hagar, the second wife of Abraham and mother of Ismail. According to Islamic tradition, she was desperately seeking water for her infant son, but she could not find any, as Mecca is located in a hot dry valley with few sources of water. Muslim traditions say that Hagar ran seven times back and forth in the scorching heat between the two hills of Safa and Marwah, looking for water. Getting thirstier by the second, Ishmael scraped the land with his feet, where suddenly water sprang out. There are other versions of the story involving God sending his angel, Gabriel, who kicked the ground with his heel and the water rose.

The name of the well traditionally comes from the phrase Zomë Zomë, meaning 'stop', which, according to legend, was a command repeated by Hagar during her attempt to contain the spring water.

According to Islamic tradition, Abraham rebuilt the Bait-ul-Allah ("House of God", cognate of the Hebrew-derived place name Bethel) near the site of the well, a building which had been originally constructed by Adam, and today is called the Kaaba, a building toward which all Muslims around the world face in prayer, five times each day. The Zamzam Well is located approximately 20 m (66 ft) east of the Kaaba.

Bottle of Zamzam water as sold in the local mosque.

IV. MOLTEN SEA



An artist's rendition of the Brazen Sea (or Molten Sea) of Solomon.

The *Molten Sea* or *Brazen Sea* (ים מוצק "cast metal sea") was a large basin in the Temple in Jerusalem made by Solomon for ablution of the priests. It is described in 1 Kings 7 and 2 Chronicles 4.

The Molten Sea stood in the southeastern corner of the inner court. The fact that it was a washbasin that was too large to enter from above lends to the idea that water would likely have flowed from it down into a sub-container beneath. The water was supplied by a conduit from Solomon's Pools. The Molten Sea was made of brass or bronze, which Solomon had taken from the captured cities of Hadarezer, the king of Zobah, 1 Chronicles 18:7-8. Ahaz later removed this laver from the oxen, and placed it on a stone pavement, 2 Kings 16:17-18. Then the Babylonians destroyed the Molten Sea, 2 Kings 25:1-16**.**

In Rabbinical Literature

The basin contained water sufficient for 150 ritual baths (mikvah). The symbolism of the brazen sea is described in detail in the Midrash Tadshe. The sea represented the world; the ten cubits of diameter corresponded to the ten Sefirot; and it was round at the top as the heavens are
round. The depth of the sea was five cubits, corresponding to the distance of five hundred years' journey between heaven and earth (compare Chagigah 13a). The band of thirty cubits around it corresponded to the Ten Commandments, to the ten words of God at the creation of the world, and to the ten Sefirot: for the world can exist only when the Ten Commandments are observed; and the ten Sefirot as well as the ten words of God were the instruments of the Creation. The two rows of colocynths (knops) below the rim were symbolic of the sun and the moon, while the twelve oxen on which the sea rested represented the zodiac ("mazzalot").

2 Chronicles 4:1-6
The Temple's Furnishings

[1]He made a bronze altar twenty cubits long, twenty cubits wide and ten cubits high. [2]He made the Sea of cast metal, circular in shape, measuring ten cubits from rim to rim and five cubits high. It took a line of thirty cubits to measure around it. [3]Below the rim, figures of bulls encircled it—ten to a cubit. The bulls were cast in two rows in one piece with the Sea.

[4]The Sea stood on twelve bulls, three facing north, three facing west, three facing south and three facing east. The Sea rested on top of them, and their hindquarters were toward the center. [5]It was a handbreadth in thickness, and its rim was like the rim of a cup, like a lily blossom. It held three thousand baths.

[6]He then made ten basins for washing and placed five on the south side and five on the
north. In them the things to be used for the burnt offerings were rinsed, but the Sea was to be
used by the priests for washing.

V. OLD JERUSALEM'S LIVING WATERS

Gihon Spring

The *Gihon Spring* was the main source of water for the City of David, the traditional original site of Jerusalem. One of the world's major intermittent springs - and a reliable water source that made human settlement possible in ancient Jerusalem - the spring was not only used for drinking water, but also initially for irrigation of gardens in the adjacent Kidron Valley which provided a food source for the ancient settlement. The spring, being intermittent, required the excavation of the Pool of Siloam which stored the large amount of water needed for the town when the spring was not flowing. Currently it is sometimes used by Jewish men as a sort of mikveh.

Three main water systems allowed water to be brought from the spring to the city under cover:

- The Middle Bronze Age channel - a fairly straight channel dating from the Middle Bronze Age, cut 20 feet into the ground, and then covered with slabs (which were then hidden by foliage). This led from the spring to the Pool of Siloam and was an aqueduct.

- Warren's Shaft - a steep tunnel, dating from slightly later than the Middle Bronze Age channel, leading from

the Well Gate at the top of Ophel above Gihon, down to the spring. This passage was for people to collect water from the spring.

- Hezekiah's tunnel - a winding tunnel carved into the rock, leading from the spring to the Pool of Siloam. Dating from the time of Hezekiah, and seemingly built in response to the threat of siege by Sennacherib, it was an aqueduct that effectively replaced the Middle Bronze Age channel.

In 1997, while a visitor center was being constructed, the spring was discovered to have been heavily fortified since the Middle Bronze Age, when archaeologists unexpectedly uncovered two monumental towers - one protecting the base of Warren's Shaft, and the other protecting the spring itself. Due to the area around the site still being inhabited, and hence not excavated, it is unknown whether any further fortifications exist (though a further tower to the south of that protecting Warren's Shaft is thought likely).

During an archaeological dig in 2009, a fragment of a monumental stone inscription securely dated to the eighth century BC was discovered. Although only fragments of Hebrew lettering survive, the fragment proves that the city had monumental public inscriptions and the corresponding large public buildings in the eighth century.

Pool of Siloam

The *Pool of Siloam* is a rock-cut pool on the southern slope of the City of David, the original site of Jerusalem, located outside the walls of the Old City to the southeast. The pool was fed by the waters of the Gihon Spring, carried there by two aqueducts.

Pool of Siloam.

Biblical History:

The Pool of Siloam is mentioned several times in the Bible. Isaiah 8:6 mentions the pool's waters, while Isaiah 22:9 refers to the construction of Hezekiah's tunnel.

Isaiah 8:6
6 "Because this people has rejected
the gently flowing waters of Shiloah
and rejoices over Rezin
and the son of Remaliah,

Isaiah 22:9
9 You saw that the walls of the City of David
were broken through in many places;
you stored up water
in the Lower Pool.

For Christians, the *Pool of Siloam* has additional special significance: it is mentioned in the Gospel of John as the location to which Jesus sent "a man blind from birth" in order to complete the healing of the man.

John 9:1-7
Jesus Heals a Man Born Blind
9 As he went along, he saw a man blind from birth. 2 His disciples asked him, "Rabbi,
who sinned, this man or his parents, that he was born blind?"
3 "Neither this man nor his parents sinned," said Jesus, "but this happened so that the works of God might be displayed in him. 4 As long as it is day, we must do the works of him who sent me. Night is coming, when no one can work. 5 While I am in the world, I am the light of the world."
6 After saying this, he spit on the ground, made some mud with the saliva, and put it on the man's eyes. 7 "Go," he told him, "wash in the Pool of Siloam" (this word means "Sent"). So the man went and washed, and came home seeing.

A substantial remodeling of a nearby pool, thought to be the *Pool of Siloam*, was constructed in the 5th century under Byzantine direction and is said to have been built at the behest of Empress Aelia Eudocia. This pool, having been somewhat abandoned and left to ruin, partly survives to the present day, surrounded by a high wall of stones on all sides (except for an arched entrance to Hezekiah's tunnel – which was only rediscovered in the 19th century).

Lower Pool

City of David, Lower Pool; Shiloach Pool.

Ancient records report that during the Second Temple period, there was a *Lower Pool*. In the autumn of 2004, Ir David Foundation workers excavating for a sewer near the present-day pool uncovered stone steps, and Ronny Reich and Eli Shukron (prominent archaeologists) were called in; it became obvious to them that these steps were likely to have been part of the Second Temple period pool. Excavations commenced and confirmed the initial supposition; the find was formally announced on August 9, 2005, and received substantial international media attention. The pool is less than 70 yards from the edge of the Byzantine reconstruction of a pool previously thought to be the Pool of Siloam. This small pool collected some of the water as it emptied there at the southern end of Hezekiah's tunnel. The water continued on through a channel into the recently discovered Pool of Siloam. The source of the water is from the Gihon Spring, located at the northern end of Hezekiah's tunnel on the eastern side of the City of David. An ancient pool (Upper Pool) existed near the Gihon Spring but was no longer used after King Hezekiah redirected the waters to the western side of the city.

The *lower pool* is not perfectly rectangular, but a soft trapezoid. There are three sets of five steps, two leading to a platform, before the bottom is reached, and it has been suggested that the steps were designed to accommodate various water levels. The pool is stone-lined, but underneath, there is evidence of an earlier version that was merely plastered (to help it retain water). Coins found within this plaster date from the time of Alexander Jannaeus (104–76 BC), while a separate collection of coins, dating from the time of the First Jewish–Roman War (AD 66–70), were also found.

How much of the pool and its surrounding structures were a result of monumental construction by Herod the Great is not yet understood (as of September 2006); nor is the relationship of this pool to the earlier one (i.e., why it was built when the earlier pool already existed). A portion of this pool remains unexcavated, as the land above it is owned by a nearby Greek Orthodox church and is occupied by an orchard known as the King's Garden.

Nehemiah 3:15

[15]*The Fountain Gate was repaired by Shallun son of Kol-Hozeh, ruler of the district of Mizpah. He rebuilt it, roofing it over and putting its doors and bolts and bars in place. He also repaired the wall of the Pool of Siloam, by the King's Garden, as far as the steps going down from the City of David.*

As a freshwater reservoir, it would have been a major gathering place for ancient Jews making religious pilgrimages to the city. The Gospel of John suggests that it was probably used as a mikvah (ritual bath), although mikvahs are usually much smaller in size; if the pool were a mikvah, it would be the largest ever found by a substantial margin. Yoel Elitzur has proposed that the pool was used for swimming rather than ritual immersion. It is thought that the current structure was originally the Shrine of the Four Nymphs (Tetranymphon), a nymphaeum built by Hadrian during the construction of Aelia Capitolina in 135 and mentioned in Byzantine works such as the 7th-century Chronicon Paschale; other nymphaeum built by Hadrian, such as that at Sagalassos, are very similar.

Sultan's Pool

Yemin Moshe in snow, with Birket es-Sultan in foreground. 1920s.

The *Sultan's Pool* is an ancient water basin by the west side of Mount Zion. Its origins likely date to the time of Herod and some think it is the Snake pool mentioned by Josephus. In Hasmonean times, water was fed into it via the lower aqueduct. The Ottoman sultans (for whom the pool is named) enlarged it into a reservoir measuring 67 m × 169 m × 12 m. It was part of the water supply network for Jerusalem from antiquity to late Ottoman Empire times. During the Crusaders times, it was known as Lacus Germani.

Al-Kas Ablution Fountain

The mosque's "Al-Kas" ablution fountain.

The Al-Aqsa Mosque main ablution fountain, known as *al-Kas* ("the Cup"), is located north of the mosque between it and the Dome of the Rock. It is used by worshipers to perform wudu, a ritual washing of the hands, arms, legs, feet, and face before entry into the mosque. It was first built in 709 by the Umayyads, but in 1327–28 Governor Tankiz enlarged it to accommodate more worshipers. Although originally supplied with water from Solomon's Pools near Bethlehem, it currently receives water from pipes connected to Jerusalem's water supply. In the 20th-century, al-Kas was provided taps and stone seating.

Proverbs 18:4

The words of the mouth are deep waters, but the Fountain of Wisdom is a rushing stream.

VI. The Lower Aqueduct of Jerusalem

Location of the Lower Aqueduct of Jerusalem: The west bank of the Hinnom Valley, above the Sultan's Pool and near the entrance to Mishkenot Sha'ananim. The aqueduct structure can be viewed both from the stepped path which runs below it and from the paved overlook above.

The aqueduct represents a true link between antiquity and modern times: From its original construction in the Hasmonean period (142-63 BCE), it continued to deliver water to Jerusalem (off and on, through various alterations) until the turn of the 20th Century! (Other parts of the system were actually supplying water to the Old City until 1967.)

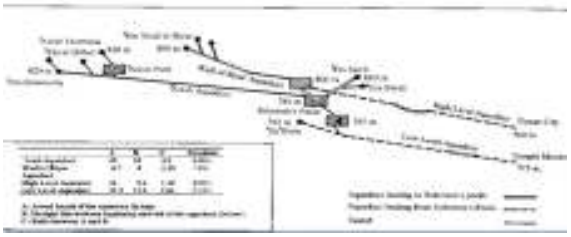
The segment of several meters that has been restored and left exposed here is a small part of a system that stretches over 60km from its furthest sources to its destination. It is vast and ingenious, representing the finest of ancient engineering, and is easily the longest and most complex aqueduct system in the Land of Israel, including even Caesarea.

What we see in the following picture is quite simple: The constructed water channel is elevated atop a buttressed stonewall, in order to maintain the required gradient; it is plastered
inside to prevent leakage and capped with stone slabs in order to minimize evaporation and
contamination of the water. (What we actually see is the aqueduct as rebuilt and repaired in the late Ottoman period, but at its core it is Hasmonean.) It can only be properly understood in relation to the larger system of which it is one part.

Aqueduct channel with plastered inside to prevent leakage and capped with stone slabs in order to minimize evaporation and contamination of the water.

An overview of Jerusalem's aqueduct system:

The system, which reached its height in the late Second Temple period (200 BCE-70 CE), consists of several independent but interconnected elements, which were built/re-built at different times. It was explored & mapped by the 19th century explorers Charles Wilson and Conrad Schick and again in recent times by Amihai Mazar and others. Think of the aqueduct system in terms of the human circulatory system: a heart, veins, and arteries.



Map of collecting pools and aqueduct channels creating the headwaters for the Jerusalem Aqueduct System.

The "heart" of the system is Solomon's Pools, three reservoirs which descend in stair-step fashion down the Artas Valley just southwest of Bethlehem [aerial view looking east].

They are monumental and quite astounding, with a combined surface area equal to more than 4 football fields and a capacity of over a quarter million cubic meters lower pool; human figure (arrow) shows scale.

The pools serve as a kind of junction or interchange for the system: Besides collecting rainwater from the plains to the west and the output of 2 local springs, the pools receive the water of the two major "veins" and then send it all out through 3 "arteries" (all discussed below). Solomon's Pools were completely refurbished by the British in 1924, when a pumping station was also installed, to send the water through iron pipes to the Old City.

One of the two "veins" feeding Solomon's Pools from the south is the Biyar Aqueduct. It is 4.7km long and descends at a gradient of 1.9% (1:53). Compared to the other aqueducts in the system, it is the shortest, straightest and steepest. It runs roughly parallel to today's Road 60, about 1 km to the east of the road. It actually collected water in three ways: (1) It conducted water from three springs in Wadi Biyar; (2) it collected rainwater from the valley via a series of dams and shafts; and (3) along much of its length it was constructed as a trench/tunnel hewn down through an aquifer layer into an aquiclude. Thus, along the seam between these two strata the aqueduct itself is draining the aquifer—like a long, continuous layer spring—with water flowing in through the tunnel walls! This technique, which the Romans borrowed from the *Persian's Qanat* underground water transportation system originating in the 1st millennium BCE, is seen nowhere else in Israel-Palestine. Built probably in the time of Herod the Great, the Biyar Aqueduct remained in service for much of the next 2000 years. It was refurbished by the British (along with Solomon's Pools) in 1924 and provided water for Jerusalem's Old City until 1967.

The other "vein" feeding Solomon's Pools is the *Arrub Aqueduct*. It gathered the waters of a group of springs lying between Bethlehem and Hebron and then, following the contours of the ridges and wadis, wound through the Judean hills for an astounding 40 km (a straight-line distance of 10 km), at an equally amazing gradient of only 0.09% (1:1100+)! For much of its length it was a simple channel hewn into the bedrock slopes, plastered and capped with stone slabs. Obstacles were overcome via two methods: (1) tunneling through some ridges and (2) elevating the channel on a stone wall/dam in order to cross wadis. The dating of the Arrub Aqueduct is not certain, but it is a good candidate for the aqueduct mentioned by Josephus for which Pontius Pilate appropriated Temple funds, thereby sparking riots in Jerusalem (Antiquities 18). It seemingly was re-built in the Mamluk period and finally went out of service in Ottoman times. (It's great length made it especially susceptible to damage or clogging, unauthorized tapping of the water, and the pilfering of cap-stones.)

One "artery" (unrelated to Jerusalem's water supply, and thus only mentioned here) is the *Herodion Aqueduct*, descending eastward from Solomon's Pools, through the Artas Valley, and ending at that desert fortress of Herod the Great; it can thus be dated to the 1st century BCE. It was long thought that its source was the spring of Artas, but further analysis and excavation now show it to be connected to (and fed by) Solomon's Pools.

Another "artery" is the Upper Aqueduct of Jerusalem. It began at Solomon's Pools, ran northward, close to the watershed line (at a higher elevation than its "Lower" counterpart); it skirted Bethlehem on the west side, closely paralleled today's Hebron Road, and apparently had as its destination "Hezekiah's Pool" in the Upper City. Its length was 14 km, with an average gradient of 0.28% (1:357). Its best-known remains are a still-visible section of stone pipe (a siphon) across a shallow valley near Rachel's Tomb. Based on inscriptions of the 10th Roman Legion on the pipe segments, the Upper Aqueduct was long dated to the Late Roman period, however it is now clear there was an earlier, parallel, high-level aqueduct section (i.e., elevated on arches), an earlier solution to the same topographical problem, i.e. transmission of the water across the valley. Several surviving piers of this high-level section were discovered and documented in the 1980s and '90s but were later removed in a building project. Thus, the Upper Aqueduct is now thought to have been Herodian in its earliest phase, and then re-built in the Late Roman period, and it may have gone out of use after the Byzantine era. Overall, it is not well-preserved. No remains have been found from the area of Jerusalem's old railway station northward, except for its confluence with the Mamilla Pool outflow channel, in the direction of the Old City (but it did not feed that pool—the elevation is not sufficient), and a segment near its terminal end, in front of the medieval fore-wall north of Jaffa Gate.

The Lower Aqueduct

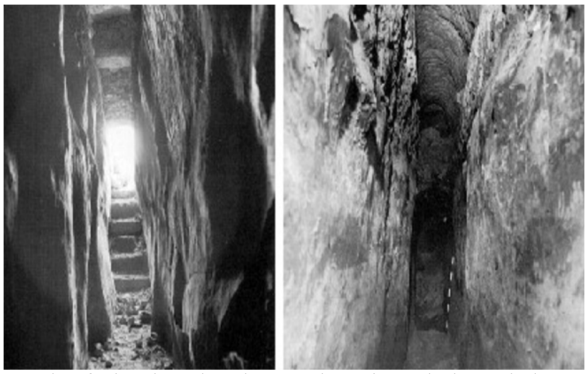
The "artery" we are dealing with here is the Lower Aqueduct of Jerusalem, which ran for 21.5 km from Solomon's Pools (765m ASL) to the Temple Mount (735m ASL). In that distance, it dropped a mere 30 meters, for a gradient of 0.14% (1:700+). Put another way, in the length of a football field, the floor of the channel dropped only 14 cm!

The original *Lower Aqueduct* has now been firmly dated to the Hasmonean period. This is based not only on ceramics and coins related to its construction but also similarities to the water systems of the Hasmonean desert fortresses, including a

distinctive type of plaster. Thus, along with one or more of the Solomon's Pools, it is (in its original form) the oldest part of Jerusalem's aqueduct system. It is mentioned a number of places in the Talmud, which says that "an aqueduct ran to [the Temple] from Eitam [a spring located near Solomon's Pools]," and a Midrash states that the same aqueduct was destroyed by the Sicarii during the Great Revolt.

The Route of the Lower Aqueduct:

From Solomon's Pools it leads northward and runs beneath the town of Bethlehem through a tunnel, which is mostly blocked today. Reaching the Armon ha-Natziv ridge (Jebel Mukkaber/E. Talpiot) it overcame this barrier by means of a tunnel stretching nearly 400m and punctuated by several vertical shafts.



The tunnel of the low-level aqueduct in Jebel Mukaber ("Armon Hanatziv") (East Talpiot) interior view near the South entrance (left) and further inside (right).

The channel emerges from the tunnel just east of and downslope from today's Haas Promenade, continues below the Sherover Promenade and then across the slopes of Abu Tor, to here—the western slopes of the Hinnom Valley. (NOTE: Many of the aqueduct remains in the area of Armon ha-Natziv, the Tayelet, and the Peace Forest have been preserved and marked: At the beginning of the Goldman Promenade one of the tunnel shafts has been exposed, with an interpretive installation; in a park across the street to the south, centered on another shaft opening, is an interesting mosaic map of the whole system.)

From here, Mishkanot Sha'ananim, the Lower Aqueduct continues northward past the Sultan's Pool, and many remains of this section can still be traced with little effort. North of the Sultan's Pool it then traverses through the Hinnom Valley on an arched stone bridge. This bridge (the Mamluk version, complete with a dedicatory inscription) was visible into the early 20th century, but was then totally obscured by modern grading and landscaping. (Note: As of 2010, two complete arches near the western end of the bridge were exposed once again in an infrastructure project; the announced plan is to uncover the entire bridge starting in 2011.)

The aqueduct continues around the slopes of Mt. Zion and then beneath (outside) the present-day city walls.

Entering the Old City west of Dung Gate, it proceeds through today's Jewish Quarter, where portions are preserved along the scarp of the Western Hill (Josephus' "Upper City") opposite the southwestern corner of the Haram al-Sharif (Herodian Temple Mount) [note tall, rock-hewn channel and later ceramic pipe].

The map shows the Jerusalem's Lower Aqueduct in its 1st century context, from the west bank of the Hinnom Valley (left), through the city (center), to its destination on the Temple Mount (upper right). Numbers indicate location of visible remains illustrated here: 1 – Reference segment at Mishkenot Sha'ananim; 2 – Mamluk Bridge across Hinnom Valley; 3 – sections outside the south wall of the Old City; 4 – Tall hewn channel opposite southwest corner of the Haram/Temple Mount.

Via the Wilson's Arch Bridge, the Lower Aqueduct led onto the Temple Mount to its ultimate (original) destination, the cisterns beneath the Haram platform. The 19th century explorers Conrad Schick and Charles Warren demonstrated that the Haram is pockmarked—akin to a Swiss cheese—with numerous water reservoirs and channels, and their mapping and numbering system still provide our only scientific knowledge of these spaces. They identified a water channel leading from Wilson's Arch [ARROW] in the direction of two particular cisterns, numbers 6 and 36 [CIRCLED].

Very nearby is the largest of the Temple Mount cisterns, the one the explorers dubbed "The Great Sea." Accompanying Warren in 1869, the English artist William "Crimea" Simpson sketched this cavernous reservoir by the light of a burning magnesium wire, and later produced this watercolor. It is still not known whether (or how) these various cisterns might be connected.

Alterations to the Lower Aqueduct:

Starting at Armon ha-Natziv and running all the way into today's Jewish Quarter there is a second, parallel channel constructed at a slightly higher level. (This is not to be confused with the "Upper Aqueduct.") In some places it is a separate, adjacent channel [PHOTO, outside Old City walls west of Dung Gate], and in others, like here in Mishkenot Sha'ananim, it exactly overrides and obscures the older Hasmonean channel. Thus, the original, lower channel was rendered forever useless. This shows clearly that the destination and uses of the water changed over time. A good guess for this rebuilding is the Byzantine period, when Jerusalem was burgeoning and the Temple Mount, by all accounts, lay unoccupied and largely forgotten. A drawing of Conrad Schick in fact shows this higher channel passing very near the top of the great vaulted cistern of the NEA Church, one possible destination for the water.

In the Early Islamic and Crusader periods, the water was probably once again directed to the Temple Mount via Wilson's Arch. At the southwest corner of the Temple Mount, the large horizontal grooves cut into the Herodian masonry probably held pipes which, in the Omayyad period (660-750 CE), conducted at least some of the water to the large palaces or pilgrim hostels situated there.

The aqueduct was restored and rebuilt by the Mamluks; their arched bridge crossing the Hinnom Valley was visible into the 20th century (see photo above), was then buried, and now (as of 2010) is being exposed once again.

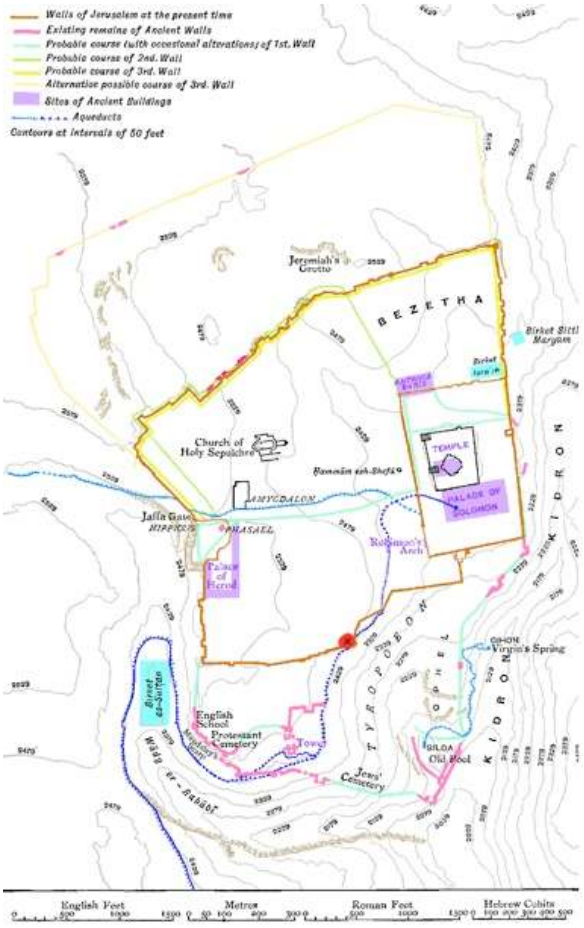
The Ottoman Turks reconstructed the Lower Aqueduct along its entire length, laying a closed ceramic pipe [PHOTO] into the existing channel in the 1500s. The pipe made the aqueduct more vulnerable to blockage, but it continued in use, off and on, into modern times.

In 1902, the Ottomans bypassed most of the ancient Lower Aqueduct (including the stretch seen here) by constructing a shunt of iron pipe on a more direct line from Bethlehem to the Old City. This pipeline never worked well, however, until the British installed pipe all the way from Solomon's Pools and built a pumping station there in 1924, a system which continued to supply water to the Old City until 1967. The partially-ruined pumping station structures, with much of the original apparatus still in place, are still visible.

VII. Mount Zion Fountain

Announcing the discovery of the *Mount Zion Fountain!*

Isaiah 55:1
Invitation to the Thirsty
"Come, all you who are thirsty,
Come to the waters;
And you who have no money,
Come, buy and eat!
Come, buy wine and milk
Without money and without cost.



Mount Zion Fountain location marked with a red dot and X above the Jerusalem Lower Aqueduct at the highest elevation entering Old Jerusalem on Mount Zion.

Entrance to Gan HaTekuma (Garden of Redemption) Archaeological Park with Mount Zion Fountain in the background. Excavations of this area were first made by Dr Nahman Avigad, 1970s/1980s.

Mount Zion Fountain outlet central structure with defensive tower walls, dating from the Ayyubid period (ca.1171-1341) surrounding the fountain. Note the original construction of Jerusalem's Lower Aqueduct has been dated back to at least the Hasmonean dynasty of classical antiquity, from 140 BCE to 37 BCE. Thus, predating the Ayyubid period fortress by over a millennium.

Mount Zion Fountain outlet viewed from atop Old Jerusalem Southern Wall. Notice the rocky cement sealing the central fountain outlet structure closed.

Mount Zion Fountain with Old Jerusalem Southern Wall in background.

View from atop Old Jerusalem Southern Wall with Red arrow pointing to the Mount Zion Fountain water drainage outlet. The black hole would have been the headwater source to feed into descending groundwater engineering works from this highest elevated point on Mount Zion.

Red arrow pointing to the Mount Zion Fountain water drainage outlet from ground level on the Eastern side.

Top of the Corner Stone of the original Mount Zion Fountain complex at ground level just outside of the Southern Wall of Old Jerusalem. Notice the Corner Stone design has the chiseled edge with rough protruding face that matches the oldest stones used in Old Jerusalem and the matches the same stone design used on the central structure of the Mount Zion Fountain.

This giant Temple Corner Stone with its most top part only revealed at ground level is a perfect example of why the discovery of the Mount Zion Fountain has lain hidden, out of sight and out of mind throughout the ages. When Jerusalem was destroyed with "Not one stone on another" they were also buried by invading siege ramps used to level Jerusalem to the ground level. These siege ramps have only been slowly unearthed to make room for development of the Modern State of Israel which has been uncovering new archeological discovers of Old Jerusalem's true ancient past only to be rediscovered in modern times today.

Mark 13:1-2

The Destruction of the Temple and Signs of the End Times

[1]As Jesus was leaving the temple, one of his disciples said to him, "Look, Teacher! What massive stones! What magnificent buildings!"

[2]"Do you see all these great buildings?" replied Jesus. "Not one stone here will be left on another; everyone will be thrown down."

Current Southern Wall of Old Jerusalem covering the Mount Zion Fountain area complex.

Current Southern Wall of Old Jerusalem above the Mount Zion Fountain and Ayyubid period structure. Additionally, the Southern Wall is also covering other archaeological locations in connection to the original function of Jerusalem's Lower Aqueduct system in this area.

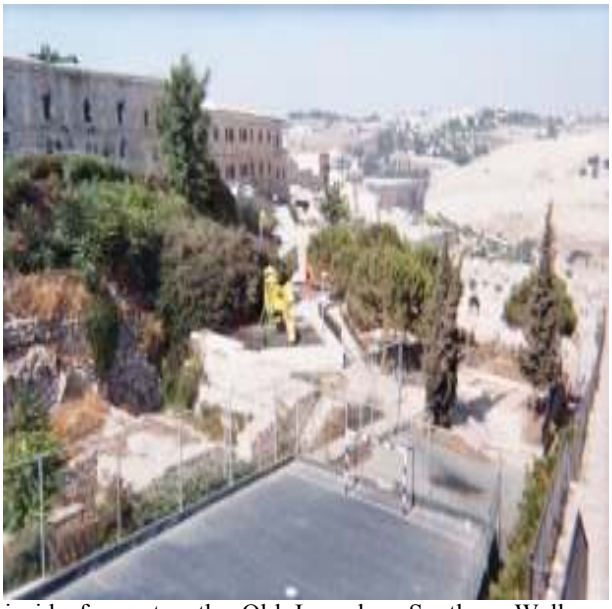
Mount Zion Archeological remains of 1st century Priestly Quarter of ancient Jerusalem as seen from outside the Southern City Wall of Old Jerusalem, Mount Zion Fountain on the inside.

Archaeologist uncover luxuries of priestly ruling class for example one house with its own cistern, mikveh and another with a bathtub. Which all require clean water hence their relative proximity to the Living Waters flowing from the Jerusalem Lower Aqueduct and Mount Zion Fountain system once in existence in this area as the source of clean water on Mount Zion.

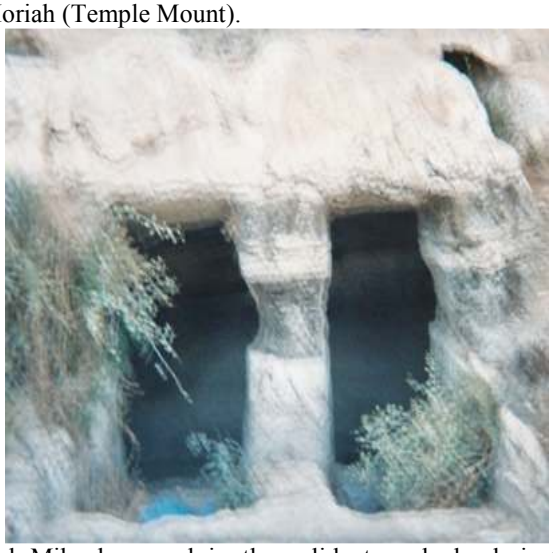
Opening to an aqueduct channel pipe outside the Southern Wall of Old Jerusalem.

Jerusalem's Lower Aqueduct channel with various outlets beneath and predating the current Southern Wall of Old Jerusalem.

View of the Gan HaTekuma (Garden of Redemption) Archaeological Park, inside Old Jerusalem, seen from atop the Southern Wall of Old Jerusalem facing North.



View inside from atop the Old Jerusalem Southern Wall facing North East down towards the entrance to the Great Vaulted Cistern of the NEA Church inside Gan HaTekuma (Garden of Redemption) Archaeological Park and further down the mountain is the Dome of the Rock area Southern end of Mount Moriah (Temple Mount).



Original Mikveh carved in the solid stone bedrock just outside the Southern Wall of Old Jerusalem in an Eastern descending flow line from the Mount Zion Fountain area at an elevation well above the final destination of Jerusalem's Lower Aqueduct to the Temple Mounts several water reservoirs, cisterns and the Al-Kas ablution fountain.

The *Mount Zion Fountain* and *Jerusalem Lower Aqueduct* system is set in stone and has been protected below ground level throughout antiquity to modern times. These living waters of purification have been a central source of renewal and promise of life to the Holy City of Jerusalem. Throughout the ages of devastation, destruction and complete decimation of Jerusalem's past, these aqueduct channels have lost their use in modern times. Thus, preventing the Orthodox practice of Spiritual Cleansing with Holy Water, a once central way of life

in Jerusalem. According to several scriptures in the Bible, revelations alluding to water flowing from Mount Zion need to be fulfilled to restore the divine ordinances of the Holy Land creating a heavenly Jerusalem, for the Living God's holy, eternal city on Earth for all Nations. After excavations are completed a restoration project can then take place restoring the living waters to Jerusalem through the original design and Jerusalem aqueduct system. Drought and low flowing spring water feeding Solomon's Pools has also resulted with the heart of the pools and veins of the headwaters to the aqueduct channels being empty. However, with modern technology we can provide a new pump house in Bethlehem to once again refill and restore the Solomon's Pools ensuring its continual use thus reviving the heart and veins of the system once again.

Psalm 46:4

4 There is a river whose streams make glad the city of God, the holy place where the Most High dwells.

Ezekiel 47:1-12

The River from the Temple

[1]The man brought me back to the entrance to the temple, and I saw water coming out from under the threshold of the temple toward the east (for the temple faced east). The water was coming down from under the south side of the temple, south of the altar. [2]He then brought me out through the north gate and led me around the outside to the outer gate facing east, and the water was trickling from the south side.

[3]As the man went eastward with a measuring line in his hand, he measured off a
thousand cubits and then led me through water that was ankle-deep. [4]He measured off
another thousand cubits and led me through water that was knee-deep. He measured off another
thousand and led me through water that was up to the waist. [5]He measured off another thousand,
but now it was a river that I could not cross, because the water had risen and was deep enough to swim in—a river that no one could cross. [6]He asked me, "Son of man, do you see this?"

Then he led me back to the bank of the river. [7]When I arrived there, I saw a great number of trees on each side of the river. [8]He said to me, "This water flows toward the eastern region and goes down into the Arabah, where it enters the Dead Sea. When it empties into the sea, the salty water there becomes fresh. [9]Swarms of living creatures will live wherever the river flows. There will be large numbers of fish, because this water flows there and makes the salt water fresh; so where the river flows everything will live. [10]Fishermen will stand along the shore; from En Gedi to En Eglaim there will be places for spreading nets. The fish will be of many kinds—like the fish of the Mediterranean Sea. [11]But the swamps and marshes will not become fresh; they will be left for salt. [12]Fruit trees of all kinds will grow on both banks of the river. Their leaves will not wither, nor will their fruit fail. Every month they will bear fruit, because the water from the sanctuary flows to them. Their fruit will serve for food and their leaves for healing."

Revelation 22:1-2

Eden Restored

[1]Then the angel showed me the river of the water of life, as clear as crystal, flowing from the throne of God and of the Lamb [2]down the middle of the great street of the city. On each side of the river stood the tree of life, bearing twelve crops of fruit, yielding its fruit every month. And the leaves of the tree are for the healing of the nations.

VIII. Conclusion

The *Temple Mount* (Hebrew: הַר הַבַּיִת, *Har HaBáyit*, "Mount of the House [of God, i.e. the Temple in Jerusalem]"), known to Muslims as the *Haram esh-Sharif* (Arabic: الحرم الشريف, *al-Ḥaram al-Šarīf*, "the Noble Sanctuary", or الحرم القدسي الشريف, *al-Ḥaram al-Qudsī al-Šarīf*, "the Noble Sanctuary of Jerusalem") and the *Al Aqsa Compound*, is a hill located in the Old City of Jerusalem that for thousands of years has been venerated as a holy site in Judaism, Islam, and Christianity alike.

The present site is a flat plaza surrounded by retaining walls (including the Western Wall) which was built during the reign of Herod the Great for an expansion of the temple. The plaza is dominated by three monumental structures from the early Umayyad period: the al-Aqsa Mosque, the Dome of the Rock and the Dome of the Chain, as well as four minarets. Herodian walls and gates, with additions from the late Byzantine and early Islamic periods, cut through the flanks of the Mount. Currently it can be reached through eleven gates, ten reserved for Muslims and one for non-Muslims, with guard posts of Israeli police in the vicinity of each.

According to Jewish tradition and scripture, the First Temple was built by King Solomon the son of King David in 957 BCE and destroyed by the Neo-Babylonian Empire in 586 BCE – however no substantial archaeological evidence has verified this or the actual location of the First Temple. The Second Temple was constructed under the auspices of Zerubbabel in 516 BCE and destroyed by the Roman Empire in 70 CE. Jewish tradition maintains it is here that a third and final Temple will also be built. The location is the holiest site in Judaism and is the place Jews turn towards during prayer. Due to its extreme sanctity, many Jews will not walk on the Mount itself, to avoid unintentionally entering the area where the Holy of Holies stood, since according to Rabbinical law, some aspect of the divine presence is still present at the site.

Among Muslims, the Mount is the site of one of the three Sacred Mosques, the holiest sites in Islam. Amongst Sunni Muslims, it is considered the third holiest site in Islam. Revered as the Noble Sanctuary, the location of Muhammad's journey to Jerusalem and ascent to heaven, the site is also associated with Jewish biblical prophets who are also venerated in Islam. Umayyad Caliphs commissioned the construction of the al-Aqsa Mosque and Dome of the Rock on the site. The Dome was completed in 692 CE, making it one of the oldest extant Islamic structures in the world. The Al Aqsa Mosque rests on the far southern side of the Mount, facing Mecca. The Dome of the Rock currently sits in the middle, occupying or close to the area where the Holy Temple previously stood.

In light of the dual claims of both Judaism and Islam, it is one of the most contested religious sites in the world. Since the Crusades, the Muslim community of Jerusalem has managed the site as a Waqf. The Temple Mount is within the Old City,

which has been controlled by Israel since 1967. After the Six-Day War, Israel handed administration of the site back to the Waqf under Jordanian custodianship, while maintaining Israeli security control. It remains a major focal point of the Arab-Israeli conflict. In an attempt to keep the status quo, the Israeli government enforces a controversial ban on prayer by non-Muslims.

The Temple was of central importance in Jewish worship in the Tanakh (Old Testament). In the New Testament, Herod's Temple was the site of several events in the life of Jesus, and Christian loyalty to the site as a focal point remained long after his death. After the destruction of the Temple in 70 CE, which came to be regarded by early Christians, as it was by Josephus and the sages of the Jerusalem Talmud, to be a divine act of punishment for the sins of the Jewish people, the Temple Mount lost its significance for Christian worship with the Christians considering it a fulfillment of Christ's prophecy, for example;

Matthew 24:1-2
[1]Jesus left the temple and was walking away when his disciples came up to him to call his attention to its buildings.
[2]"Do you see all these things?" he asked. "Truly I tell you, not one stone here will be left on another; everyone will be thrown down."

It was to this end, proof of a biblical prophecy fulfilled and of Christianity's victory over Judaism with the New Covenant, that early Christian pilgrims also visited the site. Byzantine Christians, despite some signs of constructive work on the esplanade, generally neglected the Temple Mount, especially when a Jewish attempt to rebuild the Temple was destroyed by the earthquake in 363, and as Christian worship in Jerusalem shifted to the Church of the Holy Sepulchre, and Jerusalem's centrality was replaced by Rome.

However, it is still important to understand how the fulfillment of this Biblical prophecy in the scripture of Matthew and Mark took place and its effect on the land, *"Not one stone here will be left on another; everyone will be thrown down."* Evidence of this destruction can still be seen today by the remnants of giant siege ramp works surrounding Old Jerusalem, burying the ground up to the Old City Walls from Mount Zion, surrounding the outside of the Temple Mount, down through the City of David and to the Gihon Spring and to the bottom of the Kidron Valley. The siege ramp built up to surrounding the Temple City Walls of Old Jerusalem, was similar in scope to the one built at Masada during the end of the Second Temple period, was used to "throw down every stone" and in turn encapsulated shards of the Biblical past. A past that Archeologist are still excavating today and learning more about with every layer.

None the less "tradition" teaches us to except the fact that we can't know with any degree of certainty where any of these "great buildings" once stood because of this fulfillment of the Biblical destruction attributed to "*Not one stone here will be left on another*" and the general knowledge of Jerusalem's continual destructive historical past through the millennia's have left us with a perceived unsolvable depiction of Jerusalem's authentic Biblical past.

Fortunately, finding evidence of our TRUE authentic Biblical past in Old Jerusalem, not in controversy with scripture or tradition, is still possible today, if you only know where to first start to look. Which is in the foundation of the bedrock from which every stone was cast down to and then below the ground level remains Jerusalem's Aqueduct System from whence its original construction was created providing the life source and purification necessary for Temple Life during Biblical times. Specifically, Jerusalem's Lower Aqueduct system once brought the Living Waters to Biblical Jerusalem that fed the city providing the purification rights necessary for Temple Life in the Upper City that the Gihon Spring headwaters could not provide. By restoring the Mount Zion Fountain and Jerusalem's Lower Aqueducts system, we can in turn restore the Living Waters to Jerusalem creating purification and spiritual cleansing in the Holy Land once again authored about in our Biblical texts…

Jeremiah 2:13
13 *"My people have committed two sins:*
They have forsaken me,
The spring of living water,
And have dug their own cisterns,
Broken cisterns that cannot hold water.

Zechariah 13:1
Cleansing From Sin
[1]"On that day a fountain will be opened to the house of David and the inhabitants of Jerusalem, to cleanse them from sin and impurity.

Revelation 22:17
[17]The Spirit and the bride say, "Come!" And let the one who hears say, "Come!" Let the one who is thirsty come; and let the one who wishes take the free gift of the water of life.

Zechariah 14:8
The LORD Comes and Reigns
[8]On that day living water will flow out from Jerusalem, half of it to the Eastern Sea and half of it to the Western Sea, in summer and in winter.

Revelations of a flow of living water from Jerusalem have been interpreted as the Eastern Sea being the Dead Sea and the Western Sea being the Mediterranean Sea. However, an alternative view can be based on the antiquity of archeological remains of Jerusalem's past and Judaic Law. The term "Sea" can also be interpreted as a large body of water, like in reference to the Molten Sea. The remnants of two existing examples of a large pool as a "Sea" in antiquity can both be found to the Eastern and Western sides of Old Jerusalem. The "Eastern Sea" of Old Jerusalem is the Pool of Siloam dating to the First Temple Period and its Lower Pool dating to the Second Temple Period. The "Western Sea" is the Sultan's Pool fed by the spring waters channeled from the Lower Aqueduct of Jerusalem. These large pools or "Seas" once played a vital role supporting life and offering purification to the masses of people during pilgrimages to the Holy City of Jerusalem.

Joel 3:17-18
Blessings for God's People
[17]"Then you will know that I, the Lord your God,

dwell in Zion, my holy hill.
Jerusalem will be holy;
never again will foreigners invade her.
18 "In that day the mountains will drip new wine,
and the hills will flow with milk;
all the ravines of Judah will run with water.
A fountain will flow out of the Lord's house
and will water the valley of acacias.

It is water that purifies and quenches the thirst of all nations. It is the Spring of Life harnessed from the Mount Zion Fountain that can unite the tribes of Israel and make the world whole again. The promise of a never-ending flow of water to freely give the blessing of life is what the world must seek as a God given right and promise with every life. Living in conjunction with a never-ending flow of living water of life is the divine plan that God has willed since the Garden of Eden, to create a better world in cadence with nature and all creation. The foundation has been set in stone by the Wisdom of Solomon through the Mount Zion Fountain and Jerusalem Aqueduct system that still exists today only waiting to be utilized again as in Biblical times. It is time for the trumpets to blow the sound of truth, resonating the promise of life through the Mount Zion Fountain, for the Golden Age will mark the beginning of the promise of Everlasting Life on God's Earth for all Nations.

Mount Zion Fountain marks the location at the beginning of where the headwaters would have started at the highest elevated point in order to best maximize the utilization of *Jerusalem's Lower Aqueduct* groundwater engineering plan after first reaching the Old City of Jerusalem on Mount Zion. Jerusalem's low-level aqueduct goes around Mount Zion at approximately 738-737 meters above sea level at the highest elevation before descending to Mount Moriah and the Dome of the Rock, Temple Mount area at approximately 736 meters above sea level. The original objective of the low-level aqueduct was not to just get water to its final destination at the end of a pipe to a network of water installations on the Temple Mount but to first maximize its efficiency through an extensive master groundwater engineering plan starting first from its maximum highest elevation once reaching Old Jerusalem and then flowing naturally downhill from there to the rest of the city.

Therefore, true construction of an authentic House of Prayer for All Nations that ushers forth the Spring of Life from the Mount Zion Fountain can be resurrected in peace without affliction from surrounding areas. In Fact, the Great Vaulted Cistern of the NEA Church inside Gan HaTekuma (Garden of Redemption) Archaeological Park was connected to Jerusalem's Lower Aqueduct and can be restored to work in conjunction as a "Molten Sea" which in turn would render the descending Mikveh's purity all the way to the "Eastern and Western Seas" for spiritual baths and baptisms for all pilgrims visiting the Holy City. The NEA Church also known as the New Church of the Theotokos was a Byzantine church erected in Jerusalem by the Byzantine Emperor Justinian I (c. 527-565). The church was completed in 543 but was severely damaged or destroyed during the Persian conquest of the city in 614.

None the less, once again like the aqueduct system that still exists today underground still exists the remains of the Great Vaulted Cistern under the ground level. However, it may be incorrect to call it a "cistern" when it has more characteristics of a giant mass "Mikveh" or Baptismal Bath. The main difference between a cistern and mikveh is the openings, for example cisterns typically only require one access point to retrieve the water many of the openings are also smaller and can be closed to allow only a bucket to enter to retrieve water to prevent evaporation during dry seasons. Whereas traditional mikvehs will have an entrance for the impure to enter and separate exit to leave as spiritually cleansed and renewed in order to not become atomically defiled again by leaving through the same entrance. Not only does the NEA Church Great Vaulted Mass Mikveh Baptismal Bath system have both an entrance and exit; the exit was once a spiral stair case that symbolizes your ascension in life to a purer higher state of being and never turning your back on the Lord as you re-enter life Spiritually Renewed and one with the Lord. Spiritually, the spiral represents a connectivity with the Divine, spiraling from the outer ego of the outside world, into the inner soul of cosmic awareness and enlightenment. The spiral stair case represents evolution and growth of the spirit, a symbol of change and development as one enters an everlasting life with the Lord.

Central to GOD's Divine design for his Holy Mountain is the resurrection of the Mount Zion Fountain for the flow of living waters from Jerusalem's Lower Aqueduct to spring forth in conjunction with the NEA Church Great Vaulted Mass Mikveh Baptismal bath system for Interfaith Spiritual Cleansing for Jerusalem, Israel and all pilgrims visiting the Holy Land.

Most depictions of the reconstruction of the First and Second Temple periods and the Jewish prophesied Third Temple, showcase a magnificent central Temple connected to the Holy of Holies, Inner and Outer Courts separated by walls, gates and all the details listed in the Bible. However, most of all these modern depictions lack a very key element to the central role and function of how the flow of living water actually existed, necessary for Temple life for the rights of purification for priests to enter the Temple and its surrounding City offering purification for the masses and its promise of life. The importance of this ever-lasting flow of water through the true Holy Temple and City of God is vital for its creation and promise of purification for the never-ending duration of divine life on Earth for all to take part. GOD is the creator of life, and water is the tool to sustain and renew the purification of life that GOD has created. The Life, Light and Energy that comes from our Lord God Almighty Creator is in the essence of every seed of life.

