

SpringerBriefs in Applied Sciences and Technology

PoliMI SpringerBriefs

Carlo G. Riva *Editor*

Special Topics in Information Technology



POLITECNICO
MILANO 1863

OPEN ACCESS

 **Springer**

SpringerBriefs in Applied Sciences and Technology

PoliMI SpringerBriefs

Series Editors

Barbara Pernici, Politecnico di Milano, Milano, Italy

Stefano Della Torre, Politecnico di Milano, Milano, Italy

Bianca M. Colosimo, Politecnico di Milano, Milano, Italy

Tiziano Faravelli, Politecnico di Milano, Milano, Italy

Roberto Paolucci, Politecnico di Milano, Milano, Italy

Silvia Piardi, Politecnico di Milano, Milano, Italy

Springer, in cooperation with Politecnico di Milano, publishes the PoliMI Springer-Briefs, concise summaries of cutting-edge research and practical applications across a wide spectrum of fields. Featuring compact volumes of 50 to 125 (150 as a maximum) pages, the series covers a range of contents from professional to academic in the following research areas carried out at Politecnico:

- Aerospace Engineering
- Bioengineering
- Electrical Engineering
- Energy and Nuclear Science and Technology
- Environmental and Infrastructure Engineering
- Industrial Chemistry and Chemical Engineering
- Information Technology
- Management, Economics and Industrial Engineering
- Materials Engineering
- Mathematical Models and Methods in Engineering
- Mechanical Engineering
- Structural Seismic and Geotechnical Engineering
- Built Environment and Construction Engineering
- Physics
- Design and Technologies
- Urban Planning, Design, and Policy

Carlo G. Riva
Editor


Special Topics in Information Technology



POLITECNICO
MILANO 1863

 Springer

Editor

Carlo G. Riva 

Dipartimento di Elettronica, Informazione e
Bioingegneria
Politecnico di Milano
Milano, Italy



ISSN 2191-530X

ISSN 2191-5318 (electronic)

SpringerBriefs in Applied Sciences and Technology

ISSN 2282-2577

ISSN 2282-2585 (electronic)

PoliMI SpringerBriefs

ISBN 978-3-031-15373-0

ISBN 978-3-031-15374-7 (eBook)

<https://doi.org/10.1007/978-3-031-15374-7>

© The Editor(s) (if applicable) and The Author(s) 2023. This book is an open access publication.

Open Access This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This volume is a collection of the most promising results achieved by graduated Ph.D.s in Information Technology (IT) at the Department of Electronics, Information and Bioengineering of the Politecnico di Milano. The presented contributions summarize the main achievements of their theses, successfully defended in the a.y. 2021–2022 and selected for the IT Ph.D. Award (out of more than 50 theses defended this year). As in the tradition of this Ph.D. program, these theses cover a wide range of topics in IT, reflecting the broad articulation of the program in the areas of computer science and engineering, electronics, systems and control, and telecommunications, and emphasizing the interdisciplinary nature of IT. Doctoral studies in the IT Ph.D. program pursue state-of-the-art research and the development of innovative, cutting-edge methodologies and technologies, and aim at preparing generations of young researchers that will shape future innovation, as the 12 authors of this volume will undoubtedly do. Overall, this book gives an overview of some of the newest research trends in IT developed at the Politecnico di Milano, presenting them in an easy-to-read format, amenable also to non-specialists.

Milan, Italy
July 2022

Carlo G. Riva

Contents

System and Control

Data-Informed Models for the Coupled Dispersal of Microplastics and Related Pollutants Applied to the Mediterranean Sea	3
Federica Guerrini	

Enhancing the Quality of Human-Robot Cooperation Through the Optimization of Human Well-Being and Productivity	15
Costanza Messeri	

Dynamic Sediment Connectivity Modelling for Strategic River Basin Planning	27
Marco Tangi	

Electronic

Gamma-Ray Spectroscopy and Imaging with SiPMs Readout of Scintillators: Front-End Electronics and Position Sensitivity Algorithms	41
Luca Buonanno	

Ultra-High Performance Digital Electronic Architectures for Events Management in Real Time Environments	53
Fabio Garzetti	

Computer Science and Engineering

Temporal Logic and Model Checking for Operator Precedence Languages: Theory and Applications	67
Michele Chiari	

Deep and Wide Tiny Machine Learning	79
Simone Disabato	

Learning in the Presence of Multiple Agents 93
Giorgia Ramponi

**Post-cloud Computing: Addressing Resource Management
in the Resource Continuum** 105
Michele Zanella

Telecommunications

Intelligent Networked Music Performance Experiences 119
Luca Comanducci

**Fast and Robust Estimation of Atmospheric Phase Screens Using
C-Band Spaceborne SAR and GNSS Calibration** 131
Marco Manzoni

**How Deep Learning Can Help Solving Geophysical Inverse
Problems** 141
Francesco Picetti

System and Control

Data-Informed Models for the Coupled Dispersal of Microplastics and Related Pollutants Applied to the Mediterranean Sea