Genesis 1:27
27 So God created mankind in his own image, in the image of God he created them; male and female he created them.

The righteous must unite creating a Fountain of Life on Mount Zion giving glory to the Lord and all his creation through its promise of life, purification, and spiritual renewal for all eternity. As GOD is all-pure and Divine in all ways, it is our responsibility and purpose to spiritually cleanse our soul and live life with our truest Divine-self in the spirit of all things. Living with a spiritually clean soul and world is the central way to the straight and narrow path to living a life in Heaven on Earth. You cannot be pure of heart and expect to live in Sodom and Gomorrah. It is time for the New Jerusalem to come!

1 Peter 1:15-16

[15]But just as he who called you is holy, so be holy in all you do; [16]for it is written: "Be holy, because I am holy."

Leviticus 19:1-2

[1]The LORD said to Moses, [2]"Speak to the entire assembly of Israel and say to them: 'Be holy because I, the LORD your God, am holy.'"

Isaiah 40:9

You who bring good news to Zion, go up on a high mountain. You who bring good news to Jerusalem, lift up your voice with a shout, lift it up, do not be afraid; say to the towns of Judah, "Here is your God!"

We, the one people of our Earth, created in the image of GOD in Heaven, may be united in the flow of the living springs of water of this world. As a drop of water is a part of the whole of the ocean, the same is true for a breath of air in all the wind between Heaven and Earth. Blessed be even the vapor of water in the mist of the wind to cleanse and baptize you with every breath to the center most part of your soul with Holy Water in the vapor of Air from its everlasting connection to the Holy Water that still flows from the Garden of Eden today! Water is the great purifier and Divine Union for all of God's creation. May we walk with every step in the presence of our Living Lord GOD Almighty and breathe in his cleansing righteousness with every breath.

We should be inspired to build complete Holy Temple cities built around purification and spiritual rebirth from water that extends from the threshing floor of the most central part of the Holy of Holies of the Temple of the Lord's House of Prayer of All Nations. The Temple would connect consecutive descending spiritual baths with aqueducts throughout the city that eventually flow into massive collection pools or seas of water for mass purification ceremonies, and conclude feeding into irrigation pasture land for the animals and crops to be connected and thrive from the same seed of purification and divine holiness for all. A self-sustaining ecological economy will in turn be created by this cycle of living waters of purification and spiritual rebirth from the Holy of Holies. For the Day of the Lord is upon us, may we have a righteous and straight path for our Lord to walk among us! For our body's temple is the greatest of all creations and it is our duty to our Lord God Almighty to continually cleanse away the sinful material world to awaken our truest transcendental consciousness of our being and become one with the Lord and live-in harmony with all creation as it was in the beginning.

REFERENCES

[1] The Lower Aqueduct of Jerusalem. Tom Powers. In *VIEW FROM JERUSALEM, It's EVERYBODY'S "Holy Land"*. Retrieved July 2011, from <http://israelpalestineguide.files.wordpress.com/2010/07/lower-aqueduct-wpics.pdf>.

[2] *The Aqueducts of Israel* (Journal of Roman Archaeology Supplementary Series, 46, 2002). [This work has been issued as a monograph by other publishers as well, in both Hebrew and English.]

Articles:

• "A Survey of the Aqueducts to Jerusalem" by Amihai Mazar, pp. 210-244.

• "The Low-Level Aqueduct to Jerusalem: Recent Discoveries" by Ya'akov Billig, pp. 245-252.

• "New Data for Dating the High-level Aqueduct and the Wadi el-Biyar Aqueduct, and the Herodion Aqueduct" by David Amit, pp. 253-266.

[3] *New Encyclopedia of Archaeological Excavations in the Holy Land* (Israel Exploration Society / Carta), Vol. 2, "Jerusalem," pp. 746-7, 766.

[4] *Jerusalem: A Walk Through Time*, Vol. 1 (Yad Ben Zvi, 1999), pp. 90-100.

[5] "Solomon's Pools (Berekhot Shelomo)," in *The Holy Land*: *An Oxford Archaeological Guide from Earliest Times to 1700*, 4th Edition (Oxford University Press, 1998) by Jerome Murphy-O'Connor, pp. 425-428.

[6] "The Aqueducts of Jerusalem" by Amihai Mazar, in *Jerusalem Revealed: Archaeology in the Holy City 1968-1974* (Israel Exploration Society, 1973), pp. 79-84
[7] Mount Zion Fountain, Location & Surroundings Pictures. Personal Photos by Brandon WinsGaris. Retrieved July 2003.
[8] Fountain of Life. (2013, May 30). In *Wikipedia, The Free Encyclopedia*. Retrieved July 15, 2013, from <http://en.wikipedia.org/w/index.php?title=Fountain_of_Life&oldid=557599594>.
[9] Life-giving Spring. (2013, April 15). In *Wikipedia, The Free Encyclopedia*. Retrieved July 16, 2013, from <http://en.wikipedia.org/w/index.php?title=Life-giving_Spring&oldid=550545236>.
[10] Ritual Purification. (2013, April 18). In *Wikipedia, The Free Encyclopedia*. Retrieved July 16, 2013, from <http://en.wikipedia.org/w/index.php?title=Ritual_purification&oldid=550971464>.
[11] Zamzam Well. (2013, July 3). In *Wikipedia, The Free Encyclopedia*. Retrieved July 17, 2013, from <http://en.wikipedia.org/w/index.php?title=Zamzam_Well&oldid=562758553>.
[12] Molten Sea. (2013, June 7). In *Wikipedia, The Free Encyclopedia*. Retrieved August 4, 2013, from <http://en.wikipedia.org/w/index.php?title=Molten_Sea&oldid=558783449>.
[13] Pool of Siloam. (2013, July 22). In *Wikipedia, The Free Encyclopedia*. Retrieved July 30, 2013, from <http://en.wikipedia.org/w/index.php?title=Pool_of_Siloam&oldid=565262631>.
[14] Gihon Spring. (2013, July 17). In *Wikipedia, The Free Encyclopedia*. Retrieved July 30, 2013, from <http://en.wikipedia.org/w/index.php?title=Gihon_Spring&oldid=564670840>.
[15] Sultan's Pool. (2013, March 21). In *Wikipedia, The Free Encyclopedia*. Retrieved July 30, 2013, from <http://en.wikipedia.org/w/index.php?title=Sultan%27s_Pool&oldid=546104269>.
[16] Bible Verses. (2010). New International Version. Received from www.biblegateway.com

Brandon WinsGaris Author (San Antonio, 17 October 1982). Batchelor of communications, Texas A&M-Corpus Christi, 2011.

He has a unifying voice for the Holy Land from a non-secular perspective and he explores the commonalities of Holy Water and Spiritual Cleansing within Judaism, Islam and Christianity. He's passionate about creating a multipurpose functional Fountain on Mount Zion in Old Jerusalem as an Interfaith symbol for a United Holy Land and World Peace.

In his advocacy work, Brandon provides connections and resources to help understand the functionality, historical existence and importance of resurrecting the Mount Zion Fountain in connection to the Lower Aqueduct of Jerusalem system that starts at the Etam spring and Solomon's Pools near Bethlehem which extends to the Fountain and ultimately restores the Nea Church giant vaulted subterranean water reservoir in Gan HaTekumah, or the 'Garden of Redemption' park, near Zion Gate, in Jerusalem.

His Interfaith Peacebuilding vision is to see the restoration of the Mount Zion Fountain and NEA Church giant vaulted subterranean water reservoir for the purpose of creating a functioning mass Mikveh and Baptismal bath system for Interfaith Spiritual Cleansing for Jerusalem, Israel, and all pilgrims visiting the Holy Land.

Mr. WinsGaris' activism for the Mount Zion Fountain started as a necessity to shed light on a new perspective of cornerstone discoveries in the Holy Land. His discoveries were made while attending the Hebrew University in Jerusalem, studying Biblical Hebrew for his foreign language credits.

After college, Brandon pursued a natural calling for automotive sales and quickly became a Gold Master Certified, Truck Master, Toyota New Car salesman reaching the top 1% of all Toyota salesmen in Texas and the Nation! Utilizing social media and diverse online marketing platforms with IT management skills he blazed the way consecutively winning Salesmen of the Month and several other awards and accolades with Toyota.

Over the next few years, Brandon was blessed with a very successful career as a top Toyota salesman and invested his money in real estate and development. He has now retired from automotive sales freeing up his time to spend it full-time with his family and pursue his passion for writing and advocating for a United Holy Land and World Peace.

Social Protection for Ex-Combatants as an Instrument to Reduce Poverty and Enhance Lasting Peace in the Post-Conflict of Colombia

Santiago Luna Castro

***Abstract*—** This research addresses the importance of implementing Social Protection Policies to reduce poverty and enhance a long and lasting peace in post-conflict situations. Therefore, it focuses on the policies agreed in the Final Peace Agreement (FPA), signed between the Colombian Government and the guerrilla group FARC-EP in 2016. Moreover, it provides special attention to the design and implementation of Social Protection policies in the reintegration process of former ex-combatants into civil life.

Consequently, this paper aims to answer the research question: To what extent Social Protection policies for ex-combatants are required to build lasting peace in the post-conflict in Colombia?
To answer the research question, this research follows a holistic approach and a multidisciplinary methodology. First, after collecting data from national and international sources, it offers a general diagnostic through statistics and graphs of the social conditions of Colombia in 2016 to clarify why social protection policies are required. Thus, it identifies and differentiates the targeting groups of these policies, ex-combatants and rural populations.
Subsequently, the research provides a comprehensive study of the new social protection policies designed in the FPA and explains its objectives, length and targeting groups. The

primary sources of information are the FPA, national regulations, national legislation, and UN reports.

The research also compares the experience of other Social Protection schemes in conflict situations, using literature from the International Labour Organization (ILO) and United Nations (UN). This contrast works to demonstrate why the inclusion of Social Protection in post-conflict situations is an innovative tool, proving a broader view of the possible uses of Social Protection.

Finally, the research shows the importance of Social Protection as a tool to overcome poverty and support the reintegration of ex-combatants into civil society. It concludes that implementing Social Protection policies in post-conflict conditions to ex-combatants has a short and middle-term positive impact. However, its effectiveness in the long term depends on the political willingness and the effective co-implementation of employment, formalization, and law enforcement strategies.

***Keywords*—** ex-combatants**,** peace, poverty, reincorporation, social protection.

Santiago Luna Castro is a lawyer and works as a private attorney and consultant in the implementation of social protection policies. Address: Schmollerstraße 4, Saarbücken, 666111. Germany. Phone: +49 157-8890-4361; e-mail: sanluna1906@gmail.com.

Current Federal and State Immigration Policy in the U.S. On Un-Documented and Documented Immigrants: Implications for Economic Hardship, Mental Health, Community and Family Stability

Kamps, Peter (Kamps, P.)

City of Chicago, Department of Family and Support Services

petekamps@aol.com/peter.kamps@cityofchicago.org

February 14, 2021

ABSTRACT

State legislation and federal immigration policy in the United States plays a pivotal role in terms of influencing the economic well-being and the associated behavioral health and mental health outcomes within the immigrant population. Prior federal welfare administrative changes directed at_non-citizens, designed to inhibit utilization of U.S. assistance programs such as the Temporary Assistance to Needy Families (TANF) and Medicaid programs, has had a chilling effect upon access to "safety net" programs available to the low-income documented and un-documented population at large. More importantly, these policies undermine family economic mobility and community stability. This includes restrictions in employment supports resulting in lost labor productivity. This paper will examine the preliminary effects of immigration policy, including the previously enacted "public charge" rule at the federal level. A systematic discussion of federal immigration policy, affecting immigrant access and utilization of health care resources, public benefits, housing, and employment will be examined. An analysis of the public charge federal legislation and state immigrant "safety net" policies utilizing the National Association of State Legislatures database is included. Finally, a discussion of the overall economic consequences associated with restrictive federal policy within the immigrant community, as well as the alignment of innovative state responses, will be provided.

Keywords: immigration, policy, poverty, legislation, public benefits

INTRODUCTION

Negative public perceptions of persons newly arrived within the United States and recently enacted legislation targeted towards immigrants in the arena of public benefits has played a significant role in declining immigrant enrollment rates for social welfare programs. The changing landscape towards vulnerable immigrants has significant ramifications to the economic health and well-being of these low income families as well as U.S. society as a whole.

Developments in U.S. immigration policy during the prior U.S. administration, as well as modifications in safety net program eligibility requirements for low income documented and undocumented immigrants, are likely to reverse an existing trend of a declining poor immigrant population. Federal safety net initiatives have proven effective in lifting the number of low income individuals out of poverty, including immigrants (Trisi 2019, Wheaton 2011 & Haveman, 2015).

Between 2016 and 2019, the number of immigrants below the poverty level decreased nationally by 15%. This is, in part, owing to state legislative efforts to address a widening safety net funding gap. These state initiatives have proven critical in terms of addressing household economic stressors while augmenting declining federal safety net resources.

The purpose of this paper is threefold: Utilizing federal administrative data provide high level takeaways regarding the impact of the existing public charge rule[1] in reducing immigrant access to safety net programs; provide an in-depth analysis of the relevant academic research which assesses the mental and physical health ramifications of declining immigrant enrollment. These include Medicaid, the Supplemental Nutrition Assistance Program or SNAP, Subsidized housing, and employment. This paper will also evaluate the prevalence of existing state legislation in the wake of restrictive policies affecting low income documented and undocumented immigrants.

Immigrants continue to remain under-represented in the overall U.S. poverty rate, 10.5% vs. 12.3% for the population overall. (U.S. Bureau of the Census, ACS 5 Year Estimates & MDAT 2019). Aligning with the decrease in immigrant poverty has been a decade-long decrease in the national safety net expenditure rate for low income immigrants utilizing U.S. public assistance. This has accelerated over the past several years. Programs such as Medicaid, Housing Assistance, the Supplemental Nutrition Assistance Program (food stamps), and Temporary Assistance to Needy Families (TANF) have all witnessed declines in immigrant utilization.

The number of overall documented and undocumented immigrants within the United States has remained relatively constant from approximately 22.2 million in 2016 to 22.1 million in 2019. However, the number of non-naturalized immigrants below the federally designated poverty level has declined during this time period, from 5.3 million in 2016 to 4.5 million in 2019.

Despite a reduction in poverty within this population, a declining safety net enrollment rate, and the fact non U.S. citizens utilize public assistance at a lower rate than their U.S. citizen counterparts, significant concern has been raised over the likelihood immigrants will become a burden to the U.S. taxpayer. (Zu et. al. 2015, Campbell, 2012, Ingram, 1993). This concern, along with political developments within the United States regarding immigration, led to an expansion of the so called public charge rule during the

[1] At the time of this writing, the expanded public charge rule was under review due to an executive order issued by the Biden Administration. Many of the provisions of the public charge rule were later reversed during the first several months of 2021.

Trump Administration in August, 2019. These policies have the potential of negatively influencing the overall mental health and health of the undocumented and documented immigrant population.

This paper will examine the mental health and health impact of the "public charge" rule in terms of reduced access to key safety net programming within the documented and un-documented immigrant population. Enrollment trends of immigrant utilization within three key public benefit areas, Medicaid, Food Assistance, and housing are presented. A discussion of employment patterns as well as the economic ramifications within the low income documented and undocumented population within the U.S. will be included. Finally, an analysis of state legislative response in mitigating the harsh realities of the revised federal immigration policy will also be covered.

RESEARCH METHODOLOGY

The perception within the overall U.S. population of behaviors associated with low income immigrants (i.e. lack of motivation to become gainfully employed, willingness to take advantage of public benefits) as well as implementation of an expanded public charge rule has resulted in a rapidly declining public assistance utilization rate for these vulnerable residents during the period the expanded public charge rule has been in effect.

In order to gain insight into how the adverse social construction of immigrants as well as legislative policies have affected immigrant enrollment in safety net programming, administrative data from the Current Population Survey or CPS (U.S. Department of Labor) and the American Community Survey or ACS (MDAT) was collected for the most recent years available. This data was collected at both the national level and for the largest U.S. cities for which data was available (New York City and Chicago). In addition, immigrant safety net enrollment projections over the coming years were developed for this paper. ACS and CPS data sources were selected since it provides a degree of clarity as to how immigrants are utilizing these programs within all 50 states. National administrative data also provides high level takeaways to understand the scope of the issue within the U.S. To better understand the mental health and health ramifications of declining public safety enrollment patterns, a robust review of academic literature is also included.

HISTORY OF PUBLIC BENEFITS ELIGIBILITY FOR IMMIGRANTS IN THE U.S.

The scope of current federal immigration policy regarding safety net programming originated in the Personal Work Opportunity Act (PWORA) of 1996, passed by the Clinton Administration and the Immigration Reform and Immigrant Responsibility Act (IIRIRA) of 1996. These two laws together determine immigrant access and eligibility for "safety net" programming within the United States. Prior to the passage of these two pieces of legislation, immigrants were accessing safety net programs at a much higher rate (Hagan et. al, 2003)

There are two primary categories of assistance which are covered under these pieces of legislation, cash assistance and non-cash assistance. Cash assistance programs include the Temporary Assistance for Needy Families (TANF) program and the Supplemental Security Income Program (SSI). Non cash assistance programs include Medicaid, Housing subsidies such as Section 8, as well as public housing. Other non-cash assistance programs include the Supplemental Nutrition Assistance Program (SNAP) and the Children's Health Insurance Program (CHIP). The two laws limit enrollment in any of these programs to legal residents within the U.S. This includes immigrants who have long-term residency status (LTR), a green card, as well as refugees/people granted asylum who qualify within program income guidelines. Most states within the U.S. require legal residency for immigrants for five years before they can become

eligible for these federal programs. The 1996 immigration laws only allow un-documented and other immigrants, access to emergency Medicaid services (such as payment for hospital emergency room visits) and enrollment in the Women, Infants and Children emergency food program (WIC). The Affordable Care Act (ACA), passed under the Obama Administration, similarly limits health insurance enrollment to qualified and eligible immigrants.

Another key component of the 1996 legislation was the creation of two identifiable classes of non U.S. citizens, qualified and non-qualified immigrants. This distinction was developed as the primary determining factor of whether a non U.S. Citizen could apply and receive a safety net benefit (Pew Charitable Trust 2014)

Qualified immigrants are those who are lawfully present within the United States (hold a green card), refugees, and those who are granted asylum. Non-citizens who do not meet these criteria are considered unqualified immigrants (i.e. some temporary workers, students and tourists, people granted temporary protected status, and unauthorized immigrants). Qualified immigrants may enroll and receive federal public benefits if they meet specific income and length of U.S. residency requirements. Non-qualified immigrants are generally excluded from any federally sponsored safety net program (although some states have compensated for this by developing their own support programs for low income immigrants, described later in this paper).

Research has shown the expenditure rate within the documented and undocumented immigrant population for programs such as Medicaid, the Children's Health Insurance Program (CHIP), Supplemental Security Income (SSI) and the Supplemental Nutrition Assistance Program (SNAP) is much lower for these individuals than their U.S. native counterparts (O'Shea 2018). Studies have also indicated that immigrants contribute more to government revenue overall and social capital than they expend in terms of safety net usage. (Lowrey et. al. 2018). Despite these trends, a national decrease in the immigrant utilization rate within key safety net initiatives has accelerated since the passage of the PRWORA act. This continued since the transition from the Obama administration to the Trump administration. Part of this tendency within the immigrant population may include the desire to remain gainfully employed while remaining in the U.S.

Immigrants may prioritize the ability to legally work within the United States over enrollment in safety net programming. Research has shown the ability to legally work within the United States and a pathway to citizenship was valued more highly than being able to enroll in public safety net programs (Melo, et. al, 2014). However, for un-documented and documented individuals facing prolonged economic hardship, the expanded public charge rule may result in far reaching consequences to their ability to care for themselves and their families

PUBLIC CHARGE RULE

The public charge rule defines someone who either wishes to gain entry into the United States or a non-U.S. citizen seeking long term residency status (or green card) within the U.S. and is considered likely to become a public burden to the U.S. taxpayer. Public charge is determined either through enrollment in a public cash assistance program for income maintenance or enrollment in a variety of non-cash transfer programs such as Medicaid (the primary method by which low income individuals can access health care), subsidized housing programs, or food assistance.

During 2019, the Trump administration announced a final rule which modifies the public charge policies used to determine whether an individual applying for admission or adjustment towards legal status is inadmissible to the United States. Under the expanded rule, the federal government can deny an individual entry into the U.S. or make an adjustment to their residency status (green card), adjust their legal permanent residency status (LPR) status or a green card if he or she is determined to be a public charge.

On February 21 2020, the Supreme Court of the United States overturned U.S. federal district court decisions in both New York and Illinois to block the implementation of a revised public charge rule. This allowed the new policy to take hold nationwide. The ruling has had far reaching implications for the overall health and well-being of immigrants within the United States.

Prior to the Expanded Public Charge rule, the previous policy covered the following programs in Public Charge Determinations: SSI, Temporary Assistance to Needy Families (TANF), State or Local Cash Assistance Programs and Public Assistance for long term care in an institution such as a nursing home, including Medicaid. The new policy added to the already existing programs to be considered in any public charge rule the following: Non-Emergency Medicaid (for non-pregnant adults over age 21), SNAP food assistance, and Housing Assistance.

Thus, under the expanded public charge rule, immigrant enrollment in such safety net programs as Medicaid, SNAP (food assistance), and housing assistance is viewed negatively when an immigrant is applying for green card status, thus potentially derailing a path to citizenship. One of the consequences of the new rule is a likely acceleration of declining enrollment within Medicaid within the low income immigrant population and participation in other safety net programs. This trend is already evident when examining recent administrative data from the American Community Survey.

Decreased participation in Medicaid, subsidized housing and food assistance programs would contribute to increased economic hardship, increased stress within the family unit, as well mental health issues within the low income immigrant community. More significantly, these changes would likely increase the number of immigrants below the poverty level within the United States, resulting in a loss to the economy in terms of labor productivity.

Many immigrants seeking an improved life by relocating to the U.S. are already confronted with an array of economic challenges. Close to eighty percent of non-citizens who originally entered the U.S. without LPR status have at least one characteristic that the Department of Homeland Security (DHS) could potentially weigh negatively in a public charge consideration. This includes lack of educational attainment (not having a high school diploma) or private health insurance. (Artiga et. al. 2019) In other words, when considering whether to grant an immigrant permanent residency status, a majority of these immigrants are already vulnerable to losing a chance at long-term residency.

A Kaiser Foundation study conducted during 2019 found more than 56% of non-citizens in the U.S. are lacking private health insurance and close to 40% do not have a high school diploma (Garfield et. al 2019). Close to one-third of the immigrant population who entered the United States have a family income below the 125% poverty threshold**.** ($32,750 for a family of four in 2019). All these factors are weighed when factoring in long term admissibility into the United States.

Although most states require a five year residency prior to an immigrant becoming eligible for one of these federal programs, any attempt at participation would possibly service as a “tipping point” towards exclusion, ultimately resulting in their deportation to the country of origin. Similarly, concerns regarding enrollment in public assistance could result in the denial of entry within the United States. Despite possible reversal of the expanded public charge rule, fear of enrolling within these programs is likely to

persist within the immigrant community. The three federal social welfare programs affected by the expanded public charge rule: Medicaid, Housing Assistance and the Supplemental Nutrition Assistance Program (SNAP) are described below.

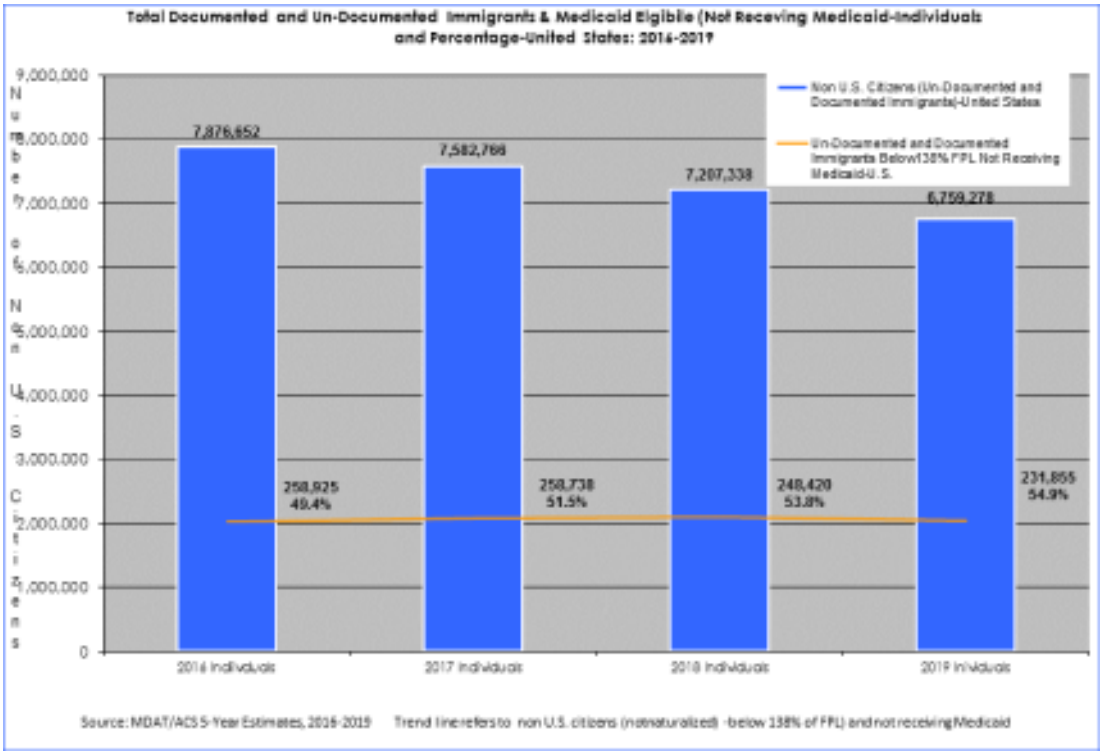
IMPACT ON MEDICAID ENROLLMENT

According to the U.S. Bureau of the Census, the number of income eligible immigrants who were below 138% of the poverty level and utilized Medicaid decreased from 2016 to 2019. Medicaid is the federal program through which low income individuals can gain access to health care services within the United States. A decrease in Medicaid enrollment for low income immigrants is likely to result in poor health outcomes, affecting the overall well-being of this population, This will also likely add to an overall increase in national health care costs due to hospitalizations and increased emergency room visits.

Bernstein, et. al. (2019), in his national immigration survey conducted prior to the final public charge rule, found that one in seven immigrant adults reported avoiding public benefit programs for fear of risking future green card status, and more than one in five adults in low-income immigrant families reported this fear. Berk (2004) in a 1996 study of 710 Latino immigrants which utilized in-person interviews found that 39% of the respondents in Houston, El Paso, Fresno and Los Angeles who were undocumented were fearful of utilizing health services due to their immigration status.

Table 1 presents the total number of undocumented and documented immigrants within the U.S. who are income eligible and not currently enrolled in Medicaid. Significantly, the proportion of eligible immigrants in the United States who were not receiving Medicaid over the past four years has increased.

Table 1-Medicaid Enrollment Trends within the U.S. Low Income Immigrant Population: 2016-2019

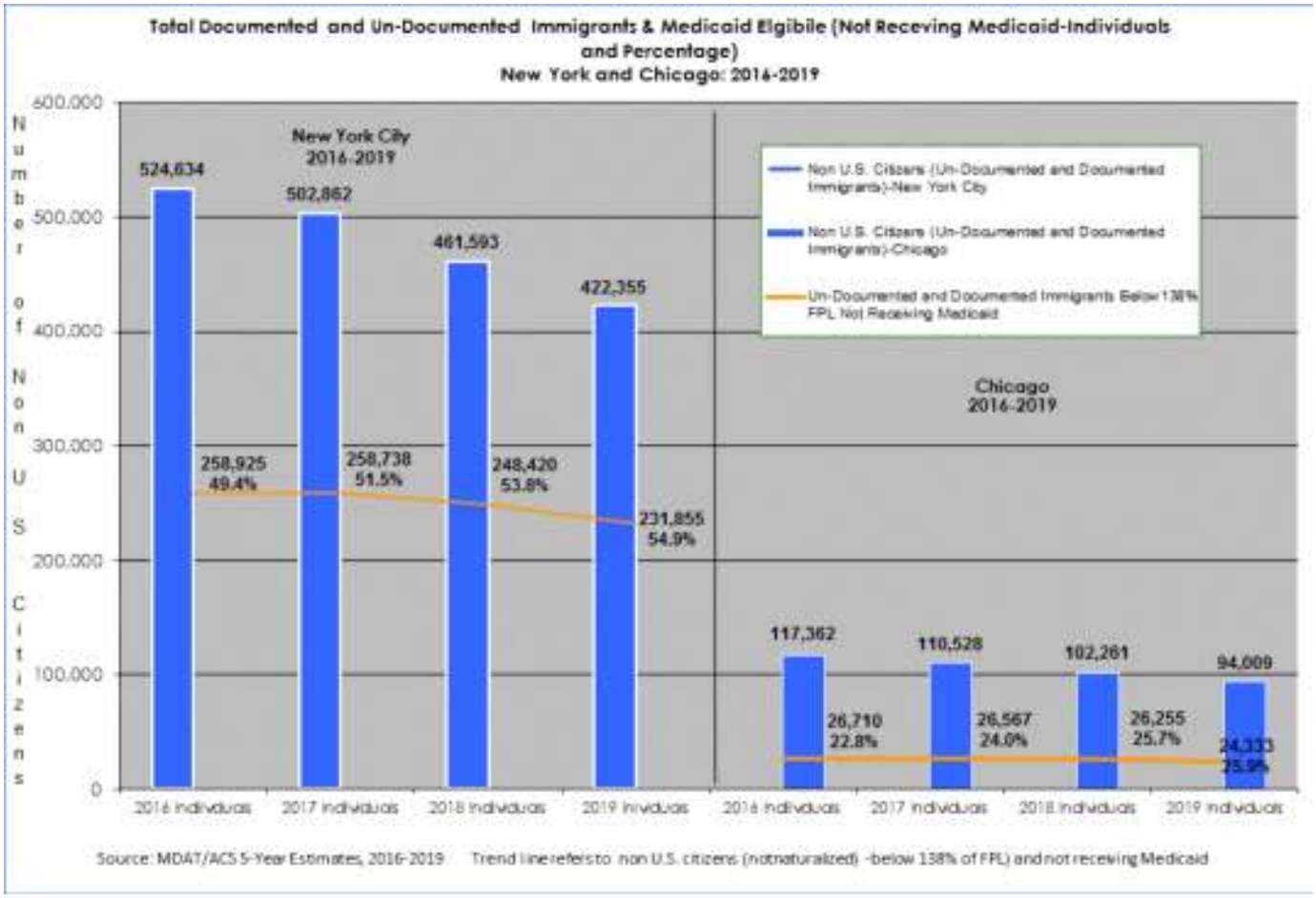


Other researchers have similarly predicted a future decrease in Medicaid immigrant enrollment due to the public charge rule (Perriera, K, 2018, Canady, 2020, Garfield, 2019, Schlosberg and Wiley, 1998). Based upon methodology developed by the Kaiser Foundation (Artiga, 2019), the projected decrease in immigrant Medicaid enrollment in New York City alone could result in 55,000 low income immigrants alone who would qualify for subsidized health insurance not enrolling.

Utilizing Current Population Survey (CPS) data from the U.S. Census Bureau, it is possible to estimate the total number of immigrants who are either un-documented or documented and eligible for Medicaid, based upon income (138% or below poverty). The CPS does not differentiate between the various sub-populations within the immigrant community, including persons seeking asylum within the United States or refugees who may immediately qualify for public safety net programs. However, national demographic data corroborates the theory of declining enrollment due, in part, to fears within the immigrant community of losing permanent residency status as well as negative public perceptions of immigrants.

The next graph illustrates the number of Non-U.S. citizens within New York City and Chicago (cities for which data was available) and the actual number of low income immigrants not enrolling in Medicaid. The bar represents total number of documented and undocumented immigrants in New York and Chicago and the yellow trend line indicates income eligible who are not enrolled in Medicaid.

Table 2-Trends in Medicaid Enrolment for Eligible Un-Documented and Documented Immigrants: 2016-2019 Chicago and New York City



The percentage of income eligible immigrants for Medicaid, not enrolled in the program, has increased since the implementation of the expanded public charge rule.

Recent research and anecdotal evidence suggests the expanded public charge rule has had have a chilling effect on immigrant families, leading them to forgo enrollment in a necessary safety net program. This is due to confusion or fear that their child's enrollment could ultimately result in deportation or denial of long term residency. Bernstein et. al (2019), in his study of immigrant safety net enrollment trends, provides reports of low income immigrant families choosing not to enroll their children in the Children's Health Insurance Program (CHIP) and themselves in Medicaid. Negative consequences associated with residency status played a significant role in the parent's decision-making. The expanded public rule has accelerated this trend over the past four years. In addition, as low income families seek out alternative health care resources, private not-for-profit health care organizations who do ask immigrant status (i.e. Federally Qualified Health Centers) are more likely to become overburdened. (Clark et. al., 2020)

The Kaiser Foundation has estimated there are at least 7.6 million children, who are noncitizens and who may be at risk for avoiding health care enrollment as a result of revised federal immigration policy. In addition, according to American Community Survey data from the U.S. Census Bureau, there were over 13.5 million Medicaid/CHIP enrollees who were noncitizens or citizens living in a household with at least one citizen who may be at risk for non-enrollment due to the expansion of the public charge rule. The Kaiser Foundation estimates at least 2.0 million to 4.7 million Medicaid and CHIP enrollees would dis-enroll nationally (15 percent to 35 percent of the eligible Medicaid non-citizen population) upon continuation of this policy.

Recent guidance from the Centers for Disease Control in the U.S., suggesting many Americans remain at home, has had a devastating economic impact on immigrants who work in the service industry, This is, in part, due to social distancing requirements. Many immigrants, both documented and undocumented, work in this occupational category, oftentimes receiving low wages. They are more likely to suffer prolonged economic hardship.

As a result of the covid-19 pandemic, the service industry has witnessed substantial job losses in the United States, making access to safety net programming, such as health care, even more critical. (Page, et. al. 2020). At the same time, the expanded public charge rule has instilled fear and confusion within the immigrant community. Fearing the loss of a path to citizenship, many immigrants have chosen to forego enrollment in key safety net programs, even if participation would not count negatively towards residency status. An example of this is the Pandemic Electronic Benefit Transfer Program under the Families First Cares Program (discussed later in this paper).

According to the U.S. Bureau of the Census, American Community Survey, 20% of all documented and un-documented immigrants who were working in the service sector in New York City, were living in households below the poverty level. Within Chicago, the percentage was similar. Approximately 19.2% of all immigrants working in the service sector were living below the poverty level during 2019. (ACS 5 Year Estimates, 2019, MDAT, U.S. Census Bureau).

In addition, lower wages, evidenced within the service industry, create a situation in which immigrants are more likely to live in multigenerational households, potentially increasing their exposure to covid-19 and making them more likely to seek out emergency rooms, rather than on-going preventative medical care. There are more than 1 in 5 documented and undocumented immigrants who are below the poverty level and living in multigenerational households in New York City. Within the city of Chicago, 1 in 7 undocumented and documented immigrants are below the poverty level and are living in multigenerational households. (ACS 5 Year Estimates, 2019, U.S. Census Bureau).

Lack of access to health care and reliance upon free clinics and federally qualified health centers, increases the probability of poor health outcomes due to the absence of primary care. (Clark, et.al, 2020). Economic hardship has also been shown to increase the probability of co-morbidities such as diabetes, obesity, as well as hypertension. This increases hospitalization rates if an individual becomes infected with covid-19.

FOOD ASSISTANCE

Prior to the expanded public charge rule, immigrants who were lawfully within the United States faced no negative consequences if they enrolled within the Supplemental Nutrition Assistance Program (SNAP), the primary means whereby low income immigrants can gain access to food assistance within the U.S. The new law reversed this policy, opening up the possibility a family can be deported if they require and enroll in federally sponsored emergency food assistance. In addition, since unauthorized immigrants often apply for safety net programs on behalf of their children who are eligible, these parents may decide not to enroll in order to keep the family unit intact. This would place the children at risk of poor health outcomes, cognitive development issues, and other negative developmental outcomes. According to the U.S. bureau of the Census, the percentage of immigrants who were either undocumented or documented and were receiving SNAP benefits as of December, 2019. has already declined over the past several years. A primary purpose of the Supplemental Nutrition Assistance Program is to lift individuals who qualify out of poverty.

As described earlier, recent studies have shown safety net programs to be effective in facilitating upward economic and social mobility (Trisi 2019, Huang et. al. 2018). Conversely, by potentially denying immigrants a path to citizenship by monitoring safety net usage, this policy may result in the acceleration of immigrants who choose not to seek out assistance, thereby contributing to financial hardship and financial instability within households. Immigrant participation in food assistance plays a critical role in reducing food insecurity, thereby promoting their overall health and well-being. Growing mistrust in government programs, owing to putative government policies towards economically vulnerable non U.S. citizens, may also contribute to a decrease in enrollment within critical safety net programs such as SNAP. This trend is substantiated by administrative data from the U.S. Census Bureau.

Between 2016 and 2019, the percentage of non-U.S. citizens below 130% of the poverty line utilizing SNAP benefits decreased by 9% A decrease in SNAP participation within the low income immigrant population is also likely to create an added strain upon charitable food organizations. New York City has already reported a substantial reduction in SNAP immigrant benefit utilization, declining 26% over the past several years. The food bank also reports a substantial increase in emergency usage since the introduction of the revised public charge rule (Food Bank for New York City, 2020).

Food insecurity has been associated negatively with child development and health. (Huang, et al., 2010) Results from a 2018 study of food insecurity among immigrants indicated children from food insecure immigrant families fare far worse in terms of health outcomes than non-immigrant families within the United States. However, Huang et. al. (2010), found children of immigrants in families experiencing economic hardship, although faring worse in terms of health outcomes, showed little difference in terms of cognitive development. Low socioeconomic status, brought on by low wages within the immigrant population, may exacerbate the health consequences of food insecurity.

Food insecurity has also been linked to poor diet, chronic diseases and health disorders (Seligman et. al. 2010). This in turn, is likely to result in increased hospital emergency room visits. According to a recent study completed on the health care expenditures from poor health in Chicago (Sorrentino, Kamps et. al, 2017), the average public hospital stay cost for a low income individual as a result of chronic heart disease was approximately $102,000, for diabetes, the annual cost was approximately $26,000 and for

stroke, the annual cost approached $44,000. Denial of these benefits to the low income immigrant population will likely result in high national medical expenditures, the burden of which will ultimately fall upon hospitals and U.S. taxpayers.

The Pandemic Electronic Benefit Transfer Program, under the Families First Covid-19 CARES Program, was created to allow states to utilize federal funds in order to provide emergency food for children who were receiving free and reduced lunches through their schools, irrespective of their immigration status. However, many immigrant families have not taken advantage of this program, due to a fear of deportation, even though this program was not subject to the expanded public charge rule (Fredricks, et. al., 2020)

Food insecurity increases a family's vulnerability to disease. The inequitable distribution of covid-19 in low income communities within the United States, as well as communities of color, can intensify the community spread of the illness. This has been evidenced in large cities within the United States where communities characterized by a high concentration of poverty and minorities have witnessed a disproportionate number of hospitalizations and mortality during the pandemic. The interaction between covid-19 and drop in economic activity will likely increase food insecurity within immigrant communities in the United States. (Smith, et. al, 2020)

HOUSING ASSISTANCE

Under the expanded public charge rule, utilization of federal subsidized housing programs including Section 8 housing choice vouchers, Section 8 Project-Based Rental Assistance, or public housing would be factored in when an immigrant is seeking long term residency status (LTR). Occupancy in such housing units which fall under the various government sponsored programs would thereby be deemed a negative factor in weighing whether an individual would be allowed to remain within the United States. As with other social welfare programs, the percentage of low income non-U.S. citizens utilizing housing assistance has already declined by a substantial margin (Current Population Survey, 2016-2018).

Chicago and New York City similarly reported a significant decrease in the number of immigrants utilizing public housing or Section 8 housing assistance. In New York City, the number of low income immigrants utilizing housing subsidies decreased from 56,904 to 35,621 between 2017 and 2018. In Chicago, the number of low income immigrants receiving housing subsidies similarly declined from 4,882 to 3,139 during the same timeframe. New federal immigration policies prohibit mixed status families-comprised of household members who are both eligible (dependent children) and ineligible (their parents or guardian) from receiving public or other housing subsidies. This includes families who live in public housing and are using Section 8. The U.S. Bureau of Housing and Urban Development estimates at least 25,000 families nationally and 55,000 children would be evicted due to the new rule.

For low income individuals, adequate housing has been associated with improved mental health and improved health outcomes for both adults and children. (Leventhal et. al.l 2005**).** Despite the beneficial aspects of adequate housing, a key component of immigrant housing policy under the prior administration has been the implementation of subsidized housing policies which exclude immigrant families. This exclusion centers upon households which contain both documented (legal) and undocumented (illegal) individuals from occupying the same housing unit. This policy could also result in an increase in homeless individuals within the United States, raising pressure on an already overburdened U.S. shelter system.

During May of 2019, the U.S. Department of Housing and Urban Development proposed a rule which would change eligibility requirements for public housing. The new housing rule banned any immigrant family from obtaining any form of subsidized housing assistance, whether this was occupancy in public housing or a housing voucher. Prior to the implementation of this rule, an immigrant family could sign a lease on a public housing unit or receive subsidized housing assistance if at least one member was within the United States legally.

The rule would deny eligibility to a family if any member is an undocumented immigrant. Currently, an immigrant family can sign a lease on a public housing unit as long as one family member is in the United States legally, usually the dependent child of undocumented immigrants. Despite the perception of immigrant abuse, housing subsidies and public housing expenditures among immigrants remain much lower than their U.S. citizen counterparts (O'Shea, et. al, 2018).

Further analysis is needed in terms of the effect of the mixed-use housing rule, potentially increasing homelessness within the immigrant community and increasing demand within an already overburdened shelter system. Of particular concern are those immigrants who are survivors of domestic violence and utilizing housing subsidies.

EMPLOYMENT

Stressors associated with transitioning to a new country along with prolonged periods of unemployment have been associated with higher mental health risks within the immigrant population. (De Castro, A. Rue, T. 2010, Barnett et. al. 2008). Federal requirements for eligibility to work within the United States also may prolong unemployment as there are significant, lengthy procedural barriers for immigrants to prove they can work legally within the U.S. In many instances, low income individuals who are new to the United States wind up in low wage, temporary positions, without benefits such as health care and sick leave. This includes seasonal employment (i.e., agricultural labor) or industrial positions.

Lack of employment can lead to economic hardship and the probability that a family will encounter a lengthened period of poverty. The inability to utilize an occupation skill set within the immigrant population has also been associated with poor mental health. (Reid et. al. 2012). Immigrant frustration with obtaining adequate or satisfying employment has also been linked with lowered physical health and mental health (De Castro et. al. 2010).

Given the vulnerability of immigrant families in terms of economic stressors, such as unstable or intermittent unemployment, safety net programs become even more critical in terms of importance. Despite a federal push towards disinvestment of public resources towards assisting vulnerable immigrant populations, states have proven resilient in addressing the growing need for income support within their midst.

STATE INITIATIVES

The U.S. government, in passing the 1996 welfare reform legislation, did allow states some flexibility in terms of funding immigrant programming. This was accomplished through a combination of state revenue and federal resources. A key component includes non U.S. citizens who have resided lawfully within the United States for less than five years. However, eligibility remains restricted to "qualified immigrants" or individuals who can prove lawful residency within the United States.

During 2009, the federal government passed the Children's Health Insurance Program Reauthorization Act (CHIPRA) to allow states to provide medical coverage through Medicaid and the Children's Health Insurance Program (CHIP). This allows female immigrants who were pregnant and their children who

were lawfully residing within the United States to receive medical coverage without the standard five year residency waiting period. Included in the CHIPRA program is a financial incentive for states to provide an increased federal match rate for children under age 19. This has proven effective in improving the overall health coverage of children and likely facilitated improvements in the overall health of the immigrant family unit.(Saloner B. et. al. 2015).

In addition, there are twenty six states as of May, 2014 who provide state-only funding for immigrant safety net programs which emulate federal assistance programs in the areas of health care access, income transfer, food assistance, as well as assistance to the elderly and disabled. The most widely available program mirrors the Temporary Assistance to Needy Families (or TANF program). This provides cash assistance to low income immigrants who meet certain income and work requirements. The second most widely available program is state sponsored health care assistance (offered by ten states and the District of Columbia). At the other end of the spectrum are only ten states who have not taken up one of the federal state options to expand eligibility for immigrants in the area of health care.

States also vary widely on what types of safety net benefits they offer to immigrants in the areas of health care, cash assistance, as well as food assistance. Some states may extend all state-funded benefits to all non U.S. citizens, irrespective of their status. Other states confine state funded benefits to a specific sub-population within the immigrant community. Further research is needed to determine if the expanded federal public charge rule may affect enrollment within state-funded only programs (such as Medicaid) due to fears within the immigrant community, This will likely shift the overall cost burden of offering Medicaid coverage for immigrants away from the federal government and on to the states themselves.

In order to mitigate increased federal scrutiny and restrictions in terms of safety net programs within the immigrant community, at least 37 states considered and/or passed sanctuary legislation during 2017. This is designed to protect the identity of undocumented and documented immigrants, and limits state cooperation with federal immigration officials. The state of Illinois enacted the Trust Act, prohibiting state enforcement of federal civil immigration laws, without restricting communication on immigrant status required under certain provisions of Title 8 within the U.S. code. New York passed similar legislation. These actions (provisions 1373 and 1644), prohibit local and state governments from enacting laws that limit communication with the Department of Homeland Security regarding the immigration or citizenship status of individuals. However, the policy does not contain any requirement for anyone to initiate or take any action to disclose an individual's immigration status.

The District of Columbia recently also enacted legislation supporting sanctuary policies, thus limiting cooperation with Immigration and Customs Enforcement personnel (I.C.E.).Sanctuary policies generally limit state and jurisdictional cooperation with federal immigration officials, particularly regarding the disclosure of immigrant status. Other states such as Texas, Georgia, Indiana and Mississippi have enacted legislation opposing immigrant sanctuary policies.

It is unclear at the time of this writing whether enrollment in any of these state sponsored public benefit programs would be weighed negatively under the expanded public charge rule when determining whether an immigrant could be granted permanent green card status. As discussed earlier, states such as Illinois and New York have passed state legislation which does not require local officials to initiate the disclosure of a particular individual's immigrant status, thereby rendering them potentially "invisible" in any public charge determination.

CONCLUSION

There continues to be a large body of demographic evidence and academic research to support a clear association between declines in immigrant safety net program participation or voluntary non-enrollment and the fear of losing long-term U.S. residency. Further restrictions as evidenced by the expanded public charge rule are likely to accelerate this trend. Longer term negative consequences for the mental, physical and economic well-being of vulnerable non U.S. citizens are likely to accrue as a result. Although the Biden Administration is likely to reverse some, if not all of the policies within the expanded public charge rule, fear and confusion is likely to remain within the immigrant community. This will likely reduce the probability of immigrant's enrollment within these key safety net programs for many years to come.

Immigrant participation in food programs such as the Supplemental Nutrition Assistance Program has been associated with improvement in birth outcomes and the cognitive development of children (Fleischhacker, S. et. al. 2019). Immigrants who participate in CHIPRA health programs have demonstratively improved health care coverage and better health outcomes in comparison to those who choose not to enroll. (Saloner B, et. al. 2018, Reich, G. 2018). Individuals who are in more stabilized housing situations and reside in economically stable neighborhoods are less likely to report anxiety and depressive symptoms (Katz et. al. 2000). Hence, allowing immigrants access to subsidized housing in safe neighborhoods offers a clear path to ensuring the integrity of the family unit.

Decreased immigrant participation in public safety net programs, and increased reliance upon earnings alone, will have significant consequences to the overall health of the U.S. economy by increasing the length an immigrant household will remain in poverty. In addition, the revised public charge rule has likely shifted the immigrant cost burden from the federal government to states which have safety net programs that emulate national welfare programs. This undermines any cost savings to the U.S. taxpayer which would have been achieved through implementation of the expanded federal policy.

With the continuation of the Covid-19 global pandemic, further research is necessary regarding immigrant fears associated with utilizing public health care resources and how this may affect the overall course of transmission within the immigrant community as well as the overall population at large. Subsequent analysis of the effects of the revised public charge rule on potentially over-burdening health care organizations also needs to be undertaken.

Finally, despite the availability of some safety net programming offered through both federal and state governments, and the likely reversal of the expanded public charge rule under the Biden Administration, substantial fear and confusion over participation in public welfare programs will likely persist. Declining enrollment in federally sponsored health, housing and food assistance programs is already evident in large U.S. cities such as New York and Chicago.

Research has clearly indicated that immigrants tend to ultimately provide an economic dividend to U.S. society, through contributions in the fields of science, technology, and the arts. Immigration policies and the actions of government officials, rather than throwing up road-blocks, could facilitate the necessary supports to encourage vulnerable immigrant families to succeed. This will undoubtedly make possible, not only a heightened economic potential for newcomers to U.S. society, but help drive their contributions to humanity within subsequent generations as well.

REFERENCES:

American Civil Liberties Union, State and Local Immigration Laws (Website, 2019)

Artiga et. al, "Estimated Impacts of Final Public Charge Inadmissibility Rule on Immigrants and Medicaid Coverage", Kaiser Foundation Issue Brief, September, 2019.

Berk, M., & Schur, C. (2004). The Effect of Fear on Access to Care Among Undocumented Latino Immigrants. Journal of Immigrant and Minority Health, 3(3), 151-156.

Bernstein, Hamutal; Gonzalez, Dulce, Karpman, Michael, Zuckerman, Stephen; "With Public Charge Rule Looming, One in Seven Adults in Immigrant Families Reported Avoiding Public Benefit Programs in 2018" (Washington, DC: Urban Institute, May 2019),

Bleich, S., & Fleischhacker, S. "Hunger or Deportation: Implications of the Trump Administration's Proposed Public Charge Rule", Journal of Nutrition Education and Behavior, 51(4), 505-509. (2019)

Campbell, A., "Policy Makes Mass Politics", Annual Review of Political Science, 15, 333-51 (2012)

Clark, E., Fredricks, K., Woc-Colburn, L., Bottazzi, M., & Weatherhead, J. (2020). Disproportionate impact of the COVID-19 pandemic on immigrant communities in the United States. PLoS Neglected Tropical Diseases, 14(7),

De Castro, A., Rue, T., & Takeuchi, D., "Associations of Employment Frustration with Self-Rated Physical and Mental Health Among Asian American Immigrants in the U.S. Labor Force", Public Health Nursing, 27(6), (2010).

FAQ ON 8 USC § 1373 and Federal Funding Threats To "Sanctuary Cities", Immigrant Legal Resource Center (April, 2017)

Fox, L., Wimer, C., Garfinkel, I., Kaushal, N., & Waldfogel, J. (2015). Waging War on Poverty: Poverty Trends Using a Historical Supplemental Poverty Measure. Journal of Policy Analysis and Management, 34(3), 567-592.

Haveman, R., Blank, R., Moffitt, R., Smeeding, T., & Wallace, G. (2015). THE WAR ON POVERTY: MEASUREMENT, TRENDS, AND POLICY. Journal of Policy Analysis and Management, 34(3), 593-638.

Huang, Y., Potochnick, S., & Heflin, C. (2018). Household Food Insecurity and Early Childhood Health and Cognitive Development among Children of Immigrants. Journal of Family Issues, 39(6), 1465-1497.

Katz, L., Kling, J., & Liebman, J. (2001). Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment. The Quarterly Journal of Economics, 116(2), 607-654.

Leventhal, T., Fauth, R., & Brooks-Gunn, J. (2005). Neighborhood Poverty and Public Policy: A 5-Year Follow-Up of Children's Educational Outcomes in the New York City Moving to Opportunity Demonstration. Developmental Psychology, 41(6)

Lowrey, Annie, "Are Immigrants a drain on Government Resources" (Atlantic Magazine, September, 2018)

Macias Gil, .R., Marcelin, J.R., Zuniga-Blanco, .B., Marquez, .C., Mathew, .T., & Piggott, D.A. (2020). COVID-19 Pandemic: Disparate Health Impact on the Hispanic/Latinx Population in the United States. The Journal of Infectious Diseases, 222(10), 1592-1595.

Melo, G., Colson, G., & Ramirez, O. (2014). “Hispanic American Opinions toward Immigration and Immigration Policy Reform Proposals. Applied Economic Perspectives and Policy”, 36(4), 604-622 (2014)

National Conference of State Legislatures, Immigration Resources (Website)

O’shea, Tim et. al, “Immigrants and Public Benefits: What Does the Research Say?” November 2018, Bipartisan Policy Center, Tim O’shea et. al, November, 2018.

Page, K., Venkataramani, M., Beyrer, C., & Polk, S. (2020). Undocumented U.S. Immigrants and Covid-19. The New England Journal of Medicine, 382(21), 1

Perreira, K., Yoshikawa, H., & Oberlander, J. (2018). A New Threat to Immigrants’ Health — The Public-Charge Rule. The New England Journal of Medicine, 379(10), 901-903.

Pew Charitable Trust Brief, “Mapping Public Benefits For Immigrants Within the United States” September, 2014

Reich, G. (2018). Hitting a Wall? The Trump Administration Meets Immigration Federalism. Publius: The Journal of Federalism, AdvanceArticle(3), 1-395.

Reid, A., “Under-use of Migrants' Employment Skills Linked to Poorer Mental Health” Australian and New Zealand Journal of Public Health, 36(2), (2012)

Rothman, .S., Gunturu, .S., & Korenis, .P. (2020). The mental health impact of the COVID-19 epidemic on immigrants and racial and ethnic minorities. QJM: An International Journal of Medicine, 113(11), 779-782.

Saloner, B., Koyawala, N., & Kenney, G. (2015). Coverage for low-income immigrant children increased 24.5 percent in states that expanded CHIPRA eligibility.. Health affairs (Project Hope), 33(5), 832-9.

Schneider, A., & Ingram, H., “Social Constructions of Targeted Populations: Implications for Politics and Policy” American Political Science Review, 87(2), 334-47, (1993)

Seligman et. al, “Food Insecurity Is Associated with Chronic Disease among Low-Income NHANES Participants”, Journal of Nutrition, Feb. 2010

Smith, M.D., & Wesselbaum, .D. (2020). COVID-19, Food Insecurity, and Migration. The Journal of Nutrition, 150(11), 2855-2858.

Sorrentino, Kamps et. al, “Practicing Food Justice, A Comparative Evaluation of Recent Developments in Chicago and Philadelphia”, Macalester College, June, 2017

Trisi, Danilo et. al, Economic Security Programs Cut Poverty Nearly in Half Over Last 50 Years, Center for Policy and Budget Priorities, November, 2019.

U.S. Bureau of the Census, ACS 5 Year Estimates (MDAT) 2015-2019

U.S. Department of Labor/U.S. Bureau of the Census, Current Population Survey, March Supplement: 2016-2018.

Vanjani, R., Martino, S., & Wunsch, C. (2020). Health Equity During COVID-19. Journal of General Internal Medicine, 35(10), 3067-3068.

Wheaton, L., Giannarelli, L., Schiferl, M., & Zedlewski, S. (2011). How Do States' Safety Net Policies Affect Poverty?. Poverty & Public Policy, 3(4),

Zhu, L., & Xu, P. (2015). The Politics of Welfare Exclusion: Immigration and Disparity in Medicaid Coverage. Policy Studies Journal, 43(4), 456-483.

Machine Learning Methods for Flood Hazard Mapping

S. Zappacosta, C. Bove, M. Carmela Marinelli, P. di Lauro, K. Spasenovic, L. Ostano, G. Aiello, M. Pietrosanto

***Abstract*—**This paper proposes a neural network approach for assessing flood hazard mapping. The core of the model is a machine learning component fed by frequency ratios, namely statistical correlations between flood event occurrences and a selected number of topographic properties. The classification capability was compared with the flood hazard mapping River Basin Plans (Piani Assetto Idrogeologico, acronimed as PAI) designed by the Italian Institute for Environmental Research and Defence, ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale), encoding four different increasing flood hazard levels. The study area of Piemonte, an Italian region, has been considered without loss of generality. The frequency ratios may be used as a standalone block to model the flood hazard mapping. Nevertheless, the mixture with a neural network improves the classification power of several percentage points, and may be proposed as a basic tool to model the flood hazard map in a wider scope.

***Keywords*—**Flood modeling, hazard map, neural networks, hydrogeological risk, flood risk assessment.

I. Introduction

A flood event refers to a situation where the river flow and water level rise unexpectedly [1], [2]. Forecasting or modeling climatic-hydrologic factors triggering such phenomena is a complex task, nevertheless the dependence of flood intensity from the geographical location and the climatic and geological conditions can be investigated [3], [4].

Floods are one of the most dangerous natural disasters, either in terms of material damages and in terms of human losses [5], [6]. Therefore, assessing the flood hazard is a key task for any risk management program [7], [8].

Risk analysis is the study of the underlying uncertainty of the consequences of a specific hazard in terms of statistical expected damages [9]. Indeed, risk is evaluated through the correlation between the frequency of a natural disastrous event and the intensity of the corresponding damages. Present paper focuses on floods, inspecting the expected degree of loss after such events. As far as the flood is concerned, the $Risk$ at a given position pos may be modelled by:

$$Risk(pos) = F(Hazard(pos), Vulnerability(pos), Elements(pos))$$

where F is a generic function, monotonically non-decreasing on every component.

This paper focuses on the hazard component, $Hazard(pos)$, encoding flood occurrence probability. The goal is to design an artificial intelligence-based solution to estimate that probability is depending on some morphological features observed in a specific geographic position pos. It turns out that such an estimate is time-invariant in the short run, since morphological features are supposed to vary with very low time-frequency.

ISPRA has deployed a flood hazard risk map called PAI. The PAI map encodes three different levels, according to an increasing time-frequency, namely 1, 2, and 3. The value of hazard in any position is assessed on the basis of historical recorded events.

To estimate the flood probability map, a combination of data-driven techniques with remote sensing and geographical information systems was mixed with a machine learning module. Namely, the frequency ratio approach [10], on twelve topographic properties was combined with a multilayer perceptron [10]-[12]. Each class of conditioning factor was weighted using the frequency ratio approach and entered as input for feed-forward neural network to optimize all the parameters.

Piemonte, an Italian region with a wide range of topographic features, has been selected as a study area, and used to show model performances.

II. Materials

The Italian region Piemonte was considered as a study area, with a spatial resolution of 10x10 meters. Every layer involved with the model may be represented as a matrix with shape:

- width: 20.367 pixels
- height: 26.642 pixels

Because of the boundaries (see Fig. 1) only a portion of the 542.617.614 pixels is admissible. Namely, we have the following size

- inner: 253.648.261 pixels
- outer: 288.969.353 pixels

The following twelve conditioning factors are used to assess flood hazard.

Static Data

PAI - Hydraulic Hazard Scenarios

The Extract Plan for Hydrogeological Asset (PAI), designed by ISPRA provides the hydrogeological hazards (see Fig. 2 for a detailed view of the Piemonte region). It is divided into the part of landslides and the part of floods. The flood plan represents the hazard maps and describes scenarios with different probability of occurrence:

Marco Pietrosanto is with Omnitech Research Group, Italy (e-mail: marco.pietrosanto@omninext.it).

- P3: the highest probability of occurrence, has smaller extensions than the others, but shorter return period, from 20 to 50 years
- P2: the average probability of occurrence, has greater extensions than P3 but has a higher return period from 100 to 200 years
- P1: the lowest probability of occurrence, has a maximum extension and is associated with extreme and catastrophic events with a return period between 300 and 500 years.

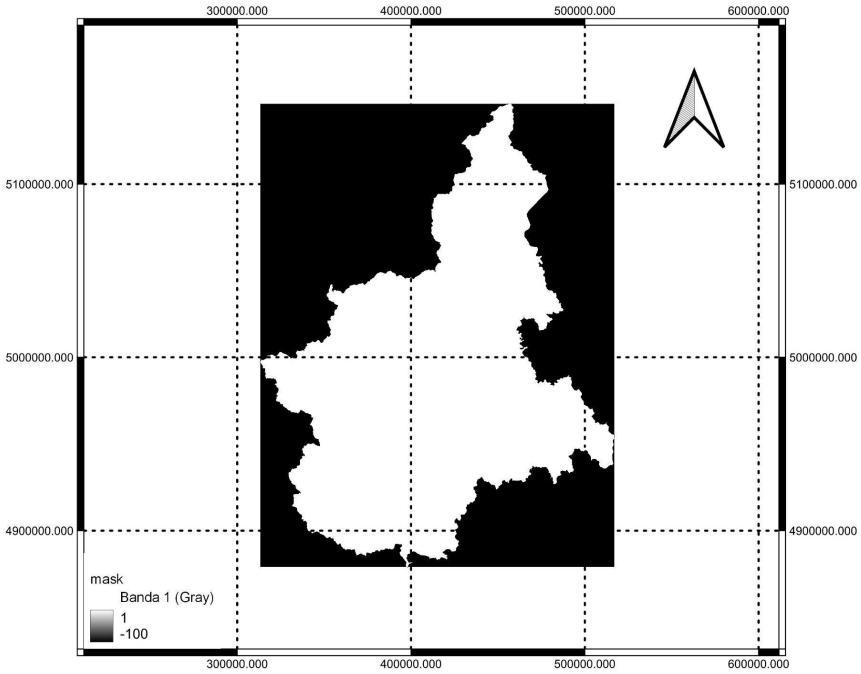


Fig. 1 Mask raster of the study are of Piemonte (admissible pixels are white)

Areas P3 and P2 fall within P1 areas, since they encode monotonic probabilities. Therefore, the levels do not add up, they represent the highest hazard level. This layer will be used to build the target of every model, in such a way to have supervised training.

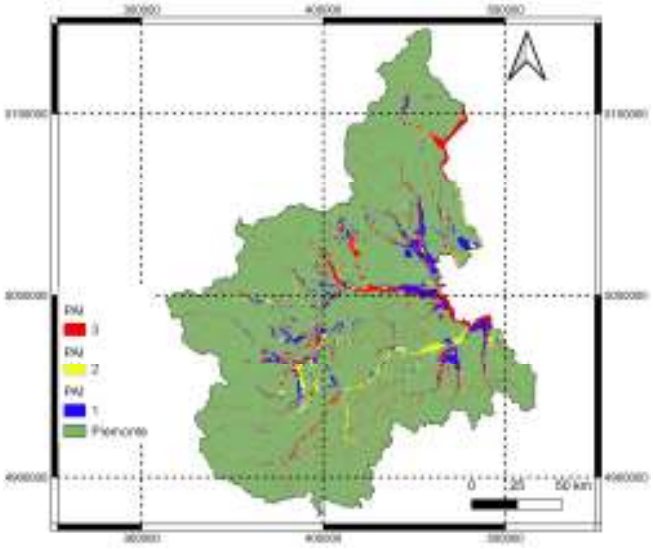


Fig. 2 Hydro-geological hazards raster of the study are of Piemonte

Digital Terrain Model and Derived Parameters

The main data are the DTM (Digital Terrain Model), it represents the topography of the study area with a pixel resolution of 10x10 meters. Most of the static data (such as: Elevation, Slope, Aspect, Curvature, STI, TWI, TRI, STI and Distance from rivers) derive from it.

All topographic parameters are extracted from the DTM, this leads to the fact that the accuracy of the DTM affects the result of the other parameters.

Elevation

The main factor for studying a flood is the altitude, in fact a flood is highly unlikely to occur in areas at high altitudes. Streams will always move from higher elevations to lower elevations and will always gather in flat areas. In fact, the topography is directly connected to the flows. See Fig. 6 for the corresponding raster of the study area of Piemonte.

Aspect

The aspect is the direction of the slope. An aspect of 0 means that the slope faces North, 90 faces East, 180 South and 270 West.

Flat areas will return nan values. In fact, an area with a slope equal to 0 will not have an exposure direction. See Fig. 8 for the corresponding raster of the study area of Piemonte.

Slope

Another factor that influences the flow path is the slope, in fact the water will always move on the higher slope topography, channeling into the *impluvium*.

As the slope increases, the water infiltration time decreases and therefore a greater quantity of water enters the network of *impluvia* and channels causing an accumulation of water and therefore a probable flood [13]. The slope is also responsible for the speed of the runoff. Higher slopes will have higher flow rates and also greater erosion and transport force, in relation to the substrate. This can affect the flux density. The hydraulic action dissolves the materials along the sliding surface changing the flow solution. Furthermore, the same hydraulic action impacts the sides and bed of the channel, displacing the materials and making them available for transport as part of the flow load. Lighter materials are transported in suspension, while heavier materials are transported by traction and jumping on the bed, eroding the channel by abrasion [14]. See Fig. 7 for the corresponding raster of the study area of Piemonte.

Curvature

In relation to the slope there is the exposure of the slope. It is divided into 3 classes, concave, convex and flat. These classes in turn are divided into 9 categories which reflect the shape of the slope profile (see Fig. 4). These shapes influence the path of water flows during the events that trigger the floods.

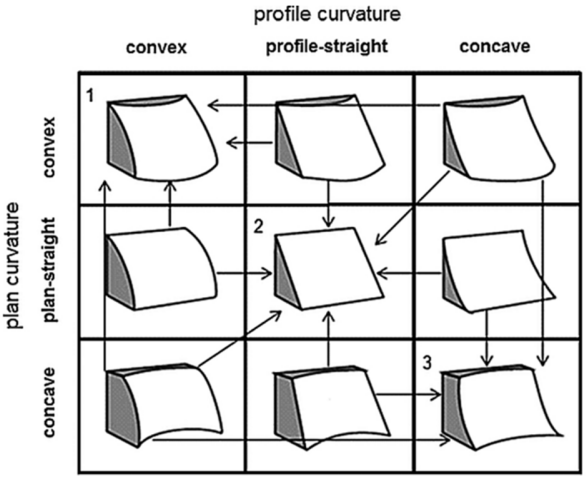


Fig. 3 The nine categories of slope shapes with respect to convexity and concavity [15]

In order to quantify the parameters that affect the power of the flows in relation to the slope two indices were calculated (namely, Stream Power Index (SPI) and Topographic Wetness Index (TWI)) according to the formulas described below [16]. See Fig. 9 for the corresponding raster of the study area of Piemonte.

Stream Power Index

The SPI is defined by: $SPI = \ln(CA \cdot \tan \alpha)$ where CA is the catchment area[1] and α is the slope angle. This index characterizes the areas subject to erosion. SPI can be used to describe potential flow erosion at the given point of the topographic surface. As catchment area and slope gradient increase, the amount of water contributed by upslope areas and the velocity of water flow increase, hence SPI and erosion risk increase. See Fig. 12 for the corresponding raster of the study area of Piemonte.

Topographic Wetness Index

The TWI is defined by: $TWI = \ln(B / \tan \alpha)$ where B is the upstream contributing area per unit width orthogonal to the flow direction and α is the slope angle. This index is a useful estimate of where the water will accumulate in an area with elevation differences. In the case of a flat area, accumulation numbers will be very large, so TWI will not be a relevant variable. See Fig. 15 for the corresponding raster of the study area of Piemonte.

Sediment Transport Index

The Sediment Transport Index (STI) is defined by: $STI = (m + 1) \cdot (CA / 22.13)^m \cdot \sin(\alpha / 0.0896)^n$ where CA is the specific catchment area, α is the slope, m is the contributing area exponent (usually set to 0.4) and n is the slope exponent (usually set to 1.4). It can be used to describe the erosive and depositional processes [16]. See Fig. 13 for the corresponding raster of the study area of Piemonte.

Topographic Ruggedness Index

The Topographic Ruggedness Index (TRI) is defined by: $TRI = \sqrt{\sum_{ij=-1}^{1}(X_{ij} - X_{00})^2}$ where X_{ij} are the values of the 8 cells surrounding the central cell X_{00} as shown by Fig. 4 [17]. It was calculated automatically via the GDAL library in OSGeo4W. It is used to quantify the heterogeneity of the territory and obtained by calculating the difference in elevation between the adjacent cells of the DTM. See Fig. 4 for the corresponding raster of the study area of Piemonte.

-1,-1	0,-1	1,-1
-1,0	0,0	1,0
-1,1	0,1	1,1

Fig. 4 Indices of the 9 cells

Distance from River

The last parameter derived from the DTM is the distance from rivers. It was developed in 2 stages:

1. the extraction of rivers and waterways from the DTM, using the *r.stream.extract* algorithm of the Grass78 library with QGis software.
2. the creation of the distances, in meters, from the lines extracted using the Euclidean distance in QGis.

This parameter is fundamental as floods often occur near rivers or bodies of water. Therefore, the proximity to it represents a higher danger. See Fig. 6 for the corresponding raster of the study area of Piemonte.

The lithological map and the type of soil are important parameters for identifying areas subject to flooding. In fact, the

[1] The catchment area (CA) is an area of a closed figure formed by a contour segment at a given point of the topographic surface and two flow lines coming from upslope to the contour segment ends. The unit of measurement is m^2.

type of soil affects its drainage processes based on the degree of permeability and texture.

Soil

The soil type was elaborated by taking into account the percentages of sand, silt and clay and plotting the data in the ternary diagram (see Fig. 5).

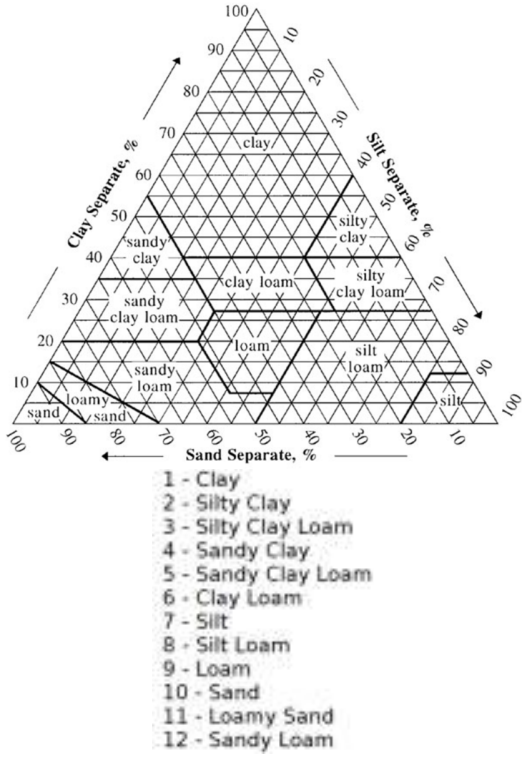


Fig. 5 Ternary diagram for the classification of soils [19]

The information was taken from [20] and processed in QGis via Soil Texture Classification from the SAGA library. See Fig. 17 for the corresponding raster of the study area of Piemonte.

Lithology

We considered the following lithologies:

1. Alluvial deposits in the plain and valley areas (Rissian age-current)
2. Morainic deposits (Quaternary)
3. Banks and clayey levels (Villafranchiano)
4. Fine to medium sands (Asti sands)
5. Clays and clayey marls (Upper Miocene-Middle Pliocene)
6. Layers of marl (Middle Miocene)
7. Marly siltstones (Upper Oligocene-Miocene)
8. Sandstones (Arenaria di Ranzano, Oligocene)
9. Alternations of clays (Cretaceous-Eocene)
10. Serpentinites (Piemonte area, Jurassic-Cretaceous)
11. Calcescisti (Piemonte area, Jurassic-Cretaceous)
12. Dolomite and limestone (Mesozoic Units)
13. Gneiss minutes (Massifs of Dora-Maira, Sesia-Lanzo)
14. Gneiss occhiadini (Massifs Monte Rosa and Val d`Ossola)
15. Granites (Migmatites, ercinic and late-alpine)
16. Other

These lithologies can be grouped into 4 types (see Fig. 16 for the corresponding raster of the study area of Piemonte):

1. clastic rocks
2. intrusive rocks
3. metamorphic rocks
4. sedimentary rocks

In clastic rocks are included:

- Clay banks and levels: alternations of clays in banks and in lists with sandy and gravelly levels of fluvial-lacustrine origin (Villafranchian age)
- Alternations of clay: alternations of clays with calcarenites (arenaceous limestones), elements of dolomitic limestones from Trias, encrinites from Lias, marly limestones and locally granite pebbles (Cretaceous-Eocene age).
- Morainic deposits: accumulation of rock fragments, even in the order of one meter, sand and clay of erosive origin by the movement of the ice that it deposits after its melting (Quaternary age).
- Ranzano sandstones: arenaceous-conglomeratic formation formed by several sedimentary bodies of turbidite origin (Oligocene age)
- Clays and clayey marls: alternation of clays and clayey marls (from 65 to 75% of clayey part and the rest limestone, middle-lower Miocene age)
- Fine to medium sands: alternation of fine (125μ to 250μ) and medium (250 medie to 500μ) sands. Incoherent or locally cemented yellow sands generally homogeneous with the presence of ripple (upper Plieocene age)
- Alluvial deposits: eluvius-colluvial deposits (gravels, sands and poorly classified clays) with relict rocks, due to the erosive action and transport of water (Rissian age)
- Marly siltstones - Termô Fôrà complex: siltstones, sands and conglomerates. Alternations of silts and marly-siliceous limestones (Miocene age)

In intrusive rocks are included:

- Granites: coarse-grained acidic plutonic rocks made up of quartz (20-60%), alkaline feldspar, plagioclase and mafic minerals. It is formed by fractional crystallization within the earth's crust between 10 and 30 km deep (Paleozoic age)

In metamorphic rocks are included:

- Serpentinites: metamorphic rocks made up almost exclusively of minerals from the serpentine group such as antigorite, lizardite and chrysotile, deriving from hydration processes of mafic and ultramafic rocks. Serpentinization is a low-temperature metamorphic process that tends to transform anhydrous minerals (pyroxene and olivine) into hydrated minerals. Depending on the type of rock, particular minerals such as brucite and magnetite can be formed (Triassic-Jurassic age)
- Gneiss minutes: rocks formed at great depth, due to the changed conditions of pressure and temperature, in the areas of convergence of two plates. Estimated metamorphic conditions of 1.4-1.6 Gpa and 500-600 °C (Carboniferous-Permian age)
- Gneiss Occhiadini: medium-large-grained rocks, with a

generally granitic composition, characterized by large oval-shaped feldspar crystals (porphyroclasts) called "augen". These are rocks with a medium-high metamorphic degree (Triassic - Jurassic age)

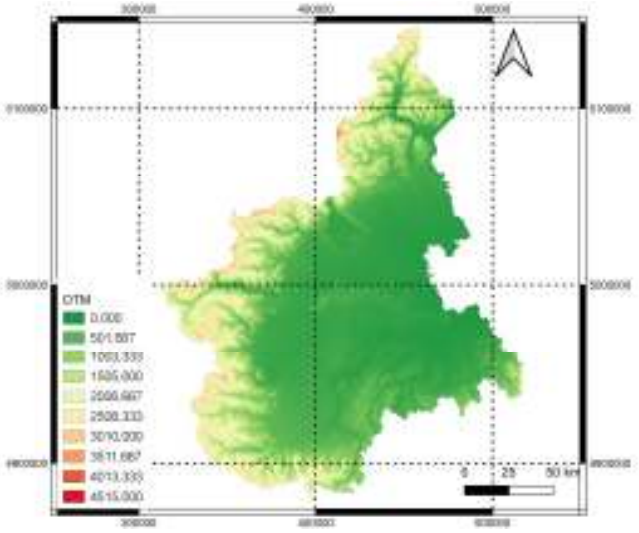


Fig. 6 Elevation raster

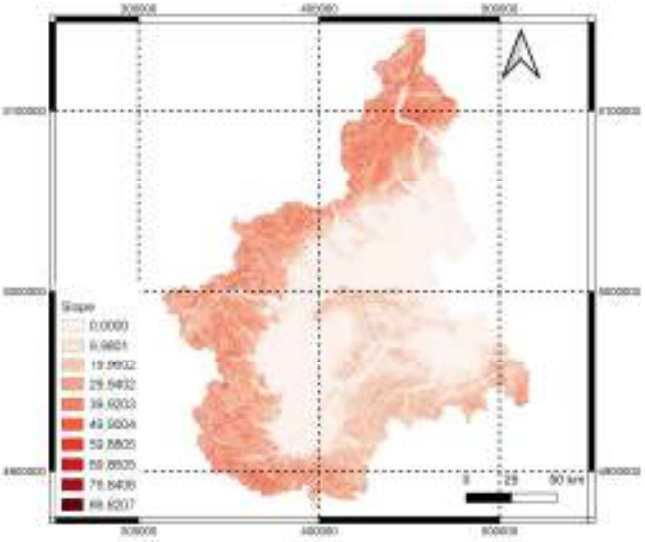


Fig. 7 Slope raster

- Calcareous schist: low grade metamorphic rock, formed by calcite, micas, biotite and quartz. The formation of the calcescisti with green stones or the Piemonte area is the set of rocks that originally constitute the seabed of the Ligurian-Piemontese Ocean. The deposits that by metamorphism gave rise to the calcescists were impure calcareous sediments (from clayey limestones to calcareous clays) deposited in the deep sea (lower Jurassic age)

In sedimentary rocks are included:

- Layers of marls: a terrigenous sedimentary rock composed of a clay and a limestone fraction, there are different types of marl depending on the percentages of limestone and clay. The terrigenous fraction derives from the washout of the continental areas adjacent to the sedimentation area or from the change of factory (from carbonate factory to pelagic factory), while the limestone fraction from calcium carbonate precipitation (middle Miocene age).

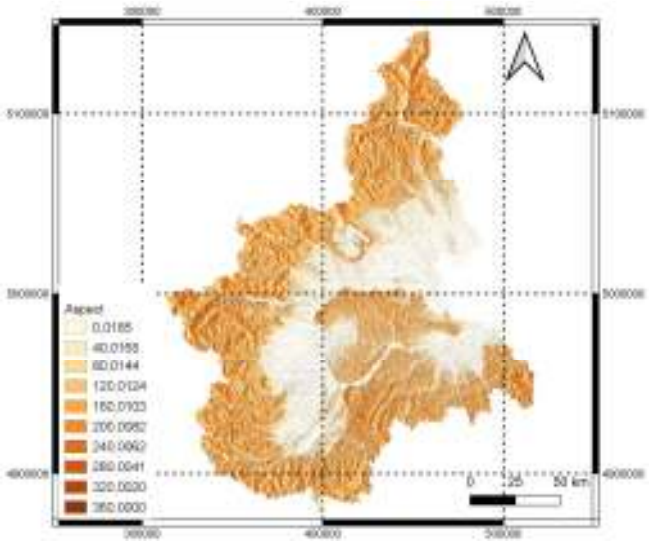


Fig. 8 Aspect raster

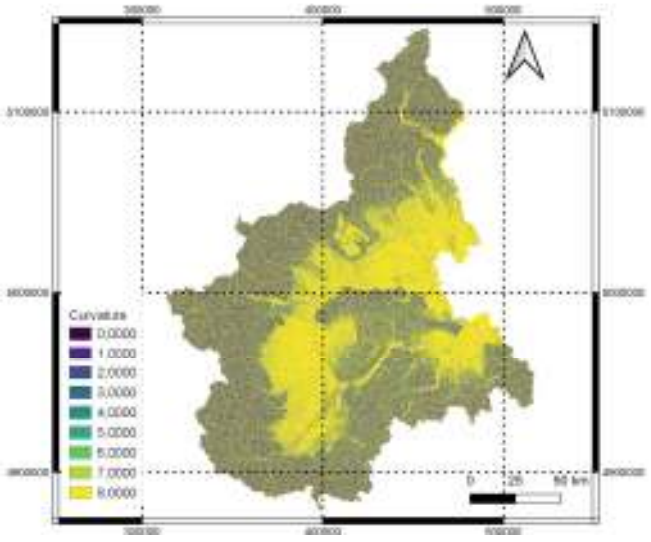


Fig. 9 Curvature raster

- Dolomite and limestone: sedimentary rocks with precipitation of $CaCO_3$ which is replaced by Mg at different percentages depending on the conditions of the water in which the sediment is found (upper Triassic age)

Land Cover/Land Use (LULC)

Another important fact is the use of the land. It is the land cover inventory and is divided into 44 classes (to improve visualization the classes are grouped into five categories in Fig. 11). They are described as vegetated, bare, wooded, cultivated areas, bodies of water, industrial areas, built up areas and road surfaces, and especially urban areas. The latter are

imported not only as they represent areas in which waterproofing is maximum with the increase in surface runoff and the collection of water. It is also important because it indicates the presence of people. This information was taken from the Corine Land Cover 2018 of the Copernicus system. See Fig. 11 for the corresponding raster of the study area of Piemonte.

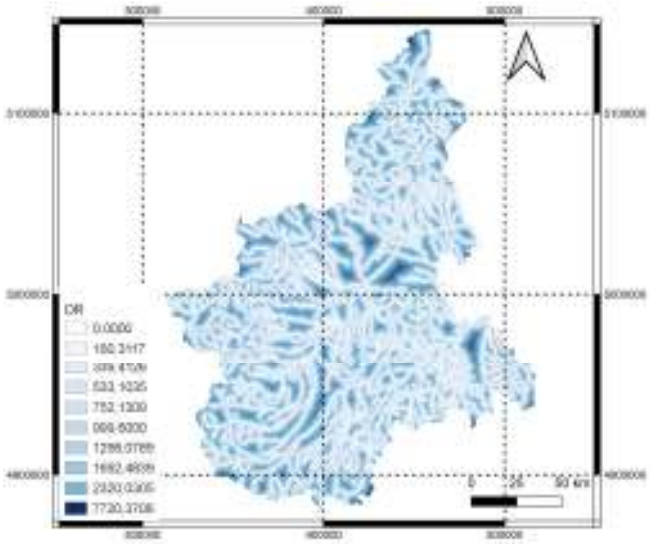


Fig. 10 Distance from River raster

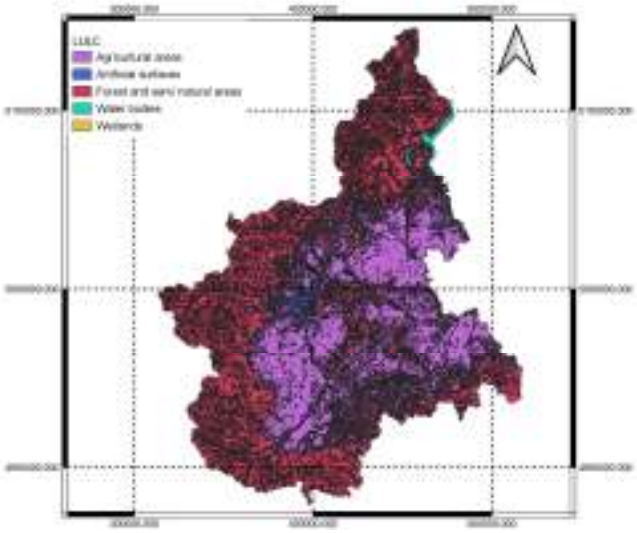


Fig. 11 Land Use/Land Cover raster

Frequency Ratio Mapping

For each static feature, we aimed to assess the relative importance of each range of values with respect to the flood chance. In order to achieve this, the frequency ratio approach was used, that is commonly used in landslide [18] and creates a mapping where each range of values in a feature can be substituted with its odds ratio of the event not occurring by chance.

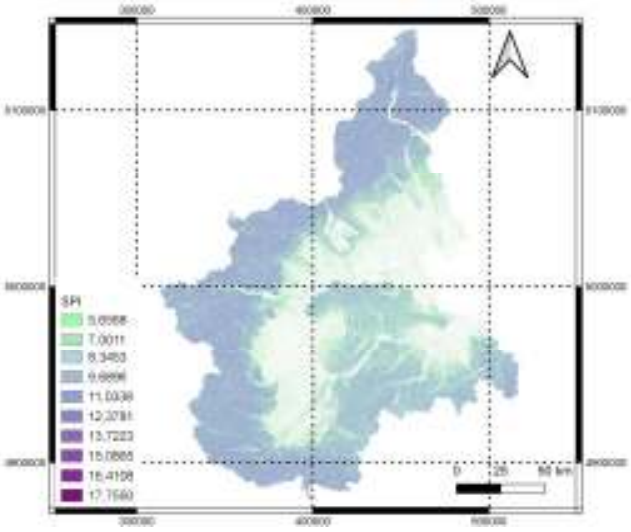


Fig. 12 SPI raster

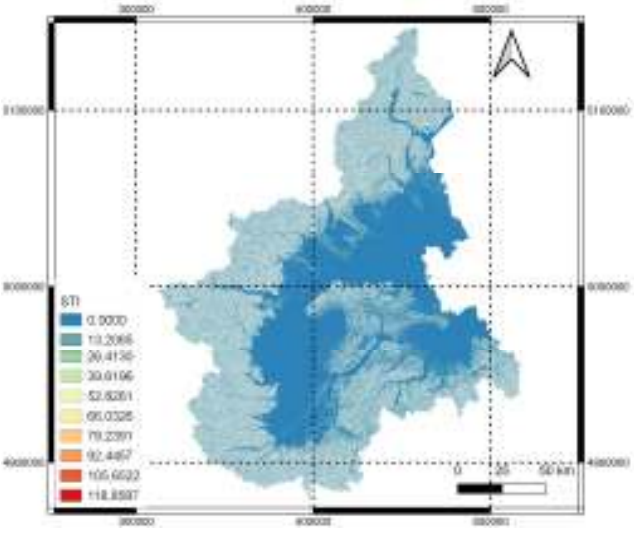


Fig. 13 STI raster

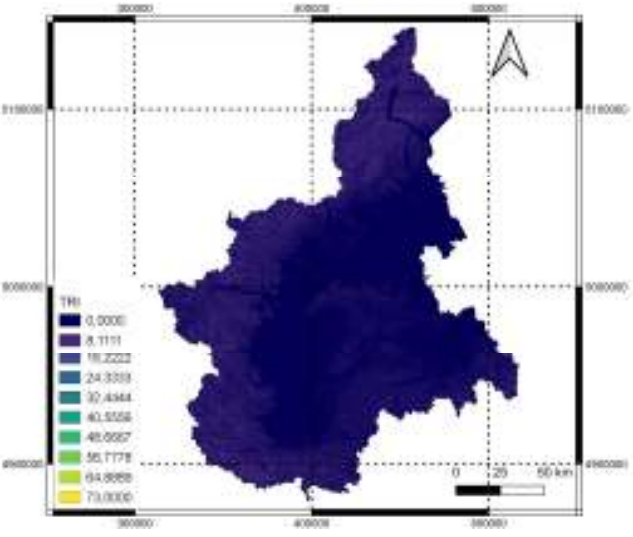


Fig. 14 TRI raster

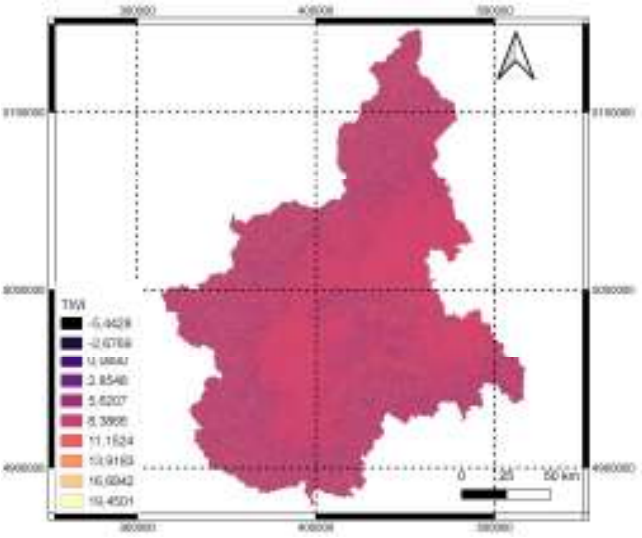


Fig. 15 TWI raster

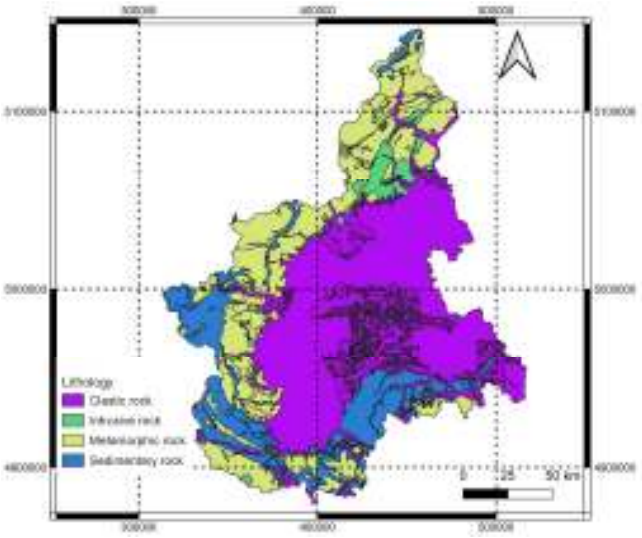


Fig. 16 Lithology raster

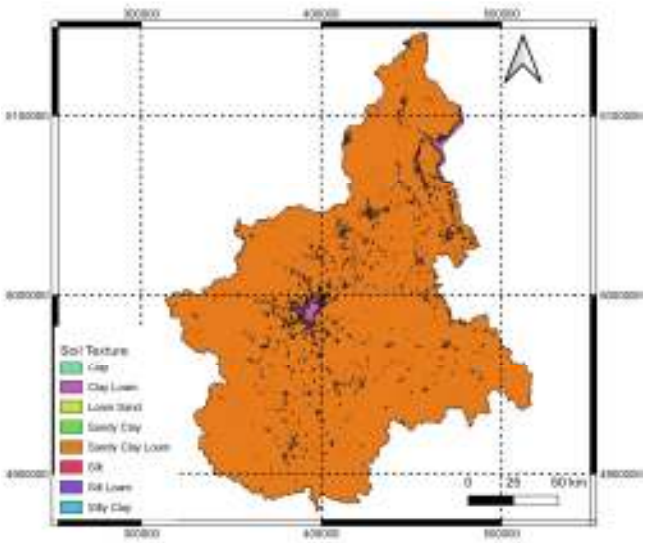


Fig. 17 Soil raster

Frequency ratio is defined as the ratio between two probabilities:

$$FR^i = \frac{P(event,i)}{P(event)\cdot P(i)} \simeq \frac{N^{event,i}/N^i}{N^{event}/N}$$

The first equality is obtained by using Bayes theorem and by considering our data as a pixel matrix with as many layers as the number of our features. This calculation can be done for each range of finite bins i, for each feature layer. Namely,

- $N^{event,i}$ is the number of pixels of class *i* that are also pixels containing a flood event
- N^i is the number of pixels of class i of the feature layer, that is the size of the i-th bin
- N^{event} is the number of pixels containing a flood event
- N is the total number of pixels considered (in the study area the whole Piemonte region)

Methods

In the study area of Piemonte the raster has a size of 26642 x 20367 pixels.

The raster has the following ratio:

- 288.969.353 outer pixels
- 253.648.261 inner pixels

As stated in the introduction the estimate of the flood hazard map of an area is achieved by getting the static data layers mapped to their FR counterparts as input features.

Train/Validation/Test split

The inner pixels have been randomly divided into three splits:

- training: 177.553.785 (70% of the dataset)
- validation: 50.729.654 (20% of the dataset)
- test: 25.364.822 (10% of the dataset)

The splitting is implemented to keep the ratio between the target values of the PAI layer (0, 1, 2 and 3) constant.

TABLE I
SPLIT OF THE PIEMONTE DATASET

Target value	Training	Validation	Test
0	166.283.309	47.509.517	23.754.757
1	8.163.717	2.332.491	1.166.244
2	1.293.817	369.662	184.831
3	1.812.942	517.984	258.990

In order to avoid any use of the test set in training the neural network, the train dataset is the only dataset used to compute frequency ratios.

Single Binary Output Models

Since the target hazard mapping has four different levels three different binary models have been implemented. The architecture is the same, whereas the outputs are the following

- model 1-2-3: detector of any flood event
- model 2-3: detector of medium and high frequency flood events
- model 3: detector of high frequency flood events

The model of a single hazard level detector is a multi-layer perceptron with a single hidden level with 16 neural units (see

Fig. 18 for the diagram of the architecture).

The activations of the hidden layer are rectified linear, whereas the output unit has a sigmoid activation, representing the probability of a flood event.

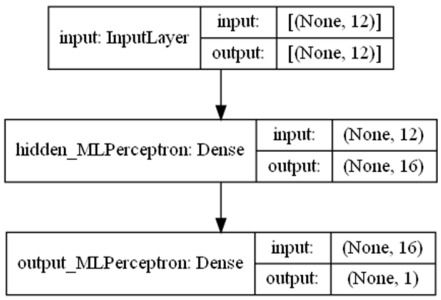


Fig. 18 Architecture of a single hazard level detector

The total number of trainable parameters is $225 = (12 + 1) \cdot 16 + (16 + 1) \cdot 1$ since biases are included as well.

Models' Combination

Since the target hazard mapping has four different levels, the three different binary models have been combined into one. The architecture is similar, whereas the outputs are the four probabilities of each flood hazard class (see Fig. 20 for the diagram of the architecture).

The last layer of this neural network has a softmax activation, σ, defined by:

$$\sigma(i) = \frac{e^{u_i}}{\sum_k e^{u_k}}$$

The special form of such activation normalizes the output (the sum over the units u_i is 1), therefore the response may be interpreted as probability distribution over predicted classes.

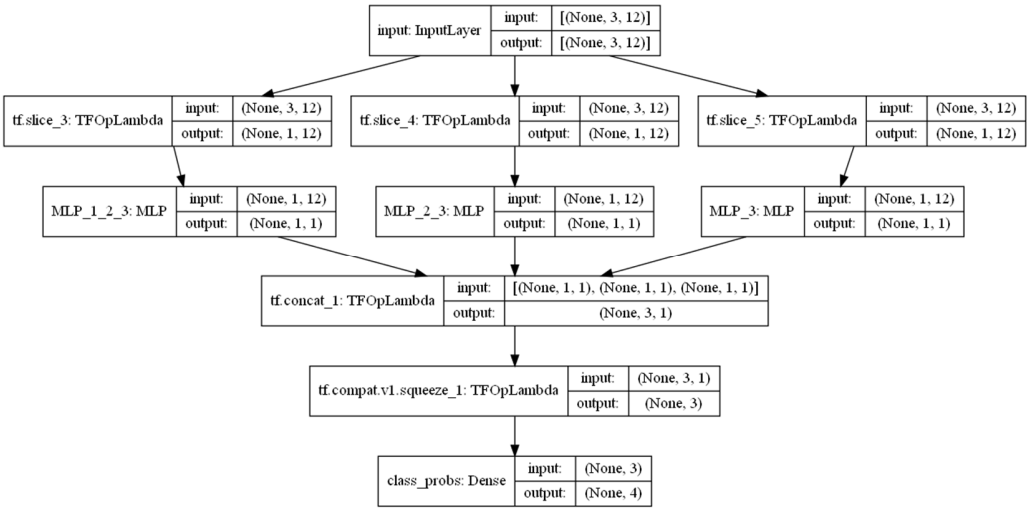


Fig. 19 Architecture multi flood hazard levels classifier

The total number of trainable parameters is $691 = 225 \cdot 3 + (3 + 1) \cdot 41$ since biases are included as well.

Performance Assessment

A baseline model, with no trainable parameters, whose decision is taken by the sum of the frequency ratios has been set as a benchmark for the performance of the model.

Since the flood hazard classes are strongly imbalanced, the loss function used for training has been endowed with weights $\{w_0, w_1\}$ defined by:

$$w_0 = \frac{\#pos + \#neg}{2 \cdot \#neg}$$
$$w_1 = \frac{\#pos + \#neg}{2 \cdot \#pos}$$

where $\#pos$ and $\#neg$ denote the positive and negative occurrences, respectively. Namely we have the following version of the binary cross-entropy:

$$CrossEntropy(pred, labels) = -\sum_{i\,in\,batch} w_{label_i} \cdot label_i \cdot log(pred_i)$$

We consider the following metrics:

- confusion matrix: collecting all the possible combinations of the classification, true positive TP, true negative TN, false positive FP, false negative FN
- F_1 score: a standard score used in statistical analysis of binary classification to measure the model's accuracy. It is defined as

$$F_1 = \frac{2}{\frac{1}{p}+\frac{1}{r}} = \frac{TP}{TP + \frac{1}{2} \cdot (FP + FN)}$$

where $p = \frac{TP}{TP+FP}$ and $r = \frac{TP}{TP+FN}$ denote the precision and recall, respectively.

- the model output for the two binary classes (flood vs. no flood)

The results of single event detectors are represented in Figs. 20-22.