Federica Guerrini

Abstract Microplastic pollution is a ubiquitous environmental threat, in particular to the oceans. In the marine environment, microplastics are not just passively transported by sea currents, but often get contaminated with organic pollutants during the journey. The uptake of chemicals onto microplastics can worsen the adverse effects of microplastics to marine organisms; however, investigation on this urgent phenomenon is hampered by the impossibility of monitoring and tracking such small plastic fragments during their motion at sea. This work aims at addressing the need for an effective modelling of the advection–diffusion processes jointly involving microplastics and the pollutants they carry to further our understanding of their spatiotemporal patterns and ecological impacts, focusing on the Mediterranean Sea. Here we present the conceptual design, methodological settings, and modelling results of a novel, data-informed 2D Lagrangian–Eulerian modelling framework that simultaneously describes (i) the Lagrangian dispersal of microplastic on the sea surface, (ii) the Eulerian advection–diffusion of selected organic contaminants, and (iii) the gradient-driven chemical exchanges between microplastic particles and chemical pollutants in the marine environment in a simple, yet comprehensive way. Crucial to the realism of our model is exploiting the wide variety and abundance of data linked with drivers of Mediterranean marine pollution by microplastics and chemicals, ranging from national censuses to satellite data of surface water runoff and GPS ship tracking, other than the use of oceanographic reanalyses to inform microplastics’ motion at sea. The results of our method applied to a multi-year simulation contribute to a first basin-wide assessment of the role of microplastics as a vehicle of other pollutants of concern in the marine environment. The framework proposed here is intended as a flexible tool to help advance knowledge towards a comprehensive description of the multifaceted threat of marine plastic pollution and an informed support to targeted mitigation policies.

F. Guerrini (✉)

Department of Electronics, Information, and Bioengineering, Politecnico Di Milano, Via Ponzio 34/5, 20133 Milano, Italy

e-mail: fede.guerr@gmail.com; federica.guerrini@polimi.it

© The Author(s) 2023

C. G. Riva (ed.), *Special Topics in Information Technology*,

PoliMI SpringerBriefs, https://doi.org/10.1007/978-3-031-15374-7_1

Keywords Microplastic pollution · Numerical modelling · Lagrangian particle tracking · Eulerian advection–diffusion · Coupled Lagrangian–Eulerian · Plastic-Related Organic Pollutants

1 Introduction

The pervasiveness of plastic waste in the environment, and particularly in the oceans, is eliciting global concern for its role of additional stressor in the functioning of ecosystems that are already heavily impacted by the local effects of the global climate crisis and of human activities. It has been estimated that, in the global oceans, most of the mass of polluting plastics can hardly be seen by eye, not only because it may be found on the seafloor, but also because it is present in form of smaller fragments, the so-called microplastics (<5 mm, [1]). Microplastics may be difficult to detect, but not less risky: because of their high surface-to-volume ratio, microplastics can effectively release pollutants (starting from the chemical compounds added during the production of their parent materials) and absorb other contaminants, such as hydrophobic organic chemicals, when transported through polluted marine environments by wind and ocean currents. The capacity of microplastics to exchange such Plastic-Related Organic Pollutants (PROPs, [9]) with the environment may exacerbate their toxicological impacts to marine organisms, potentially involving entire marine food webs, even up to humans.

Like in the case of other atmospheric and oceanographic pollution phenomena, further understanding of microplastics' spatial patterns has often been supported by numerical modelling and simulations, crucially relying on the most up-to-date information to produce novel knowledge. In recent years, ocean data collection and dissemination has experienced an unprecedented growth as a consequence of the deployment of new technologies and infrastructures [3], other than political will, allowing to dramatically improve our understanding of the global oceans and their dynamics. Nonetheless, on-field data is insufficient to describe the interactions between marine microplastics and the seascape they travel through, as these involve widely different spatial scales, with chemical interactions occurring at the surface of each microplastic particle, which is in turn entrained by large-scale ocean dynamics. On the other hand, there is abundance of data related to the release of plastic waste and chemical pollution to the marine environment: for instance, presence of people and the related activities, both in-land and offshore, like fishing; in addition, also the characteristics of the land, such as hydrography, may modulate the inflows of plastic and contaminants to the oceans. We thus argue that scientific computing and numerical simulations informed by such drivers of pollution can be of uttermost importance to investigate large-scale dispersal patterns and/or other spatial processes that could be hard to measure and track from field evidence. Crucial to the realism of such modelling efforts is the design and use of relevant data inputs, able to describe the drivers of pollution at sea.

To improve our understanding of marine plastic pollution and related phenomena, this work aims at investigating the important and quantitatively misregarded topic of the chemical exchanges occurring between microplastics and the seawater they travel through by developing and performing an *in silico* experiment with a novel data-informed modelling methodology, here applied to the Mediterranean Sea. This semi-enclosed, highly urbanized basin is considered as one of the global ecoregions most affected by anthropogenic pressure [16] including accumulation of plastic waste [5] and chemical pollutants [2]. This work is thus intended at providing a data-informed tool to support mitigation and control actions targeted to our study region.

2 Methods

In our model, the dynamics of microplastics and PROPs at the sea surface is modelled through two widespread, yet often independently used, fluid dynamics modelling frameworks to solve the advection–diffusion–reaction equation (as from [6]), i.e. the Lagrangian for microplastic particle tracking, and the Eulerian for the advection–diffusion of PROPs. The liaison between the two approaches is obtained here by modelling the step-by-step chemical exchanges occurring between particles and the seawater surrounding them, and by account for the simultaneity of the phenomena represented by these processes by intertwining them through a suitable algorithm.