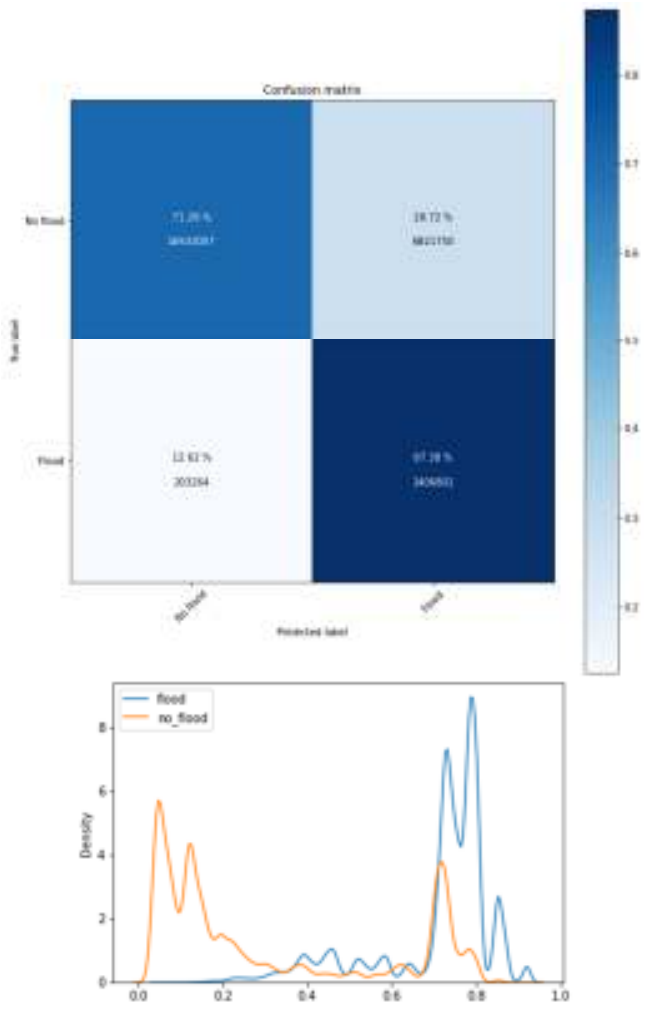


Fig. 20 Baseline detector of any flood event

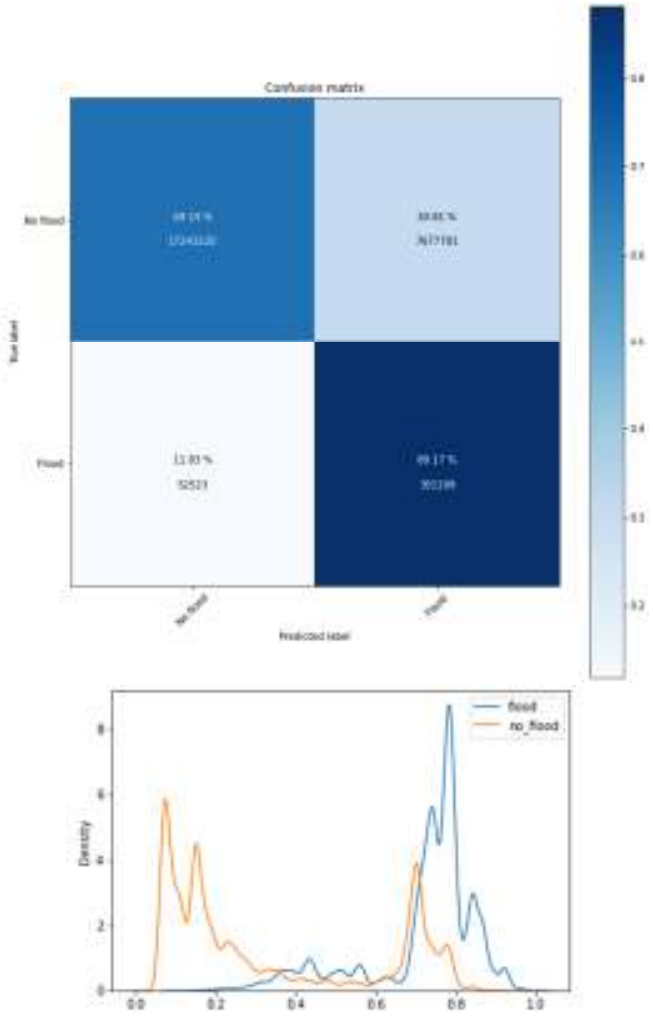


Fig. 21 Baseline detector of medium and high frequency flood events

TABLE II
BASELINE DETECTOR METRICS

	True negative	True positive	False negative	False positive	F_1
1-2-3	16933007 (71.28 %)	1406801 (87.38 %)	203264 (12.62 %)	6821750 (28.72 %)	0.286
2-3	17243220 (69.19 %)	391298 (88.17 %)	52523 (11.83 %)	7677781 (30.81 %)	0.092
3	16539679 (65.88 %)	244298 (94.33 %)	14692 (5.67 %)	8566153 (34.12 %)	0.054

Table II collects the indices of the three baseline models.

III. RESULTS

Frequency Ratio Mapping

Using the historical information by the train dataset the following mappings for the features are obtained. Tables III-XV collect the values of the frequency ratios obtained for the three binary models:

- 1-2-3: any flood events
- 2-3: medium and high frequency flood events
- 3: high frequency flood events

TABLE III
FREQUENCY RATIOS FOR ELEVATION

Bins	Frequency ratios		
	1-2-3	2-3	3
-10 m - 0 m	--	--	--
0 m - 50 m	--	--	--
50 m - 100 m	7.259	8.642	0.000
100 m - 200 m	2.751	3.010	3.061
200 m - 300 m	1.533	1.640	2.201
300 m - 500 m	0.663	0.223	0.381
500 m - 700 m	0.180	0.077	0.133
700 m - 1100 m	0.062	0.000	0.000
1100 m - 1500 m	0.003	0.000	0.000
1500 m - 2000 m	0.000	0.000	0.000
2000 m -5000 m	0.000	0.000	0.000

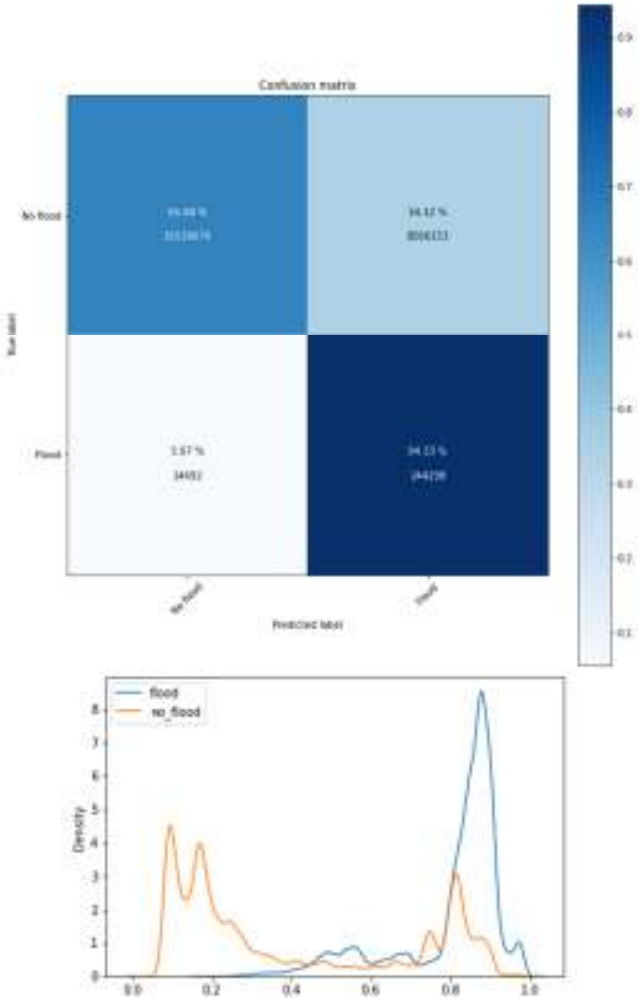


Fig. 22 Baseline detector of high frequency flood events

TABLE IV
FREQUENCY RATIOS FOR ASPECT

Bins	Frequency ratios		
	1-2-3	2-3	3
no direction	2.651	2.706	2.664
-22.5° - 22.5°	0.309	0.270	0.269
22.5° - 67.5°	0.287	0.253	0.274
67.5° - 112.5°	0.396	0.363	0.388
112.5° - 157.5°	0.301	0.282	0.322
157.5° - 202.5°	0.331	0.350	0.386
202.5° - 247.5°	0.207	0.192	0.209
247.5° - 292.5°	0.244	0.216	0.207
292.5° - 337.5°	0.214	0.173	0.180

TABLE V
FREQUENCY RATIOS FOR SLOPE

Bins	Frequency ratios		
	1-2-3	2-3	3
0° - 1.69°	2.653	2.708	2.666
1.69° - 4.49°	1.824	1.739	1.840
4.49° - 7.85°	0.307	0.199	0.240
7.85° - 11.9°	0.093	0.056	0.067
11.9° - 16.3°	0.039	0.026	0.035
16.3° - 21.1°	0.022	0.013	0.018
21.1° - 26.4°	0.015	0.009	0.013
26.4° - 33.1°	0.010	0.006	0.009
33.1° - 45.7°	0.005	0.002	0.003
> 45.7°	0.024	0.044	0.001

TABLE VI
FREQUENCY RATIOS FOR SPI

Bins	Frequency ratios		
	1-2-3	2-3	3
-1 - 0	2.653	2.708	2.666
0 - 2	--	--	--
2 - 4	--	--	--
4 - 6	1.921	1.876	1.959
6 - 8	0.753	0.696	0.747
8 - 10	0.044	0.029	0.032
10 - 12	0.011	0.005	0.007
12 - 14	0.005	0.001	0.002
14 - 16	0.003	0.000	0.000
16 - 18	0.002	0.000	0.000
18 - 20	0.000	0.000	0.000

TABLE VII
FREQUENCY RATIOS FOR TWI

Bins	Frequency ratios		
	1-2-3	2-3	3
< 0	0.027	0.030	0.004
0 - 0.96	0.037	0.028	0.031
0.96 - 5.47	0.331	0.303	0.328
5.47 - 6.08	0.041	0.028	0.033
6.08 - 6.62	0.049	0.031	0.042
6.62 - 7.07	0.052	0.029	0.038
7.07 - 7.53	0.049	0.025	0.034
7.53 - 7.99	0.053	0.021	0.030
7.99 - 8.45	0.041	0.018	0.025
8.45 - 9.057	0.032	0.012	0.017
9.057 - 9.89	2.192	2.227	2.203
> 9.89	0.188	0.168	0.179

TABLE VIII
FREQUENCY RATIOS FOR STI

Bins	Frequency ratios		
	1-2-3	2-3	3
< 0.0	0.022	0.015	0.017
0.0 - 0.5	2.503	2.555	2.516
0.5 - 1.0	1.133	1.105	1.154
1.0 - 1.5	0.867	0.843	0.869
1.5 - 2.0	0.149	0.114	0.125
2.0 - 3.0	0.575	0.515	0.562
3.0 -4.0	0.592	0.549	0.593
4.0 - 5.0	0.107	0.073	0.090
5.0 - 7.5	0.068	0.039	0.050
7.5 - 10.0	0.053	0.021	0.028
10.0 - 15.0	0.042	0.013	0.019
15.0 - 50.0	0.020	0.004	0.005
50.0 - 100.0	0.008	0.000	0.000
100.0 - 500.0	0.000	0.000	0.000

TABLE IX
FREQUENCY RATIOS FOR TRI

Bins	Frequency ratios		
	1-2-3	2-3	3
< 0	2.738	2.806	2.744
0.0 - 0.05	--	--	--
0.11 - 0.16	2.157	2.109	2.248
0.16 - 0.21	--	--	--
0.21 - 0.32	2.037	1.970	2.086
0.32 - 0.48	2.013	1.948	2.040
0.48 - 0.69	1.374	1.237	1.344
0.69 - 1.17	0.227	0.149	0.176
> 1.17	0.021	0.015	0.016

TABLE X
FREQUENCY RATIOS FOR CURVATURE

Bins	Frequency ratios		
	1-2-3	2-3	3
1	0.213	0.188	0.208
2	0.015	0.008	0.011
3	0.019	0.012	0.014
4	1.173	1.121	1.166
5	0.117	0.075	0.090
6	1.207	1.186	1.214
7	0.027	0.016	0.022
8	0.022	0.014	0.019
9	1.632	1.656	1.639

TABLE XI
FREQUENCY RATIOS FOR DISTANCE FROM RIVER

Bins	Frequency ratios		
	1-2-3	2-3	3
0 m - 281 m	2.849	3.210	2.797
281 m - 561 m	1.338	1.569	1.731
561 m - 842 m	0.711	0.638	0.826
842 m - 1122 m	0.434	0.288	0.390
1122 m - 1403 m	0.291	0.118	0.177
1403 m - 1683 m	0.257	0.058	0.097
1683 m - 1964 m	0.274	0.070	0.120
1964 m - 2244 m	0.276	0.036	0.062
2244 m - 2525 m	0.244	0.014	0.024
2525 m - 2806 m	0.223	0.000	0.000
2806 m - 3086 m	0.243	0.000	0.000
3086 m - 3367 m	0.250	0.000	0.000
3367 m - 3647 m	0.171	0.000	0.000
3647 m - 3928 m	0.102	0.000	0.000
3928 m - 4208 m	0.077	0.000	0.000
4208 m - 4489 m	0.071	0.000	0.000
4489 m - 4769 m	0.025	0.000	0.000
4769 m - 5050 m	0.105	0.000	0.000
5050 m - 5330 m	0.141	0.000	0.000
5330 m - 5611 m	0.001	0.000	0.000
5611 m - 5892 m	0.000	0.000	0.000
5892 m - 6172 m	0.000	0.000	0.000
6172 m - 6453 m	0.000	0.000	0.000
6453 m - 6733 m	0.000	0.000	0.000
6733 m - 7014 m	0.000	0.000	0.000
7014 m - 7294 m	0.000	0.000	0.000
7294 m - 7575 m	0.000	0.000	0.000
7575 m - 7855 m	0.000	0.000	0.000
> 7855 m	0.000	0.000	0.000

TABLE XII
FREQUENCY RATIOS FOR SOIL

Bins	Frequency ratios		
	1-2-3	2-3	3
Clay	15.069	0.000	0.000
Silty Clay	3.315	1.606	1.397
Silty Clay Loam	--	--	--
Sandy Clay	0.000	0.000	0.000
Sandy Clay Loam	0.986	0.986	1.001
Clay Loam	0.694	0.190	0.074
Silt	1.623	1.572	0.664
Silt Loam	1.618	1.922	1.437
Loam	--	--	--
Sand	--	--	--
Loamy Sand	3.502	0.000	0.000
Sandy Loam	--	--	--

The bins have been chosen without any assumption on the region, in order to let them be general enough, or at least to be applied all over the Italian territory. On the other hand, given the large size of the train dataset and the time span of the observations, it is likely to assume that the mapping will be constant.

TABLE XIII
FREQUENCY RATIOS FOR LITHOLOGY

Bins	Frequency ratios		
	1-2-3	2-3	3
Alluvial deposits in the plain and valley areas	2.023	4.770	0.000
Morainic deposits	2.581	2.578	2.593
Banks and clayey levels	0.008	0.002	0.004
Fine to medium sands	0.005	0.000	0.000
Clays and clayey marls	0.013	0.007	0.010
Layers of marl	0.029	0.036	0.026
Marly siltstones	0.015	0.000	0.000
Sandstones	0.013	0.021	0.013
Alternations of clays	0.004	0.000	0.000
Serpentinites	0.019	0.004	0.006
Calcescisti	0.000	0.000	0.000
Dolomite and limestone	0.000	0.000	0.000
Gneiss minutes	0.002	0.000	0.000
Gneiss occhiadini	0.004	0.000	0.000
Granites	0.001	0.000	0.000
Other	0.004	0.002	0.004

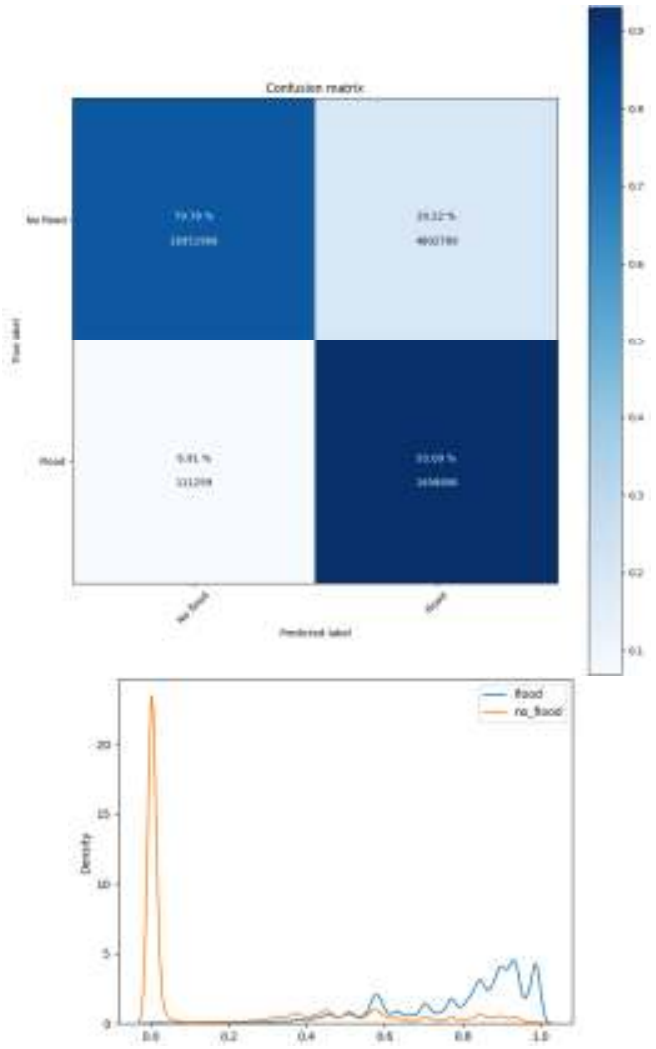


Fig. 23 Neural network detector of any flood event

TABLE XIV
FREQUENCY RATIOS FOR LAND USE/LAND COVER

Bins	Frequency ratios		
	1-2-3	2-3	3
Continuous urban fabric	1.066	0.157	0.265
Discontinuous urban fabric	1.004	0.361	0.559
Industrial or commercial units	1.765	1.350	2.057
Road and rail networks and associated land	2.441	1.972	3.062
Port areas	--	--	--
Airports	0.000	0.000	0.000
Mineral extraction sites	4.531	9.515	10.585
Dump sites	0.453	0.000	0.000
Construction sites	0.000	0.000	0.000
Green urban areas	3.020	0.946	0.000
Sport and leisure facilities	2.520	3.835	5.886
Non-irrigated arable land	2.173	2.348	2.858
Permanently irrigated land	--	--	--
Rice fields	2.520	0.598	0.849
Vineyards	0.094	0.005	0.008
Fruit trees and berry plantations	2.222	0.020	0.034
Olive groves	--	--	--
Pastures	1.450	0.232	0.397
Annual crops associated with permanent crops	0.370	0.000	0.000
Complex cultivation patterns	0.796	0.559	0.847
Land principally occupied by agriculture, with significant areas of natural vegetation	1.378	1.582	1.716
Agro-forestry areas	--	--	--
Broad-leaved forest	0.340	0.422	0.322
Coniferous forest	0.006	0.000	0.000
Mixed forest	0.024	0.024	0.000
Natural grasslands	0.014	0.004	0.000
Moors and heathland	0.000	0.000	0.000
Sclerophyllous vegetation	--	--	--
Transitional woodland-shrub	0.859	1.424	0.616
Beaches, dunes, sands	13.786	21.337	11.531
Bare rocks	0.000	0.000	0.000
Sparsely vegetated areas	0.000	0.001	0.000
Burnt areas	0.000	0.000	0.000
Glaciers and perpetual snow	0.000	0.000	0.000
Inland marshes	8.268	18.343	0.000
Peat bogs	--	--	--
Salt marshes	--	--	--
Salines	--	--	--
Intertidal flats	--	--	--
Water courses	13.050	32.225	14.881
Water bodies	0.139	0.380	0.517
Coastal lagoons	--	--	--
Estuaries	--	--	--
Sea and ocean	--	--	--
Nodata	1.715	3.410	0.000

Flood Hazard Map

The results of single event detectors are represented in Figs. 23-25.

The percentages of both the true negatives and the true positive are improved, and the F_1 score is doubled with respect to the corresponding baseline models (see Table II). Absolute low values of F_1 are connected with a large size of false positive cases. That is not so critical as would be the opposite case of a higher number of false negatives, since overestimation of the flood hazard may bring to a safer map. On the other hand, underestimating flood hazard leads to a very unsafe map.

As stated before, the binary detectors can be combined to build a multi-level classification model. The binary detectors are used to initialize the combined model. The confusion matrix coming out at the end of the training is shown in Fig. 26.

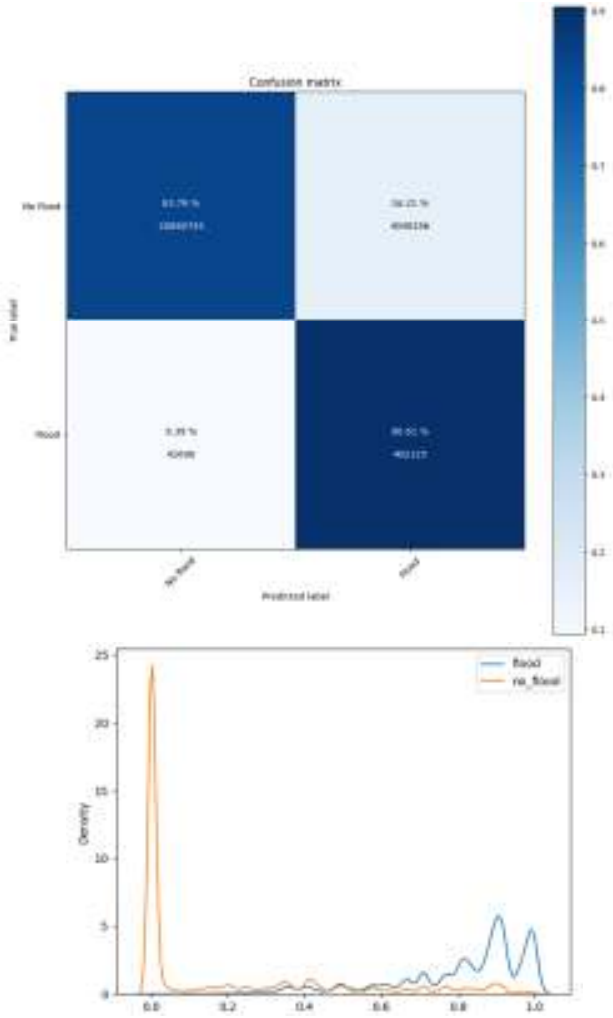


Fig. 24 Neural network detector of medium and high frequency flood events

Table XV collects the indices of the three neural network models.

Principal Component Analysis

The errors of the combined model were further analyzed. The first two principal components were considered of the test dataset, since the variance explained is large enough (see Fig. 27).

We focused on the no flood events of the test dataset, without loss of generality. The scatter plot of the two components has been used to depict how features are linearly separable according to the predicted flood hazard risk (Figs. 27-34). In order to improve the visualization figures have been coupled, Figs. 28, 30, 32, 34 represent the points where the model fails, whereas in Figs. 29, 31, 33, 35 every point is represented.

The predicted classes are represented by the following colors:

- Flood hazard level 0: orange
- Flood hazard level 1: green
- Flood hazard level 2. red
- Flood hazard level 3: blue

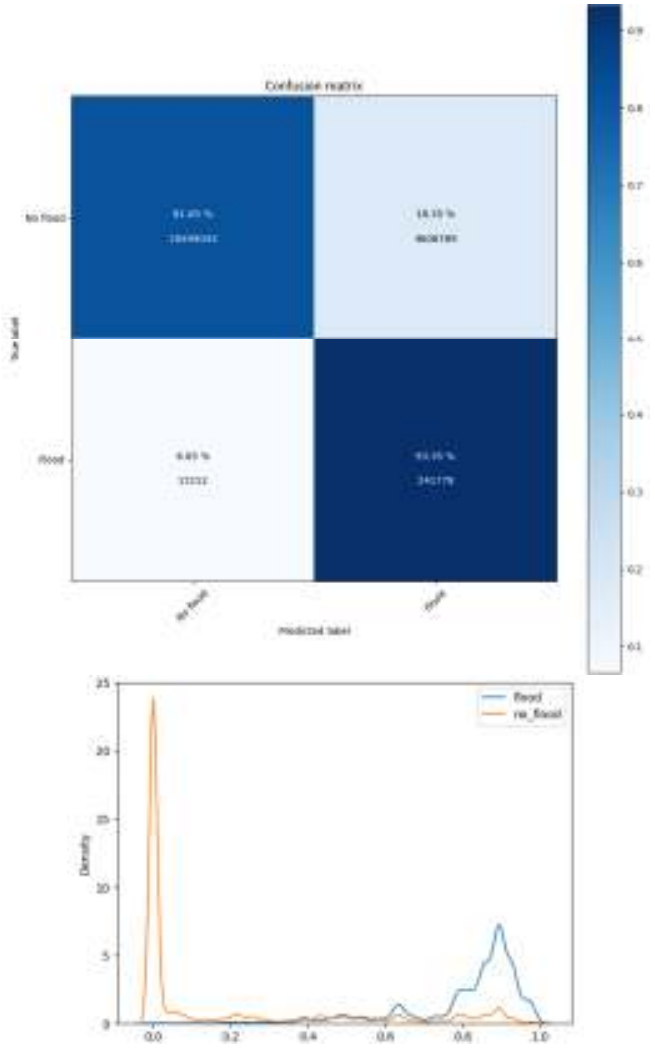


Fig. 25 Neural network detector of high frequency flood events

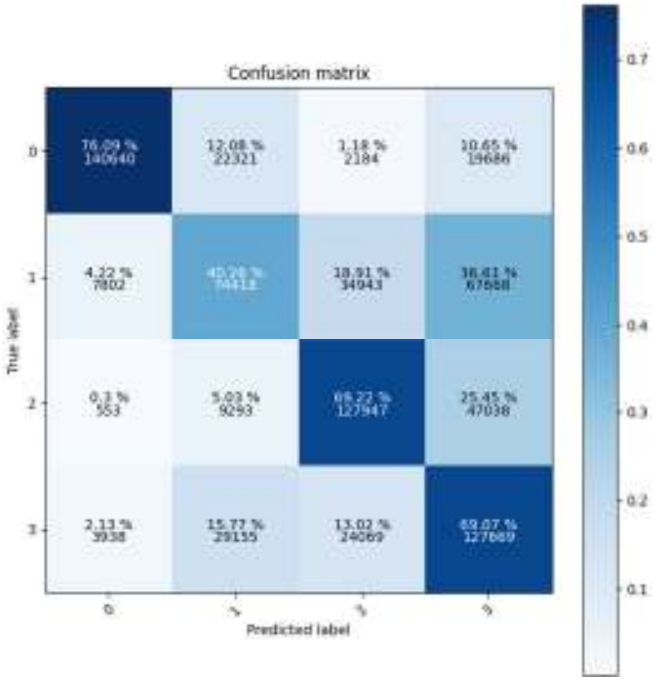


Fig. 26 Confusion matrix for the multi-level classification model

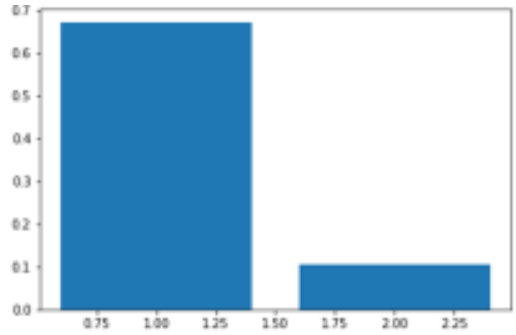


Fig. 27 Explained variance for the first 2 principal components (test dataset)

TABLE XV
NEURAL NETWORK MODELS METRICS

	True negative	True positive	False negative	False positive	F_1c
1-2-3	18951968 (79.78 %)	1498806 (93.09 %)	111259 (6.91 %)	4802789 (20.22 %)	0.379
2-3	20880745 (83.79 %)	402125 (90.61 %)	41696 (9.39 %)	4040256 (16.21 %)	0.165
3	20499043 (81.65 %)	241778 (93.35 %)	17212 (6.65 %)	4606789 (18.35 %)	0.095

It turns out that level 0 lies on the lower left corner, and level 2 on the upper right corner. Level 1 and level 3 both lie on the center, and the overlapping is an explanation for the model confusion between these two levels.

IV. CONCLUSION

This work proposed a different modelling strategy for flood hazard maps, combining classical (FR, Frequency Ratio) and modern (Neural Networks) approaches. The binary models all perform well, surely helped by the feature space remodulation induced by the FR transformation, but their combination trained to produce a multi-class classification suffered from an inherent degree of indistinguishability given by the data, as shown by the PCA analysis. These results indicate that an efficient flood hazard modeling should be enriched by more information coming from independent sources (remember that most of our data were derived from the DTM) like vegetation indices (NDVI, dNBR) or anthropic data.

The classification power of the benchmark model based only on FRs has been improved by the mixture with a neural network. That is a first step in enabling automation of flood hazard mapping.

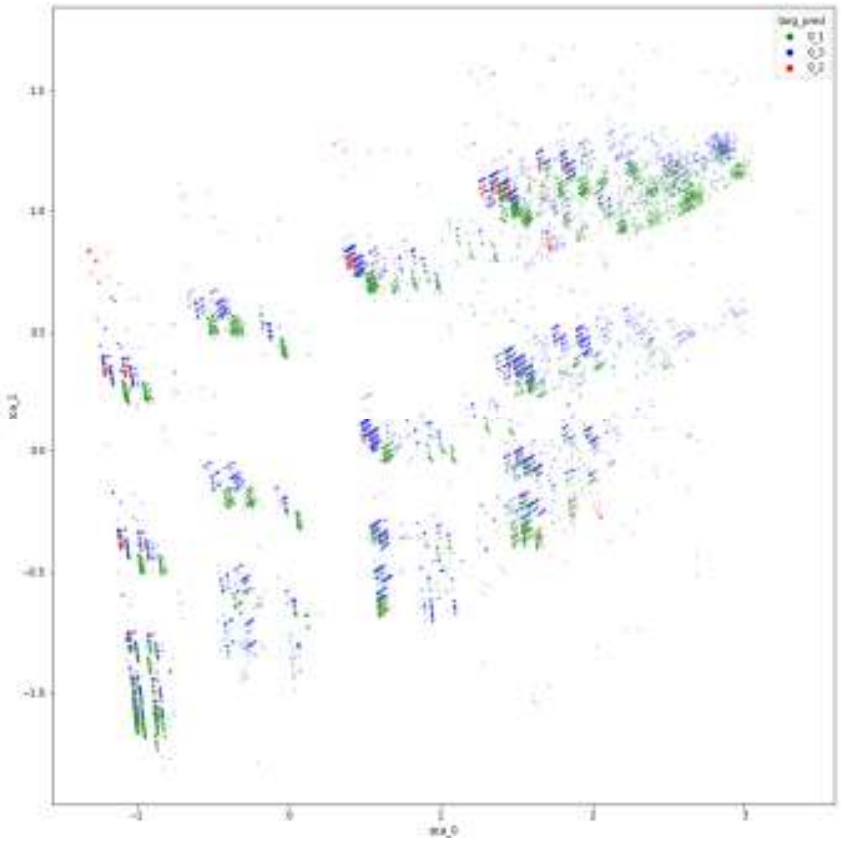


Fig. 28 Scatter plot of the first 2 principal components for level 0, without true positives

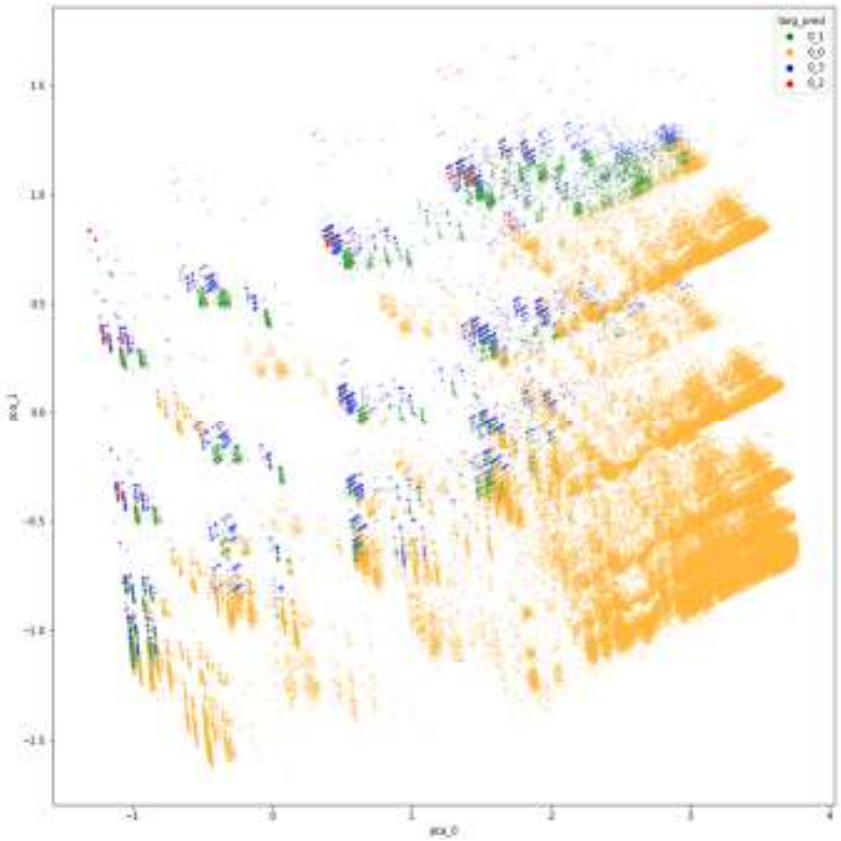


Fig. 29 Scatter plot of the first 2 principal components for level 0, with true positives (orange)

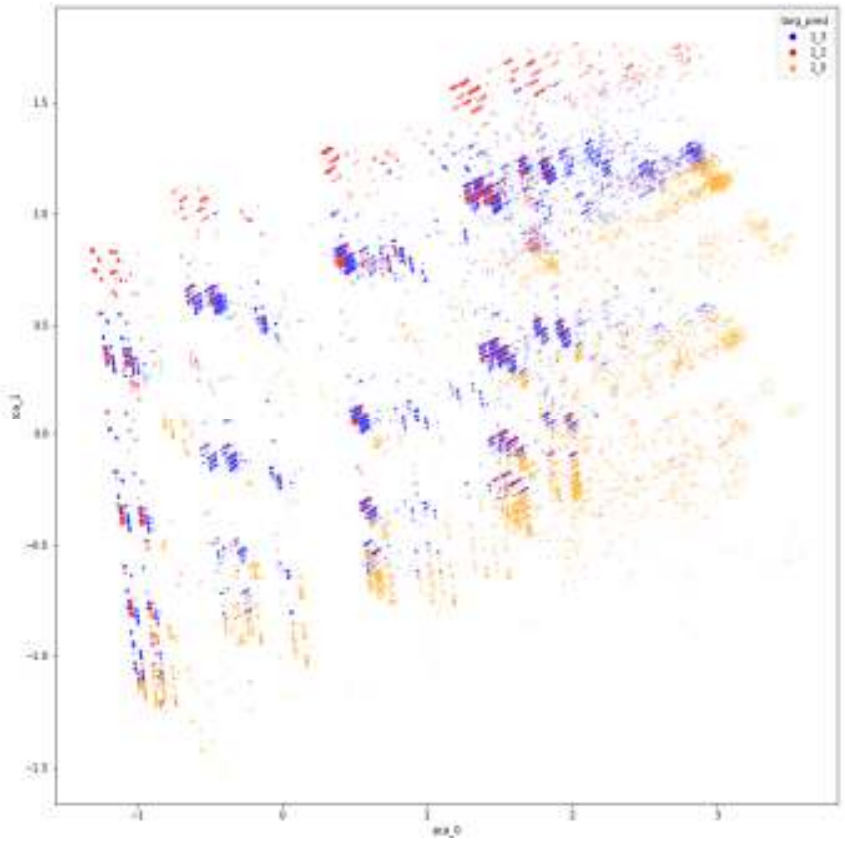


Fig. 30 Scatter plot of the first 2 principal components for level 1, without true positives

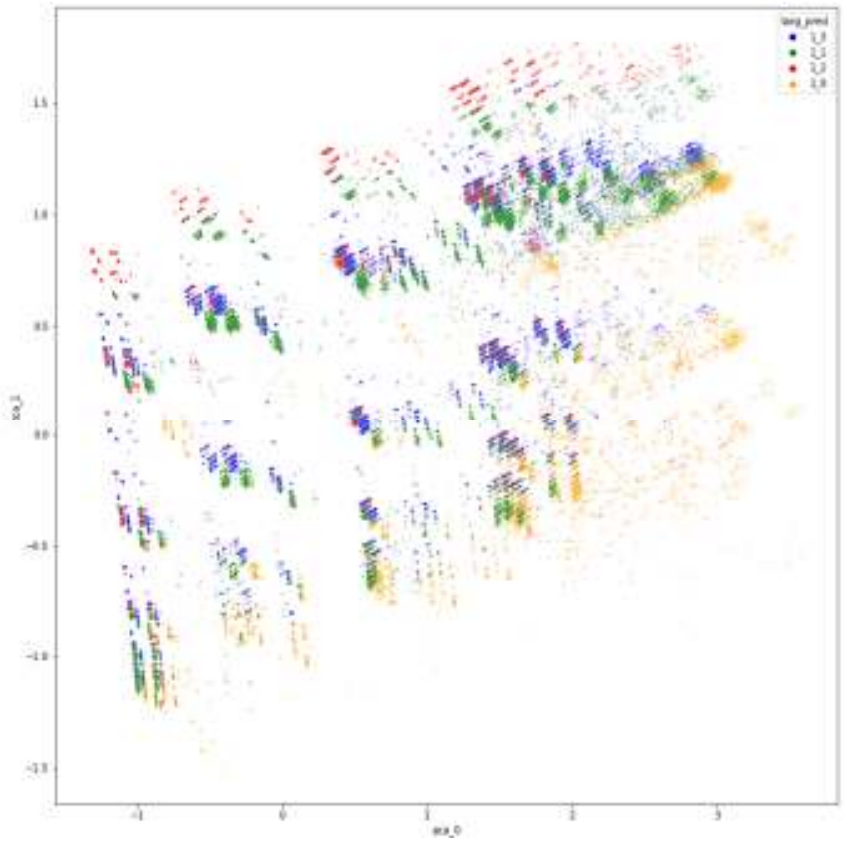


Fig. 31 Scatter plot of the first 2 principal components for level 1, with true positives (green)

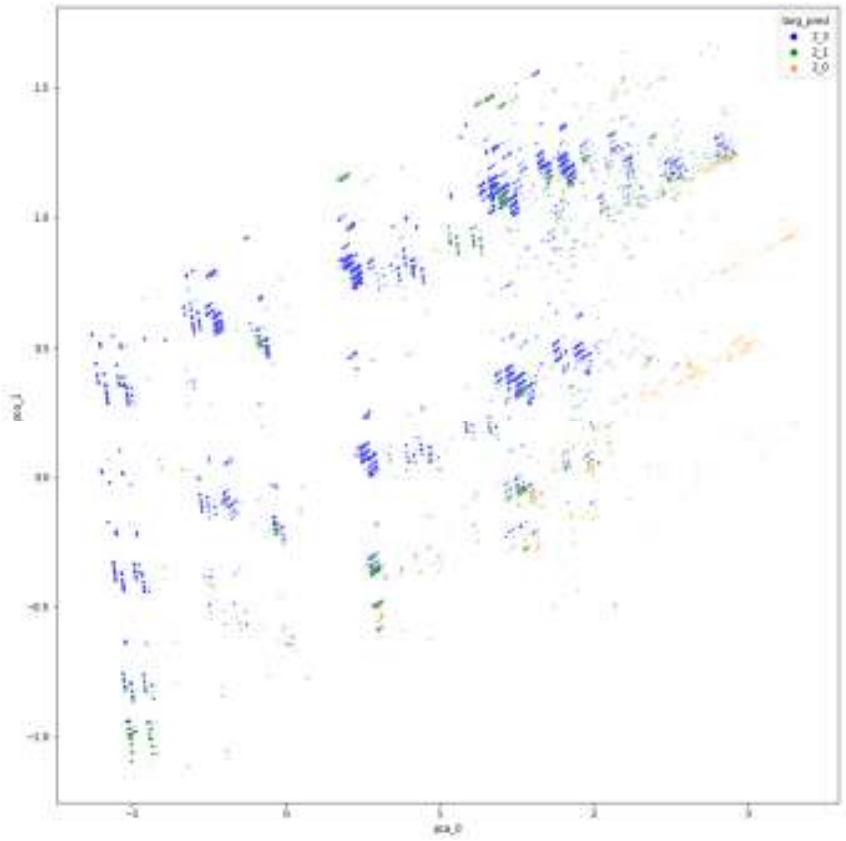


Fig. 32 Scatter plot of the first 2 principal components for level 1, without true positives

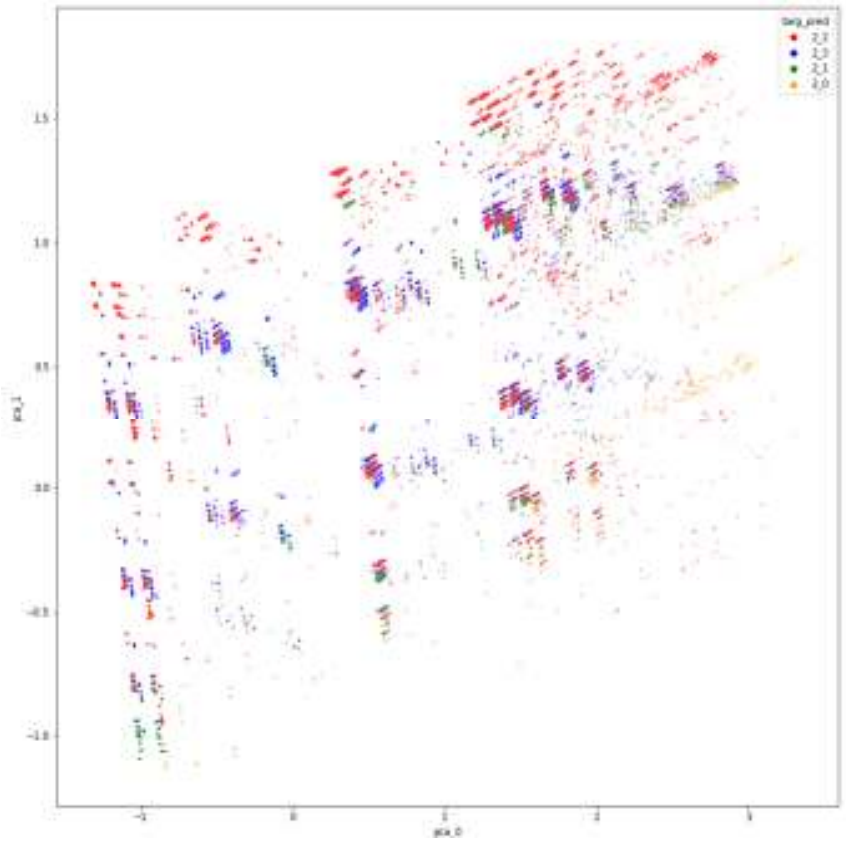


Fig. 33 Scatter plot of the first 2 principal components for level 1, with true positives (red)

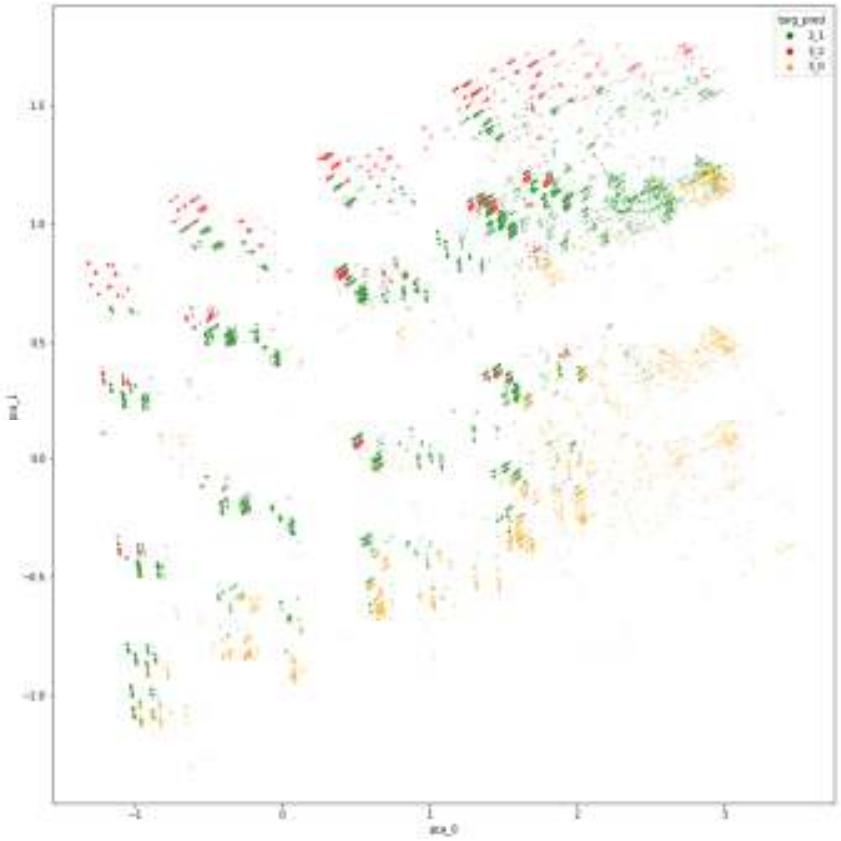


Fig. 34 Scatter plot of the first 2 principal components for level 3, without true positives

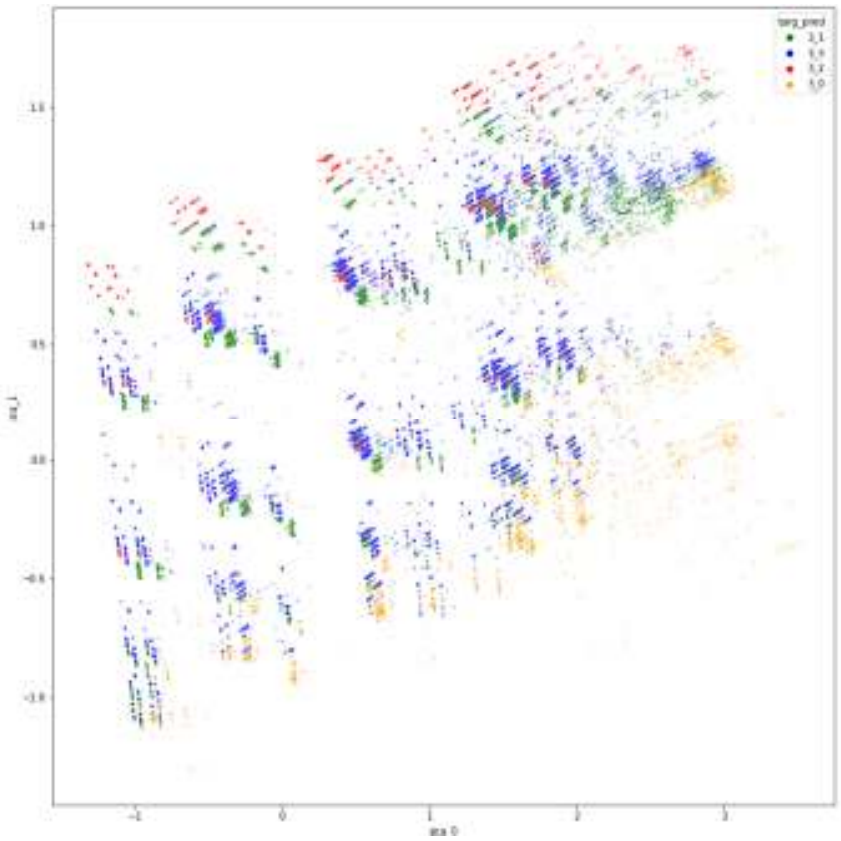


Fig. 35 Scatter plot of the first 2 principal components for level 3, with true positives (blue)

VI. Future Work

There are many research directions to be explored in the field of flood modeling.

The first problem to tackle is to overcome linear separation. A wider feature space (with less correlation between data) could surely benefit the whole model. A larger input including the neighborhoods of the central pixel could be mixed with convolutional layers to improve the classification capacity.

As far as classification of multiple flood hazard levels is concerned, a different architecture could be investigated without the combination of single binary detectors.

References

[1] D. Judi, C. Rakowski, S. Waichler, Y. Feng and M. Wigmosta, «Integrated Modeling Approach for the Development of Climate-Informed, Actionable Information,» Water, vol. 10, n. 6, p. 775, 13 June 2018.

[2] S. A. Schumm and R. W. Lichty, «Time, space, and causality in geomorphology,» American Journal of Science, vol. 263, n. 2, pp. 110-119, February 1965.

[3] K. Li, S. Wu, E. Dai and Z. Xu, «Flood loss analysis and quantitative risk assessment in China,» Natural Hazards, vol. 63, n. 2, pp. 737-760, September 2012.

[4] G. Testa, D. Zuccalà, F. Alcrudo, J. Mulet and S. Soares-Frazão, «Flash flood flow experiment in a simplified urban district,» Journal of Hydraulic Research, vol. 45, pp. 37-44, 2007.

[5] E. van Beek and J. van Alphen, «From flood defence to flood management - prerequisites for sustainable flood management,» in Floods, from defence to management: Proceedings of the 3rd International Symposium on Flood Defence, Leiden, 2006.

[6] W.-H. Teng, M.-H. Hsu, C.-H. Wu and A. S. Chen, «Impact of Flood Disasters on Taiwan in the Last Quarter Century,» Natural Hazards, vol. 37, n. 1-2, pp. 191-207, 2006.

[7] C. Kousky, «Financing Flood Losses: A Discussion of the National Flood Insurance Program,» vol. 21, n. 1, pp. 11-32, 2018.

[8] D. R. Dassanayake, A. Burzel and H. Oumeraci, «Methods for the Evaluation of Intangible Flood Losses and Their Integration in Flood Risk Analysis,» Coastal Engineering Journal, vol. 57, n. 1, pp. 1-35, 2015.

[9] V. Meyer, S. Scheuer e and D. Haase, «A multicriteria approach for flood risk mapping exemplified at the Mulde river, Germany,» Natural Hazards, vol. 48, n. 1, pp. 17-39, 2009.

[10] M.-J. Lee, J.-e. Kang and S. Jeon, «Application of frequency ratio model and validation for predictive flooded area susceptibility mapping using GIS,» in 2012 IEEE International Geoscience and Remote Sensing Symposium, July, 2012.

[11] A. K. Kar, A. Lohani, N. Goel and G. Roy, «Rain gauge network design for flood forecasting using multi-criteria decision analysis and clustering techniques in lower Mahanadi river basin, India,» Journal of Hydrology: Regional Studies, vol. 4, pp. 313-332, September 2015.

[12] M. Ahmadlou, A. Al□Fugara, A. R. Al□Shabeeb, A. Arora, R. Al□Adamat, Q. B. Pham, N. Al□Ansari, N. T. T. Linh and H. Sajedi, «Flood susceptibility mapping and assessment using a novel deep learning model combining multilayer perceptron and autoencoder neural networks,» Journal of Flood Risk Management, vol. 14, n. 1, March 2021.

[13] H. Mojaddadi Rizeei, B. Pradhan, H. Nampak, N. Ahmad and A. Halim bin Ghazali, «Ensemble machine-learning-based geospatial approach for flood risk assessment using multi-sensor remote-sensing data and GIS,» Geomatics, Natural Hazards and Risk, 1 3 2017.

[14] M. E. Ritter, The physical environment: An introduction to physical geography, 2003.

[15] A. Piloyan and M. Konečný, «Semi-Automated Classification of Landform Elements in Armenia Based on SRTM DEM using K-Means Unsupervised Classification,» Quaestiones Geographicae, vol. 36, n. 1, pp. 93-103, March 2017.

[16] I. D. Moore and G. J. Burch, «Physical Basis of the Length-slope Factor in the Universal Soil Loss Equation,» vol. 50, n. 5, pp. 1294-1298, 1986.

[17] S. J. Riley, S. D. Degloria e R. Elliot, «A Terrain Ruggedness Index that Quantifies Topographic Heterogeneity,» Intermountain Journal of Science, vol. 5, pp. 23-27, 1999.

[18] A. R. Rasyid, N. P. Bhandary and R. Yatabe, «Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia,» Geoenvironmental Disasters, vol. 3, n. 1, p. 19, December 2016.

[19] Soil Science Division Staff, «Soil survey manual,» United States Department of Agriculture, Washington, D.C., 1993.

[20] «Soil Grids,» (Online). Available: https://soilgrids.org/.

Diagnosis and Structural Assessment of the Belice Bridge in Guatemala City

Jorge L. Campo, Rolando E. Torres, Silvia J. De Paz

Abstract**—**"Belice Bridge" is the most important access bridge to Guatemala City by the Atlantic route. This route is the connection between Puerto Santo Tomas de Castilla and Guatemala City. The bridge is made of high yield strength and low alloy structural steel A242, which resists surface corrosion. It has a length of 240 meters, divided into three sections as it follows: the side sections of 72 meters each and the central section of 96 meters. Belice Bridge was inaugurated in 1958. Around 900 vehicles per day cruised it on that year. In the mid 60's, this number exceeded 1,500 vehicles. Nowadays, more than 65,000 vehicles cross it every day. The objective of the present study is to establish the dynamic properties of the bridge and a baseline to measure the probable deterioration or damage in its superstructure in the future. For these purposes, analytical and analytical-experimental methods were used, which allowed determining the dynamic characteristics of such construction. These methodologies were proposed due to the fact that only conventional practices of bridge inspection and visual damage detection had been carried out. In order to find the values of its dynamic characteristics (oscillation modes and vibration frequencies) and have reference values for future measurements, the bridge was networked with accelerometers in nine joints of the armor. These accelerometers were used to record and measure vibrations produced at three different states of charge: rest (only vibrations produced by wind), 18-ton-truck load circulating at 30 and 50 km / h, and finally with normal traffic flow. The results obtained were correlated with the theoretical values of the analysis by the finite element method, and the percentage of variation was determined for the first twelve modes of vibration. The modal parameters of the Frequency Domain Decomposition (FDD) and the Stochastic Identification Subspace (SSI) were also compared to define the Modal Assurance Criteria (MAC), where the theoretical frequency threshold value 0.6578Hz was determined. This value presented to be greater than the experimental value 0.64Hz, which is considered a satisfactory result. The variation of the values of the dynamic parameters is of essential importance to know the structural state since it evidences changes in the characteristics of the constituent materials or the geometric properties of its elements, including changes in boundary conditions and connectivity thereof.

Keywords**—** Diagnosis, field testing, health monitoring, load combinations, vibration modes.

Jorge L. Campo is with the Seismic and Structural Department, Engineering, Mathematics, and Physical Sciences Research Institute (in^3), Mariano Gálvez University of Guatemala, 01002 GUATEMALA (phone: (+502) 24111800 ext. 1358, email: jcampo@umg.edu.gt)

Rolando E. Torres is with the Seismic and Structural Department, Engineering, Mathematics, and Physical Sciences Research Institute (in^3), Mariano Gálvez University of Guatemala, 01002 GUATEMALA (phone: (+502) 24111800 ext. 1358, email: rtorres@umg.edu.gt)

Silvia J. De Paz is with the School of Civil Engineering, Mariano Gálvez University of Guatemala, 01002 GUATEMALA (phone: (+502) 24111800 ext. 1352, email: sdepazm@miumg.edu.gt)

Food Market Policy Priorities in Georgia

Eteri Kharaishvili

***Abstract*—** Introduction: The article assesses the development trends of the Georgian food market. It is demonstrated that the trend of demand for food will be maintained and the development of food markets will play an important role in solving the problem of food security. The paper assesses the impact of the COVID - 19 pandemics on food production and delivery systems.

Methodology: In order to identify similarities, differences and trends various research methods are used in the paper: data selection and grouping, analysis, synthesis, induction and other methods. Using the "Impact-Urgency" matrix we have identified the factors hindering food markets and the export of Georgian agri-food products, using the "Priority Scheme" we have identified the priorities of the economic policy stimulating the export of agri-food products.

Main Findings: Priorities for the formation of the market policy of Georgian agri-food products are: development of market financing policy; Providing access to the location of food markets; Develop programs to stimulate the export of priority agri-food products; Harmonization with European and other food market standards and legislative base; Develop mechanisms for the development of the market design; Creating an information system about food markets, etc.

Conclusions: It is concluded that by using effective food market policy mechanisms, it is possible, on the one hand, to diversify the export of Georgian agri-food products to traditional but growing food markets, and, on the other hand, to expand the geographical area of export markets. Recommendations about policy priorities for the development of local and global food markets for Georgian agri-food products are proposed.

***Keywords*—** food market, agri-food product, market policy, policy priorities.

Eteri Kharaishvili is with the Ivane Javakhishvili Tbilisi State University, Georgia (e-mail: eteri_a@yahoo.com).

Educational Migration and Returnee Employment in Georgia's Labour Market

Manana Lobzhanidze

***Abstract*—** The article analyzes the current trends in the educational migration in the world. It is substantiated that the problem of educational-professional mobility in Georgia is quite topical and is an important challenge for the country in terms of assessing intellectual potential. The importance of research on the problem of educational migration is also conditioned by the fact that the latter is considered as one of the priority directions for the formation and self-realization, the acquisition of independent living, learning and working experience in an unfamiliar environment for young people. The article hypothesizes that the integration of skilled young people into the community after their return will lead to the introduction of new knowledge and technologies, ensure sustainable and inclusive economic growth in the home country, create conditions for approximation to EU standards and effective management of migration processes. The paper examines the factors affecting educational migration and identifies the consequences of the positive impact of these factors on the socio-economic development of the country. Based on the study of youth motivation, value orientation and life strategies in Georgia, the factors hindering educational migration have been identified and recommendations for effective management of educational migration have been developed by assessing their impact on socio-economic outcomes. The article concludes that youth is the most important part of Georgia's labor resources and a strategic resource for socio-economic development. The share of youth in the population significantly changes not only its gender-age structure, but also the professional-qualified composition of labor potential. The paper analyzes youth employment trends and assesses the effects of the pandemic on youth employment. It is concluded that the demand formed in the labor market in Georgia is incompatible with the level of higher and vocational education of young people. Based on the analysis of the educational migration trends of the graduates from Georgian higher education institutions, conclusions and recommendations on the mechanisms for improving the employment of returnees from educational migration are proposed. Systematic and economic-institutional approaches, as well as general and statistical methods (analysis, synthesis, comparison, etc.) are used in the research process to analyze and evaluate presented problems. A qualitative study has been conducted to assess the employment rates of returnees from educational migration in the Georgia's labor market.

***Keywords*—** educational migration, returnee employment, labor market, qualified migrants, migration management.

Manana Lobzhanidze is with the Ivane Javakhishvili Tbilisi State University, Georgia (e-mail: manana.lobzhanidze@tsu.ge).

Opportunities for Effective Conflict Management Caused by Global Crises

Marine Kobalava

***Abstract*—** The article analyzes current global crises in the world, explains the causes of crises, substantiates that in the main cases the process accompanying the crisis are conflict situations.

The paper argues that crises can become predictable if threats are identified and addressed by a company, organization, corporation, and others. Accordingly, mechanisms for the neutralization of conflict potential are proposed, the need to develop a communication strategy and create and redistribute information flows is justified. Conflict situations are assessed according to the types of crisis and it is considered that the conflict can become a prerequisite for the crisis. The paper substantiates the need to differentiate theories of crises and conflicts.

Based on the evaluative judgment, conflict management measures are proposed taking into account institutionalization, conflict resolution norms and rules.

The paper identifies the potential for conflicts created in the context of global crises and suggests local ways and mechanisms for their effective management. The involvement of the company's Public relations (PR) and relevant communication from the qualified staff is considered important.

Conclusions are drawn on the problems of effective conflict management caused by global crises and recommendations for conflict resolution have been proposed.

***Keywords*—** global crises, conflict situations, conflict identification, conflict management, conflict potential.

Marine Kobalava is with the Ivane Javakhishvili Tbilisi State University, Georgia (e-mail: marine.kobalava@tsu.ge).

Economic Policy Promoting Economically Rational Behavior of Start-up Entrepreneurs in Georgia

Gulnaz Erkomaishvili

***Abstract*—** Introduction: The pandemic and the current economic crisis have created problems for entrepreneurship and therefore for start-up entrepreneurs. The paper presents the challenges of start-up entrepreneurs in Georgia in the time of pandemic and the analysis of the state economic policy measures. Despite many problems, the study found that in 54.2% of start-ups surveyed under the pandemic, innovative opportunities were growing. It can be stated that the pandemic was a good opportunity to increase the innovative capacity of the enterprise. 52% of the surveyed start-up entrepreneurs managed to adapt to the current situation and increase the sale of their products/services through remote channels; As for the assessment of state support measures by start-up entrepreneurs, a large number of Georgian startups do not assess the measures implemented by the state positively.

Methodology: The research process uses analysis and synthesis, quantitative and qualitative, interview/survey, groupings, relative and average values, graphic representation, comparison, data analysis and other methods.

Main Findings: Studies have shown that for the start-up entrepreneurs, the main problem remains: Inaccessible fundings, workers qualifications gap, inflation, taxes, regulation, political instability, inadequate provision of infrastructure, amount of taxes and other factors.

Conclusions: To support start-up entrepreneurs, the state should take the following measures: create an attractive environment for investment; availability of low-interest credits, creation of an insurance system; infrastructure development, increase the effectiveness of tax policy (simplicity of the tax system, clarity, optimal level of rates); promote export growth (develop a strategy to penetrate international markets, establish a broad marketing network, etc.).

***Keywords*—** start-up entrepreneurs, startups, start-up entrepreneurs support programs, start-up entrepreneurs support economic policy.

Gulnazi Erkomaishvili is with the Ivane Javakhishvili Tbilisi State University, Georgia (e-mail: guliko_5@yahoo.com).

Innovations and Agricultural Development Potential in Georgia

Tamar Lazariashvili

***Abstract*—** Introduction: The growth and development of the economy in the country depend on many factors, the most important of which is the use of innovation. The article analyzes the innovations and the potential of agricultural development in Georgia, presents the problems in the field, justifies the need to introduce innovations, shows the policy of innovation development, evaluates the positive and negative factors of the use of innovations in agriculture.

Methodology: The article uses general and specific research methods, namely, analysis, synthesis, induction, deduction, comparison and statistical ones: selection, grouping, observation, trend. All these methods used together in the article reveal the main problems and challenges and their development trends.

Main Findings: The introduction of innovations for the country has an impact if there is established state support system for business development and the State creates an effective environment for innovation development. As a result, the appropriate establishment gives incentives to increase budget revenues, create new jobs, increase export turnover and improve the overall economic situation in the country. Georgia has sufficient resource potential to create and develop new businesses in agriculture by introducing innovations and contribute to the further socio-economic development of the country.

Political and economic stability, the existing legislation in the country, infrastructure, the proper functioning of financial institutions and the qualification of the workforce are crucial for the development of innovations. These criteria determine the political and economic ratings of all countries of the world, which are of great importance to foreign investors in the implementation of innovations.

Conclusion: Enactment of agro-insurance will increase the interest and confidence of financial institutions in the farming sector, financial resources will be accessible to the farmers that will facilitate the stable development of the sector in the country. The size of the agro-insurance market in the country should be increased and the new territories should be covered. The State must have an obligation to ensure the risk of farmers and subsidize insurance companies. Based on an analysis of the insurance market the conclusions on agro-insurance issues and the relevant recommendations are proposed.

The introduction of innovations in agriculture will have a great impact on the Georgian economy: it will improve the technological base, establish enterprises equipped with modern equipment and methodologies, retrain existing enterprises, promote to improve skills of workers and improve management systems. Based on the analysis, conclusions are made about the prospects for the development of innovation in agriculture and relevant recommendations are proposed.

***Keywords*—** agriculture, development potential, innovation, optimal environment.

Tamar Lazariashvili is with the Ivane Javakhishvili Tbilisi State University, Georgia (e-mail: tamar.lazariashvili@tsu.ge).

Industrial Policy Directions in Georgia

Nino Grigolaia

***Abstract*—** Introduction - The paper discusses the role of industrial policy in the development of the economy in the country. The main challenges on the way to the implementation of industrial policy are analyzed: the long-term period of industrial policy, the risk of changes in priorities, the limited scope and external shocks.

Methodology - Various research methods are used in the paper. The methods of induction, deduction, analysis, synthesis, analogy, correlation and statistical observation are used.

Main Findings - Based on the analysis of the current situation in Georgia, the obstacles to the country's industrialization and its supporting factors are identified. Also, the challenges of the country's core industrial policies are revealed. Specific industry development strategies, ways of state support and main directions of new industrial policies are identified.

Conclusion - The paper concludes that the development of the industrial sector is critical for the future growth and development of the Georgian economy, which will accelerate the industrialization and structural transformation processes, reduce the trade deficit, increase the exports and create more jobs in the country. The listed changes will guarantee the improvement of the socio-economic situation of the population.

Accordingly, it is revealed that the study of industrial policy in Georgia is still actual. Based on the analysis, relevant conclusions in the field of industrialization of the country are developed and recommendations are proposed.

***Keywords*—** industrialization , industrial policy, industrialization of the economy, Georgia priorities.

Nino Grigolaia is with the Tbilisi Open University, Georgia (e-mail: ninogrigolaia@hotmail.com).

Natural Monopolies and Their Regulation in Georgia

Marina Chavleishvili

***Abstract*—** Today, the study of monopolies, including natural monopolies is topical. In real life, the pure monopolies are natural monopolies. The natural monopolies are used widely and are regulated by the state. In particular, the prices and rates are regulated. The paper considers the problems associated with the operation of natural monopolies in Georgia, in particular, their microeconomic analysis, pricing mechanisms and legal mechanisms of their operation. The analysis was carried out on the example of the power industry.

The rates of natural monopolies in Georgia are controlled by the Georgian National Energy and Water Supply Regulation Commission. The paper analyzes the positive role and importance of the regulatory body and the issues of improving the legislative base what will support the efficient operation of the branch.

As a result of the analysis, certain assessments and conclusions were made:

 Natural monopolies are quite a complex and versatile economic element and it is important to specify and duly control their frame conditions.

 It is important to determine the pricing policy of natural monopolies. The rates should be transparent, should show the level of life in the country and should correspond to the incomes.

 The analysis confirmed the significance of the role of the Antimonopoly Service in the efficient management of the natural monopolies. The law should adapt to the reality and should be applied only to regulate the market.

 The present-day differential electricity tariffs varying depending on the consumed electrical power need revision.

 The effects of the electricity price discrimination are important, segmentation in different seasons in particular. Consumers use more electricity in winter than in summer what is associated with extra capacities and maintenance costs. If the price of electricity in winter is higher than in summer, the electricity consumption will decrease in winter, the consumers will start to consume the electricity more economically what will allow reducing extra capacities.

Thus, the practical realization of the views given in the paper will contribute to the efficient operation of the natural monopolies. Consequently, their activity will be oriented not on the reduction, but on the increase of increments of the consumers or producers. Overall, the optimal management of the given fields will allow improving the wellbeing throughout the country.

***Keywords*—** monopolies, natural monopolies, regulation, Antimonopoly Service.

Marina Chavleishvili is with the Ivane Javakhishvili Tbilisi State University, Georgia (e-mail: marina.chavleishvili@tsu.ge).

Evaluating Factors Impacting Functioning Management Control Systems Becoming Dysfunctional Beyond Intra-Organizational Boundaries

Martin Kartomo

***Abstract*—** Though Management Control Systems (MCS) research has evolved beyond intra-organizational boundaries, there is limited understanding of the impact of a functioning MCS being functional beyond intra-organizational boundaries. The purpose of this research is to investigate factors that have an impact on functioning management Control Systems (MCS)becoming (dys-)functional beyond its intra-organizational boundaries. To bridge the theoretical gap, a systematic literature review is conducted to identify inter-and extra-organizational factors that are purposely suggested or unintendingly mentioned by MCS researchers to evaluate functioning MCS becoming (dys-)functional. A conceptual map is rationalized and constructed from five contingent inter-and extra-organizational MCS frameworks illuminating under-investigated MSC research areas and allowing new research avenues based on academically known factors. A multiple case study followed by a co-researcher discussion group with the purpose of identifying academically unknown factors for evaluating MCS (dys-)functionality beyond its intra-organizational boundaries. The study's result will help bridge the gap between what academics know and not know of evaluating MCS being functional beyond intra-organizational boundaries with the opportunity to develop better, more complete theories. Furthermore, it will help organizations to evaluate the impact of their activities beyond intra-organizational boundaries.

***Keywords*—** management control systems, management control systems evaluation, management controls, control system.

Martin Kartomo is with the European Business University, Netherlands (e-mail: martin_kartomo@yahoo.com).

Disclosure Deregulation of Quarterly Reporting

Jochen Zimmermann,Vanessa Behrmann,Lars Hornuf

Behrmann Vanessa is with the University of Bremen, Germany (e-mail: vanessa.behrmann@uni-bremen.de).

Abstract

In this article, we investigate the deregulation efforts resulting from the revised Transparency Directive and analyze whether a reduction in the minimum content requirements for quarterly reporting increases information asymmetries and decreases firm value. Using a novel dataset of firms that are listed on the Frankfurt Stock Exchange, our results reveal that over the period from 2012 to 2019 lower quarterly reporting levels on average increase information asymmetry and reduce firm value. We find that this effect is stronger for second-tier stocks and firms with a low media coverage. Our results are robust to potential selection effects regarding firms' choice of quarterly reporting content levels.

This version: September 15, 2021

JEL Classification: G14, G32

Keywords: Quarterly Reporting, Disclosure Deregulation, Financial Reporting, Interim Management Statement, Transparency Directive

1. Introduction

The regulatory consensus on disclosure is crumbling. While the academic literature has widely documented beneficial effects of increased disclosures (e.g., Leuz and Verrecchia, 2000; Botosan and Plumlee, 2002; Brown and Hillegeist, 2007; Butler et al., 2007; Daske et al., 2008; Ernstberger et al., 2012; Fu et al., 2012), authorities have begun to retrench disclosure requirements amid criticisms of cost and complexity (Kraft et al., 2018; Kajüter et al., 2019). Recently, investors in the European Union (E.U.) face reduced reporting frequencies, and in Singapore diminished amounts of mandatory information. In this article, we answer the question whether a reduction in the minimum content requirements for quarterly reporting increases information asymmetries and decreases firm value.

Regulators have consistently aimed at improving the attractiveness of stock markets, to which low information asymmetry decisively contributes. Yet, there is virtually no evidence how *deregulation* affects information asymmetries. Fu et al. (2012) addressed voluntary retrenchments of quarterly reporting for firms in the United States (U.S.) in the 1950s and 1960s and find that higher reporting frequency reduces information asymmetry and the cost of equity. More recently, a working paper by Knappstein et al. (2021) has documented that the deregulation of quarterly disclosures relates to higher information asymmetry in the short run. Using an event study method, they do, however, not tackle the question whether information asymmetry persists over time. Our article extends previous findings by investigating the medium to long run effects of deregulation and the type of firms that are particularly affected. More precisely, we investigate how reduced minimum content requirements for quarterly reporting affect information asymmetry, liquidity, and firm value over a period of four years after deregulation.

In our empirical analysis, we consider effects from the transposition of the Transparency Directive Amending Directive (2013/50/EU) into German law, which took place in 2015. The regulatory change resulted in the repeal of mandatory quarterly reporting, and the Frankfurt Stock Exchange (FSE) consequently mandated only descriptive quarterly management statements for firms listed in the Prime Standard. While some firms held on to full quarterly reporting, others reduced the information provided for investors. This is the quasi-natural experiment that we take advantage of in our empirical analysis.[1]

[1] Please note that while firms' *ex-post* choice to adopt a certain level of disclosure is endogenous, the *ex-ante* mandatory quarterly reporting constitutes an exogenous policy treatment, which is why we refer to the empirical setting as quasi-natural experiment. Some firms that would otherwise have chosen a lower level of disclosure where exogenously forced to comply with higher levels of disclosure before the year 2015.

Overall, our findings confirm that fewer disclosure on average increases information asymmetry and reduce firm value. The evidence shows that the increase in information asymmetry is the result of selection effects regarding firms' choice of quarterly reporting content levels, while the reduction in firm value can be traced back to the policy change *per se*. Moreover, the analysis of the firms' information environment reveals substantial heterogeneity. Second tier stocks and firms with a low media coverage display stronger adverse effects for liquidity and firm value. Thus, we extend previous research on disclosure regulation by showing that information asymmetry increases and firm value decreases for second tier stocks and firms with a low media coverage, but not for firms with an already very good information flow. We also add to the understanding of the heterogeneity of regulatory action, namely that quarterly reporting is more relevant for some stocks but not for others: second-tier stocks and stocks with a low media coverage suffer from deregulation, whereas reducing quarterly reporting is irrelevant when the information flow is already high.

The remainder of the paper is structured as follows. In the next section, we provide an overview of the regulation of quarterly financial reporting in Germany. Section 3 reviews the literature and develops our hypotheses. Section 4 describes our data and method. Section 5 reports results of our analysis and section 6 concludes the paper.

2. Institutional Framework

Quarterly reporting has a short tradition in Germany. German firms have begun publishing quarterly reports voluntarily only in the 1990s.[2] In 1998, quarterly reports became a regulatory requirement at the FSE for firms listed in the market segment "Neuer Markt." At this time, quarterly reports were, however, not mandated by law. The Transparency Directive (2004/109/EC) ultimately harmonized the disclosure requirements for firms listed in a regulated market throughout the European Economic Area (EEA). In 2007, the transposition into German law required firms in the regulated market to publish a quarterly Interim Management Statement. The FSE expanded the quarterly reporting requirements to the publication of a full quarterly report for firms listed in the market segment "Prime Standard." Besides an Interim Management Report, a full quarterly report has to include a condensed financial statement in accordance with IAS 34, which consists of a condensed statement of financial positions, including a comprehensive income statement, a cashflow statement, a change in equity statement, and explanatory notes. The Interim Management Report has to be prepared in

[2] For example, Volkswagen AG already published quarterly reports voluntarily in the 1990s.

accordance with German commercial law and contains information on significant opportunities and risks and their effect on the firm's future development as well as information on significant transaction with related parties.

Following criticism by the industry that the quarterly reporting requirements based on the Transparency Directive were too high and constitute a burden for small and medium-sized firms (European Commission, 2013), the revised Transparency Directive (2013/50/EU) and its transposition into national law in 2015 resulted in a significant change in regular financial reporting. Disclosure requirements were drastically reduced by eliminating the obligation to publish an Interim Management Statement. The FSE responded with a deregulation of quarterly reporting and now requires only Interim Management Statements for firms listed in the Prime Standard.

Since 2016, firms listed in the Prime Standard therefore have various options for quarterly financial reports: To fulfill the minimum requirements, a descriptive presentation of the current business situation is sufficient. In addition, firms can voluntarily exceed the minimum requirements by reporting certain interim financial statement elements such as a condensed statement of financial positions, a comprehensive income statement, a cashflow statement, a change in equity statement, and explanatory notes. If all interim financial statement elements are included, the quarterly report constitutes an interim financial statement in accordance with IAS 34. Firms can voluntarily publish a complete quarterly financial report by adding an Interim Management Report to the interim financial statement. Thus, the deregulation has left firms with a wide range of reporting options, ranging from a descriptive quarterly report to a complete quarterly financial report. This deregulation gave rise to major differences in quarterly reports among firms listed in the Prime Standard, and constitutes the variation we use to empirically investigate the capital market effects of firms' disclosure decisions.

3. Hypotheses

Insider information leads to information asymmetries in the capital markets, which results in market inefficiency and, in the worst case, in market failure (Glosten and Milgrom, 1985; Kyle, 1985). When less informed investors have to expect to trade with better informed investors, they are no longer able to make efficient investment decisions (Kim and Verrecchia, 1994; Easley and O'Hara, 2004). As a result, they either exit the market or demand a price premium to compensate for the risk of a possible loss. Thus, adverse selection results in reduced liquidity because if becomes more difficult to trade shares quickly at low costs and with little price impact (Kyle, 1985; Glosten and Milgrom, 1985; Leuz and Verrecchia, 2000; Leuz and

Wysocki, 2016). When the information acquisition of investors is viewed as exogenous, the relevance of financial reporting increases because it can reduce information asymmetries and the resulting adverse selection through the disclosure of information (Diamond, 1985; Bushman, 1991; Lundholm, 1991; La Porta et al., 2000).

The theoretical link between financial reporting and liquidity can also be extended to firms cost of capital and firm value (Amihud et al., 2005). High bid-ask spreads and low illiquidity impose higher trading costs, for which investors demand a compensation in equilibrium, which in turns increases the required return and cost of capital (Amihud and Mendelson, 1986; Garleanu and Pedersen, 2004). The adverse selection problem also transfers to primary markets, because investors are less willing to pay for a security at the time at which the firm issues share, resulting in a lower firm value (Baiman and Verrecchia, 1996; Verrecchia 2001). In addition to the liquidity channel, there is a direct link between disclosure and cost of capital as well as firm value. Because of incomplete information, some investors are not able to consider all firms in the economy, resulting in inefficient and incomplete risk sharing. Increased financial reporting of less known firms enriches a firms' investor base and improve risk sharing in the market, leading to lower cost of capital and higher firm value (Leuz and Wysocki, 2016). There is also a direct link between disclosure and cost of capital arising from estimation risk (Brown, 1979; Barry and Brown, 1984; 1985). As the quality and quantity of firm-specific disclosure increases, uncertainty of future cash-flows decreases thereby moving the cost of capital closer to the risk-free rate and lowering betas (Lambert et al. 2007).

Because disclosure reduces the cost of capital and increases firm value, firms are expected to disclose information voluntarily, especially if the benefits outweigh potential costs arising through disclosure. However, the argument that all firms have incentives to full disclosure no longer holds if the costs of disclosure are sufficiently high (Dye, 1990). Direct and indirect costs of disclosure arise as a result of the preparation and certification of reports as well as through the publication of sensitive information to competitors. This in turn makes the optimal level of disclosure firm specific (Leuz and Wysocki, 2016). Accordingly, a mandatory disclosure rule cannot be justified across the board by stating that the benefits of the disclosure generally outweigh the costs. Arguments justifying the necessity of regulation of financial reporting can be summarized in three theoretical aspects.[3]

First, the existence of externalities in the private production of information can lead to an over- or underproduction of information. By mandating the socially optimal level of disclosure,

[3] For a detailed overview regarding the economics of mandated or regulated disclosure see Leuz and Wysocki, 2008.

regulation can mitigate this problem because investors' incentives to collect and use private information is reduced. In this case, where information acquisition is viewed as endogenous, investors incentives to profit from private information are reduced (Kim and Verrecchia, 1991; McNichols and Trueman, 1994). However, incentives to produce private interim information even exist if interim and annual reports are mandatory, because the timeliness of mandatory reports cannot be great enough to avoid insider information (Mahoney, 1995). Moreover, market forces themselves might limit an overproduction of information because prices tend to reveal at least some of investors' private information, which in turn reduces the incentives to acquire private information (Grossman, 1977; Grossman and Stiglitz, 1980). Thus, the gap between over- and underproduction of information is small, and therefore it is difficult to say whether the resulting externalities are socially desirable.

Second, a mandatory disclosure regime implies a commitment to reveal information in good and bad times (Mahoney, 1995; Rock, 2002). Reduced information asymmetries and less uncertainty are the results (Verrecchia, 2001). Even if firms can voluntarily make such commitments, mandatory disclosure results in lower-cost commitment. In the case of a voluntary commitment, investors anticipate that there is still a residual probability that firms withhold at least some information. Investors thus demand a compensation and raise the demanded return (Leuz and Wysocki, 2008).

Third, producing a sufficient level of disclosure privately can be difficult for firms for many reasons. Corporate insiders can act opportunistically and forego profitable investment decisions in favor of their own private benefit (Shleifer and Wolfenzon, 2002). The mandatory character of regulated disclosure offers access to penalties and remedies, which are often not available or at least limited in private contracts. Thus, costs from fraud and agency conflicts can be mitigated through mandatory disclosure. Moreover, mandatory disclosure reduces social losses because it limits the ability for controlling insiders to waive a disclosure commitment voluntarily in order to consume private benefits (Shleifer and Vishny, 1997; La Porta et al., 2000).

From a theoretical point of view, there are a number of benefits that justify a mandatory disclosure regime. However, indirect and direct costs must be considered, which is why regulation is a trade-off of imperfect alternatives (Djankov et al., 2003; Shleifer, 2005).

Previous research has mostly studied the capital market effects—information asymmetries, cost of capital, and firm value—of different disclosure levels by studying the adoption of international financial reporting standards (IFRS) or the extent of voluntary disclosure (Leuz and Verrecchia, 2000; Botosan, 1997; Botosan and Plumlee, 2002). For the case of quarterly reporting, most empirical studies have examined the capital market effects of *increased*

disclosure (Butler et al., 2007; Fu et al., 2012; Kajüter et al., 2019), and there are only few articles studying the capital market effects of *reduced* levels of financial disclosure (noteworthy exceptions are Fu et al., 2012 and Knappstein et al., 2021). While Fu et al. (2012) evidence that an increase in reporting frequency is associated with lower information asymmetries, they cannot find a significant change in information asymmetries for firms that decreasing their reporting frequency. In a sample of firms listed at the Singapore Stock Exchange (SGX) in 2003, Kajüter et al. (2019) find no relation between the introduction of mandatory quarterly reporting and information asymmetry. Knappstein et al. (2021) find for firms listed at the FSE that reduced quarterly reporting affects information asymmetry as measured by bid-ask-spreads and price impact. However, they only look at a short period between the publication of the quarterly report and the publication of the subsequent half-year report. Brown and Hillegeist (2007) have examined the potential mechanisms by which disclosure affects information asymmetry for firms that have an Association for Investment Management and Research disclosure quality scores for the period from 1986 to 1996. They find a negative effect of disclosure quality on information asymmetry only when private information is reduced in the market.

In line with the ambiguous effects of disclosure on information asymmetries, the effects on cost of equity and firm value are also mixed. The seminal study by Botosan (1997) finds a negative correlation between voluntary disclosure in annual reports and cost of equity of firms with low analyst coverage. Fu et al. (2012) find that an increased reporting frequency leads to lower cost of equity. In contrast, Botosan and Plumlee (2002) show that a voluntary higher reporting frequency, such as quarterly reporting, results in higher cost of equity. Recently, Kajüter at al. (2019) find a negative association between mandatory quarterly reporting and firm value for small firms.

The mixed results of empirical studies indicate that the theoretical debate of capital market effects of higher mandatory disclosure is still unsolved and requires further empirical evidence. The current deregulation trend in quarterly reporting provides another piece of evidence that complements the numerous empirical studies on tightening regulations. Overall, accuracy and quantity of information provided to investors play a decisive role to reduce information asymmetries (Diamond and Verrecchia, 1991; Kim and Verrecchia, 1994; Easley and O'Hara, 2004). Lang and Stice-Lawrence (2015) argue that a lower extend of information in annual reports and a reduced comparability leads to negative economic effects. As a result of the deregulation in 2015, firms can respectively reduce both the accuracy and quantity of information in quarterly reports. We expect firms' quarterly reports to be on average less

accurate, provide less information to investors, and are difficult to compare, requiring an increase in private information in the market. Therefore, we hypothesize:

H1a: Lower quarterly disclosures increase information asymmetry

H1b: Lower quarterly disclosures reduce firm value.

Besides the general effects of lower quarterly disclosure on information asymmetries and firm value, it is unclear how capital market effects of a lower disclosure level depend on the information environment of a firm. Because the quarterly disclosure level is not a comprehensive measure of the overall information environment of a firm, other factors such as analyst-following and firm size are also important factors influencing investors' information set (Botosan, 1997). In theory, capital market effects under a mandatory quarterly reporting regime may by firm specific for the following reason: Mandatory disclosure can be beneficial for unknown or smaller firms with a smaller investor base and an insufficient information environment. Disclosure of these firms improves risk sharing and enables them to increase their investor base, which in turn improves their information environment. Consequently, they get easier access to capital, which in turn reduces cost of capital and increases firm value (Leuz and Wysocki, 2008). Under mandatory disclosure, it is also easier for less-known firms to prove commitment to the capital market and therefore, it is easier to raise capital (Ferrel, 2004; Leuz and Wysocki, 2008). However, the existence of externalities in the production of information makes it difficult to draw a precise conclusion of whether a large investor base reduce or increase private information in the market (Kim and Verrecchia, 1991). The better the information environment of a firm, the less important may be the quarterly reporting level because alternative information channels provide investors with information, which enriches the overall information available in the market (Brown and Hillegeist, 2007).

Under the assumption that larger firms have a better information environment because of greater analyst-following and greater media coverage (Collins et al., 1987; Bhushan, 1989), the marginal effects of lower quarterly reporting levels on information asymmetries and firm valuation are larger for small firms operating in a generally inferior information environment. However, there are also theoretical arguments making quarterly reporting less beneficial or even costly for small firms. Fixed costs for disclosure can make quarterly reporting burdensome for small firms (Leuz and Wysocki, 2016). The shortage of financial analysts implies that information reaches investors in an unfiltered manner, which increases the abilities and incentives of investors to acquire private information, resulting in higher information asymmetries (Lang and Lundholm, 1996). Because of the concentrated ownership structure of

small firms (Demsetz and Lehn, 1985), information in quarterly reports only provide new information for a few investors, because large blockholder typically have timely access to information (Kajüter et al., 2019).

Empirical research suggests a moderating role of firm size regarding the impact of disclosure on capital market effects (Botosan, 1997; Kajüter at al., 2019; Knappstein et al., 2021). Using a self-constructed disclosure index, Botosan (1997) finds a negative effect of disclosure on cost of equity for firms with low analyst-following, making quarterly reporting more important for small firms because of their generally lower amount of analyst-following. Brown and Hillegeist (2007) show that a firm's media coverage is important for the intensity by which the capital market react to information. This is because conference calls and press releases have become more relevant than quarterly reports in recent decades, suggesting that benefits of quarterly reporting found in earlier sample periods, such as in the U.S. from 1951 to 1973 (Fu et al., 2012), may no longer hold in current market environments. While Kajüter et al. (2019) find that large firms in Singapore experience informational benefits of quarterly reporting, they also evidence that quarterly reporting constitutes a net burden for small firms, which reduces firm value. In contrast, Knappstein et al. (2021) find that quarterly reporting is more important for small firms, because the effect of a decrease in quarterly reporting in Germany resulted in higher information asymmetries in the short-term especially for small firms.

Following the assumption that lower quarterly disclosure leads to a worse information provision and increases private information in the market, we expect that for firms operating in an inferior information environment, the marginal effect of quarterly reporting is higher because of the low information flow through alternative information channels. We therefore hypothesize:

H2: The increase in information asymmetry and the reduction in firm value are stronger for firms operating in a poor information environment.

4. Data and Method

4.1 Data

We identified all firms that have been listed in the Prime Standard of the FSE from 2012 to 2019 and consider them in our empirical analysis. For price and accounting information, we use data from Thomson Reuters Datastream.[4] To investigate firms' quarterly reporting level,

[4] We use XETRA as basis for price information.

we hand-collected quarterly reports from firm websites. For our analysis of firms' information environment, we collected yearly information of index listings.[5] Our final sample consists of 361 firms over the period 2012 to 2019 resulting in 2.005 firm-year observations.[6] In addition to index listings, we collect the number of press releases and newswires for each firm from Lexis Nexis to investigate firms' media coverage.[7]

4.2 Variables

4.2.1 Dependent Variables

We test hypothesis 1 by analyzing the effect of lower quarterly disclosures on information asymmetry (H1a) and firm value (H1b). Hypothesis 1a is identified by two common liquidity measures—the bid-ask spread and price impact. In line with Daske et al. (2008) and Ernstberger et al. (2012) we calculate daily relative bid-ask spreads as the mean of daily differences between bid and ask prices divided by their average ((ask-bid)/(ask+bid)/2)).[8] We use the relative spread because total spreads possibly increase as the price of the securities increases (Gros and Wallek, 2015). Our second variable is price impact, which measures illiquidity by the capacity to trade stocks without affecting the price. We follow Daske et al. (2008) and Fu et al. (2012) and calculate price impact as the yearly median of the Amihud (2002) illiquidity measure.

We test hypothesis 1b using two common measures of firm value. Our first measure is the market-to-book ratio (MTB), which is the ratio between the market value of a firm's equity to the book value of its equity. Assuming that lower information risk leads to lower cost of equity and/or higher expectations of future cash inflows, a higher ratio suggests a higher valuation. The second measure is Tobin's Q, which captures the ratio between the market value of a firm to the replacement costs of all assets. A higher ratio captures a higher willingness of investors to provide funds for shares, thereby suggesting a higher firm value.

We test hypothesis 2 using our liquidity and firm valuation measures as dependent variables to identify possible differences of quarterly reporting levels for different information environments.

[5] Historical index listings come from www.dax-indices.de.

[6] Overall, 256 observations were dropped because of missing observations in the dependent or explanatory variables.

[7] Missing observations reduced the sample size to 1.822 observations.

[8] Results remain unchanged when we follow Fu et al. (2012) and regress our raw spread measure on the daily absolute return for each firm year and use the estimated intercept term.

4.2.2 Explanatory Variables

To test hypotheses 1a and 1b, we have classified quarterly reports of firms listed in the FSE Prime Standard in four categories. More precisely, we have generated the variable *QLevel*, which takes the value 1 (low quarterly reporting level) to 4 (high quarterly reporting level). Figure 1 gives an overview of the precise classification of *QLevel*. Prior to 2016, the disclosure of quarterly financial reports was mandatory for all sample firms and quarterly reports prior to 2016 were thus all classified into the highest quarterly reporting level. After mandatory quarterly financial report has been abolished, the content of quarterly reports has been regulated in a more basic form. From 2016 onwards, only a descriptive presentation of the financial position and performance and an explanation of material events and transactions and their effect on the financial position of the business is required. However, firms could freely choose the level of quarterly reporting beyond this minimum requirement and could in principle stick to the highest quarterly reporting level.

[Figure 1 About Here]

To test hypotheses 2, we have measured a firms' information environment using two different variables. First, we examined the additional listing of Prime Standard Firms in a selection index. Based on market capitalization and stock exchange turnover, Prime Standard firms are entitled to be additionally listed in a selection index such as DAX, MDAX, SDAX or TecDAX and thus benefit from higher visibility and public interest. According to Botosan (1997), the listing in a selection index constitutes a valid proxy for the quality of the overall information environment because of the corresponding higher analyst-following. We therefore create the dummy variable *NoIndex*, which is coded as 1 for firms that are not part of an index and 0 otherwise. To identify hypotheses 2, we consider an interaction term of our variable of interest *QLevel* with the dummy variable *NoIndex*. Second, we measure the media coverage with the information flow through press releases and newswires after the abolishment of mandatory quarterly financial reports in 2015, which allows us to test the information environment as outlined in hypothesis 2. We count the sum of press releases and newswires for each firm from 2016 to 2019. We generate the variable *Low_MC*, which equals one if the respective firm operates in a news environment with below average media coverage and zero otherwise. Consequently, we examine the interaction term of *QLevel* and *Low_MC*.

4.2.3 Control Variables

In line with prior research (Leuz and Verrecchia, 2000; Daske et al., 2008; Fu et al., 2012), we use firm size, share turnover, and return volatility as control variables when testing hypothesis 1a. *Size* is the natural logarithm of the average market value of firms' equity for the prior calendar year. Share turnover is calculated as the natural logarithm of the yearly median value of daily share trading volume divided by the market capitalization on that day. Return volatility is the natural logarithm of the standard deviation of daily returns during the year.

We follow Daske et al. (2008) and Meser et al. (2012) and use firm size, leverage, return on asset (*ROA*), and firm growth as control variables when testing hypothesis 1b. In line with Fu et al. (2012), we calculate leverage as the total liabilities divided by the sum of total liabilities and beginning-of-year market value of equity.[9] *ROA* is defined as a firm's net income divided by its total assets. We measure firm growth as the natural logarithm of one plus the percentage change in book value of equity.[10] All of our control variables are winsorized at the top and bottom 1% to account for outliers.[11]

4.3 Method

4.3.1 Baseline Model

As a starting point, we estimate a pooled OLS regression explaining information asymmetry and firm value with quarterly reporting levels. Because the relation between our dependent and explanatory variables might depend on the individual information environment of a firm, we include index fixed-effects. Our basic regression model is as follows:

$$DV_{\textit{Information Asymmetry/Firm Value }it} = \alpha + \beta_1 QLevel_{it} + \beta_2 \mathbf{Controls}_{it} + \varepsilon_{it} \quad (1)$$

where *DV* is the dependent variable measuring information asymmetry or firm value; *QLevel* is our variable of interest measuring the level of quarterly reporting; **Controls** is a vector of control variables.

In a next step, we follow Fu et al. (2012) and control for time-invariant unobserved variation. A Hausman test (1987) indicates that random effects would be inconsistent, which is why we estimate a fixed effects regression with clustered standard errors at the firm level.

[9] Our results remain unchanged if we calculate leverage according to Daske et al. (2008) as ratio of total liabilities to total assets.
[10] Our results remain unchanged if we use asset growth as alternative.
[11] Results remain unchanged if we use unwinzorized values.

4.3.2 Causal Inference

Quarterly financial reporting in the Prime Standard is no longer mandatory after the year 2015 and the observed level of quarterly reporting reflects firms' deliberate choices. Because only those firms might decide to reduce the content level of quarterly financial reporting that benefit the most from it, the pooled and panel regression estimates of *QLevel* potentially suffer from endogeneity in the form of a selection bias. We address this concern in two ways: (1) we implement an instrumental variable (IV) approach and (2) we generate an artificial control group of firms.

For the instrumental variable approach to be successful, we need to find a valid instrument. A valid instrument is strongly correlated with the endogenous variable (*relevance*), but uncorrelated with the error term (*exogeneity*). We follow Fu et al. (2012) and use a Year Index as instrumental variable, which is calculated as the calendar year of the respective quarterly report minus 2019, the year when our sample period ends. Thus, the Year Index is zero for observations from the year 2019 and seven for the year 2012. We assume that the FSE's action to abolish mandatory quarterly financial reporting from 2016 was to some degree unexpected and therefore led to decreasing quarterly reporting levels over time. This implies a positive relation between the quarterly reporting level *QLevel* and the Year Index (*relevance*). At the same time, the year index should not systematically affect liquidity and firm value (*exogeneity*).

While the exogeneity assumption of the instrument cannot be statistically tested, we investigate the first-stage regression to test for the relevance of our instrument. According to Bound et al. (1995), partial R^2 are a useful indicator of the quality of the instrument. The partial R^2 for our information asymmetry model is 39.7% and for our firm value model 39.5%, which indicates that we do not have a weak instrument. In addition to the partial R^2-statistics, we follow Stock and Yogo (2005) and test for the relevance of our instrumental variable by performing a Wald test. Our results lead us to reject the null hypothesis of weak instruments for both our information asymmetry and firm value models. In all cases, the resulting F-statistic are greater than 10, which indicates that the instruments have an influence on the endogenous variable that can be distinguished from zero. Thus, both tests suggest that we do not have a weak instrument problem. We estimate Two Stage Least Squares (2SLS) simultaneous with heteroscedasticity-robust standard errors (Papies et al., 2017).

Because the conjecture of Fu et al. (2012) that the year index does not systematically affect liquidity and firm value cannot be tested, we consider another way to establish causality. Following Fu et al. (2012) and Knappstein et al. (2021), we generate an artificial control group using propensity score matching. The matched control sample compares changes in information

asymmetry and firm value for firms that reduce their quarterly reporting level with otherwise similar firms that did not. Put differently, firms that have not changed their quarterly reporting level serve as control group. We require the matched observations to have the same year-and-industry combination as the treatment observation, which allows us to capture intertemporal changes in industry- and market-wide factors. Furthermore, we match firms that are closest to the treatment observation in terms of firm size, and exactly corresponds to the industry and year of the treatment observation.[12]

In line with Daske et al. (2008) we use a difference-in-difference approach to examine the change in information asymmetry and firm value. We compare differences between treatment and control firms that have not changed their quarterly reporting level, respectively for the period before and after the change in mandatory quarterly reporting levels. In line with Fu et al. (2012), we consider observations three years before and up to three years after the change in quarterly reporting level. Because some firms lower the quarterly reporting level only in 2018 or 2019, the post period is not three years for all of our observation. In particular, we have three post periods for 157 firms that changed their quarterly reporting level in 2016 or 2017. In 2018, another 25 firms changed their quarterly reporting level, implying two post-treatment periods for these firms. Finally, 20 firms changed their quarterly reporting level in 2019, implying one post-treatment period for these firms.[13] We estimate the following difference-in-difference regression:

$$DV = \alpha + \beta_1 Post_{it} + \beta_2 Treat_{it} + \beta_3 Post_{it} \times Treat_{it} + \beta_4 \mathbf{Controls}_{it} + \varepsilon_i \quad (2)$$

in which *DV* is our dependent variable measuring information asymmetry or firm value. *Post* equals one for the periods after the change in quarterly reporting level starting in 2016 and zero otherwise. *Treat* equals one if the firm changed the quarterly reporting level and zero otherwise. The interaction term produces our coefficient of interest, which represents the incremental effect of quarterly reporting on bid-ask spread, price impact, MTB and Tobin's Q that is not the result of time trends, firm size, and industry-factors. **Controls** is a vector of control variables.