This theoretical framework, as represented in Fig. 1, has to be placed in the Mediterranean Sea context by informing it through relevant data. The dispersal of microplastics and of PROPs is driven by the underlying circulation patterns: this aspect is here addressed by using oceanographic reanalyses, providing state-of-the-art reconstruction of ocean dynamics. Furthermore, sources of microplastics and chemical pollution need to be modulated in their spatial distribution and temporal behaviors. Despite the recent attention catalyzed by this global issue and the variety of surveys aimed at quantifying the emissions of plastic waste or of chemicals into the marine environment, we can confirm, after some extensive research, that available on-field data is still insufficient to inform large-scale models: for this reason, we have to link them directly to their determinant factors, as will be discussed below.

3 Coupling the Lagrangian and the Eulerian Frameworks

We identify three separate Sub-Processes (SPs) to be solved when considering the coupled dynamics of microplastic particles and their related pollutants in the marine environments: Lagrangian particle transport by sea surface currents, SP1; the Eulerian advection–diffusion dynamics of PROPs at sea surface, SP2; and the chemical gradient-driven mass transfer of PROPs between particles and the surrounding seawater, SP3, as schematically shown in Fig. 1 (see [8, 10], for further details about the setup of each SP).

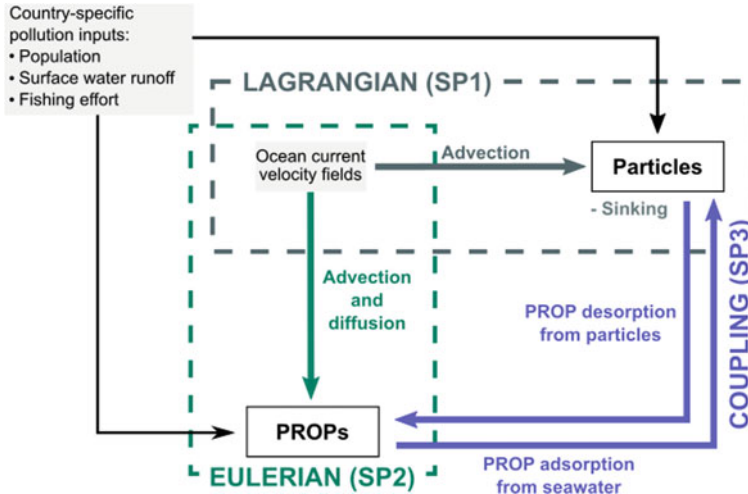


Fig. 1 Diagram of our coupled Lagrangian–Eulerian framework. The modelled entities, i.e. plastic particles and PROPs, are framed in black. Thick color-coded arrows represent Sub-Problems (grey: SP1, green: SP2, lilac: SP3). Black thin arrows represent the data used to force the model (gray box). From Fig. 1a in [10], used under CC BY Version 4.0/Colors modified from original

To integrate the SPs over space and time and account for their simultaneity, we use an operator splitting approach, that is frequent in fluid dynamics and advection–diffusion problems, where a PDE or system of PDEs contains terms representing different physical processes that may occur over different time scales.

We base our coupling algorithm on the Strang operator splitting approach [19]. We thus implement our algorithm integrating each SP in staggered time steps as follows:

1. Eulerian advection–diffusion of PROPs over the oceanographic domain (SP2) is solved over a half time-step $[0; \Delta t/2]$;
2. Lagrangian particle transport by surface currents (SP1) is simulated at a Mediterranean-wide scale over an entire time step $[0; \Delta t]$;
3. Coherently with the chemical gradient between the concentration of PROPs on each particle and in the surrounding seawater at time $t = \Delta t$, water-particle pollutant exchanges are solved at particle scale and concentration of PROPs in seawater is updated over $[0; \Delta t]$ (SP3);
4. SP2 is solved again for another half time step $[\Delta t/2; \Delta t]$, so that the numerical scheme is completed over a full time step.

The three SPs and the algorithm are then implemented in Python programming language and executed throughout the duration of the coupled Lagrangian–Eulerian simulation. In particular, details about the calculation of water-particle pollutant exchanges and the updating of seawater and particle concentrations can be found in [10].

We simulate two years (2015–2016) of coupled Lagrangian–Eulerian interactions between daily-released particles transported by Mediterranean surface currents and target PROP (phenanthrene, a polycyclic aromatic hydrocarbon frequently found in Mediterranean seawater and bound to microplastics; [14]), with daily inputs throughout the Mediterranean Sea and undergoing advection–diffusion on its surface layer. During the simulation, we track the trajectories, particle–seawater PROP fluxes, and daily PROP concentrations of each of the 100 million microplastic particles released per year of simulation, for a total of over 250 million particles released.

4 Data Sources and Processing

4.1 Oceanographic Data

The advection of plastic particles and the advection–diffusion of PROPs are forced by the zonal and meridional components of surface ocean current velocities from the Mediterranean Sea Physics Reanalysis (MED REA) [18]; an example is visible in Fig. 2a. Reanalyses provide a coherent and consistent dynamical representation of ocean dynamics, as they consist of a retrospective modelling experiment where the state of several ocean variables of interest is simulated over a long period of time through an Ocean General Circulation Model (OGCM), and re-aligned to the best available observations from satellite sensors and/or in-situ measurements.

4.2 Drivers of Plastic-Related Pollution in the Mediterranean Sea

Mismanaged waste from coastal cities, rivers collecting surface water runoff, and accidental release during fishing activities are commonly regarded as the main sources of marine plastic pollution in the Mediterranean [12]. In our simulations, we account for plastic and chemical pollution coming from each of these three sources, as shown in Fig. 2; we assume that the release of PROP follows the same indicators as the inputs of microplastics, described below. Importantly, particle release from onshore sources, i.e. along the coastline and at river mouths, depends on solid waste management within the country where the source point is located. To address this important aspect, both our coastal and riverine proxy indicators include the country-specific daily per capita generation rates of mismanaged plastic waste (MPW) [11].

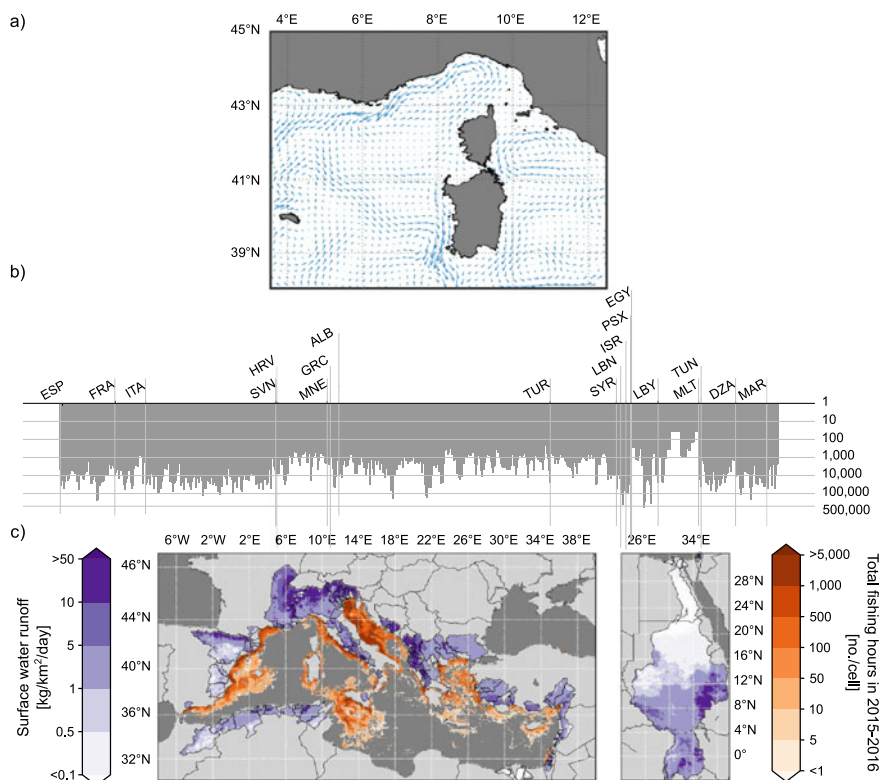


Fig. 2 Visual description of the input data: ocean reanalysis and proxy variables used to characterize particle sources in SP1. **a** Surface currents as from MED REA; example from the North-Western Mediterranean Sea; from Fig. 1a in [7], used under CC BY Version 4.0; **b** Population along the Mediterranean coast, starting from the Strait of Gibraltar on a clockwise path, CIESIN dataset; **c** Average surface water runoff in the top-100-scoring river basins selected (purple color scale) and total fishing hours in 2015–2016 (orange color scale; see text for details). Because of its wide extent in latitude, the Nile basin is reported separately. **b**, **c** are from Fig. S1 in the supplementary information of [10], used under CC BY Version 4.0

4.3 Plastic Pollution Originating from Coastal Population

Coastal particle inputs (Fig. 2b) are set proportionally to the estimated quantity of MPW produced by the populations living close to the coasts of the Mediterranean Sea. Each coastal source thus emits a number of particles proportional to the product between two factors: the country-specific estimate of per capita MPW generation rate, as described above, and the population inhabiting in a 5 km-wide strip along the coast from the Gridded Population of the World—Population Count 2015 (CIESIN, 2018) which has a spatial resolution of 2.5 min (approximately 5 km × 5 km).

4.4 Plastic Pollution Originating in River Watersheds

As for riverine sources (Fig. 2c), we include as model forcings only the top-100 strongest riverine inputs of plastic pollution among all the Mediterranean coastal watersheds to reduce the computational effort during the simulations. The proxy indicator of particle release from rivers accounts for MPW generated within coastal watersheds (from the satellite-derived HydroBASINS—HydroSHEDS products for Europe and Africa, [13]) and for seasonal variations in surface water runoff, as the hydrological regime strongly affects the presence of debris in river waters. It is obtained by summing over each basin the gridded product of (a) monthly surface runoff (from the gridded 2015–2016 monthly FLDAS dataset, [15]), (b) inhabiting population (again from [4]), and (c) country-specific estimated per-capita MPW generation rates, using the same gridding as the CIESIN population data (the finest among the three datasets involved in the operation). We then average the indicator over 2015–2016 and select as particle sources for our model the rivers ranking in the top-100, and modulate their particle releases during the simulation according to the monthly values of their proxy indicator.