[12] Results remain unchanged when we use a kernel matching procedure to identify control group firms. Requiring the same index-year combination for our matching does not affect our results.

[13] The parallel trend assumption is fundamentally not testable in settings with variation in treatment timings (Callaway and Anna, 2020). Therefore, we estimate the mean of our dependent variables bid-ask spread, price impact, MTB und Tobin's Q for the year before the deregulation of the FSE sets in. Means of treatment and non-treatment firms are not significantly different from each other.

4.3.3 Interaction Effects

To investigate hypothesis 2, we interact our variable of interest *QLevel* with the dummy variable *NoIndex*. The interaction between *QLevel* and the dummy variable *No Index* captures the impact of the overall information environment on the effect of information disclosure. We estimate the following regression:

$$DV_{Information\ Asymmetry/Firm\ Value\ it} = \alpha + \beta_1 QLevel_{it} + \beta_2 QLevel_{it} \times NoIndex + \beta_3 \mathbf{Controls}_{it} + \varepsilon_{it} \quad (3)$$

where *DV* is our dependent variable measuring information asymmetry or firm value; *QLevel* and *NoIndex* are the variables of interest and respectively capture the quarterly reporting level and whether firms are not listed in a selection index. **Controls** is a vector of control variables.

We further test hypothesis 2 with the interaction of *QLevel* and *Low_MC,* which captures the impact of capture the quarterly reporting level conditional on the extent of media coverage and press releases. We estimate the following regression:

$$DV_{Information\ Asymmetry/Firm\ Value\ it} = \alpha + \beta_1 QLevel_{it} + \beta_2 QLevel_{it} \times Low_MC + \beta_3 \mathbf{Controls}_{it} + \varepsilon_{it} \quad (4)$$

All regressions include firm fixed-effects and standard errors are clustered at the firm level.

5. Results

5.1 Descriptive Statistics

Table 1 shows the distribution of firms in our sample by the content level of quarterly reports. In the period from 2012 to 2015, when quarterly financial reporting was mandatory for Prime Standard firms, 100% of our sample firms reported on the highest quarterly reporting level. After the year 2015, when quarterly financial reporting was no longer mandatory, the content of quarterly financial reports declined steadily. In 2016, 44% of the Prime Standard firms still published quarterly reports on level four of *QLevel*, while the remaining firms relaxed their reporting quality. The level of quarterly reports persistently continued to decline further over the years, potentially because many firms first wanted to learn about the market effects when peer-firms deviated from the mandatory reporting regime. Finally, in 2019, only 24% of Prime Standard firms voluntarily published a quarterly report on level four of *QLevel*.

[Table 1 About Here]

Table 2 presents descriptive statistics. The mean values of our information asymmetry measures, bid-ask spread and price impact, are 1.00 and 15.04. The values are in line with other studies (Knappstein et al., 2021) and appears reasonable for firms listed in the prime standard. Price impact, however, is high in comparison with older samples and international standards (Daske et al., 2008; Fu et al., 2012). Our measures for firm value, *Tobin's Q* and *MTB*, have a mean value of 1.87 and 3.42. The mean balance sheet size (*Size)* is 5,033,659 EUR.[14] Our control variables *LnVolatility* and *LnTurnover* have a mean value of -4.11 and -6.98, respectively, which is comparable to the sample of Fu et al. (2012). The mean values of *LnLeverage* and *LnGrowth* are -1.02 and 0.04. *ROA* has a mean value of 0.21. All these values are within plausible ranges.

Table 3 reports Spearman and Pearson correlation coefficients for our dependent and explanatory variables. The correlation between the level of quarterly reporting and the variables measuring information asymmetry and firm value imply that information asymmetry is higher and firm value is lower for firms with a high quarterly reporting level, providing first evidence in line with our hypotheses. Our two information asymmetry and the two firm value measures are respectively positively correlated, indicating that they are measuring the same theoretical concepts. In addition, the content level of quarterly reporting is weakly negatively correlated with *LnSize* and *LnLeverage* and weakly positively correlated with *LnTurnover*, *LnVolatility*, *ROA* and *LnGrowth*, indicating that multicollinearity is not a severe problem.

[Table 2-3 About Here]

5.2 Regression Results

Table 4 presents regression results for our information asymmetry measures with bid-ask spreads as dependent variable in Panel A, and price impact as dependent variable in Panel B. The "Pooled" column reports results of the OLS regression with index fixed effects, the "fixed effects" column reports results from the OLS fixed effects panel regression, and "2SLS" reports the results of our IV approach. In Panel A, we find negative coefficients for our variable of interest *QLevel* across all regression models. Its value ranges from -0.025 to -0.038, suggesting that bid-ask spreads increase between 0.025 and 0.038 when quarterly reporting levels decrease by one level. The coefficient of *QLevel* is significant at the 5% and 1% level for our pooled OLS and fixed effects panel regression. In the IV regression the coefficient of *QLevel* is not

[14] To give an intuitive interpretation, we use unlogarithmised values of *Size* for the descriptive statistics.

significant, indicating that the increase in information asymmetry is the result of selection effects regarding firms' choice of quarterly reporting. In line with our expectation that quarterly reporting level affects bid-ask spreads negatively (Fu et al., 2012; Knappstein et al., 2021), these results provide evidence that information asymmetry increase when quarterly reporting level decreases. However, the increase in information asymmetry in the market is not the result of the deregulation *per se*, but stems from the fact that firms, which are *per se* more prone to having problems with information asymmetries, decided to report less. All regressions show statistically significant coefficients for our control variables *LnSize*, *LnTurnover* and *LnVolatility* at the 1% level, implying that they provide a significant contribution to explain bid-ask spreads. The signs of the coefficients are in line with prior literature (Daske et al., 2008; Fu et al., 2012; Ernstberger et al., 2012; Knappstein et al., 2021).

In Panel B, coefficients of *QLevel* in the OLS regressions are negative and not significant in any of the regression models. With regard to bid-ask spreads, we do find some evidence in line with our hypothesis 1a that quarterly reporting levels have a negative effect on price impact (Fu et al. 2012; Knappstein et al., 2021), which however results from the fact that inferior firms decided to report less. All control variables show statistically significant coefficients.

Overall, with the exception of the pooled and fixed effects bid-ask-spread regressions, our results in Table 4 do not indicate that a decrease in quarterly reporting level increase information asymmetry in the long-term, which might indicate that other sources of information potentially replace mandatory quarterly reporting. In contrast with theoretical predictions (Diamond, 1985; Bushman, 1991; Lundholm, 1991; Kim and Verrecchia, 2000), our findings suggest that lower quarterly reporting levels apparently do not change the amount of information provided to investors. Thus, we have to reject hypothesis 1a.

While an increase in information asymmetry is *per se* undesirable, the question remains whether more information asymmetry effectively reduces firm value. Table 5 reports the results of our firm value models with *MTB* as dependent variable in Panel A and *Tobin's Q* as dependent variable in Panel B. In Panel A, we find a positive effect between *MTB* and our variable of interest quarterly reporting level *QLevel* in all regression models. In the preferred IV regression, the coefficient is significant at the conventional 5%-level. As expected, the effect of quarterly reporting level on firm value is positive and a decrease in the quarterly reporting level by one reduces Tobin's Q between 0.986 and 2.095, which corresponds to 0.070 and 0.142 standard deviations. Hence, our results evidence a decrease in firm value if quarterly reporting level decline. Our control variables show robust signs across all regression models and are mostly statistically significant at the 1% and 5% level. While the negative effect of *LnLeverage*

and *LnGrowth* on valuation can be attributed to the possible uncertainty of firms with a high leverage and growth firms, the negative sign of *ROA* is rather unexpected. Using *Tobin's Q* as dependent variable in Panel B, again, our variable of interest *QLevel* is positive and statistically significant across all models. Most importantly, the coefficient of *QLevel* is significant at the 1%-level in the preferred IV regression. Its value ranges between 0.218 and 0.566, indicating that *Tobin's Q* decreases between 0.064 and 0.167 standard deviations if quarterly reporting decline by one. The coefficients of our control variables are mostly statistically significant at the 1% and 5% level and signs are consistent with theoretical predictions. Overall, our results for our firm value models are in line with our theoretical expectation, indicating that a lower level of quarterly reporting after all does not lead to more private information in the market. Therefore, private information apparently does not increase as a result of lower quarterly reporting levels, leading to an increase in the cost of equity and investors demanding a higher risk premium (Daske et al., 2008; Ernstberger et al., 2012). Following these results, we cannot reject hypothesis 1b and find evidence that lower quarterly reporting levels lead to lower firm values.

[Table 4-5 About Here]

5.3 Robustness

To test for the robustness of our results and to consider possible endogeneity concerns, we construct a matched control sample approach. Table 6 Panel A reports results of our information asymmetry measures bid-ask spread and price impact. While we no longer find significant results for the interaction term in the bid-ask spread regression, we now find a significant positive coefficient for the interaction term in the price impact regression, which provides some evidence that the increase in information asymmetry is not exclusively due to selection effects. Thus, firms opting for a lower quarterly reporting level show a higher price impact relative to control group firms, which indicates higher information asymmetries. This result provides evidence in line with hypothesis 1a.

Table 6 Panel B shows the results for our valuation measures *MTB* and *Tobin's Q*. The interaction term is not statistically significant. While this might stem from the fact that there is simply no impact of mandatory reporting levels on firm value, there are at least two further explanations. First, the matching procedure has almost halved the sample size, which negatively affects the power of our statistical analysis. Second, as argued by Fu et al. (2012), market participants perceive an increased reporting as a permanent change, whereas a reduction in reporting is often seen only as a temporal policy measure. Nevertheless, our results provide

some evidence for higher information asymmetries of firms that reduce their quarterly reporting level (hypothesis 1a), but do not provide additional evidence for reduced firm values (hypothesis 1b).

[Table 6 About Here]

5.4 Information Environment

In this section, we more explicitly examine whether private information is an alternative to mandatory disclosure. Table 7 reports results of regressions for our information asymmetry measures as dependent variable in Panel A and firm value measures as dependent variable in Panel B. Our variable of interest identifying the impact of the information environment is the interaction term between the quarterly reporting level (*QLevel)* and no index membership (*NoIndex)*. In Panel A, we find that the coefficient of *QLevel×NoIndex* has a negative value of -0.023 for bis-ask-spreads but is not statistically significant at conventional levels. For our second measure *price impact*, we find a negative effect of quarterly reporting level on *price impact* if firm are not part of an index. The coefficient of *QLevel×NoIndex* has a value of -1.690 and is significant on the 10% level. Overall, there is weak evidence in line to the assumption that non-index firms suffer more from a poor information environment and therefore show stronger effects of lower quarterly reporting on information asymmetry relative to non-index firms.

Panel B shows results for the dependent variables *MTB* and *Tobin's Q*. First, in the regression with *MTB* as dependent variable, we find a weakly significant positive coefficient of 1.218 for the interaction term *QLevel×NoIndex*. Our results suggest that *MTB* decreases stronger for non-index firms if quarterly reporting level decline. Regarding our second measure of firm value, regressions with *Tobin's Q* as dependent variable provide similar results. The coefficient of *QLevel×NoIndex* is positive and weakly significant on the 10% level, suggesting that Tobin's Q decreases more for non-index firms if quarterly reporting level decline. Therefore, we find that second-tier stocks that have a poorer information environment, show a consistent negative effect of quarterly reporting levels on firm value. Although the coefficients are only weakly significant at the 10%-level, the direction of the effect are as predicted and consistent across all specifications. We therefore do not reject hypotheses 2.

[Table 7 About Here]

In a next step, we examine whether private information is an alternative to mandatory disclosure. Table 8 reports results of regressions for our information asymmetry measures as dependent variable in Panel A and firm value measures as dependent variable in Panel B. Our variable of interest identifying the impact of the information environment is the interaction term between the quarterly reporting level (*QLevel)* and the extent of media coverage (*Low_MC)*. In Panel A, we find that the coefficient of *QLevel*×*Low_MC* has a value of -0.032 for *bid-ask-spreads* and is statistically significant on the 5% level. For our second measure *price impact*, we find a negative effect of quarterly reporting level on *price impact* if firms have below average media coverage. The coefficient of *QLevel*×*Low_MC* has a value of -1.059, but is not significant on conventional levels. Overall, in line to the assumption that firms with a worse media coverage suffer more from a poor information environment and therefore show stronger effects of lower quarterly reporting on information asymmetry relative to firms with a high amount of media coverage.

Panel B shows results for the dependent variables *MTB* and *Tobin's Q*. First, in our regression with *MTB* as dependent variable, we find a significant positive coefficient of 1.079 for the interaction term *QLevel*×*Low_MC*. Our results suggest that *MTB* decreases stronger for firms with a low extent of media coverage if quarterly reporting level decline. Regarding our second measure of firm value regressions with *Tobin's Q* as dependent variable find similar results. The coefficient of *QLevel*×*Low_MC* has a value of 0.266 and is positive and significant on the 10% level, suggesting that Tobin's Q decreases more for firms with a low extent of media coverage if quarterly reporting levels decline. Therefore, we find that firms with a low extent of media coverage that have a poorer information environment, show a consistent negative effect of quarterly reporting levels on firm value. We therefore do not reject hypotheses 2.

Overall, we find significant effects for firms operating in a poor information environment. First, the observed capital market effects of quarterly reporting levels on information asymmetries and firm valuation are stronger for non-index firms. Second, we find that firms with below average press coverage show negative effects for reduced levels of quarterly reporting on information asymmetry and firm valuation. Consequently, we cannot reject hypothesis 2 and find that an increase in information asymmetry and the reduction in firm value are stronger for firms operating in a relatively poor information environment.

[Table 8 About Here]

6. Conclusion

The current trend towards deregulation of quarterly reporting worldwide raises again the question about the optimal level of mandatory financial reporting. We address this issue by investigating the deregulation for Prime Standard firms of the FSE in Germany in 2015, in which the removal of mandatory quarterly financial reporting has led to a reduction in quarterly reporting levels of many firms. The evidence shows that information asymmetry increases and firm value decreases when quarterly reporting levels decrease. However, the increase in information asymmetry in the market is not the result of the deregulation as such, but most likely stems from the fact that firms, which are *per se* more prone to having problems with information asymmetries, now that they can decided to report less. The analysis of a matched control sample provides some evidence that information asymmetry also increased because of the deregulation above and beyond these selection effects. Moreover, the reduction in firm value can also be directly attributed to the policy change. In addition, we find that firms with a poor information environment drive these effects, but reducing disclosure might be even irrelevant for firms in a very good information environment. This suggests that while mandatory regulation is beneficial for the average firm but not for all market participants.

Our results have clear limitations. The classification based on the content of quarterly reports is only one possible approach to measure the extent and quality of quarterly reports. Therefore, our results may be affected by other characteristics of quarterly reports, such as their readability and complexity. Nevertheless, we consider our classification based on the content elements contained in quarterly reports to be a good proxy for the information flow resulting from quarterly reports. Moreover, firm specific missing variables such as the readability and complexity of reports will be captured by firm fixed effects in our empirical model.

The usefulness of quarterly reporting is questioned both in Asia on the Singapore Stock Exchange, Europe, and in the U.S (European Commission, 2013; SEC, 2016; Singapore Exchange, 2018). Our findings could therefore be of interest to regulators, exchange authorities, and firms worldwide. While quarterly reporting has previously been assumed to be an additional burden for small firms, it is precisely firms operating in a poor information environment that draw negative effects from lower quarterly reporting levels. From the perspective of investor protection, regular and comprehensive reporting seems to be necessary for firms after all.

References

Amihud, Yakov (2002): Illiquidity and stock returns: cross-section and time series effects. In: *Journal of Financial Markets,* 5:1, 31–56.

Amihud, Yakov; Mendelson, Haim (1986): Asset pricing and the bid-ask spread. In: *Journal of Fianancial Economics,* 84:3, 223–249.

Amihud, Yakov; Mendelson, Haim (1986); The Effects of Beta, Bid-Ask Spread, Residual Risk and Size on Stock Returns. In: *Journal of Finance*, 44:2, 479-486

Amihud, Meldeson, Pedersen, L. (2005): Liquidity and Asset Prices. In: Foundations and Trends in Finance, 1:4, 269-364.

Baiman, Stanley; Verrecchia, Robert E. (1996): The Relation among Capital Markets, Financial Disclosure, Production Efficiency, and Insider Trading. In: *Journal of Accounting Research*, 34:1, 1-22Barry, Christopher B.; Brown, Stephen J. (1984): Differential Information and the Small Firm Effect. In: *Journal of Financial Economics*, 13:2, 283-294

Barry, Christopher B.; Brown Stephen J. (1985): Differential information and security market equilibrium. In: *Journal of Financial and Quantitative Analysis,* 20:4, 407–422.

Bell, Andrew; Jones, Kevin (2015): Explaining Fixed Effects: Random Effects Modeling of Time-Series Cross-Sectional and Panel Data. In: *Political Science Research and Methods*, 3:1; 133–153.

Bertrand, Marianne; Duflo, Esther; Mullainathan, Sendhil (2004): How Much Should We Trust Differences-In-Differences Estimates? In: *The Quarterly Journal of Economics,* 119:1, 249–275.

Bhushan, Ravi (1989): Firm characteristics and analyst following. In: *Journal of Accounting and Economics,* 11:2, 255–274.

Botosan, Christine A. (1997): Disclosure level and the cost of equity capital. In: *The Accounting Review,* 72:3, 323–349.

Botosan, Christine A. (2006): Disclosure and the cost of capital: what do we know? In: *The Accounting Business Research,* 36:1, 31–40.

Botosan, Christine A.; Plumlee, Marlene A. (2002): A Re-examination of Disclosure Level and the Expected Cost of Equity Capital. In: *Journal of Accounting Research,* 40:1, 21–40.

Bound, J.; Jaeger, D. A.; Baker, R. M. (1995): Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. *Journal of the American Statistical Association,* 90:1, 443–450.

Brown, Stephen J. (1979): The Effect of Estimation Risk on Capital Market Equilibrium. In: *Journal of Financial and Quantitative Analysis*, 14:2, 215-220.

Brown, Stephen; Hillegeist, Stephen (2007): How disclosure quality affects the level of information asymmetry. In: *Review of Accounting Studies,* 12:1, 443–477.

Bushman, Robert M. (1991): Public disclosure and the structure of private information markets. In: *Journal of Accounting Research,* 29:2, 261–276.

Butler, Marty; Kraft, Arthur; Weiss, Ira S. (2007): The effect of reporting frequency on the timeliness of earnings: The cases of voluntary and mandatory interim reports. In: *Journal of Accounting and Economics,* 43:2, 181–217.

Callaway, Brantly; Sant'Anna, Pedro H.C. (2020): Difference-in-Differences with multiple time periods. In: *Journal of Econometrics*, forthcoming.

Collins, D. W.; Kothari, S. P., Rayburn, J. D. (1987): Firm size and the information content of prices with respect to earnings. In: *Journal of Accounting and Economics*, 9:2, 111–138.

Daske, Holger; Hail, Luzi; Leuz, Christian; Verdi, Rodrigo (2008): Mandatory IFRS reporting around the world: early evidence on the economic consequences. In: *Journal of Accounting Research,* 46:5, 1085–1142.

Demsetz, Harold; Lehn, Kenneth (1985): The Structure of Corporate Ownership: Causes and Consequences, in: Journal of Political Economy, 93:6; 1155-1177.

Diamond, Douglas W. (1985): Optimal Release of Information by Firms. In: *The Journal of Finance,* 40:1, 1071–1094.

Diamond, Douglas; Verrecchia, Robert E. (1991): Disclosure liquidity and the cost of capital, in: *The Journal of Finance*, 46:4, 1325-1360.

Diamond, Douglas W. (1985): Optimal Release of Information by Firms. In: *Journal of Finance*, 40:4, 1071-1094.

Djankov, Simeon; La Porta, Rafael; Lopez-de-Silanes, Florencio; Shleifer, Andrei (2003): Courts. In: *Quarterly Journal of Economics*, 118:2, 453-517.

Dye, Ronald A. (1990): Mandatory Versus Voluntary Disclosures: The Cases of Financial and Real Externalities. In: *The Accounting Review*, 65:1, 1-24.

Easley, David; O'Hara, Maureen (2004): Information and the cost of capital, in: *The Journal of Finance*, 59:4, 1553–1583.

Ernstberger, Jürgen; Stich, Michael; Vogler, Oliver (2012): Economic consequences of Accounting Enforcement Reforms: The Case of Germany. In: *European Accounting Review,* 21:2, 217–251.

European Commission (2013): Directive 2013/50/EU of the European Parliament and of the Council of 22 October 2013 amending Directive 2004/109/EC of the European Parliament and of the Council on the harmonisation of transparency requirements in relation to information about issuers whose securities are admitted to trading on a regulated market, Directive 2003/71/EC of the European Parliament and of the Council on the prospectus to be published when securities are offered to the public or admitted to trading and Commission Directive 2007/14/EC laying down detailed rules for the implementation of certain provisions of Directive 2004/109/EC. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32013L0050&from=EN

Ferrell, Allen (2004): The Case for Mandatory Disclosure in Securities Regulation Around the World. Harvard Law School Working Paper.

Fu, Renhui; Kraft, Arthur G.; Zhang, Huai (2012): Financial reporting frequency, information asymmetry, and the cost of equity, in: *Journal of Accounting and Economics*, 54:2-3, 132-149.

Gârleanu, Nicolae; Pedersen, Lasse H. (2004): Adverse Selection and the Required Return. In: *Review of Financial Studies*, 17:3, 643-665.

Gigler, Frank; Hemmer, Thomas (1998): On the frequency, quality, and informational role of mandatory financial reports. In: *Journal of Accounting Research,* 36:1, 117–147.

Glosten, Milgrom (1985): Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. In: *Journal of Financial Economics*, 14:1, 71-100.

Greene, William H. (2008): Econometric Analysis. 6th Edition, Pearson Prentice Hall, Upper Saddle River.

Grossman, Sanford J. (1977): The Existence of Futures Markets, Noisy Rational Expectations and Informational Externalities. In: *Review of Economics Studies*, 44:3, 431-449

Grossman, Sanford J.; Stiglitz, Joseph E. (1980): On the Impossibility of Informationally Efficient Markets. In: *American Economic Review*, 70:3, 393-408

Healy, Paul M.; Palepu, Krishna G. (2001): Information asymmetry, corporate disclosure, and the capital markets: A review of the empirical disclosure literature. In: *Journal of Accounting and Economics,* 31:1, 405–440.

Hausman, Jerry A. (1978): Specification Tests in Econometrics. In: *Econometrica*, 46:6, 1251–1271.

Kajüter, Peter; Klassmann, Florian; Nienhaus, Martin (2019): The Effect of Mandatory Quarterly Reporting on Firm Value. In: *The Accounting Review,* 94:3, 251–277.

Kim, Oliver; Verrecchia, Robert E. (1991): Trading Volume and Price Reactions to Public Announcements. In: *Journal of Accounting Research,* 29:2, 302–321.

Kim, Oliver; Verrecchia, Robert E. (1994): Market liquidity and volume around earnings announcements. In: *Journal of Accounting and Economics,* 17:1-2, 41–67.

Knappstein, Janina; Muschalik, Marco; Schmidt, André (2021): Less is more? Decreases in Disclosure and Information Asymmetry. Working Paper: Ruhr-Universität Bochum 2021. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3278236.

Kyle, Albert S. (1985): Continuous Auctions and Insider Trading. In: *Econometrica*, 53:6, 1315–1335.

La Porta, Rafael; Lopez-de-Silantes, Florencio; Shleifer, Andrei; Vishny, Robert (2000): Investor Protection and Corporate Governance. In: *Journal of Financial Economics*, 58:1-2, 3-27.

Lambert, Richard; Leuz, Christian; Verrecchia, Robert E. (2007): Accounting Information, Disclosure, and the Cost of Capital. In: *Journal of Accounting Research*, 45:2, 385-420.

Lang, Mark; L. Stice-Lawrence, Lorien: Textual Analysis and International Financial Reporting: Large Sample Evidence. Working Paper, University of North Carolina, 2015.

Lang, Mark H.; Lundholm, Russell J. (1993): Cross-Sectional Determinants of Analyst Ratings of Corporate Disclosures. In: *Journal of Accounting Research,* 31:2, 246-271.

Lang, Mark H.; Lundholm, Russell J. (1996): Corporate Disclosure Policy and Analyst Behavior. In: *The Accounting Review,* 71:4, 467-492.

Leuz, Christian; Verrecchia, Robert E. (2000): The Economic Consequences of Increased Disclosure. In: *Journal of Accounting Research,* 38:1, 91-124.

Leuz, Christian; Wysocki, Peter (2008): Economic Consequences of Financial Reporting and Disclosure Regulation: A Review and Suggestions for Future Research, Working paper, University of Chicago, University of Miami, 2008.

Leuz, Christian; Wysocki, Peter (2016): The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research, in: *Journal of Accounting Research*, 54:2, 525-622.

Lundholm, Russel J. (1991): Public Signals and the Equilibrium Allocation of Private Information. In: *Journal of Accounting Research,* 29:2, 322–349.

Mahoney, Paul G. (1995): Mandatory Disclosure as a Solution to Agency Problems. In: *The University of Chicago Law Review*, 62:3, 1047-1112.

McNichols, Maureen; Trueman, Brett (1994): Public disclosure, private information collection, and short-term trading. In: *Journal of Accounting and Economics,* 17:1, 69–94.

Meser, Michael; Veith, Stefan; Zimmermann, Jochen (2015): Disclosure; Enforcement, and Capital Market Properties: A Longitudinal Analysis for Germany. In: Schmalenbach Business Review, 67:2, 254-287.

Papies, Dominik; Ebbes, Peter; Van Heerde, Harald J. (2017): Addressing Endogeneity in Marketing Models. In: Advanced Methods for Modeling Markets, 2017, 581-627.

Rock, Edward (2002): Securities Regulation as Lobster Trap: A Credible Commitment Theory of Mandatory Disclosure. In: 23 *Cardozo Law Review*, 23:2, 675-704.

Securities and Exchange Commission (SEC) (2016): Business and Financial Disclosure Required by Regulation S-K. Release No. 33-10064. Washington, DC: SEC.

Shleifer, Andrei (2005): Understanding Regulation. In: *European Financial Management*, 11:4, 439-451.

Shleifer, Andrei; Vishny Robert W. (1997): A Survey of Corporate Governance. In: *Journal of Finance*, 52:2, 737-783.

Shleifer, Andrei; Wolfenzon, Daniel (2002): Investor Protection and Equity Markets. In: *Journal of Financial Economics*, 66:1, 3-27.

Singapore Exchange (2018): Consultation Paper: Review of the Quarterly Reporting Framework, 11 January 2018. Available at: https://www.cfasociety.org/singapore/Lists/MAS%20Consultation%20Paper/Attachments/13/Consultation%20Paper%20on%20Quarterly%20Reporting%20Framework.pdf.

Stock, J.; Yogo, M. (2005): Testing for weak instruments in linear IV regression. In *Identification and Inference for Econometric Models*: Essays in Honor of Thomas Rothenberg, ed. D. W. K. Andrews and J. H. Stock, 80–108. New York: Cambridge University Press.

Van Buskirk, Andrew (2012): Disclosure frequency and information asymmetry. In: *Review of Quantitative Finance and Accounting,* 38:1, 411–440.

Verrecchia, Robert E. (2001): Essays on Disclosure. In: *Journal of Accounting and Economics*, 32:1-3, 97-180.

Figure 1: Classification of Quarterly Reports

Definition of Quarterly Report	Content Level	Content Elements
Quarterly Financial Report	QLevel 4	Interim Financial Statement in accordance with IAS 34 and Interim Management Report
Quarterly Report	QLevel 3	Interim Financial Statement in accordance with IAS 34
Quantitative Interim Management Statement	QLevel 2	At least a condensed statement of financial positions, of comprehensive income and of cashflows
Descriptive Interim Management Statement	QLevel 1	Less than a condensed statement of financial positions, of comprehensive income and of cashflows

Classification of Quarterly Reports were hand-collected by analyzing firms quarterly reports from 2016-2019. Quarterly reports were obtained from firm websites. Prior to 2016, the disclosure of quarterly financial reports (QLevel 4) was mandatory for our sample firms.

Table 1: Quarterly Reporting of Prime Standard Firms

year	*N*	*QLevel=1 (%)*	*QLevel=2 (%)*	*QLevel=3 (%)*	*QLevel=4 (%)*	*Mean QLevel*
2012-2015	1.005	0.00%	0.00%	0.00%	100.00%	4.00
2016	239	8.37%	40.17%	7.11%	44.35%	2.87
2017	239	13.81%	44.77%	7.11%	34.73%	2.61
2018	257	15.18%	52.53%	3.89%	28.40%	2.45
2019	265	18.49%	53.96%	3.40%	24.15%	2.32

The sample includes 2.005 firm-year observations for the period 2012-2019. Quarterly reporting level data were hand-collected from firm websites. *QLevel*=1(%) indicates the percentage of firms with a descriptive quarterly report. *QLevel*=2(%) indicates the percentage of firms publishing a quarterly report with at least a condensed statement of financial positions, of comprehensive income and of cashflows. *QLevel*=3(%) indicates the percentage of firms publishing a quarterly report according to IAS 34. *QLevel*=4(%) indicates the percentage of firms publishing a quarterly report according to IAS 34 and a management report according to German Commercial Law.

Table 2: Descriptive Statistics

	N	*Mean*	*Std. Dev.*	*P25*	*P50*	*P75*
SPREAD	2.005	1.0019	1.4685	0.1906	0.5244	1.4016
PI	2.005	15.0368	191.5106	0.0183	0.2762	4.1536
MTB	2.005	3.4174	14.7238	1.1188	1.7928	3.0283
TobinsQ	2.005	1.8721	3.3804	1.0352	1.2832	1.8165
Size	2.005	5,033,659	13,400,000	122,744.7	457,292.3	2,391,526
LnTurnover	2.005	-6.9826	1.3620	-7.8058	-6.7855	-5.9747
LnVolatility	2.005	-4.1126	0.3026	-4.3108	-4.0904	-3.8914
LnLeverage	2.005	-1.0222	0.7817	-1.3737	-0.8258	-0.4578
ROA	2.005	0.2145	0.1181	0.0072	0.0369	0.0678
LnGrowth	2.005	0.0437	0.2672	-0.0171	0.0539	0.1235

The sample includes 2.005 firm-year observations for the period 2012-2019. All variables are based on a calendar year. IA_{Spread} is daily bid-ask spreads calculated as the mean of daily differences between bid and ask prices divided by the mean of daily bid-ask spreads. $IA_{Price\ Impact}$ is price impact calculated as the yearly median of the Amihud (2002) illiquidity measure. V_{MTB} is market-to-book ratio, calculated as ratio between the market value of a firm's equity and the book value of its equity. $V_{Tobin's\ Q}$ is calculated as the ratio between the market value of a company and the replacement costs of all assets. *Size* is the average market value of equity at the beginning and end of the prior calendar year. *LnTurnover* is computed as the log of the median of the euro value of all shares traded during the day divided by the market capitalization on that day. *LnVolatility* is the log of the standard deviation of daily returns during the year. *LnLeverage* is total liabilities divided by sum of total liabilities and beginning-of-year market value of equity. *ROA* is return on assets. *LnGrowth* is the log of one plus the percentage change in book value of equity. The values of IA_{Spread}, $IA_{Price\ Impact}$ are multiplied by 100 for expositional purpose.

Table 3: Pearson (Upper Diagonal) and Spearman (Lower Diagonal) Correlation Coefficients

	[1]	*[2]*	*[3]*	*[4]*	*[5]*	*[6]*	*[7]*	*[8]*	*[9]*	*[10]*	*[11]*
QLevel [1]	1	0.0551	0.0278	0.0037	-0.0279	-0.0843	0.0247	0.0861	0.1082	0.0398	0.0056
IA Spread [2]	0.0286	1	0.7256	0.0582	0.0179	-0.5537	-0.5101	0.2998	-0.0253	-0.2305	-0.1457
IA PI [3]	0.0616	0.9822	1	0.0348	-0.0127	-0.1151	-0.1150	0.0887	0.0421	-0.0750	-0.0880
V MTB [4]	-0.1382	-0.0651	-0.0953	1	0.8770	0.0691	-0.1704	0.0824	-0.2691	-0.1471	-0.1644
V TobinsQ [5]	-0.1328	-0.0389	-0.0616	0.9397	1	0.0803	-0.1926	0.0874	-0.4686	-0.1337	-0.0998
LnSize [6]	-0.0989	-0.8814	-0.8954	0.1375	0.0853	1	0.3778	-0.3993	0.0934	0.1500	0.0451
LnTurnover [7]	0.0370	-0.7338	-0.7509	-0.0320	-0.0320	0.4630	1	0.0795	0.1430	0.0772	0.0826
LnVolatility [8]	0.0678	0.3907	0.3543	0.0175	0.0353	-0.3899	0.0550	1	-0.1320	-0.2855	-0.1266
LnLeverage [9]	-0.0587	-0.1531	-0.1481	-0.5751	-0.7458	0.1164	0.1582	-0.1179	1	0,0389	0.0585
ROA [10]	0.0379	-0.1618	-0.1502	0.1951	0.2863	0.0866	0.0508	-0.2281	-0.3594	1	0.4663
LnGrowth [11]	0.0240	-0.1151	-0.1253	0.0284	0.0577	0.0500	0.1011	-0.0822	-0.0414	0.5235	1

The sample includes 2.005 firm-year observations for the period 2012-2019. All variables are based on a calendar year. *QLevel* is the level of firms quarterly reports (based on hand-collected data from firms quarterly reports). IA_{Spread} is daily bid-ask spreads calculated as the mean of daily differences between bid and ask prices divided by the mean of daily bid-ask spreads. $IA_{Price\ Impact}$ is price impact calculated as the yearly median of the Amihud (2002) illiquidity measure. V_{MTB} is market-to-book ratio, calculated as ratio between the market value of a firm's equity and the book value of its equity. $V_{Tobin's\ Q}$ is calculated as the ratio between the market value of a company and the replacement costs of all assets. *LnSize* is the natural log of the average market value of equity at the beginning and end of the prior calendar year. *LnTurnover* is computed as the natural log of the median of the euro value of all shares traded during the day divided by the market capitalization on that day. *LnVolatility* is the natural log of the standard deviation of daily returns during the year. *LnLeverage* is the natural log of total liabilities divided by sum of total liabilities and beginning-of-year market value of equity. *ROA* is return on assets. *LnGrowth* is the natural log of one plus the percentage change in book value of equity. The values of IA_{Spread}, $IA_{Price\ Impact}$ are multiplied by 100 for expositional purpose.

Table 4: Information Asymmetry Models

Panel A: Bid-Ask-Spread as dependent variable

$IA_{Spread} = \alpha+\beta_1 QLevel_{it}+\beta_2 Size_{it}+\beta_3 LnVolatility_{it}+\beta_4 LnTurnover_{it}+\varepsilon_i$

Variable	*Pooled*	*Fixed Effects*	*2SLS*
QLevel	-0.025**	-0.038***	-0.037
	(-1.96)	(-2.67)	(-1.50)
LnSize	-0.438***	-0.316***	-0.441***
	(-10.12)	(-10.61)	(-10.19)
LnTurnover	-0.551***	-0.395***	-0.552***
	(9.00)	(-10.65)	(-9.03)
LnVolatility	1.147***	0.695***	1.150***
	(7.09)	(6.63)	(7.13)
Fixed Effects	*Index*	*Firm*	*Index*
R^2 *(overall)*	0.506	0.436	0.506
R^2 *(within)*		0.234	
R^2 *(between)*		0.373	
F-statistic/Wald Chi^2	32.95	49.60	2659.16
N	2005	2005	2005

Panel B: Price Impact as dependent variable

$IA_{Price\ Impact} = \alpha + \beta_1 QLevel_{it} + \beta_2 Size_{it} + \beta_3 LnVolatility_{it} + \beta_4 LnTurnover_{it} + \varepsilon_{it}$

Variable	*Pooled*	*Fixed Effects*	*2SLS*
QLevel	-0.106	-1.131	1.054
	(-0.09)	(-1.49)	(0.44)
LnSize	-20.714**	-7.198***	-20.514**
	(-2.39)	(-4.58)	(-2.39)
LnTurnover	-26.646**	-13.259***	-26.597**
	(-2.12)	(-2.94)	(-2.12)
LnVolatility	63.731*	32.782**	63.450*
	(1.92)	(1.96)	(1.92)
Fixed Effects	*Index*	*Firm*	*Index*
R^2 *(overall)*	0.039	0.023	0.039
R^2 *(within)*		0.012	
R^2 *(between)*		0.036	
F-statistic/Wald Chi^2	1.84	6.99	61.33
N	2005	2005	2005

The sample includes 2.005 firm-year observations for the period 2012-2019. Information asymmetry is measured by IA_{Spread} or $IA_{Price\ Impact}$ respectively. Panel A and Panel B report results when the dependent variable is *MTB* and *Tobin's Q*, respectively. The "*Pooled*" column reports the results based on OLS regressions. The "*Fixed Effects*" column reports the results based on OLS regression with firm fixed effects. The "*2SLS*" column reports results based on simultaneous estimated two stage least square approach for panel data with firm fixed effects. All variables are based on a calendar year. *Qlevel* is the level of firms' quarterly reports (based on hand-collected data from firms' quarterly reports). IA_{Spread} is daily bid-ask spreads calculated as the mean of daily differences between bid and ask prices divided by the mean of daily bid-ask spreads. $IA_{Price\ Impact}$ is price impact calculated as the yearly median of the Amihud (2002) illiquidity measure. *LnSize* is the natural log of the average market value of equity at the beginning and end of the prior calendar year. *LnTurnover* is computed as the natural log of the median of the euro value of all shares traded during the day divided by the market capitalization on that day. *LnVolatility* is the natural log of the standard deviation of daily returns during the year. The values of IA_{Spread}, $IA_{Price\ Impact}$ are multiplied by 100 for expositional purpose. The table reports OLS, Fixed Effects and 2SLS estimates and (in parentheses) t-statistics based on robust standard errors for the OLS and 2SLS estimates and clustered standard errors on the firm level for Fixed Effects estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 5: Firm Value Models

Panel A: Market-to-Book ratio as dependent variable

$V_{MTB} = \alpha+\beta_1 QLevel_{it}+\beta_2 Size_{it}+\beta_3 LnTurnover_{it}+\beta_4 ROA_{it}+\beta_5 LnGrowth_{it}+\varepsilon_i$

Variable	*Pooled*	*Fixed Effects*	*2SLS*
QLevel	1.016**	0.986*	2.095**
	(1.97)	(1.83)	(2.33)
LnLeverage	-5.033***	-9.800	-5.163***
	(-3.84)	(-1.59)	(-3.84)
ROA	-11.492**	-22.106	-12.070**
	(-2.06)	(-1.03)	(-2.11)
LnSize	2.500***	2.324	2.689***
	(2.51)	(1.54)	(2.54)
LnGrowth	-6.576**	-5.379***	-6.453**
	(-1.94)	(-3.39)	(-1.94)
Fixed Effects	*Index*	*Firm*	*Index*
R^2 *(overall)*	0.143	0.110	0.138
R^2 *(within)*		0.132	
R^2 *(between)*		0.112	
F-statistic or Wald Chi^2	10.25	3.12	55.05
N	2005	2005	2005

Panel B: Tobin's Q as dependent variable

$V_{TobinsQ} = \alpha+\beta_1 QLevel_{it}+\beta_2 Size_{it}+\beta_3 LnTurnover_{it}+\beta_4 ROA_{it}+\beta_5 LnGrowth_{it}+\varepsilon_i$

Variable	*Pooled*	*Fixed Effects*	*2SLS*
QLevel	0.218**	0.247*	0.566***
	(2.09)	(1.76)	(2.74)
LnLeverage	-2.047***	-2.995**	-2.089***
	(-7.72)	(-2.18)	(-7.61)
ROA	-3.031***	-4.054	-3.218***
	(-2.96)	(-1.10)	(-3.08)
LnSize	0.647***	0.646	0.708***
	(3.14)	(1.49)	(3.17)
LnGrowth	-0.653	-0.423	-0.614
	(-1.32)	(-1.42)	(-1.27)
Fixed Effects	*Index*	*Firm*	*Index*
R^2 *(overall)*	0.286	0.240	0.275
R^2 *(within)*		0.242	
R^2 *(between)*		0.372	
F-statistic or Wald Chi^2	46.37	3.31	246.67
N	2005	2005	2005

The sample includes 2.005 firm-year observations for the period 2012-2019. Valuation is measured by V_{MTB} or $V_{Tobin's\ Q}$ respectively. Panel A and Panel B report results when the dependent variable is *MTB* and *Tobin's Q*, respectively. The "*Pooled*" column reports the results based on OLS regressions. The "*Fixed Effects*" column reports the results based on OLS regression with firm fixed effects. The "*2SLS*" column reports results based on simultaneous estimated two stage least square approach for panel data. All variables are based on a calendar year. *Qlevel* is the level of firms' quarterly reports (based on hand-collected data from firms' quarterly reports). V_{MTB} is market-to-book ratio, calculated as ratio between the market value of a firm's equity and the book value of its equity. $V_{Tobin's\ Q}$ is calculated as the ratio between the market value of a company and the replacement costs of all assets. *LnSize* is the natural log of the average market value of equity at the beginning and end of the prior calendar year. *LnLeverage* is the natural log of total liabilities divided by sum of total liabilities and beginning-of-year market value of equity. *ROA* is return on assets. *LnGrowth* is the natural log of one plus the percentage change in book value of equity. The table reports OLS, Fixed Effects and 2SLS estimates and (in parantheses) t-statistics based on robust standard errors for the OLS and 2SLS estimates and clustered standard errors on the firm level for Fixed Effects estimates. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 6: Matching Model

Panel A: Information Asymmetry Measures

$IA_{Spread,\ Price\ Impact} = \alpha+\beta_1 Post_{it}+\beta_2 Treat_{it}+\beta_3 Post \times Treat_{it}+\beta_4 Size_{it}+\beta_5 LnVolatility_{it}+\beta_6 LnTurnover_{it}+\varepsilon_i$

Variable	IA_{Spread}	IA_{PI}
Post	-0.029	-5.785*
	(-0.33)	(-1.19)
Treat	-0.066	-6.682**
	(-0.85)	(-1.99)
Post × Treat	0.065	7.369**
	(0.70)	(2.03)
LnSize	-0.196***	-2.426***
	(-8.56)	(-3.41)
LnTurnover	-0.358***	-5.419***
	(-10.39)	(-4.51)
LnVolatility	0.762***	16.597***
	(6.40)	(3.38)
R^2	0.689	0.184
F-statistics	44.91	7.45
N	1269	1269

Panel B: Valuation Measures

$V_{MTB,\ TobinsQ} = \alpha+\beta_1 Post_{it}+\beta_2 Treat_{it}+\beta_3 Post \times Treat_{it}+\beta_4 Size_{it}+\beta_5 LnLeverage_{it}+\beta_6 ROA_{it}+\beta_7 LnGrowth_{it}+\varepsilon_i$

Variable	V_{MTB}	$V_{TobinsQ}$
Post	-5.027	-0.939
	(-1.17)	(-1.20)
Treat	-4.292	-0.611
	(-0.93)	(-0.73)
Post × Treat	2.968	0.502
	(0.91)	(0.83)
LnSize	2.022	0.428*
	(1.37)	(1.67)
LnLeverage	-9.998*	-2.781*
	(-1.68)	(-2.60)
ROA	-30.625	-5.378
	(-1.28)	(-1.55)
LnGrowth	-20.789	-2.712
	(-1.32)	(-1.20)
R^2	0.174	0.266
F-statistics	1.24	5.34
N	1269	1269

The sample includes 1.335 observations for treatment firms and control firms 3 years before and after the event year during the period 2013-2019. Control firms are matched on industry, size, and year. The results are based on the model $DV=\alpha+\beta_1 Post_{it}+\beta_2 Treat_{it}+\beta_3 Post\times Treat_{it}+\beta_4 CV_{it}+\varepsilon_i$. *DV* are the information asymmetry or valuation measures. *Treat* is the dummy variable coded as 1, for firms that change their quarterly reporting level and 0 for control firms; After is the dummy variable coded as 1 for three years after the change in quarterly reporting level, and 0 for three years before the change. *CV* are the control variables. Information asymmetry is measured by IA_{Spread} or $IA_{Price\ Impact}$. Valuation is measured by V_{MTB} or $V_{Tobin's\ Q}$. Panel A and Panel B report results for our information asymmetry measures and valuation measures, respectively. IA_{Spread} is daily bid-ask-spreads calculated as the mean of daily differences between bid and ask prices divided by the mean of daily bid-ask-spreads. $IA_{Price\ Impact}$ is price impact calculated as the yearly median of the Amihud (2002) illiquidity measure. V_{MTB} is market-to-book ratio, calculated as ratio between the market value of a firm's equity and the book value of its equity. $V_{Tobin's\ Q}$ is calculated as the ratio between the market value of a company and the replacement costs of all assets. *LnSize* is the natural log of the average market value of equity at the beginning and end of the prior calendar year. *LnTurnover* is computed as the natural log of the median of the euro value of all shares traded during the day divided by the market capitalization on that day. *LnVolatility* is the natural log of the standard deviation of daily returns during the year. *LnLeverage* is the natural log of total liabilities divided by sum of total liabilities and beginning-of-year market value of equity. *ROA* is return on assets. *LnGrowth* is the natural log of one plus the percentage change in book value of equity. The values of IA_{Spread}, $IA_{Price\ Impact}$ are multiplied by 100 for expositional purpose. The table reports OLS estimates and (in parentheses) t-statistics based on clustered standard errors on the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 7: Interaction Effect of Index Membership

Panel A: Information Asymmetry Measures

$IA_{Spread,\ Price\ Impact} = \alpha + \beta_1 QLevel_{it} + \beta_2\ QLevel \times Index + \beta_3 Size_{it} + \beta_4 LnVolatility_{it} + \beta_5 LnTurnover_{it} + \varepsilon_i$

Variable	IA_{Spread}	$IA_{Price\ Impact}$
QLevel	-0.049***	-0.710
	(-3.97)	(-1.18)
QLevel × *NoIndex*	-0.023	-1.690*
	(-1.23)	(-1.63)
LnSize	-0.310***	-7.451***
	(-10.24)	(-4.49)
LnTurnover	-0.388***	-13.567***
	(-10.08)	(-2.93)
LnVolatility	0.696***	32.71*
	(6.66)	(1.96)
Fixed Effects	*Firm*	*Firm*
R^2 *(overall)*	0.437	0.023
R^2 *(within)*	0.236	0.013
R^2 *(between)*	0.372	0.037
F-statistics	50.60	15.49
N	2005	2005

Panel B: Valuation Measures

$$V_{MTB,\ Tobin's\ Q} = \alpha + \beta_1 QLevel_{it} + \beta_2 QLevel \times NoIndex + \beta_3 Size_{it} + \beta_4 LnTurnover_{it} + \beta_5 ROA_{it} + \beta_6 LnGrowth_{it} + \varepsilon_i$$

Variable	V_{MTB}	$V_{Tobin's\ Q}$
QLevel	0.816*	0.202*
	(1.85)	(1.81)
QLevel × *NoIndex*	1.218*	0.307*
	(1.79)	(1.72)
LnLeverage	-9.862	-3.011**
	(-1.60)	(-2.18)
ROA	-21.688	-3.945
	(-1.02)	(-1.09)
LnSize	2.385	0.662
	(1.56)	(1.51)
LnGrowth	-5.382***	-0.423
	(-3.38)	(-1.42)
Fixed Effects	*Firm*	*Firm*
R^2 *(overall)*	0.116	0.249
R^2 *(within)*	0.133	0.244
R^2 *(between)*	0.118	0.258
F-statistics	2.64	3.06
N	2005	2005

The sample includes 2.005 firm-year observations for the period 2012-2019. Firm Value is measured by V_{MTB} or $V_{Tobin's\ Q}$ respectively. Panel A and Panel B report results for our information asymmetry measures and valuation measures, respectively. *QLevel* is the level of firms' quarterly reports (based on hand-collected data from firms' quarterly reports). The variable *NoIndex* is coded as one for firms without Index listing. All variables are based on a calendar year. V_{MTB} is market-to-book ratio, calculated as ratio between the market value of a firm's equity and the book value of its equity. $V_{Tobin's\ Q}$ is calculated as the ratio between the market value of a company and the replacement costs of all assets. *LnSize* is the natural log of the average market value of equity at the beginning and end of the prior calendar year. *LnTurnover* is computed as the natural log of the median of the euro value of all shares traded during the day divided by the market capitalization on that day. *LnVolatility* is the natural log of the standard deviation of daily returns during the year. *LnLeverage* is the natural log of total liabilities divided by sum of total liabilities and beginning-of-year market value of equity. *ROA* is return on assets. *LnGrowth* is the natural log of one plus the percentage change in book value of equity. The table reports Firm Fixed Effects estimates and (in parentheses) t-statistics based on clustered standard errors on the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

Table 8: Interaction Effect of Media Coverage

Panel A: Information Asymmetry Measures

$$IA_{Spread,\ Price\ Impact} = \alpha + \beta_1 QLevel_{it} + \beta_2\ QLevel \times Low_MC + \beta_3 Size_{it} + \beta_4 LnVolatility_{it} + \beta_5 LnTurnover_{it} + \varepsilon_i$$

Variable	IA_{Spread}	$IA_{Price\ Impact}$
QLevel	-0.069**	-1.534**
	(-2.25)	(-2.03)
QLevel × *Low_MC*	-0.032**	-1.059
	(-2.14)	(-1.27)
LnSize	-0.318***	-7.216***
	(-10.65)	(-4.60)
LnTurnover	-0.394***	-13.249***
	(-10.65)	(-2.94)
LnVolatility	0.696***	32.794**
	(6.64)	(1.96)
Fixed Effects	*Firm*	*Firm*
R^2 *(overall)*	0.425	0.012
R^2 *(within)*	0.234	0.013
R^2 *(between)*	0.360	0.035
F-statistics	40.37	5.70
N	2005	2005

Panel B: Valuation Measures

$$V_{MTB,\ Tobin's\ Q} = \alpha + \beta_1 QLevel_{it} + \beta_2 QLevel \times Low_MC + \beta_3 Size_{it} + \beta_4 LnTurnover_{it} + \beta_5 ROA_{it} + \beta_6 LnGrowth_{it} + \varepsilon_i$$

Variable	V_{MTB}	$V_{Tobin's\ Q}$
QLevel	0.465	0.137
	(0.84)	(0.84)
QLevel × *Low_MC*	1.079*	2.663*
	(1.91)	(1.86)
LnLeverage	-9.838	-3.003**
	(-1.60)	(-2.19)
ROA	-22.307	-4.096
	(-1.02)	(-1.12)
LnSize	2.286	0.638
	(1.56)	(1.47)
LnGrowth	-5.364***	-0.420
	(-3.39)	(-1.42)
Fixed Effects	*Firm*	*Firm*
R^2 *(overall)*	0.113	0.245
R^2 *(within)*	0.133	0.242
R^2 *(between)*	0.115	0.256
F-statistics	2.66	2.82
N	2005	2005

The sample includes 2.005 firm-year observations for the period 2012-2019. Firm Value is measured by V_{MTB} or $V_{Tobin's\ Q}$ respectively. Panel A and Panel B report results for our information asymmetry measures and valuation measures, respectively. *QLevel* is the level of firms' quarterly reports (based on hand-collected data from firms' quarterly reports). The variable *Low_MC* is coded as one for firms with an extent of press releases below the average. All variables are based on a calendar year. V_{MTB} is market-to-book ratio, calculated as ratio between the market value of a firm's equity and the book value of its equity. $V_{Tobin's\ Q}$ is calculated as the ratio between the market value of a company and the replacement costs of all assets. *LnSize* is the natural log of the average market value of equity at the beginning and end of the prior calendar year. *LnTurnover* is computed as the natural log of the median of the euro value of all shares traded during the day divided by the market capitalization on that day. *LnVolatility* is the natural log of the standard deviation of daily returns during the year. *LnLeverage* is the natural log of total liabilities divided by sum of total liabilities and beginning-of-year market value of equity. *ROA* is return on assets. *LnGrowth* is the natural log of one plus the percentage change in book value of equity. The table reports Firm Fixed Effects estimates and (in parentheses) t-statistics based on clustered standard errors on the firm level. ***, **, and * indicate statistical significance at the 1%, 5%, and 10% levels (two-tailed), respectively.