4.5 Plastic Pollution from Fishing Activities

The release of particles from fishing-related sources (Fig. 2b) can be intuitively assumed proportional to the fishing effort at each location within the Mediterranean Sea. We retrieve daily fishing effort (expressed in fishing hours/day) from the Global Fishing Watch dataset (<https://globalfishingwatch.org/>) provided as rasters with a spatial resolution of $1/100^\circ$. This dataset originates from the collection and processing of data from several sources, including vessel positioning from their Automatic Identification Systems (AIS, required by the International Maritime Organization on large ships), and satellite imagery.

5 Results

In Fig. 3, the average PROP pollution in the Mediterranean Sea resulting from running the coupled Lagrangian–Eulerian simulation is unpacked into its two components: PROP dissolved in seawater (Fig. 3a) and PROP bound to particles (Fig. 3b). Interestingly, even under our conservative assumption that particles are released without any PROP, particle bound PROP concentrations do not exactly match the average distribution of PROP in seawater, meaning that microplastics are far from mere passive samplers in constant equilibrium with their surroundings. In fact, both the chemical properties of microplastics and particle history (including source type and location, other than journey at sea) influence the particle-bound concentration of the target

PROP, as this depends on the chemical gradients experienced by the particle during transport, and on the resulting mass transfer of PROP.

Polluted particles may in fact be transported far away from the source they were released from and finally be found polluted by PROPs at various and different levels, depending upon the PROP uptake/release kinetics between particles and seawater during the trip and on the timescales of particles' motion at sea (see Fig. 3 in [10]).

During their journey at sea, particles travelling through highly-contaminated waters might thus convey the PROP they carry to less-polluted areas. Lagrangian tracking permits to apportion the microplastic-mediated chemical pollution resulting from the coupling with Eulerian advection–diffusion of PROPs at sea to specific source types and countries of origin of microplastic particles, as in Fig. 4.

Importantly, our approach leverages, on the one hand, numerical schemes to model a complex, large-scale phenomenon, and Big Data to inform its forcings on the other. Thanks to real data, the modelled concentrations of PROPs and particles and the related PROP fluxes are not purely theoretical, but can be regarded instead as a first, verisimilar depiction of the chemical impacts of microplastic pollution in the Mediterranean seascape and of their sources.

6 Conclusions and Perspectives

The results of our novel Lagrangian–Eulerian simulations provide evidence that the presence of microplastics in Mediterranean seawater can affect the dispersal patterns of PROPs there, as microplastics can, during their transport by currents, effectively act as carriers of contaminants at sea, and are thus far from mere passive samplers in constant equilibrium with seawater PROP concentrations. This important knowledge provides insights about processes whose observation or measurement is complicated, or even impossible, to perform in an operational manner on field. Data from modelling experiments, albeit simplified like ours, can thus provide valuable information about recurrent pathways for microplastics at sea, their pollution level, and about areas with potentially higher pollution by particles and/or PROPs, and can do this on large spatial scales while ensuring high spatial resolution and continuous time series.

Comprehensive information about these phenomena can be crucial to deal with a set of pivotal tasks for marine conservation, such as the identification of marine regions exposed to higher ecological risks (as in [7]) and to provide valuable insights for planning basin-wide mitigation policies. Furthermore, the generality and adaptability of the Lagrangian–Eulerian modelling approach proposed here make it possible to be successfully deal not only with other geographic areas, potentially up to a global use, but also with similar problems arising when moving microplastic particles modify the properties of the surrounding marine environment.

It is important to highlight here that the major limitation faced by holistic modelling approaches like ours resides, at present, not just in the need to embrace more of the complexity of the issue of marine microplastic pollution and its related contaminants, but most importantly in the lack of suitable field data for detailing

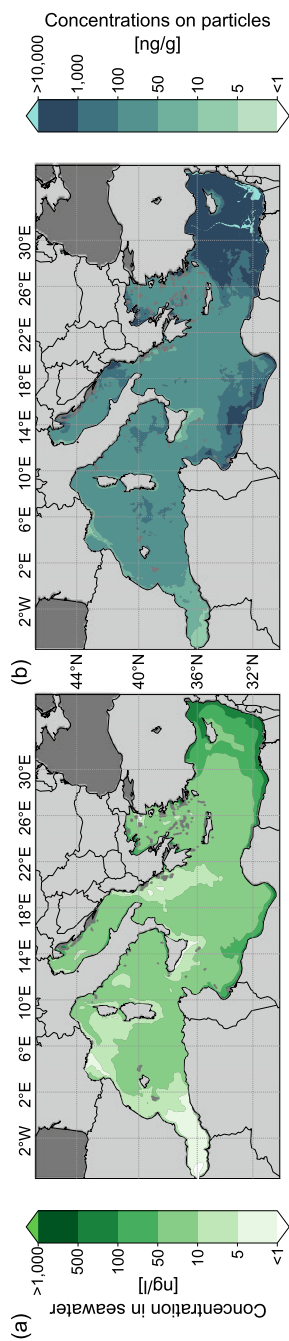


Fig. 3 Concentrations of the target PROP **a** in seawater and **b** on particles as resulting from our coupled Lagrangian–Eulerian model, temporally averaged simulation over 2015–2016. From Fig. 2 in [10], used under CC BY Version 4.0

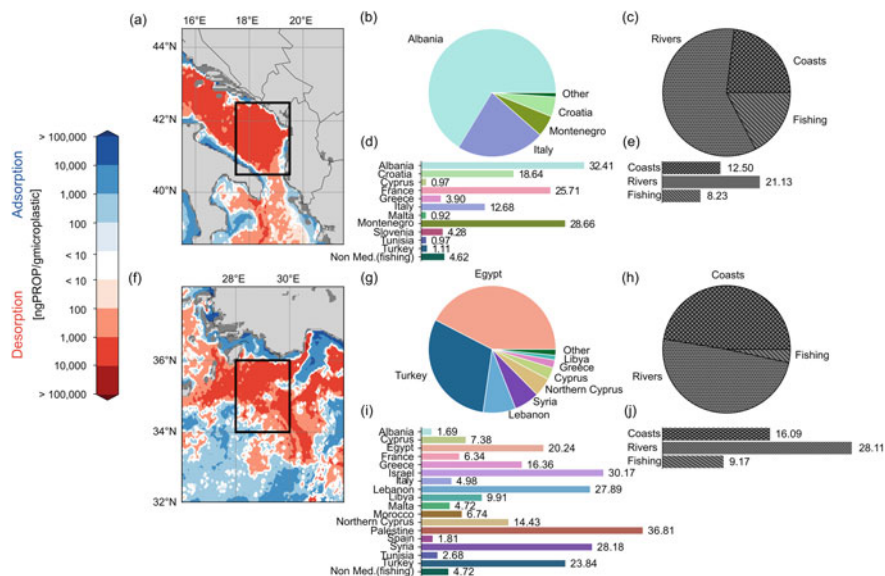


Fig. 4 Sourcing of desorbing particles and related PROP apportionment obtained from simulation for two exemplificative desorption areas, located in the southern Adriatic Sea (black frame in (a)) and in the northern Levantine Sea (black frame in (f)); **b** Apportionment of the particles desorbing in the southern Adriatic Sea by country of origin; **c** Source types of desorbing particles; **d** Average mass of PROP desorbed by each particle (in ngPROP /g_{particles}) by country of origin; **e** Same as in (d), by source type; **g–j** As in (b)–(e), for the desorption area in (f). (a) and (f) are from Fig. 5a, f in [10], used under CC BY Version 4.0

the inputs and for validating the model. In fact, even if we kept our Lagrangian and Eulerian parts of the model relatively simple and took assumptions based on available data, we still could not properly validate our model (although we could still perform some promising comparisons, see [9, 10]). Model validation needs a coherent, spatially and temporally-wide dataset including measurements of the concentrations of microplastics and PROPs in seawater and of microplastic-bound PROPs, that is currently unavailable (as also discussed in [10]). These knowledge gaps heavily influence quantitative modelling assessments of this kind, in term of both modelled seawater and particle-bound concentrations. Therefore, we argue that further model complexity will have to be supported by the availability of suitable data for validation.

Marine plastic pollution is a multifaceted and multiscale issue, as it presents physical, chemical, and biological implications for marine compartments at small scales while involving vast geographic domains. To counteract what can be regarded as a planetary boundary threat [17], we need tailored instruments to provide various layers of knowledge in an accessible manner. To this end, laboratory experiments and observations are indeed fundamental to capture specific aspects of plastic and plastic-related pollution; yet this zoo of data is still insufficient to provide a holistic description of this pervasive phenomenon. Quantitative methods, such as data informed

numerical modelling, can link and exploit our current understanding of the different processes and complex interaction characterizing marine plastic pollution to ultimately help to shed some light on aspects that are hard to investigate in the field. Actions against marine pollution have to be timely, bold, and knowledge-based. In this perspective, comprehensive, data-informed approaches can help not only to identify and protect the mechanisms that interconnect marine ecosystems but also to understand the processes and synergies that are disrupting them.