The disposable identities; enabling trust-by-design to build sustainable data driven value

1st Lorna Goulden
0000-0003-4087-124
Twinds Foundation
Herent, Belgium

2nd Kai M. Hermsen
0000-0002-7084-0668
Twinds Foundation
Herent, Belgium

3rd Jari Isohanni
0000-0002-7154-2515
Twinds Foundation
Herent, Belgium

4th Mirko Ross
Twinds Foundation
Herent, Belgium

5th Jef Vanbockryck
0000-0002-1323-2690
Twinds Foundation
Herent, Belgium

Abstract—This article introduces the Disposable Identities, with reference use cases and explores possible technical approaches. Proposed approach, when fully developed as an open source toolkit, enables developers of mobile or web apps to employ a self-sovereign identity and data privacy framework, in order to rebuild trust in digital services by providing greater transparency, decentralized control and GDPR compliance. With a user interface for the management of self sovereign identity, digital authorizations and associated data driven transactions, the advantage of Disposable Identities is that they may also contain verifiable data such as the owner's photograph, official or even biometric identifiers for more proactive prevention of identity abuse. These Disposable Identities designed for decentralized privacy management, can also be time, purpose and context bound through a secure digital contract; with verification functionalities based on tamper-proof technology. *Keywords*—dentity, trust, self-sovereign, disposable identity, privacy toolkit, decentralised identity, verifiable credential, cybersecurity, data driven business, PETs, GDPRdentity, trust, self-sovereign, disposable identity, privacy toolkit, decentralised identity, verifiable credential, cybersecurity, data driven business, PETs, GDPRI

I. Introduction

Many of our daily interactions demand that we expose identity data to the parties that we interact with. These interactions are considered as privacy trade-offs, where people share personal data in return for receiving something of perceived value [1]. During these trade-offs more information than needed is often exposed to other parties. As so our personal data is collected just for basic identification and authentication purposes [2]. How are we able to protect individuals towards data breaches, keep our sensitive data safe and build mutual trust in data sharing in a world that is more and more digital? For example when retrieving a parcel from a collection depot or 'click & collect' service point, one may be required to show a driver's license or passport to validate one's identity. However this also exposes additional personal data such as age, gender, place of birth, nationality and in some cases even Social Security Number - which are not relevant in the context of parcel delivery. Disposable Identities makes it possible to claim a parcel with the minimum data required. This is achieved by showing a verified digital proof of recipient, no name, address, age or any other personal identifier would be exposed, and still being assured that no-one else can claim delivery without permission. Since 2018 the rate of cybercrime has continued to rise at an alarming rate, this is after the General Data Protection Regulation (GDPR) came into force. The key sources of breaches range from:

- External malicious acts, for example, hacking or scams;
- Sending emails/documents to incorrect recipients;
- Loss or theft of unsecured devices, such as, mobile phones and laptops;
- Inadequate security measures of data available over the Internet, for example, improperly secured databases. [3]

And while not all data breaches stem from malicious intent, with negligent security or human error are also playing a contributory factor, the impact on organisations can be devastating, with severe business disruption, decline in customer trust, as well as potentially huge GDPR fines. Cybersecurity Ventures predicts that global cybercrime costs will grow by 15 percent per year, reaching $10.5 trillion USD annually by 2025 [4]. Yet the impact on individuals cannot be underestimated as (often deeply intimate) personal data and unique identity credentials have become a highly valued currency being traded on the dark web - often to be utilized in future cybercriminal attacks. According to data gathered by Finanso.se, 56% of Europeans have experienced at least one type of fraud in the last two years, one-third of them became victims of identity theft, making it the second most-common type of fraud in Europe [5].

Also the Coronavirus Disease 2019 (COVID-19) pandemic has greatly affected how individuals and organisations conduct and organise their daily activities, by taking our daily lives into more digital direction. This has changed the global economy and related activities dramatically. [6] Many organisations and individuals have been driven online by the restrictive measures put into place to curb the pandemic. Some of the measures have led to individual privacy being increasingly breached around the world.[7, 8] A recent assessment by INTERPOL of the impact of COVID-19 on cybercrime showed an alarming rate of cyberattacks, as well as a significant target shift from individuals to governments and critical health infrastructures [9].

978-1-7281-5684-2/20/$31.00 ©2021 European Union

In one such case of a massive private data leak, from the Dutch health service's coronavirus test system, repeated warnings about major security issues were not deemed serious enough to act upon, even after it was discovered that large sets of sensitive personal data were being sold on Telegram. [10] And as the Global Head of Linklaters' Privacy and Data Protection Practice comments "There is also a danger of GDPR fatigue amongst businesses and the Covid-19 crisis is impacting budgets which could limit resources to ensure compliance going forward" [3].

COVID-19 will eventually be brought under control by vaccines [11], but how can we ensure that individuals are able to both regain and maintain better control of their individual privacy when the world is moving forward? Especially when we cannot fully trust the organisations and institutions tasked with safeguarding our privacy to adequately fulfill this responsibility?

With the proposed Disposable Identities organisations can enable a much higher level of trust-by-design and opt to provide more respectful, as well as fully GDPR compliant, data privacy practices. We believe that this will in turn provide a stronger foundation from which to build more trusted and thus more sustainable value interactions. When the proposed approach is used parties can be assured that only information that is truly needed for the current process is being shared – in line with the GDPR data minimization principle. With an additional layer of transparency and control ensuring that a Disposable Identity is bound by time (temporally defined access), purpose (for an intention or goal) and context (under specific circumstances or conditions), all fully controlled by the owner of the identity, which makes it:

(a) Self-Sovereign [12]
(b) GDPR compliant [13]
(c) Compatible with MyData [14]

Technical approach for Disposable Identities could be built upon advanced Privacy Enhancing Technologies (PETs), including technologies to identify a natural person, cryptographic data projection, tamper-proof storage of data. To enable a wide range of use case scenarios locally and remotely stored data would be needed, and local data should be verifiable without network connection. Relevant technologies for disposable identities could include eIDAS compliant identity providers, decentralised identities, blockchain and verifiable identities. However research of the technical approach needs to be done in future work. [15, 16].

In this paper we will present The Disposable Identity, and its relation to real world identity and identification process (Chapter 2.). In Chapter 3. we link the real world and self-sovereign identity into a disposability approach, then Chapter 4. presents possible use-cases for disposable identities and Chapter 5. discusses about recognised challenges in the presented approach.

II. The Disposable Identities

A. What's is identity?

Social and human sciences have argued for a long time how to define identity, and how identity changes during time, also definition of an identity varies depending on which point-of-view in research we choose. In this research identity is considered as something unique which can be used to recognise natural person in real world use cases. As so identity defines a natural person. When the identity of a natural person or organisation (legal entity) is considered, it is constructed from a set of identifiers and linked to personal data (Figure 1). These identifiers are something that - as a linkable set - can be used to reconstruct or even validate a person's identity, or can similarly be used to construct an author's pen name, a pseudonym or alias. As these identifiers can be associated with various personal data properties - being transient or permanent - the act of collecting enough identifiers into a set, making this set unique, can identify an individual. [17, 18]

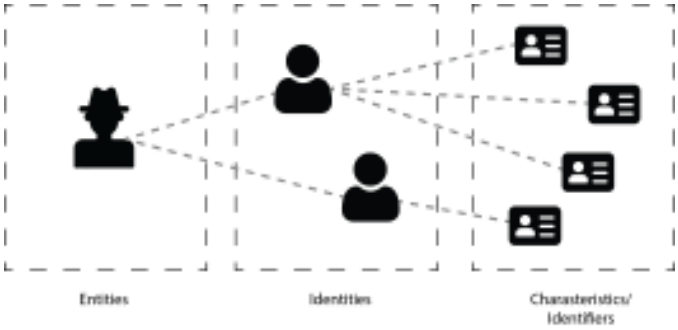


Fig. 1. Identities relationship to identifiers (adapted from [19])

It must not be assumed however that a natural person may only have one 'real world' identity, this concept is a relatively recent legal construct, and many European countries are even using, or are beginning to experiment with, official pseudonymous names and numbers in their official identification systems. For example under the UK Common Law on privacy it is even legal to own a number of aliases as long as they are not used to commit fraud or are not in conflict with criminal law. Such an obtained identity would even prove to have standing in a legal situation, and the identity holder may practice "the right to be known by a chosen identity". The artist Heath Bunting used this as a fundamental for his Status Project work from 2004, where he created a database with over 10,000 entries used to map new 'status' identities, going so far as to construct and sell entirely legal new British natural person identities consisting of both digital and physical identifiable data. [20]

Identifiable data can thus be self-selected or issued by an authority. Some of this data are government or authority controlled "official" documents of identification; passport, visa, driver's licence, some are non-permanent; such as name, address, telephone number or email address, and some of them are inherent and permanent; such as biometric identifiers. [19]

B. Identification process

Identification process is a process where the user presenting an identity is verified to be who they claim to be. During this process identity applicant presents identity evidence for the party they are engaging with, this evidence can be documents or some other material/data - a purpose and context-specific subset of identifiers. If presented evidence is sufficient, applicant is able to claim the identity. As so identification is the process of linking a unique set of identifiers into a natural person, this process can either fail or succeed. Identification process should not be confused with authentication, which is the ability to prove that a natural person is authorised to use specific resource. Example of identification is a natural person presenting their driver license during parcel retrieval, if the identification process is successful the next step is to check if the person really can retrieve the parcel e.g. authenticate person with logistics systems holding information about who can claim parcel. [21, 22, 2]

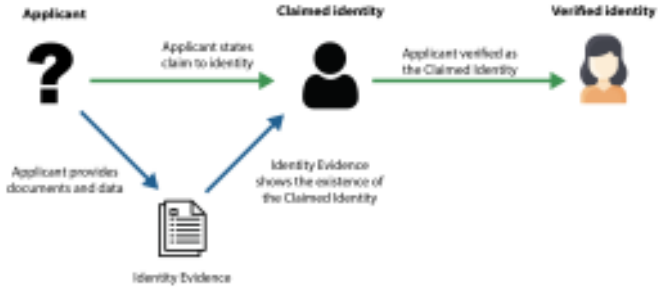


Fig. 2. Illustration of identification process, adapted from [22]

.

Depending on risk associated with the identification process, the identification process is looking to meet defined confidence during identity claim. This confidence is different for example in case of parcel retrieval or in legal actions [22]. The process of identification is not trivial or easy as it relies on documents and data that are subject to inaccuracy and potential fraud. [2]. This makes the task of a person or system performing identification process to encounter challenges like verifying authenticity of presented material.

Identification process is also used for example to prove a gained qualification, such as level of education, ability to drive a car, or fly an airplane. And it is also used, together with authentication, to provide access in both physical, digital and even hybrid domains. For example providing access to a restricted building, boarding an airplane, or visiting a healthcare professional (physical), sending email, trading on Ebay, or social media (digital) and studying at a university, having a bank account, or taking an Uber (hybrid). Yet, it is primarily in the digital domain that we observe a heightened risk of identity theft and subsequent fraud, at the same time also an increasing need to provide a balanced “middle ground” between accountability and privacy through the use of pseudonymous identities.

As seen in the previous chapter a lot of user related information, a subset of identifiers, is needed during the identification process. As this data is presented to other party during identification there is possibility of privacy breach, this breach can be for example party storing the data in their systems in unsecure way or person conduction verification using data for personal purposes. On the other hand, party conducting verification needs to be sure that authentic and high quality data is presented to them. For both parties transparency might be lacking in the process which weakens trust between parties. [2]

C. Empowering the user with a Disposable Identity

Identity, as we defined it in the last section, is built up from various characteristics {a, b, c, d, e, ...} (Figure 3.). One of these characteristics, for example face, iris or fingerprint data, can be used as a base for identity matching. This is because it has the highest possibility of being unique [23] and harder to fake. So when we are verifying the identity of a person we basically make assertions such as ”if you have features of him/her/them you must be him/her/them”.

Within the Disposable Identities, this identity matching is part of the core, but relies on identity data that is verified by an external trusted source, e.g. an eIDAS-compliant provider [24]. These identity providers can be used to create new base identity which is from verified and trusted origin.

After an creation of base identity, a user can collect and combine a chosen set of identity related characteristics {a, d, e, ...} and associate this data with a Disposable Identity. So Disposable Identity is a unique identifier generated by a natural person to prove their identity on a specific use-case, it provides its owner an ”alias”, so that their identity and data will be secured from unwanted tracking and information theft.

Users can also attach other types of personal data to a disposable identifier depending on the specific purpose or context at hand. This could for example be an official COVID-19 test result or verified vaccination status or any other relevant documents or data reflecting changing events that they wish or need to share in a privacy-secure manner.

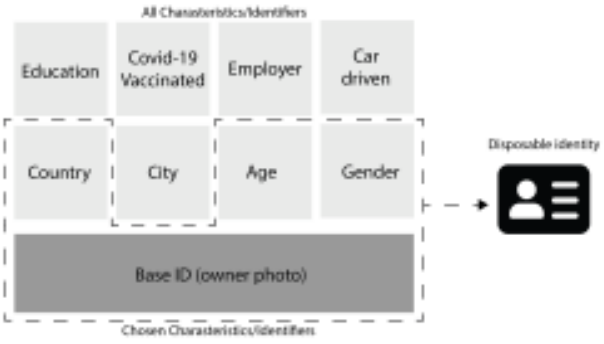


Fig. 3. Collection of characteristics chosen as Disposable Identity

When a user has collected their characteristics into their core user profile, they are able to create multiple self-sovereign Disposable Identities from a chosen subset of characteristics,

as illustrated in Figure 3. Each Disposable Identity, each characteristic and each piece of associated data - can be time-bound and made valid and available for a specified period of time, purpose, or context. With the control of these availability parameters always remaining in the hands of the identity (and data) owner.

One aspect of Disposable Identities is that identifiers person user has collected into their user profile can be verified (Figure 3.). Verification of identifiers can be done in various levels, like user verified, peer verified, authority verified, etc. This makes some identifiers more trustworthy than others. With this approach some process can use Disposable Identities with lower trust level and some process can require Disposable Identity with higher level of trust.

An additional aspect of the application of Disposable Identities for personal data exchange - or data dialogues - is that there are at least two entities involved; the giver and the receiver - as the simple use case below illustrates. This Disposable Identities solution "Same Place Same Time" was conceived as a privacy secure alternative to the restaurant contact tracing forms enforced by many countries across Europe during the COVID-19 pandemic [25]. This hastily installed measure led to major data breaches and saw many people giving false data due to an inherent lack of trust in this system of data capture, ultimately negatively impacting the effectiveness of the measure [26]. With Disposable Identities in place the restaurant would be able to initiate a secure data dialogue with every customer that entered their premises by presenting them with an embedded bar-code (Figure 4 to scan - whereby the subsequent data transferred is neither accessible to staff nor to other customers. The customers Disposable Identity would provide access to the 'legally demanded' personal data for a limited time (until the window of infection risk has passed) for a single purpose (logging entry and exit time to the establishment) and specific context (to receive a verified test appointment and quarantine information only if and when a positive infection has been logged at the same place at the same time).

Fig. 4. Illustration "Same Place Same Time" use case by M. Darienzo, L.Goulden, V.Jabbour and R.Stefanich

D. Possible technical approach for Disposable Identities

In this section a possible high-level technical approach (Figure 5) is presented for Disposable Identities. In the core of this approach is a trusted identity provider (like a national electronic identification service compliant with eIDAS), which can be used to create new verified identity. Chosen identity provider delivers the verified personal identity data, which can then be linked to multiple disposable identifiers containing chosen identifiers. In turn, other characteristics (personal data) can be linked to such identifiers (data-set).

The entire data-set or data-set subset can then be verified by a trusted actor by means of anchoring the disposable identifiers to a blockchain, or some other tamper proof technology. The owner of the Disposable Identities can prove 'untampered' ownership via cryptography and an anchoring mechanism using blockchain technology. Disposable Identities can also be verified by established representatives or authorities at any time, to leverage its status from self-verified to (trusted)-other-verified,

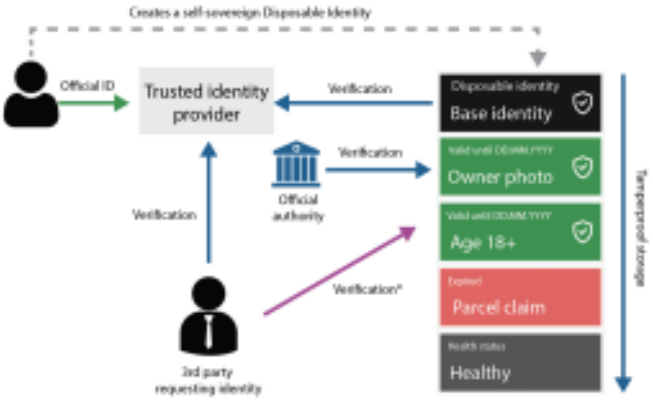


Fig. 5. Disposable identity process

Without this level of verification, the value of the identity-related characteristics may be deemed too low for certain applications. Chosen characteristics can be verified by official bodies (like customs, police, or driving license authority) or by private companies (a photograph of a user) if a use-case - in which a disposable identity is used - requires such established verification. However, some use cases may require less formal verification, as we are already familiar with in the physical world, where identity can be verified by being 'vouched for' by a mutually trusted person from a close social or familial circle (such as a neighbor, close friend, daughter or parent).

E. How to manage privacy with Disposable Identity?

One substantial part of the definition of privacy is the individuals' ability to maintain control of their personal information [27]. This is in direct relation to individuals openness to reveal such information to others in specified circumstances and with assumed protections [28].

When an individual is revealing identity-related information, they are doing so to receive some value within this privacy

trade-off. One theory suggests that individuals are willing to make this trade-off by performing a privacy calculus. In this calculus an individual looks to minimize the negative consequences and maximize the positive ones. [29]

The Disposable Identity is designed to enable more sustainable privacy trade-offs in a wide range of applications, where identity+data dialogues, that is the exchange of (personal) data value between two identified entities - remains transparent and in control of the owner of the identity and their data. This achievement is realised by allowing ex-ante and user-controlled time-spaced visibility of selected identifiers and by keeping other privacy-related data secure and compliant within the highest privacy requirements of GDPR [13] and of the latest European Data Governance Act [30].

A Disposable Identity is based on four important properties:

(a) Disposability (time-spaced, purpose-specific, context-driven and user controlled) – for it is often historical data, which holds the potential for increased risk (e.g., outdated or forgotten accounts abused for fraudulent activities)[31]
(b) Verifiability – unverified IDs are an invitation to adversaries to be used for their fraudulent activities [32]
(c) Privacy – for it is a human right and otherwise citizens will either not use said services, or if they use it may adapt or mask their behavior by what has also been termed “surveillance-related chilling effects” [33]
(d) Transparency – by enabling personal data to remain more visibly in the hands of individuals—we are minimizing the privacy risks posed by governments, corporations, and hackers because personal information is not concentrated in a single place [34]
(e) Trustworthiness – the above points combined with tamper-proof PET’s both engender and guarantee the trustworthiness of data, so it can be transformed into greater value [35]

Together these properties build up the core value proposition of disposable identities which can be used to create links between - or to “intertwine” - virtual (data) and real word artefacts. The advantages of this approach are discussed in the next chapter.

III. Advancing the central link between the virtual and real world - to have an impact in both

When we are creating new ways of presenting an individuals’ identity, we need to be able to effectively link the virtual world with the real world, in such a way that both parties trust the data presented. Recent advances on identification which are related to decentralisation and self-sovereignty are discussed and linked to disposable identities in this section.

A. Self-sovereignty

When self-sovereignty is considered then one of the most commonly cited descriptions is by Christopher Allen[36], stating: ”The user must be central to the administration of identity. That requires not just the interoperability of a user’s identity across multiple locations, with the user’s consent, but also true user control of that digital identity, creating user autonomy. To accomplish this, a self-sovereign identity must be transportable; it can’t be locked down to one site or locale. A self-sovereign identity must also allow ordinary users to make claims, which could include personally identifying information or facts about a personal capability or group membership. It can even contain information about the user that was asserted by other persons or groups.”

A Self Sovereign Identity (SSI), as a digital record or container of identity transactions that you control [37], combined with the potential to spin off an unlimited amount of temporary, local, and purpose-only smart contracts (identities+data) and privately exchanging them with specific other parties (identity+data dialogues); is the core principle behind Disposable Identities. It allows nothing less than (cryptographically) proving ownership and allowing contract-based sharing of identity data, from which we believe ultimately greater trust, higher relevance and more sustainable value can be generated. And all of this with strict permission from the user, in full compliance with the GDPR and without the unwanted feature of leaving ‘data-traces’ that third parties may access or ‘scrape’ to further exploit. When these data exchanges are accompanied by GDPR compliant consent mechanisms – the data contract – encoded as a Verifiable Credential [38] and the legitimate usage of that data can be cryptographically proven by all directly involved parties.

B. Decentralized Identifiers

Decentralised Identifiers (DIDs) are the latest innovation in the identifiers domain. With DIDs, any subject (person, organisation, etc...) can create the basis for a verifiable and decentralised identity. An owner of a DID can decide which cryptographic keys have control over their DID. DIDs are decoupled from centralised registries, identity providers, and certificate authorities. Although some services might be used for the discovery of the information related to a DID. This is a common (and standardised) method that enables the owner of a DID to prove control over it, without requiring permission from any other party. DIDs are Uniform Resource Identifiers (URIs) which create a link between a DID subject and a DID document. A DID document contains cryptographic material, verification methods or services, which provide a set of mechanisms enabling a DID controller to prove unequivocal ownership and control of the DID. [39]

C. Disposable Identity advancing SSI and decentralisation

The benefits of Disposable Identities can be derived from the widely defined benefits of SSI from past research. In short, the main benefits are: interoperability, ownership, pseudonymity, scalability, recovery, portability and security. [40]

However, Disposable Identities introduce three additional features to the SSI’s self-control: Time-bounding, purpose limitation and context specification. This way a Disposable Identity can be made valid for a limited time period, for a very specific use, under certain conditions.he user can also choose to change the access conditions or usage criteria at any time through an identity+data management interface (Figure

6.). This adds a privacy (GDPR-compliant) layer that benefits both users (data subjects) and receivers (controllers).

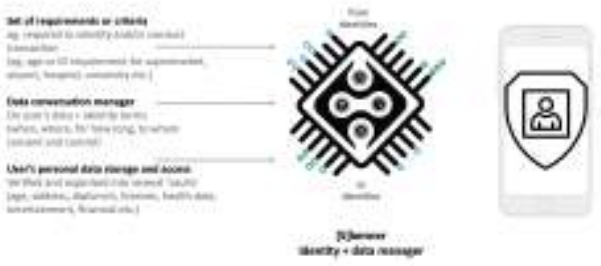


Fig. 6. Illustration of Disposable Identity interfaces

Decentralisation is at the core of Disposable Identities, as users are able to generate and present their Disposable Identities without being controlled by any central system. However, in use cases where an official identity attribute (or characteristic) is required, there is still a reliance on certified third parties (for example identity providers that are eIDAS compliant).

IV. Possible use-cases for Disposable Identities

Some of the possible use-cases for Disposable Identities are outlined in this section. They are based on the features of self-sovereignty, time-boundness, purpose limitation, context driven and decentralisation.

A. I'm over 18

During business interactions, there might be a need for a person to verify that they are over a required age. In this particular case, a person has created a base identity using a trusted identity provider. Furthermore, they could link a verified photo - from a second identity data provider - or a biometric identifier to the same disposable identifier to create a verified link to the person (physical) to the base identity (digital). When this user needs to prove they fulfill an age requirement, and show their photograph, one purpose-specific disposable identity can be presented to the requesting business party, enabling this party to verify the disposable identity and age requirement of the user (without having to reveal any additional personal data such as name, address or even date of birth). Visual interface for Disposable Identity with characteristics over 18, has been illustrated in Figure 7.

B. Social media

In social media many ideas and opinions are articulated and yet within this openness, trust is an extremely fragile component. The people - or other entities - that present themselves behind social media profiles are not using verified identities, as social media platforms have not (yet) come to agreement on this matter - and regulation has yet to be applied except in cases where a major crime is being investigated.

Fig. 7. Example of possible visual interface of disposable identity

Even for famous people or well-known authorities, social media platforms offer no means to prove that individuals are actually who they claim to be. However, in some cases there might be a clear need to verify that certain articulated claims within this social media 'virtual world' hold some merit in the real world. For this purpose the articulating person might provide a disposable identity based on verified credentials linked to their social media identity.

C. Linking an anonymous ID to a real person

There might be a need in certain use cases to confirm that there is a real person behind an anonymous identifier. In such cases, the verified personal identification data could be linked to the anonymous identifiers. An issue arises when such a link conflicts with the privacy intended by the anonymous identifier. A disposable identity could solve this conflict by providing advanced zero knowledge proof cryptographic verification tools to all actors, enabling them to verify the personal identity data related to the individual owning the anonymous identifier, without revealing any of their personal data.

V. Challenges

The real-world challenges regarding disposable identities can largely be considered equal to the ones that have been traditionally recognised in the SSI space. Having this in mind, if a person has features similar enough to one who calls himself "John", then this person might or might not be recognised as John. This reasoning applies when John has created his base identity and photo via an verified identity provider. How can others recognise that the person depicted in the photo is the actual John who owns the digital identity? This is a challenge when bridging between the digital world and the physical world.

Some of these challenges may occur when the Disposable Identity is being transferred to a receiver, as shown in Figure 7.

An example of a solution might be to use QR codes to bridge the physical and the digital world, whereby the owner of a disposable identity generates a QR code that directly transfers a disposable identifier to a receiving party to open a private and secure (data) communication channel. The receiving party could then, via this secured channel, receive the disposable identity data - for example containing said identifying photo - related to the earlier shared disposable identifier. Within this transfer, GDPR-compliant data contracts are also transmitted that include consent from the owner, delimiting the time period, purpose and context of use. Both parties keep a record of such - digitally signed - data contracts to digitally prove compliance. One novel aspect of these GDPR-compliant data contracts is that future privacy requests for adaptation or revoking access remain decentralised and in direct - and transparent - control of the user.

Thus, the receiver can match the digital identity of the person with the physical person standing in front of them. This will not, however, remove the risk and additional care required by both parties to avoid unauthorised or malicious capture of information through 'shoulder surfing', but ultimately the digital transfer remains secured - through its very nature of mutual consent, cryptographic contract and transparent control.

VI. Future work

A Disposable Identity toolkit for application developers is currently being developed by the non-profit Twinds Foundation for release under an open source license. This technical toolkit is designed to address the previously mentioned challenges as well as being piloted with targeted use cases to ensure an effective balance between the core Privacy Enhancing Technology integration and the user experience and acceptance of what will essentially be an entirely new paradigm of interaction for identity and (personal) data transactions. The Twinds Foundation vision is to proactively contribute to re-building and growing trust as an exclusive, value creating, asset which is considered vital if we are to build a more sustainable bridge between the physical and digital domains.

Ultimately, as with the platform economy, wide user adoption will be essential, and while this can be initially facilitated by targeting the developers community to lower the barriers of entry for application developers, as the SSI domain requires a higher degree of domain knowledge - which the SDK can help to bridge. From there, the value for both business and government actors must be demonstrated through targeted use cases in order to further scale the user base via validated value creation built on trust, and in close alignment with European regulatory requirements.

By open sourcing all work and by working in concert with official identity providers, the Twinds Foundation envisions an even further increase in said value by using the Disposable Identity toolkit in trustworthy application contexts. Applications that firmly require an official proof from a trusted identity provider will open up a whole world of new opportunities for a more digitally empowered society with more trusted authorities and institutions.

In future, more research work into machine-to-machine (M2M) use-cases will also be conducted, in order to explore if disposable identities might contribute to solving key security issues or to add more trust in IoT use-cases. There is also further consideration to be given to the potential role of multiple disposable identities in Gartner's Secure Access Service Edge (SASE) security model [41] and Zero Trust strategies [42].

Furthermore, the Twinds Foundation constitutes a platform for the development of cryptography-based technology, the exchange of viewpoints on its application and acts as a forum to facilitate public debate on the topic of SSI. The Twinds Foundation believes in the power of ecosystems, for they are considered more knowledgeable and have the ability to scale faster. And more importantly, they bring a diversity of perspectives to the tables, ensuring that the range of challenges yet to be addressed will be considered not only from a technical but also a social, cultural, political and also environmental perspective, ultimately generating a greater level of trust in the Twinds Foundation solutions.

The authors of this paper have decided to incorporate these ideas and developments as a non-profit foundation under Belgian law, one that is based on open source principles, collaboration, transparency and mutual exchange.

VII. Conclusion

In this paper the authors have presented an approach for disposable identities as self-sovereign, temporary identities which are totally controlled by their owner. The Disposable Identity is based on the latest approach to digital identities, such as SSI, decentralisation and verifiable credentials. Disposable identities leverage these approaches but are also temporal, purpose and context bound. This way the owner of the identity can always be sure that their identity is safe, and their privacy protected in an approach that fully meets the requirements of the GDPR. The proposed disposable identity toolkit is not only based on novel Privacy Enhancing Technologies. It also aims to establish a new interaction paradigm, a new organizational model and a new way of doing business where building mutual trust to promote higher value data transactions is at its core.

References

[1] Rainie Lee and Maeve and Duggan. *Privacy and information sharing*. 2016.

[2] A. Slomovic. "Privacy Issues in Identity Verification". In: *IEEE Security Privacy* 12.3 (2014), pp. 71–73. DOI: 10.1109/MSP.2014.52.

[3] Linklaters LLP. *Linklaters analysis shows a 66% increase in data breach notifications across major European markets*. June 24, 2020. URL: https://www.linklaters.com/en/about-us/news-and-deals/news/2020/june/linklaters-analysis-shows-increase-in-data-breach-notifications (visited on 02/02/2021).

[4] S. Morgan. *Cybercrime To Cost The World $10.5 Trillion Annually By 2025*. Nov. 13, 2020. URL: https://cybersecurityventures.com/cybercrime-damage-costs-10-trillionby-2025/ (visited on 02/02/2021).

[5] Jastra Ilic. *One in Five Europeans Have Experienced Identity Theft Fraud in the Last Two Year*. Aug. 11, 2020. URL: https://finanso.se/one-in-five-europeans-have-experienced-identity-theft-fraud-in-the-last-two-years/ (visited on 02/02/2021).

[6] Yaron Gvili. *Security Analysis of the COVID-19 Contact Tracing Specifications by Apple Inc. and Google Inc.* Cryptology ePrint Archive, Report 2020/428. https://eprint.iacr.org/2020/428. 2020.

[7] *Coronavirus: Privacy in a pandemic*. https://www.bbc.com/news/technology-52135916. Accessed: 2020-01-27.

[8] Iniobong Ekong, Emeka Chukwu, and Martha Chukwu. "COVID-19 mobile positioning data contact tracing and patient privacy regulations: exploratory search of global response strategies and the use of digital tools in Nigeria". In: *JMIR mHealth and uHealth* 8.4 (2020), e19139.

[9] INTERPOL General Secretariat. *INTERPOL report shows alarming rate of cyberattacks during COVID-19*. Aug. 4, 2020. URL: https://www.interpol.int/News-and-Events/News/2020/INTERPOL-report-shows-alarming-rate-of-cyberattacks-during-COVID-19 (visited on 02/02/2021).

[10] Thomson Reuters. *Personal data stolen from Dutch coronavirus track-and-trace programme*. Jan. 29, 2021. URL: https://www.reuters.com/article/us-health-coronavirus-netherlands-datapr-idUSKBN29Y1H3 (visited on 02/08/2021).

[11] Sarun Charumilind, Matt Craven, Jessica Lamb, Adam Sabow, and Matt Wilson. *When will the COVID-19 pandemic end?* https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/when-will-the-covid-19-pandemic-end. Accessed: 2020-01-27.

[12] Alexander Mühle et al. "A survey on essential components of a self-sovereign identity". In: *Computer Science Review* 30 (2018), pp. 80–86. ISSN: 1574-0137. DOI: https://doi.org/10.1016/j.cosrev.2018.10.002. URL: http://www.sciencedirect.com/science/article/pii/S1574013718301217.

[13] *2018 reform of EU data protection rules*. European Commission. May 25, 2018. URL: https://ec.europa.eu/commission/sites/beta-political/files/data-protection-factsheet-changes_en.pdf (visited on 06/17/2019).

[14] Antti Poikola et al. "MyData–an introduction to human-centric use of personal data (3rd". In: (2020).

[15] Nesrine Kaaniche, Maryline Laurent, and Sana Belguith. "Privacy enhancing technologies for solving the privacy-personalization paradox: Taxonomy and survey". In: *Journal of Network and Computer Applications* 171 (2020), p. 102807. ISSN: 1084-8045. DOI: https://doi.org/10.1016/j.jnca.2020.102807. URL: https://www.sciencedirect.com/science/article/pii/S1084804520302794.

[16] Jing Liu and Zhentian Liu. "A survey on security verification of blockchain smart contracts". In: *IEEE Access* 7 (2019), pp. 77894–77904.

[17] Audun Jøsang and Simon Pope. "User centric identity management". In: *AusCERT Asia Pacific information technology security conference*. Citeseer. 2005, p. 77.

[18] Craig Thompson and Dale Thompson. "Identity Management". In: *Internet Computing, IEEE* 11 (June 2007), pp. 82–85. DOI: 10.1109/MIC.2007.60.

[19] Audun Jøsang et al. "Trust requirements in identity management". In: *Proceedings of the 2005 Australasian workshop on Grid computing and e-research-Volume 44*. Citeseer. 2005, pp. 99–108.

[20] Ida Hirsenfelder. *Heath Bunting: Identity bureau. against residual self-image*. 2012.

[21] Jan E Stets and Alicia D Cast. "Resources and identity verification from an identity theory perspective". In: *Sociological Perspectives* 50.4 (2007), pp. 517–543.

[22] Government Digital Service UK's National Technical Authority on Information Assurance Cabinet Office. *Good Practice Guide No. 45, Identity Proofing and Verification of an Individual*. Oct. 1, 2015. URL: https://ntouk.files.wordpress.com/2015/06/gpg-45-validating-and-verifying-the-identity-of-an-individual-issue-2-4-ncsc-web.pdf (visited on 03/07/2021).

[23] David Menotti et al. "Deep representations for iris, face, and fingerprint spoofing detection". In: *IEEE Transactions on Information Forensics and Security* 10.4 (2015), pp. 864–879.

[24] Nils Engelbertz et al. "Security analysis of eidas–the cross-country authentication scheme in europe". In: *12th {USENIX} Workshop on Offensive Technologies ({WOOT} 18)*. 2018.

[25] Information Commissioner's Office. *Maintaining records of staff, customers and visitors for contact tracing purposes*. URL: https://ico.org.uk/global/data-protection-and-coronavirus-information-hub/coronavirus-recovery-data-protection-advice-for-organisations/maintaining-records-of-staff-customers-and-visitors-for-contact-tracing-purposes/ (visited on 02/10/2021).

[26] Thomson Reuters. *German restaurants object after police use COVID data for crime-fighting*. July 31, 2021. URL: https://www.reuters.com/article/us-health-coronavirus-germany-privacy-idUSKCN24W2K6 (visited on 02/08/2021).

[27] Alan F Westin. "Privacy and freedom". In: *Washington and Lee Law Review* 25.1 (1968), p. 166.

[28] Richard O Mason. "Four ethical issues of the information age". In: *MIS quarterly* (1986), pp. 5–12.

[29] Verena Wottrich, Eva Reijmersdal, and Edith Smit. "The privacy trade-off for mobile app downloads: The roles of app value, intrusiveness, and privacy concerns". In:

Decision Support Systems 106 (Dec. 2017). DOI: 10.1016/j.dss.2017.12.003.

[30] *REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act)*. European Commission. Nov. 25, 2020. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0767&from=EN (visited on 01/17/2021).

[31] Thorin Klosowski. *Delete Your Old Accounts (If You Can)*. https://www.nytimes.com/wirecutter/blog/delete-your-old-accounts/. Accessed: 2020-01-28.

[32] Kristian Hollins. *Comparative international approaches to establishing identity in undocumented asylum seekers*. https://www.lowyinstitute.org/sites/default/files/documents/Hollins\%20-\%20Comparative\%20international\%20approaches\%20to\%20establishing\%20identity\%20in\%20undocumented\%20asylum\%20seekers_WEB_0.pdf. Accessed: 2020-01-28.

[33] Chantal Mutimukwe, Ella Kolkowska, and Åke Grönlund. "Information privacy in e-service: Effect of organizational privacy assurances on individual privacy concerns, perceptions, trust and self-disclosure behavior". In: *Government Information Quarterly* 37.1 (2020), p. 101413.

[34] Marijn Janssen et al. "Transparency-by-design as a foundation for open government". In: *Transforming Government: People, Process and Policy* (2017).

[35] David J Hand. "Aspects of data ethics in a changing world: Where are we now?" In: *Big data* 6.3 (2018), pp. 176–190.

[36] Christopher Allen. *RThe Path to Self-Sovereign Identity*. Apr. 25, 2016. URL: http://www.lifewithalacrity.com/2016/04/the-path-to-self-soverereign-identity.html (visited on 01/17/2021).

[37] Andrew Tobin and Drummond Reed. "The inevitable rise of self-sovereign identity". In: *The Sovrin Foundation* 29.2016 (2016).

[38] World Wide Web Consortium et al. "Verifiable Credentials Data Model 1.0: Expressing Verifiable Information on the Web". In: *https://www. w3. org/TR/vc-data-model/?# core-data-model* (2019).

[39] W3C. *Decentralized Identifiers (DIDs) v1.0*. Jan. 28, 2021. URL: https://www.w3.org/TR/did-core/ (visited on 02/02/2021).

[40] Marcos Allende López. "SELF-SOVEREIGN IDENTITY". In: (2020). URL: http://dx.doi.org/10.18235/0002635.

[41] Michael Wood. "How SASE is defining the future of network security". In: *Network Security* 2020.12 (2020), pp. 6–8.

[42] Larry Nace. "Securing Trajectory based Operations Through a Zero Trust Framework in the NAS". In: *2020 Integrated Communications Navigation and Surveillance Conference (ICNS)*. IEEE. 2020, 1B1–1.

Infertility Awareness: Knowledge and Attitude of Medical & Non-Medical Moroccan Young People

Sana El Adlani, Yassir Ait Ben Kaddour, Abdelhafid Benksim, Abderraouf Soummani, Mohamed Cherkaoui

***Abstract*—** Background: Infertility in all countries of the word is on an increase, it's why the World Health Organization included an investigation into young people's fertility. In this sense, it's important to increase efforts to improve the knowledge about fertility for the young population. The aim of this study is to describe the difference between knowledge and attitude of medical and non-medical Moroccan young people. Materials and Methods: 100 medical Moroccan students (group 1) participated in the study, between 18 and 30 years, by a simple random sampling method, during 2020 and using a previously validated questionnaire. The answers were confronted to the result of our same study among 355 non-medical Moroccan young people (group 2) in 2019. Statistical analyses were performed using Statistical Package for the Social Sciences (version 10). Result: Medical students had a significantly higher level of knowledge about infertility than non-medical young people. However, both groups were aware of the impact of lifestyle on infertility. The knowledge state of the first group about infertility management was higher than the second group. Moreover, all non-medical Moroccan young people believed that it is easier to conceive if the couples had already their first baby, whereas, among medical students, only 53% had confirmed this belief. The results showed that 65% of medical students had proposed to try fertility treatments more than one time if treatment fails. Besides, the first advice of the second group was polygamy and adoption. Conclusion: Following the result of our study, the investigation of young people is the measure to optimize reproductive health. So, it's crucial that the government increase efforts to improve the knowledge about infertility not only for medical universities but for all scholar programs.

***Keywords*—** attitude, infertility, knowledge, medical, non-medical, young people.

Sana El Adlani, Yassir Ait Ben Kaddour, Abderraouf Soummani are with the Department of Gynecology and Obstetrics, University Hospital Mohammed VI, Marrakech, Moroccoo (e-mail: seladlani@gmail.com).

Sana El Adlani, Abdelhafid Benksim, Mohamed Cherkaoui are with the Laboratory of Human Ecology, Department of Biology, Faculty of Sciences Semlalia, Cadi Ayyad University, Marrakech, Morocco.

Sana El Adlani, Abdelhafid Benksim are with the High Institute of Nursing and Technical Health, Marrakech, Morocco.

Turin, from Factory City to Talents Power Player: the Role of Private Philanthropy as Agent of Innovation in the Revolution of Human Capital Market in the Contemporary Socio-Urban Scenario

Renato Roda

***Abstract*—** With the emergence of the so-called "Knowledge Society", the implementation of policies to attract, grow and retain talents, in an academic context as well, has become critical – both in the perspective of didactics and research and as far as administration and institutional management are concerned. At the same time, the contemporary philanthropic entities/organizations, which are evolving from traditional types of social support towards new styles of aid, envisaged to go beyond mere monetary donations, face the challenge of brand-new forms of complexity in supporting such specific dynamics of the global human capital market. In this sense, it becomes unavoidable for the philanthropic Foundation, while carrying out their daily charitable tasks, to resort to innovative ways to facilitate the acquisition and the promotion of talents by academic and research institutions. In order to deepen such specific perspective, this paper features the case of Turin, former "factory city" of Italy's Northwest, headquarters - and main reference territory- of Italy's largest and richest private formerly bank-based philanthropic Foundation, the Fondazione Compagnia di San Paolo. While it was assessed and classified as "medium" in the city Global Talent Competitiveness Index (GTCI) of 2020, Turin has nevertheless acquired over the past months a status of impact laboratory for a whole series of innovation strategies in the competition for the acquisition of excellence human capital. Leading actors of this new city vision are the Foundations with their specifically adjusted financial engagement and a consistent role of stimulus towards innovation for research and education institutions.

***Keywords*—** human capital, post-Fordism, private foundation, war on talents, .

I. The McKinsey Dossier

"SUPERIOR talent will be tomorrow's prime source of competitive advantage. Any company seeking to exploit it must instill a talent mindset throughout the organization, starting at the top"- so anticipated Chambers, Foulon, Handfield-Jones, Hankin and Michaels in their famous 1998 paper that for the first time introduced the notion of "War for Talent" to define the -back then- incumbent scenario of a very strong competition between companies to attain, engage and retain those human resources (the "talents" from the title) able to quickly represent a significant competitive edge for the businesses who hired them [1]. In 2007, ten years after introducing the expression, the consulting firm McKinsey came back to the subject with a new contribution in the very same McKinsey Quarterly, where it was stated "The War of Talent never ended. Executives must constantly rethink the way their companies plan to attract, motivate and retain the employees" [2]. Today, twenty years after the first article, what was only a predictive hypothesis based on the study of a sample of large American companies has become a shared reading key for a series of labour market phenomena increasingly evident in the contemporary context. What allowed the continuous and insisted research of brains to become one of the constants of the human capital market? What kept the fires of the Talent War high after they were foretold at the end of the last Millennium? The theory of Chambers et alii found a fertile ground -only partially foreseen by their talented authors- in a series of dynamics of change that are involving the whole planet at the dawn of the XXI century [3-4-5-6].

The first of these certainly is -in the countries with highest GDP per capita- the shift toward a service economy that automatizes or delocalises the manufacturing companies and the whole production industry, replacing it with economic, social and political processes based on knowledge, research and innovation and entrusting to the former growth and competitivity.