References

1. Arthur, C., Baker, J., Bamford, H.: Proceedings of the International Research Workshop on the Occurrence, Effects, and Fate of Microplastic Marine Debris. NOAA Technical Memorandum NOS-OR&R-30 (2008)
2. Berrojalbiz, N., Dachs, J., Ojeda, M.J., Valle, M.C., Castro-Jiménez et al.: Biogeochemical and physical controls on concentrations of polycyclic aromatic hydrocarbons in water and plankton of the Mediterranean and Black Seas. *Global Biogeochem. Cycles* **25**(4), 1–14 (2011). <https://doi.org/10.1029/2010GB003775>
3. Brett, A., Leape, J., Abbott, M., Panel, H.L., Ocean, S., Blue, E.: Ocean data need a sea change to help navigate the warming world. *Nature* **581**, 181–183 (2020). <http://odis.iode.org>
4. Center for International Earth Science Information Network—CIESIN: Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11. NASA Socioeconomic Data and Applications Center (SEDAC) (2018). <https://doi.org/10.7927/H49C6VHW>
5. Cózar, A., Sanz-Martín, M., Martí, E., González-Gordillo, J.I., Ubeda, B., Ágálvez, J., Irigoien, X., Duarte, C.M.: Plastic accumulation in the Mediterranean Sea. *PLoS ONE* **10**(4), 1–12 (2015). <https://doi.org/10.1371/journal.pone.0121762>
6. Fischer, H.B., List, E.J., Koh, R.C.Y., Imberger, J., Brooks, N.H.: *Mixing in Inland and Coastal Waters*. Academic Press, USA (1979)
7. Guerrini, F., Mari, L., Casagrandi, R.: Modeling plastics exposure for the marine biota: risk maps for fin whales in the Pelagos Sanctuary (North-Western Mediterranean). *Front. Marine Sci.* **6**, 1–10 (2019). <https://doi.org/10.3389/fmars.2019.00299>
8. Guerrini, F., Mari, L., Casagrandi, R.: Microplastic as a vector of chemical contamination in the marine environment: a coupled Lagrangian-Eulerian approach. *Global Oceans 2020: Singapore—U.S. Gulf Coast* (2020). <https://doi.org/10.1109/IEEECONF38699.2020.9389405>
9. Guerrini, F., Mari, L., Casagrandi, R.: The dynamics of microplastics and associated contaminants: data-driven Lagrangian and Eulerian modelling approaches in the Mediterranean Sea. *Sci. Total Environ.* **777**, 145944 (2021). <https://doi.org/10.1016/j.scitotenv.2021.145944>
10. Guerrini, F., Mari, L., Casagrandi, R.: A coupled Lagrangian-Eulerian model for microplastics as vectors of contaminants applied to the Mediterranean Sea. *Environ. Res. Lett.* **17**(2) (2022). <https://doi.org/10.1088/1748-9326/ac4fd9>
11. Jambeck, J.R., Geyer, R., Wilcox, C., Siegler, T.R., Perryman, M., Andrady, A., et al.: Plastic waste inputs from land into the ocean. *Science* **347**(6223), 764–768 (2015). <https://doi.org/10.1126/science.1260879>
12. Kaandorp, M.L.A., Dijkstra, H.A., van Sebille, E.: Closing the Mediterranean marine floating plastic mass budget: inverse modeling of sources and sinks. *Environ. Sci. Technol.* (2020). <https://doi.org/10.1021/acs.est.0c01984>
13. Lehner, B., Grill, G.: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems. *Hydrol. Processes* **27**(15), 2171–2186 (2013). www.hydrosheds.org

14. León, V.M., García, I., González, E., Samper, R., Fernández-González, V., et al.: Potential transfer of organic pollutants from littoral plastics debris to the marine environment. *Environ. Pollut.* **236**, 442–453 (2018). <https://doi.org/10.1016/j.envpol.2018.01.114>
15. McNally, A.: NASA/GSFC/HSL (2018) FLDAS Noah Land Surface Model L4 Global Monthly Climatology 0.1 x 0.1 degree (MERRA-2 and CHIRPS)
16. Micheli, F., Halpern, B.S., Walbridge, S., Ciriaco, S., Ferretti, F., Frascchetti, S. et al.: cumulative human impacts on Mediterranean and black sea marine ecosystems: assessing current pressures and opportunities. *PLoS ONE* **8**(12) (2013). <https://doi.org/10.1371/journal.pone.0079889>
17. Rockström, J., Gupta, J., Lenton, T.M., Qin, D., Lade, S.J., Abrams, J., et al.: Identifying a safe and just corridor for people and the planet. *Earth's Future* **9**(4), 1–7 (2021). <https://doi.org/10.1029/2020EF001866>
18. Simoncelli, S., Fratianni, C., Pinardi, N., Grandi, A., Drudi, M., Oddo, P. et al.: Mediterranean Sea Physical Reanalysis (MEDREA 1987–2015) (Version 1). In E.U. Copernicus Marine Service Information (2014). https://doi.org/10.25423/medsea_reanalysis_phys_006_004
19. Strang, G.: On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.* **5**(3), 506–517 (1968)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Enhancing the Quality of Human-Robot Cooperation Through the Optimization of Human Well-Being and Productivity



Costanza Messeri 

Abstract In human-robot interaction frameworks maximizing the team efficiency is crucial. However, it is also essential to mitigate the physical and cognitive workload experienced by the shop-floor worker during the collaborative task. In this chapter we first investigate the impact of the robot interaction role (whether being leader or follower during cooperation) on both the human physiological stress and production rate. Based on that, a game-theoretic approach is proposed to model the trade-off between the maximization of the human performance and the minimization of the human cognitive stress. Then, we describe a closed-loop robot control strategy that, based on the proposed game-theoretic model, enables the robot to simultaneously minimize the human cognitive stress and maximize his/her performance during cooperation, by adjusting its role. Eventually, a real-time task allocation strategy is proposed to both ensure the minimization of the human physical fatigue and the effectiveness of the production process. This method relies on a new sophisticated musculoskeletal model of the human upper-body. All these methodologies have been experimentally tested in realistic human-robot collaborative scenarios involving several volunteers and the ABB IRB 14000 dual-arm “YuMi” collaborative robot.

Keywords Human-robot collaboration · Stress and fatigue estimation · Task allocation and optimization

1 Introduction

Historically, industrial manipulators had been designed to easily automate the execution of repetitive or dangerous tasks and worked in dedicated cages (denoted as robotic cells) for safety reasons. The birth of the technological revolution known

Supported by organization Politecnico di Milano.