II. The Knowledge Society

Contrary to opinion popular in the 1950s and 1960s, it is now certain that robots will not replace humans, however, it is equally certain that levels of automation and technological application in the production of consumer goods and services will increase to such an extent as to enable greater collective and individual productivity and to generate processes of extreme innovation. The mass introduction of new technologies will not eliminate all the jobs that it will replace, even if for some sectors the balance will not always be

Renato Roda is with the Compagnia di San Paolo, Italy (e-mail: renato.roda@compagniadisanpaolo.it).

positive. It is necessary to take into account how an increasing number of individual talents, of synergistic groups constituted by capable and competent people, and eventually of networks of specialists that work in punctual and rigorous collaboration will have to ensure the functioning of the automatized companies and of the robotic plants that operate there. Analyses carried out in the United States and in the European Union [7] have now shown how the investment in knowledge capital has a significant impact on productivity growth in the various labour fields, with an average of 2% up to 34%. It is therefore a question of understanding how and to what extent advanced countries are able to respond to the needs of the knowledge society: how to equip the human capital required by the Knowledge Society in a dimension in which more and more is the information introduced in economic processes for the development of digital technologies and the WWW and for the related and consequent profound changes in life. This is particularly true in the world of labour, where not only individual activities but also whole career paths run out in always shorter life cycles. It is not by chance how, in the last decades, we have begun to evaluate the employability of a worker by referring to statistical concepts that take into account the cultural competences in the humanistic and mathematical/scientific fields. However, there is still a long way to go since we have just begun to understand how many and what are the real needs of a world of innovation that requires and imposes new forms of initiative, flexibility, entrepreneurial skills and competences, ability to work as a team, willingness to ongoing training. All this is part of an even greater connection of interrelated changes, which in the years have progressively transformed relations with space and time at the geopolitical level: we are referring to the so-called phenomenon of Globalization. With the delocalization of production processes and the spread of new information technologies, human capital can be freed from the physical presence of its holders, who can move freely around the globe [8]. Talents, as well as millions of other goods on the market, have expanded their reference basin and the hunt for human capital of excellence no longer know political borders. Episodes such as the timely selection of hosted refugees -on the basis of their high level of education and skills- made by Germany during the Syrian crisis show the attention of the most advanced countries for a new, now fundamental, asset of a worldwide competition. To win this specific race to human capital, as the Merkel/Syrian episode taught us, you need to act fast and to be willing to face potential internal disapproval due to the more or less selective opening of borders to migrants.

‘‘In this sense, the situation in Italy -setting of the present contribution- is certainly problematic. Scholars are divided on the subject. The most successful narrative, even among the general media, is the one proposed by those who denounce the enactment in the country -with an increasing tendency in the post-crisis years- of a real brain drain, to be intended as a forced emigration of the best young talents and the highest human capital with consequent impoverishment of the Italian labour market [9]. But there is no lack of those who tend to relativize the phenomenon more [10], attributing such chronic talent loss alarms to an excessively simplistic interpretation of Italian mobility data, that would not in fact prove an effectively higher incidence of people with very high human capital within an actually increasing emigration away from the country [11]. On one point, however, most analyses seem to converge: although the problem of the so-called brain drain may be partly exaggerated by the media narrative and the sensationalism of some articles, in Italy the mechanism of the circulation of talents suffers a condition of imbalance, where to a consistent export of human capital does not correspond an equivalent ability to attract and maintain young talents within the borders of the country [12]. Moreover, the fallacy of brain circulation [13] in Italy appears to be twofold, in the sense that not only the country fails to attract and integrate foreign talents in adequate numbers, but does not even appear able to stimulate the return of the country’s brains emigrated abroad [14]. The most macroscopic examples in this sense can be observed in the highly qualified labour market par excellence, namely the academic one. Lack of future prospects, insufficient allocation of funds, low level of meritocracy are the reasons for emigration most often mentioned in the interviews with young academics who have moved away from the country [15]. But, more generally, the criticisms addressed to the Italian academic world by researchers who have emigrated to work in foreign universities and laboratories are numerous and quite punctual. To the point of making more than obvious that the hypotheses of policies dealing with the problem of the circulation of brains and fighting the war of talents on equal terms cannot fail to pass through a decisive and punctual confrontation with the academic world. In this contribution we intend to precisely analyse a case in which an open and structured dialogue with university institutions has allowed the creation of an experimental engine designed not only to facilitate the external acquisition and internal selection of new talents but also to increase its potential and opportunities for brain circulation.

Resuming: in a world increasingly crossed over by moving flows of goods, the "brains" travel side by side with hundreds of thousands of other goods from territory to territory, and perceptibly affecting their respective economies and levels of development. Italy, however, suffers from the difficulty of activating a full and effective brain circulation, failing to offer the right conditions to attract and integrate foreign talents and to stimulate the return of Italian emigrated brains.

III. SETTINGS

This paper addresses the theme of talents mobility and competition at an even wider level than the one anticipated by Mckinsley: that is, by referring to an entire metropolitan territory. In the following pages we will observe how the transition from industrial to service city has driven a large town of north-west Italy, Turin, to have an increasing need of qualified human capital in line with contemporary innovative necessities.

The turning point, in such sense, was the presence on the urban territory of the largest and most solid Italian ex-banking Philanthropic Foundation. The strategically coordinated action of this and other institutions has made it possible to create an actual integrated system of collection, engagement, growth

and retention of brains, which is here suggested as best practice for similar situations in similar contexts. What happened in Turin over the course of two centuries – even considering a few local peculiarities - appears emblematic of several transformation processes of many Western World cities. The first major change was the one which led Turin from being the capital of the small Savoy state - rigidly structured in a military and bureaucratic sense - to become for a few years the capital of the newly unified Italian kingdom. Subsequently, once this role was lost, there was a change towards a heavily industrialized town, led by an entrepreneurial bourgeoisie that used the mechanical knowledge of military workshops to create an environment favorable to the birth of metalworking companies, to which the Alpine valleys and the countryside surrounding the city provided the manpower needed to start mass Taylorist production. The history of Turin up to the nineties of 1900 is therefore that of a factory-city on the model of the "one-company-town", centered on the FIAT car factory and on the automotive industry. Over time a large and very long-lasting industrial related sector grew around the automobile manufacturing: at the beginning of the 90's still more than 60% of Turin's employees were working in the automotive sector. Therefore, the FIAT of Turin (and to a lesser extent but -for a certain period of time- not less significant the Olivetti of close-by Ivrea that manufactured typewriters and calculating machines) was for a long time an aggregating and propulsive pivot of local economy and, at the same time, a real identification factor of the civil society, affecting habits and rhythms of life, orienting consumption choices, organizing services and welfare, creating a collective imagination and shared mindset of rare solidity. The constant development of industry, particularly in the decades after the Second World War, also strongly influenced the demographic growth in Turin due to the migration, especially from the South of Italy, of a workforce that would easily find a job at FIAT or in the related industries: the population increased from about 600.000 inhabitants in the war years to 1.000.000 in the early 1960s, and then rose again ten years later to the peak of about 1.200.000 residents [16].

In fact, the years 1967-72 represent the period of largest inflow of new workers who moved in masses to Turin mainly to cover blue collar tasks required by the exponential growth of jobs - often only relatively skilled. If from the mere point of view of salaries, the abnormal growth of the workforce allowed many individuals from the most deprived areas of the whole country to secure a moderate well-being, on the social level this originated enormous difficulties. The lack of sufficient housing and services exacerbated tensions and ignited even bitter struggles to denounce widespread forms of exploitation and to claim social rights that could not be satisfied by simple employment without guarantees of job stability. In any case, even the most traumatic phases of the workers' struggles during the troubled 70's (in the "years of lead", the terrorism did not spare the factories) and the oil crisis of 1973 were overcome [17].

Turin then became more and more a one-company-town, since FIAT gradually absorbed and incorporated the other minor players of the automotive industry and monopolized to its own advantage the whole related industrial sector: this allowed to maintain high levels of employment and demographic stability, so that the population remained above one million until the early 1990s. Such is the threshold after which, consistenly with the global progressions of change, the post-fordist transition process begins, affecting the entire Turin metropolitan area with a series of profound consequences: firstly, a strong and increasing downsizing of the manufacturing industry, which for Turin obviously meant the automotive industry; secondly, the decline in the total of employees, reflected in a subsequent decline in the population, which from 2000 onwards fell well below 900.000 inhabitants, approaching 800.000 [18].

Even more relevant are the radical productive and territorial reorganization phenomena affecting the city. In particular, the whole urban area undergoes a strong repositioning in the new Knowledge Economy networks, that engages research and innovation institutions, high-tech industries, companies for business services, the universities of the territory (Polytechnic and State University) cultural institutions, and especially banking foundations.

In 2006, the Winter Olympic Games, with the vast investments spurred by the event, allowed the construction of new infrastructures and contributed, in the medium term, to largely improve the perception of the city [19]. It is a period of renewed development, but abruptly interrupted by the world financial crisis of 2008. In any case, if around the turn of the millennium Turin still appeared as a manufacturing metropolis, even though reduced to a local dimension and devoid of a real international projection, by the middle of the first decade of the XXI century the situation appears completely changed.

Even the crisis of the '10s, while holding back excessive ambitions of growth and containing the illusions of greatness, has not altered the transformation towards the knowledge society just described. In other words, the critical elements, which hampered the evolutionary path and which can be summarized into a slow -even compared to other areas of Northern Italy- recovery course and in a difficult rise of regional and city GDP, did not change the new occupational face of the city: Turin is now definitively oriented in a dimension more attentive to sustainability, urban quality, environmental protection, the provision of services according to new priorities connected, for example, to transport systems, to healthy housing, to new management centers, to research and innovation. All this requires new talents, increasingly targeted skills, specific knowledge, but also flexibility and considerable capacity for relations, innovation and adaptation.

The main player in this investment in high quality human capital is, as already mentioned, the Fondazione Compagnia di San Paolo [20]. One of the oldest private European foundations, the Compagnia was founded as a charitable institution shortly after the middle of the Sixteenth Century, continuing over the centuries a variously forwarded philanthropic and credit activity. In the early 1990s, credit

activities were transferred to a newly created joint stock company, the "Istituto Bancario San Paolo S.P.A." while the activities of social and public interest remained with the ancient institution -now completely secularized- which then re-assumed the original name of "Compagnia di San Paolo" (then "Fondazione Compagnia di San Paolo" since 2020).

In 2020, the Foundation adopted the United Nations Agenda 2030 as the primary reference for its activities, organizing itself to contribute to the achievement of the Sustainable Development Goals. In particular, the research, acquisition, growth and retention of talents are attributable to the Missions Enhance Research and Accelerate Innovation of Fondazione Compagnia di San Paolo, which in turn are referred to the SDGs number 4 "Quality Education", 8 "Decent Work and Economic Growth" and 9 "Industry, Innovation and Infrastructure".

The value of the Company's portfolio, referring to market values, rose from €6.9 billion in 2009 to around €7.2 billion in 2019. Contending for a similar solidity of assets, the Company was able to provide more than 168 million euros for charity programs in 2020.

The Compagnia di San Paolo Foundation has as its main reference territory the Piedmont, Liguria and Valle d'Aosta regions and is deeply linked to the city of Turin, which shares, as main headquarters, with the other important former banking foundation of the city, the Cassa di Risparmio di Torino. The presence of such a large funding and support potential in a city which, while fairly large, does not represent the political, economic or tourist capital of the nation creates what we can define as a unique urban laboratorial scenario, bent on modernization and innovation. The forces in the field, in fact, allow the practical experimentation of solutions that elsewhere can apply only a theoretical approach: all this has been made possible in the latest period also thanks to the development of the market of the human capital of excellence. This also happens as a result of the choice of Fondazione San Paolo to go beyond the schemes that usually regulate the relationship between Philanthropic Foundations and subsidized institutions, and more specifically, academic institutions. Recent studies [21] on the specific relationship between private philanthropic institutions and universities and academic institutions in the United States (the country where in proportion private philanthropy supports more the tertiary education) have highlighted how the forms of support that the Foundations offer to the universities prove to be mainly of three definite types: the construction of particular buildings (laboratories, libraries, classrooms, etc.) that enrich the campus of universities and that often bear the name of the donor; allocation of funds that finance more or less defined chairs or research calls according to the will of the donor; contributions that support the day-to-day administration of campuses, linking the donation to specific faculties or research groups or leaving the funding in the availability of universities without any particular constraints of use. As we can see, the cases in which an active and profitable collaboration is developed between private philanthropy and universities are not even taken into account by the studies: the donor offers funds and indications or constraints of use, but does not intervene with other types of resources -primarily human capital available- for a coordinated design of common solutions. Instead, this is exactly what the Compagnia di San Paolo Foundation is trying to do with the two universities of Turin, entering into agreements in which, alongside the usual allocation of economic resources for research and -often- for daily administration, other types of support are offered, for example in the form of managerial, organizational and technical skills or as a link for the development of forms of networking between different institutions that can share assets to optimize expenses and investments. A particularly meaningful example of this is represented by the system of institutions and policies set up by the two types of institution to attract and retain high human capital talents in the academic, professional and entrepreneurial system of the city of Turin.

The starting situation was not rosy: in the report related to the seventh and latest edition of the Global Competitiveness Index of Talents 2020 (GTCI 2020) [22-23], for years one of the most important references for the ability to attract and retain talents of nations and cities, Turin emerges with a rather mediocre profile. The city qualifies in fact at the 99th place - the last of the 4 Italian cities included- out of 155 cities considered in the report.

However, the most worrying are the scores obtained in relation to the 5 dimensions pillars analyzed by the GTCI: activate talents (*enable*), attract talents (*attract*), grow talents (*grow*), retain talents (*retain*) and mid-level professional profiles (*vocational and technical skills*). As you can see in Figure 1, if we exclude the excellent result recorded in relation to the "retain" dimension, where Turin reaches on an 80/100 score, all the other indicators mark results much more dramatically limited. In particular, the almost non-existent ability to attract talents of the city, in relation to whose indicator the city gets a disarming score of 12/100.

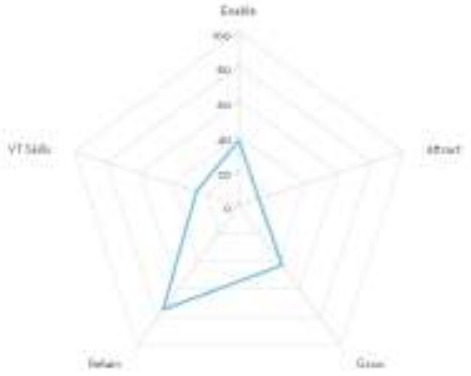


Fig. 1 The five dimensions pillars of talent competitiveness of Turin according to GTCI 2020

Overall, we can describe the profile of Turin that emerged from the analysis of the GTCI as that of a city in great struggle with the first phase of the involvement process of talents (enable and attract), that improves slightly in relation to the step of the grow and that it is much secured on the final retain

phase. In short, a scenario not devoid of noticeable potential, but that faces severe difficulties in starting a virtuous circle of engagement of human capital excellence.

IV. Exit Strategy

Considering, therefore, a starting condition that is not brilliant in relation to other similar realities in the European and world context, the Fondazione Compagnia di San Paolo has utilized its instrumental entities, the already existing operational structures, and the agreements with academic institutions to build an integrated system for the identification, selection, placement, and retention of talents. All of this, obviously, in relation to the needs of the territory and its expectations for growth and development.

The fundamental idea adopted was to promote a virtuous circuit for the circulation of talents within the city of Turin by defining of a series of actual paths of acquisition, training, and support of human capital of excellence, through the interconnection of existing or fashioned ad hoc institutions in the urban fabric. With the result of creating a unified system able to attract talents to the territory and to manage and capitalize on their work and careers. Thanks to this solution it becomes possible not only to shepherd local talents through increasing levels of training that are competitive even compared to the international offer, but also to attract brains from abroad, by engaging them at different stages of their training. Each entity involved in the system will in turn be provided with a network of national and international relations that will expand the potential for collaboration, exchanges and partnerships. In essence, to "capture" part of the global flow of excellency talents and thereby enrich the human capital of our academic institutions and our industry, the abstract concept of circulation has been translated into a real circulatory system of brains deeply connected with the territorial characteristics of Turin. The city emerges enriched in this way by the creation of a web of institutions that are already born with the intent to interact within it and that extend their ties to the international level.

The first level of this circuit designed to "capture" and retain talents consists of the two major universities of Turin, the Politecnico (Polytechnic, engineering studies) and the Università degli Studi (State University humanistic, medical and scientific studies), both linked to the Fondazione for many years by robust and extensive convention agreements. In recent years, however, these agreements have been rethought in order to reward excellence and innovation in research, and to transfer its output to implementation opportunities capable of generating new business and work. In order to do this, specific and measurable KPI have been adopted, such as the number of grants obtained by the university, the ministerial awards (recognition of excellence of the departments, the results in the Quality Evaluation of research carried out at national level) or performance in the latest years of Horizon 2020 and -in the future- Horizon Europe. In exchange for the support received by the universities, they are asked to pursue a series of objectives (missions) in a defined time frame and to make full use of the resources made available. These objectives are characterised above all by a focus on innovation, both in university teaching of first and second level and in doctoral courses. The progression towards these targets must be subject to a constant and rigorous evaluation, through strict real-time monitoring using online reporting systems that allow the Foundation to verify the use of resources and the partial objectives achieved [24]. In this, among other things, the Foundation seeks to reduce the percentages of funding diverted to the ordinary administration of the universities, compensating for the modest contribution to research that comes through institutional source.

The second level of the formative path consists of a series of participating institutions that offer high tertiary training of recognized excellence. Each of these bodies oversees a set of different disciplinary fields, in order to open to the best students at both universities a continuation of university studies on specific ranges of multidisciplinary specialization. In addition to teaching, these institutions provide students with targeted and facilitated links with the labour market, also with a view to internationalization. The disciplinary fields included in this particular type of teaching excellence are the technical-engineering (at the LINKS Foundation) the legal and social sciences (at the Collegio Carlo Alberto) and the medical-biotechnology (at the IIGM-Italian Institute for Genomic Medicine). Emphasizing the continuity of the training courses, the first two listed bodies are partnerships between the Compagnia di San Paolo Foundation and the Politecnico (LINKS Foundation) and the Università degli Studi (Collegio Carlo Alberto) respectively, while the IIGM, born as a partnership between the two universities and the Foundation, is now managed by the Compagnia di San Paolo alone, while still maintaining strong scientific and research ties with university departments.

More specifically, let us look at each of the three institutions mentioned.

The LINKS Foundation is an instrumental body of the Compagnia di San Paolo and -as we said- the Politecnico of Turin, where around 160 researchers operate [25]: it presents itself as a dynamic and constantly evolving organism at the heart of Turin's ecosystem of research and innovation. Comprised within a consolidated international network, it pursues the objective of contributing to the technological and socio-economic progress through advanced processes of applied research, often accomplishing its own study paths by winning a considerable number of national and, especially, European calls.

LINKS develops state-of-the-art technologies and innovative procedures applied to real case scenarios with a multidisciplinary and problem-solving entrepreneurial approach. In particular, it aims to face the challenges posed by digital transformation and its impact on the environment and society. A "link" between the academic world and corporations, start-ups and institutions, LINKS is active as a reference research center for organizations that want to increase their competitiveness and for talents interested in

developing their skills in an increasingly innovative dimension.

Similarly, the Fondazione Collegio Carlo Alberto [26], an institution of ancient tradition that has recently implemented and renewed its function, taking on the role of catalyst of research and training processes with a particular focus on social sciences (economics, law, political science, sociology and statistics).

The CCA's commitment is explicitly expressed as a mission of long-term investment in academic human capital and in the effort to attract the best talents from research institutions all over the world to Turin. The institution in fact hosts a faculty consisting of 7 chairs and 13 assistant professors and post-doc Fellows, all recruited from the global market of academic professionalism.

The final goal is to enhance the talents acquired by Turin so that they can compete at the same level in an open and globalized world. To fulfil its mission, the Fondazione Collegio Carlo Alberto works on three fronts: Research, Education and Outreach.

The private non-profit Foundation Italian Institute for Genomic Medicine (IIGM) - (formerly the Human Genetics Foundation - Turin, created in 2007 by the Compagnia di San Paolo, together with the two Universities of Turin) is a leading scientific institute in the field of Genetics, Genomics and Translational Medicine [27]. IIGM supports the development of excellence research and advanced training in the field of genetics, with particular attention to precision oncology. Today there are 54 researchers and collaborators (including 34 women) working on six lines of research: Molecular and Esposomic Epidemiology, Epigenomics, Cancer Genomics and Bioinformatics, Immunoregulation, Cancer Immunotherapy, Statistical Inference and Computational Biology, Genomic instability and tumor immunity. Since July 2019 IIGM has transferred its operational headquarters to the Institute of Candiolo - Fondazione del Piemonte per l'Oncologia - IRCCS, with which an ad hoc convention was signed in December 2018, to promote the integration of the two entities in a system perspective and to reach the shared goal of achieving scientific excellence. The purpose of the Italian Institute for Genomic Medicine is to create scientific programs able to interact and integrate with different public and private entities and become a benchmark at regional, national and international level for the world of research and business.

The third and final level of the innovative structure created by the Compagnia di San Paolo in Turin is represented by the LIFTT Foundation [28] and plays a fundamental role to complete the idea of a training process aimed at young Italian and foreign talents and adapted to the highest standards. Founded in June 2019 as an operating capital company, thanks to the initiative of the Compagnia di San Paolo Foundation and the Politecnico di Torino through the pre-existing LINKS Foundation, LIFTT comes to life in the Turin research ecosystem and operates within the framework of the agreements between the Compagnia di San Paolo Foundation and the five reference universities: Politecnico di Torino, Università di Torino, Università del Piemonte Orientale, University of Genoa and University Federico II of Naples. Most importantly, the other influent private philanthropic foundation of Turin, the Cassa di Risparmio Foundation has also been participating with equity capital, thus fully realizing the starting scenario that we had indicated, namely that of a medium-sized city that can nevertheless benefit from the presence of considerable and appropriately distributed philanthropic resources. Open also to private capital and since January of 2019 transformed into S.p.A., LIFTT operates as an active investor -providing swift financing- in the fields of biotech and healthcare, digital, mobility, education, industrial. But LIFTT is not only a project financier, it provides young entrepreneurs and start-ups with a whole range of resources, including the specific skills and human capital needed to accompany the first steps and the growth of the most promising innovative ideas proposed to it. Moreover, some of the professionals involved in this offer of wide-ranging advice and support come from the ranks of the very same Compagnia di San Paolo Foundation, thus achieving the goal of philanthropy that is not limited to mere economic transfer but provides expertise and professional experience of excellence.

It is not easy to fully measure the effectiveness of such a complex and articulated system, which includes a variety of modes and levels of intervention that make it difficult to compare information. However, we can report some of the figures that describe the results achieved so far by the individual institutions involved. The overall picture is very positive. For example, according to data from the European Commission, the LINKS Foundation appears to be in eighth place in Italy (out of a total of 263 entities assessed) in number of European projects funded, and in first place in number of researchers/ contributions ratio produced, thanks to an impressive total of more than 1600 publications issued since the birth of the institution. There are also more than 900 Italian industrial companies with which LINKS has partnered over time, a figure that, however, will exceed the 2000 units, if we also consider the European partners. Moreover every year the agency obtains approximately 10 million euros of revenues, thanks to the winning og projects call. The results obtained by the Collegio Carlo Alberto are even more focused on the excellence of their human capital. The institution - whose faculty includes, between full time and part-time researchers, about 60 individuals, plus 70 students- produces on average 150 publications per year, which often reach magazines of the highest rank. Over the past 10 years, as many as 5 ERC grants have been awarded to CCA researchers, and over the past 5 years 13 research projects have won international calls and been funded. The average success rate among the projects proposed by the Carlo Alberto researchers is 23.6%, about twice the European average, and much higher than the German and English ones.

The effectiveness of the work carried out by LIFTT can be approximated by evaluating the most successful projects that the institution has supported and promoted. Particularly, in the education sector, LIFTT has invested in Maieutical Labs, an

innovative SME specialized in Edtech, which carries out consultancy, prototyping and the creation of digital learning platforms. With a double-digit annual growth, 80.000 active users and over 12 million exercises provided each month, its virtual "Academies" today represent the largest digital platform of humanities in Italy.

Another flagship investment is Usophy, the Online University Library that allows university students -but not only- an unlimited and on-demand access to a catalog of e-book format textbooks, through a subscription. A simpler and more accessible way to study, thanks to the use of study tools that facilitate the search for information, their categorization and storage.

In short, what we described is a 360-degree action that, starting from education and training, allows talents to grow and stand out - also through managerial coaching provided by LIFTT - and then become an active part of industrial processes and business trends in the most advanced sectors within innovative and high-potential markets. An integrated system that features among the positive externalities also an important contribution to the commitment to retain in Turin's territory the maximum possible number of talents of excellence engaged, through the application of policies that make the whole experiment a good national and international reference practice. The initiative at various levels of the Fondazione Compagnia di San Paolo makes of Turin an important laboratory for the promotion and use of talents on a scale that goes well beyond its geographical and demographic limits. All of this happens thanks to the Compagnia's strategical agreement with other training, research, financial support and technology transfer agencies of the territory -and through instrumental bodies and foundations encouraged and supported by the Company itself [29-30]. Of great importance is also the choice of the Philanthropic Foundation to modify, articulate and enrich the traditional forms of support put in place to facilitate the commitment of the main players with respect to the objectives set. In addition to the mere financial transfer -which however is not lacking and is significant- the Company has put in place other types of resources such as specific skills, experience, and know-how built over the years, demonstrating that in order to attract excellent human capital it is necessary to spend the human capital already available.

.

REFERENCES

[1] Chambers, E.G., Foulon, M., Handfield-Jones, H., Hankin, S.M. e Michaels, E.G., 1998: War for Talent, McKinsey Quarterly 3(3), pp. 44-57

[2] Guthridge M., Komm A. B., Lawson E., 2008: Making talent a strategic priority, McKinsey Quarterly 1, pp. 49-59

[3] Michaels E., Handfield-Jones, H., Axelrod B., 2001, The War of Talent, Harvard Bsiness Press, Brighton, MA (trad. it. La guerra dei talenti. Come sedure e trattenere I manager di qualità, Milano 2002

[4] Brow P., Tannock S., 2009: Education, meritocracy and the global war for talent, Jurnal of Education Policy, 24, 4, pp. 377-392

[5] Hiltrop J.M., 1999: The Quest for the Best: Human Resource Practices to Attract nd Retain Talent, European Management Journal 17, 4, pp. 422-430

[6] Pazzaglia C. 2020: Ingaggiare, Coltivare, Trattenere. Le sfide della guerra dei talenti. Persone & Conoscenze. (4):52-56

[7] OECD 2013: Supporting Investment in Knowledge Capital, Growth and Innovation, Paris OECD

[8] Meyer, J. B., Kaplan D., and Charum, J, 2001: Scientific nomadism and the new geopolitics of knowledge. International Social Science Journal. 53(168): 309-321

[9] Tirabassi M., Del Prà A., 2014: La Meglio Italia: Le mobilità italiane del XXI secolo. Torino, Accademia University Press.

[10] Tintori G., Romei V., 2017: Emigration from Italy After the Crisis: The Shortcomings of the Brain Drain Narrative, in Lafleur J.M., Stanek M. (editors), South-North Migration of the EU Citizens in Times of Crisis, IMISCOE research series

[11] Beltrame L., 2007: "Realtà e retorica del brain drain in Italia. Stime statistiche, definizioni pubbliche e interventi politici" [Reality and rhetoric of brain drain in Italy. Statistical estimates, public definitions and political interventions]. Ed. Dipartimento di Sociologia e Ricerca Sociale.
Quaderni. Università degli Studi di Trento.

[12] Biondo, A. E., Monteleone, S., Skonieczny, G., & Torrisi, B. 2012: The propensity to return: Theory and evidence for the Italian brain drain. Economics Letters, 115(3), 359-362.

[13] Saint-Blancat, C.. 2019: Italy: brain drain or brain circulation?. International Higher Education, (96), 10-11.

[14] Dumont J.C., Spielvogel G., and Widmaier S., 2010: "International Migrants in Developed, Emerging and Developing Countries: An Extended Profile", OECD Social, Employment and Migration Working Papers No.114.

[15] Galan, A., Agasisti, T. 2014: From brain drain to knowledge transfer: Experiences of the Italian academe.

[16] https://www.tuttitalia.it/piemonte/72-torino/statistiche/popolazione-andamento-demografico/

[17] Musso S. 2002: Lo sviluppo e le sue immagini. Un'analisi quantitativa. Torino 1945-1970, in Levi F., Maida B. (editors), La città e lo sviluppo. Crescita e disordine a Torino 1945-1970, Milano, pp. 39-70

[18] https://www.rapporto-rota.it/images/rapporti/docs/2006/8_Olimpiadi.pdf;

[19] Massiani J. 2018: I promessi soldi L'impatto economico dei mega eventi in Italia: da Torino 2006 a Milano 2015 , 2, Torino 2006, fra analisi economica e autoconvincimento, Venezia, pp. 59-86

[20] https://www.compagniadisanpaolo.it/wp-content/uploads/Fondazione-Compagnia-di-San-Paolo-Piano-Strategico-2021-2024.pdf

[21] Murray, F. 2013: Evaluating the role of science philanthropy in American research universities. Innovation Policy and the Economy, 13(1), 23-60.

[22] https://gtcistudy.com/

[23] https://adeccogroup.it/gtci-2020-attraction-talenti-era-ai/

[24] https://www.compagniadisanpaolo.it/it/progetti/le-convenzioni-universitarie-2019-2021/

[25] https://linksfoundation.com/

[26] https://www.carloalberto.org/

[27] https://www.iigm.it/

[28] https://www.liftt.com/

[29] Garbellano S., Tesio V. 2013: Quando I talenti fanno la differenza. Una ricerca del Politecnico di Torino sulla gestione del personale in tempi di crisi, Harvard Business Review 7/8, Luglio-Agosto, pp. 82-91

[30] Margiotta U, 2018: La formazione dei talenti come nuova frontiera. The Talent's Education as a New Deal, Formazione e Insegnamento 16, 2, pp. 9-

Employee Engagement: The Key to Teaching Self Efficacy

Navya Jayachandran, Abey Edwin, H. Vipin

***Abstract*—** The purpose of this paper seeks to find out the effect of employee engagement on organizations with specific reference to the higher education sector. Primary as well as secondary data have been used to carry out the research. The study has been carried out on faculties as well as students of different higher education institutions. Employee engagement has been a great deal of interest in recent years. Today organizations are highly complex, more heterogeneous and boundary-less in nature. Organizations are diverse in nature comprising workgroups from different cultures, races, nations, regions and religions. Each individual has its own abilities, feelings, knowledge, skills, perceptions and attitudes. The organizations are competing for talented people who have high performance and high competence in the workplace.

***Keywords*—** employee engagement activities, employee satisfaction, teaching effectiveness, higher education.

Navya Jayachandran is with the CET School of Management, India (e-mail: navya.jc98@gmail.com).

A Mixed-method Study of Psychological Empowerment in Child Protection Practitioners

Amy R. Bromley

***Abstract*—**Child protection practitioners are a vital part of systems designed to protect children from abuse and neglect. Reforms in Anglo-American systems have shown a trend towards compliance-culture that reduces practitioner autonomy and empowerment, increasing staff turnover and negatively impacting outcomes for children. This explanatory mixed-methods study examined psychological empowerment in a national sample of child protection practitioners in Australia (n=109) using the Psychological Empowerment Instrument followed by semi-structured interviews (n=19). The results show that practitioners experience the sub-dimensions of psychological empowerment differently, perceiving themselves to have high levels of competence and satisfaction in their work, but limited opportunities for self-determination and low levels of impact on decision-making in their organizations. The qualitative data revealed that practitioners do not trust systemic reforms and have experienced them as ineffective, politically driven, and bureaucratic. The increased compliance demanded from these reforms have left practitioners feeling that their expertise is not valued, leading many to leave their organizations. The practitioners who remain employed in child protection identified their use of advocacy, curiosity, and child-centered values as ways of protecting their psychological empowerment. The findings highlight the ways psychological empowerment can be promoted within child protection systems, improving staff retention and building expertise.

***Keywords*—**Child protection, implementation, psychological empowerment, systems theory.

A. R. Bromley is with Flinders University, Adelaide, Australia (e-mail: amy.bromley@flinders.edu.au).

Microfluidic Impedimetric Biochip and Related Methods for Measurement Chip Manufacture and Counting Cells

Amina Farooq, Nauman Zafar Butt

***Abstract*—** This paper is about methods and tools for counting particles of interest, such as cells. A microfluidic system with interconnected electronics on a flexible substrate, inlet-outlet ports and interface schemes, sensitive and selective detection of cells specificity, and processing of cell counting at polymer interfaces in a microscale biosensor for use in the detection of target biological and non-biological cells. The development of fluidic channels, planar fluidic contact ports, integrated metal electrodes on a flexible substrate for impedance measurements, and a surface modification plasma treatment as an intermediate bonding layer are all part of the fabrication process. Magnetron DC sputtering is used to deposit a double metal layer (Ti/Pt) over the polypropylene film. Using a photoresist layer, specified and etched zones are established. Small fluid volumes, a reduced detection region, and electrical impedance measurements over a range of frequencies for cell counts improve detection sensitivity and specificity. The procedure involves continuous flow of fluid samples that contain particles of interest through the microfluidic channels, counting all types of particles in a portion of the sample using the electrical differential counter to generate a bipolar pulse for each passing cell—calculating the total number of particles of interest originally in the fluid sample by using MATLAB program and signal processing. It's indeed potential to develop a robust and economical kit for cell counting in whole-blood samples using these methods and similar devices.

***Keywords*—** impedance, biochip, cell counting, microfluidics.

Amina Farooq is with the LUMS - Lahore University of Management Sciences, Pakistan (e-mail: amna6230@gmail.com).

Association of Hyperuricemia with Length of Stay, Hospital Re-admissions and in-Hospital Mortality in Patients Admitted with an Acute Heart Failure Exacerbation

John Leso, Wafic Itani, Miask Harutyunyan, Hiren Patel, Mikhail Torosoff

***Abstract*—** , Wafic Itani, Misak Harutyunyan, Hiren Patel, Mikhail Torosoff

Background: In acute congestive heart failure (CHF) management, uric acid levels increase with IV diuresis, but an acute therapeutic uric acid lowering was not found to decrease BNP levels or improve LV ejection fraction. Hyperuricemia (HU) has been noted associated with increased mortality in patients with chronic CHF. However, an association between HU and outcomes of acute CHF has not been well investigated.

Methods: To study association between HU and length of stay (LOS), hospital re-admission(s) and in-hospital mortality in acute heart failure, we conducted a single tertiary academic medical center retrospective cohort study of 586 consecutive patients with known uric acid level. Acute heart failure diagnosis was ascertained according to the GWG-HF definitions. Uric acid values were divided in quartiles as less than 3.39 (UAQ1), 4-6.79 (UAQ2), 6.8-10.19 (UAQ3) and 10.2 mg/dL or above (UAQ4). The normal reference range for our laboratory is 3.6-8.0 mg/dL in males and 2.6-6.8 mg/dL in females.

Results: Of the study patients, 39% (226) were females, 68+/-15 years old, 40% (236) with history of diabetes, 70% (410) with treated hypertension, 47% (275) with history of CAD, and 37% (218) with history of chronic kidney disease (CKD), including those on hemodialysis. HU was not associated with increased prevalence of diabetes, hyperlipidemia, hypertension, or prior known CAD or CKD.

Uric acid levels were significantly increased in patients who expired (12.3+/-3.3 vs. 7.6+/-2.9 in survivors, p=0.004). Furthermore, there were strong trends towards longer LOS (UAQ1 7.8+/-7.9 vs. UAQ2 8.6+/-10.9 vs UAQ3 9.5+/-10.5 vs. UAQ4 8.5+/-6.3 days, p=0.7) and more frequent readmissions associated with increasing levels of uric acid (UAQ1 32%, UAQ2 31%, UAQ3 38%, UAQ4 47%, P=0.06).

Conclusions: In patients admitted with acute heart failure, elevated uric acid levels portend poor prognosis and are associated with increased mortality. Increased uric acid levels may also aid in identifying acute CHF patients who require increased hospital resource utilization by prolonging LOS and increasing re-admission rates.

***Keywords*—** heart failure, Hyperuricemia, mortality, re-admission.

John Leso is with the Albany Medical Center, United States (e-mail: lesoj@amc.edu).

Cross-Cultural Adaptation and Validation of the Child Engagement in Daily Life in Greek

Rigas Dimakopoulos, Marianna Papadopoulou, Roser Pons

***Abstract*—** Background: Participation in family, recreational activities and self-care is an integral part of health. It is also the main outcome of rehabilitation services for children and adolescents with motor disabilities. There are currently no tools in Greek to assess participation in young children. Purpose: To culturally adapt and validate the Greek version of the Child Engagement in Daily Living (CEDL). Method: The CEDL was cross-culturally translated into Greek using forward-backward translation, review by the expert committee, pretest application and final review. Internal consistency was evaluated using the Cronbach alpha and test-retest reliability using the intra-class correlation coefficient (ICC). Parents of children aged 18 months to 5 years and with motor disabilities were recruited. Participants completed the CEDL and the children's gross motor function was classified using the Gross Motor Function Classification System (GMFCS). Results: Eighty-three children were included, GMFCS I-V. Mean ± standard deviation of the CEDL domains "frequency of participation" "enjoyment of participation" and "self-care" were 58.4±14.0, 3.8±1.0 and 49.9±24, respectively. Internal consistency of all domains was high; Cronbach alpha for "frequency of participation" was 0.83, for "enjoyment of participation" was 0.76 and for "self-care" was 0.92. Test-retest reliability (ICC) was excellent for the "self-care" (0.95) and good for "frequency of participation" and "enjoyment of participation" domains (0.90 and 0.88, respectively). Conclusion: The Greek CEDL has good reliability. It can be used to evaluate participation in Greek young children with motor disabilities GMFCS levels I-V.

***Keywords*—** participation, child, disabilities, child engagement in daily living.

Rigas Dimakopoulos is with the University of West Attica, Greece (e-mail: rigasd@yahoo.com).

Ring Chromosome 14 Syndrome - Characterization, Dental Considerations, and Management

Chris S. Ivanoff, Athena E. Ivanoff, George P. Einstein, Orien L. Tulp, Timothy L. Hottel

***Abstract*—**

Purpose: To investigate the manifestations and challenges faced by children suffering from Ring Chromosome 14 Syndrome, a rare, genetic condition caused by an abnormal ring formation of chromosome (14). The disease was characterized by comparing patients for: phenotypic similarities; anthropometric criteria; physical appearance; craniofacial, dental and occlusal features, and; oral manifestations including caries and periodontal status. Subjects were evaluated for developmental condition and intellectual disability, and other special needs such as epilepsy and autism-related disorder that could impact the dental management of these patients. The information gathered from this study is intended to raise awareness for this extremely rare disease, both to allow health professionals to gain familiarity with the signs and symptoms and to encourage genetic testing early on. With only 50 cases reported worldwide, the disease is largely under-diagnosed and not fully studied, as the dental management of these patients also remains remarkably unaddressed. The investigation provides tips and recommendations that dentists may find helpful to manage r(14) children more effectively, as many are prone to seizures, behavioral outbursts, and unable to speak**.**

Methodology: A voluntary survey was administered to the caretakers of 13 r(14) patients who as of 2019 were registered in the (NORD-National Organization for Rare Diseases) global data bank (Ring 14 USA Outreach). The patients were assessed for age, gender, geographic distribution, phenotype, physical appearance, maxillofacial characteristics, presence of oral conditions and abnormalities, malocclusion, epileptic seizures, cognitive abilities, speech, muscle tone, nutrition, autism, and other developmental and behavioral points of interest.

Results: Of the 13 patients queried, 7 were male and 6 were female. The age of the patients ranged from 5 to 19 years of age. Ten patients were of European ancestry and three were Hispanic, all residing across the U.S. The majority of patients were diagnosed as infants, shortly after commencement of uncontrollable seizures. All the patients were microcephalic and presented with Class II malocclusions. More frequent occlusal anomalies and conditions included diastemae of the anterior teeth, congenitally missing teeth, crowding, and drooling. The majority of subjects was unable to speak, suffered from intractable seizures, and frequently exhibited behavioral outbursts.

Conclusion: A child with r(14) may present a considerable challenge to the dentist and staff, but the dental problems of r(14) children are for the most part like any other patient and can often be handled by the dentist. Depending on the severity of symptoms, some children with r(14) may be as treatable in the dental office as any other child.

***Keywords*—**Characterization, Dental Considerations, Management, Ring Chromosome 14 Syndrome.

Chris S. Ivanoff is with the Lincoln Memorial University College of Dental Medicine, Knoxville, TN 37923 USA (phone: 630-280-4372; e-mail: chris.ivanoff@lmunet.edu).

Athena E. Ivanoff was with the Einstein Medical Institute, University of Science, Arts, and Technology, Colleges of Medicine and Graduate Studies, North Palm Beach, FL 33408 USA (e-mail: athena.ivanoff@usat.edu).

George P. Einstein is with the Einstein Medical Institute, University of Science, Arts, and Technology, Colleges of Medicine and Graduate Studies, North Palm Beach, FL 33408 USA (e-mail: george.einstein@gmail.com).

Orien L. Tulp is with the University of Science, Arts, and Technology, Colleges of Medicine and Graduate Studies, Montserrat, BVI (email: o.tulp@usat.edu).

Timothy L. Hottel is with the Lincoln Memorial University College of Dental Medicine, Knoxville, TN 37923 USA (e-mail: timothy.hottel@lmunet.edu).

Gulcin's Method of Cessation Smoking

A.G. Sahan

***Abstract*—** Gulcin's Method of Cessation Smoking
AG.Sahan, Batı Medical Center, Turkey

Aim: Helping to cessation smoking

Introduction:

Lots of method try to quit for smoking . One of the most important responsibilities of a countries' health care system is to treat nicotine dependence. The cost and effectivity of these methods are not equally effective. It also is important to match the treatment method to the self and cultural differences.However, cessation smoking is still a big health problem(1)

Method:

89 smoker patients admitted this study at 2018-2020. These patients have admitted the hospital different diagnosis as acute bronchitis, bronchopneumonia, asthma attack, asthma bronchiale etc. During the physical examination, the patients have been listened their own lung sounds with a stethoscope by physicians, explain mechanism of rhonchi, wheezing and stridor sounds and explain how smokers' effect of ventilations' differences are become normal.

Result:

49 out of 89 people quit smoking within the first month after patients listened to their own breathing sounds. They were still non-smokers at the visit 6 months later. Smoking cessation rate was quite high when compared to other methods. The fact that 55% of the patients quit smoking reveals that Gulcin's method is a smoking cessation method that should be considered.

Discussion:

Cessation of smoking is difficult and laborious process. Although many different methods are used in smoking cessation methods, smoking cessation rates are still not sufficient.

***Keywords*—** Gulcin's method, cessation of smoking, preventive medicine, Novel method.

Abdulkadir Geylani Sahan is with the Suleymaniye Foundation, Turkey (e-mail: agsahan@gmail.com).

Functional Outcome and Quality of Life of Conservative versus Surgical Management of Adult Potts Disease: A Prospective Cohort Study

Mark Angelo Maranon, David Endriga

***Abstract*—** Objective: The aim of the study is to determine the differences in functional outcome and quality of life of adult patients with Potts disease who have undergone surgical versus non-surgical management. Methods: In this prospective cohort study, 45 patients were followed up for 1 year after undergoing pharmacologic treatment alone versus a combination of anti-Kochs and surgery for Potts disease. Oswestry Disability Index (ODI) and Short Form-36 (SF-36) were obtained on initiation of treatment, after three months, six months and one year. Results: ASIA scores from the onset of treatment and after 1 year significantly improved ($p<0.001$) for both non-surgical and surgical patients. ODI scores significantly improved after 6 months of treatment for both surgical and non-surgical patients. Both surgical and non-surgical patients showed significant improvement in their SF-36 scores, but scores were noted to be higher in patients who underwent surgery. Conclusions: Significant improvement with regards to functional outcome and quality of life was noted from both surgical and non-surgical patients after 1 year of treatment, with earlier improvements and better final scores in SF 36 and ODI in patients who underwent surgery.

***Keywords*—** tuberculosis, spinal, potts disease, functional outcome.

Mark Maranon is with the POC, Philippines (e-mail: maranonmicomd@gmail.com).

Ustekinumab Associated with Vasculitis as Adverse Events - Signal Review

Abdulrahman Alomair

***Abstract*—** Background:

Ustekinumab is a human IgG1κ monoclonal antibody that binds with high affinity and specificity to the p40 protein subunit of the human cytokines interleukin (IL)-12 and IL-23. Indicated to treat psoriasis and improving quality of life in patients with moderate to severe plaque psoriasis who are candidates for phototherapy or systemic therapy (1-). Vasculitis is an inflammation of the blood vessels, which can present when the body's immune system attacks its own blood vessel including arteries, veins and capillaries. It can happen because of an infection, a medicine, or another disease. The cause is often unknown (2-). Saudi Food and Drug Authority (SFDA) has initiated a signal review based on detected signal from the World Health Organization as result for signal weekly scanning.

Objectives:

The aim of this review is to evaluate the risk of Vasculitis associated with Ustekinumab.

Methods:

Signal Detection team at SFDA performed a signal review using NPC database, and World Health Organization (VigiBase) (1-), along with literature screening to retrieve all related information would be beneficial to assess the causality between Vasculitis and Ustekinumab use. The search conducted on 11/22/2020.

Results:

Local Cases: No local cases have reported for Vasculitis with the use of Ustekinumab.

Global Cases: The WHO database (VigiBase) searched for all individual case safety reports (ICSRs) reported with Vasculitis (Preferred Term) and Ustekinumab (Active Ingredient) yielded to 35 ICSRs. Our initial review revealed that 26 cases were not sufficiently documented for proper medical assessment. However the rest 9 cases have been evaluated resulted in five cases reported with positive dechallenge and thee negative rechallenge, after applying the WHO causality assessment two cases reported with possible association, three cases with unlikely association, and four cases were unassessable due to lack of information.

Literature: Two case report has been found supporting the association between Ustekinumab and vasculitis: 55 year-old woman with a history of psoriasis presented with a new, painful, purpuric, and ulcerative eruption on her lower legs four weeks after starting Ustekinumab. The patient diagnosed with drug-induced vasculitis (4-). And, 28-year-old female patient with history of colonic Crohn's disease and due to this disease, a diverting ostomy was made and Ustekinumab 90 mg (every 8 weeks) was begun. Three years later, patient presented with painful and palpable purpura in the lower abdomen and both legs, patient stopped Ustekinumab and started prednisone then patient symptoms resolved. After this period, a new dose of ustekinumab administered, with the reappearance of the same cutaneous lesions. (5-)

Datamining: Information Component (IC) developed by WHO Uppsala Monitoring Centre it been used to calculate the reporting ratio between drug and event. Higher value of IC means a strong statistical association between certain medications with the event in comparison to other medications. The combination of Ustekinumab and Vasculitis has observed slightly more than expected (IC 0.3) when compared with other medications in VigiBase.

Conclusions: Several cases of Vasculitis have been reported with the use of Ustekinumab and two were supporting the association according to WHO-UMC causality assessment. However, the weighted cumulative evidence identified in this review is sufficient to support a causal association between Vasculitis and Ustekinumab. Therefore, health care professionals should be aware of this possible risk and may consider monitoring any signs Vasculitis in patients treated with Ustekinumab.

***Keywords*—** ustekinumab, vasculitis, signal, adverse events.

Abdulrahman Alomair is with the Saudi Food and Drug Authority, Saudi Arabia (e-mail: aaomair@sfda.gov.sa).

Creation and Implementation of a New Palliative Care Drug Chart, Via A Closed-Loop Audit

Asfa Hussain,Chee Tang,Mien Nguyen

***Abstract*—** The safe usage of medications is dependent on clear, well-documented prescribing. Medical drug charts should be regularly checked to ensure that they are fit for purpose.

The purpose of this study was to evaluate whether the Isabel Hospice drug charts were effective or prone to medical errors. The aim was to create a comprehensive palliative care drug chart in line with medico-legal guidelines and to minimise drug administration and prescription errors.

50 medical drug charts were audited from March to April 2020, to assess whether they complied with medico-legal guidelines, in a hospice within East of England. Meetings were held with the larger multi-disciplinary team (MDT), including the pharmacists, nursing staff and doctors, to raise awareness of the issue. A preliminary drug chart was created, using the input from the wider MDT. The chart was revised and trialled over 15 times, and each time feedback from the MDT was incorporated into the subsequent template. In the midst of the COVID-19 pandemic in September 2020, the finalised drug chart was trialled. 50 new palliative drug charts were re-audited, to evaluate the changes made.

Prescribing and administration errors were high prior to the implementation of the new chart. This improved significantly after introducing the new drug charts, therefore improving patient safety and care. The percentage of inadequately documented allergies went down from 66% to 20% and incorrect oxygen prescription from 40% to 16%. The prescription drug-drug interactions decreased by 30%.

It is vital to have clear standardised drug charts, in line with medico-legal standards, to allow ease of prescription and administration of medications and ensure optimum patient-centred care. This closed loop audit demonstrated significant improvement in documentation and prevention of possible fatal drug errors and interactions.

***Keywords*—** palliative care,drug chart,medication errors,drug-drug interactions,COVID-19,patient safety.

Asfa Hussain is with the Lister Hospital, United Kingdom (e-mail: asfa_hussain@hotmail.co.uk).