C. Messeri (✉)

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via Ponzio 34/5, 20133 Milano, Italy
e-mail: costanza.messeri@polimi.it

© The Author(s) 2023

C. G. Riva (ed.), *Special Topics in Information Technology*,

PoliMI SpringerBriefs, https://doi.org/10.1007/978-3-031-15374-7_2

as ‘Industry 4.0’ has radically changed the global industrial scenario and also the field of robotics, making the newborn collaborative robot (cobot, in short) one of the enabling technologies of the smart factory. This is a lighter and smaller manipulator, specifically designed to help the shop-floor worker accomplish several industrial tasks, by safely working in close proximity to him/her. The human-robot cooperation (HRC) fosters the efficiency and flexibility of the industrial production systems, since it allows to effectively combine the human advanced cognitive skills and adaptability with the cobot superior accuracy and repeatability. Therefore, whereas cobots can relieve shop-floor workers from alienating or tiring tasks, humans can indulge in high-level decision-making operations that can be hardly automated. Nowadays cobots are widely employed in small- and medium-sized enterprises (SMEs) for collaborative manufacturing tasks [3].

The problem of improving the effectiveness of the HRC, i.e. minimizing the task cycle time or maximizing the productivity, has always played a pivotal role in the related scientific research. By the way, other aspects of HRC are recently emerging as crucial features to improve the quality of the cooperation. An example is the prevention or mitigation of the cognitive and physical distress that might be experienced by the shop-floor worker [1]. Indeed, humans by nature have a limited capability to endure fatigue and are also vulnerable to workload. Furthermore, high levels of physical and cognitive fatigue negatively affect the human capabilities, increasing not only the risk of work-related musculoskeletal disorders (MSDs) or burn-out phenomena, but also undermining the team performance [6]. Recently, some studies have investigated the possibility of adapting certain robot behavioral features according to the operator’s mental or physical distress, to help him/her accomplish the collaborative task, or to reach a smoother interaction [5, 7, 9]. However, the existing literature in HRC has focused either on the mitigation of the human physical and cognitive workload or on the optimization of the human productivity, without jointly addressing the problem of simultaneously optimizing both human performance and well-being. Moreover, non-invasive methods to estimate the fatigue accumulated by the human during cooperation are at their early stages in the literature.

In this chapter we first analyze how the robot interaction role (being leader or follower during cooperation) influences the human stress and productivity. Based on that, a game-theoretic (GT) model of the trade-off between stress and performance is proposed. Then, based on the real-time monitoring of this trade-off, we describe a novel control strategy that enables the cobot to simultaneously optimize online the human well-being and performance, by suitably varying its role. Moreover, we propose a strategy to estimate in real-time and non-invasively the work-related fatigue experienced by the human co-worker during HRC. Based on that, we describe a method that dynamically decides which task activities should be better allocated to the cobot and which ones to the human to minimize his/her musculoskeletal workload. The remainder of this chapter is structured as follows. Section 2 analyzes how the robot interaction role influences the stress and productivity of the human co-worker. Section 3 describes the GT-based robot strategy to simultaneously optimize the human performance and well-being. Eventually, Sect. 4 addresses the dynamic task allocation strategy to mitigate the physical fatigue accumulated by the human.

2 Understanding the Impact of the Robot Interaction Role on the Human Physiological Stress and Productivity

A pivotal feature of HRC is represented by the robot interaction role, i.e. being leader or follower during cooperation. Whereas in the first case the robot is responsible for driving the task progression, in the second case, it complies with the human decision. To our best knowledge, previous works in the literature have mainly analyzed the feasibility of the leader-follower paradigm in HRC. However, existing researches do not provide a complete analysis on how the robot leader-follower interaction modes might impact on the human co-worker in terms of stress and performance. To tackle this gap we carried out an exploratory study aiming at examining whether and how the leader-follower robot interaction strategy influences the productivity and physiological (cognitive) stress of the human co-worker. The ultimate goal of this research is to lay the foundations for determining how the robot should behave when interacting with the generic human co-worker in order to obtain the best compromise between the mitigation of his/her cognitive stress and the maximization of his/her production performance.

To perform this evaluation, we setup a task consisting in a tower-building exercise, which was jointly performed by the human and the robot (ABB YuMi cobot). They both had to place a box on the top of a stack moving alternately, with the sole constraint that a new block must be of the same color of the previous one. The leader agent was in charge of initiating the task and driving, at each step, the selection of the color of the next box. Conversely, the follower agent had to react by complying with the partner's choice. To carry out the analysis, a between-subjects design was adopted: the overall population of 33 volunteers¹ was divided into two similar and homogeneous groups. The subjects belonging to the first group were leaders during the interaction with the robot, whereas those of the second group had to follow the task progression decided by the robot. In both cases the robot performed the same motion trajectory to reach each box and the volunteers were asked to perform the task as fast as possible. We evaluated the physiological stress induced on each subject based on the analysis of the most reliable stress indicators according to the related literature, i.e., LF/HF and RMSSD [8]. These indicators are extracted by analyzing the heart-rate variability (HRV) of the subject's electrocardiographic (ECG) signal. The latter was acquired on each user via 3 disposable electrodes placed on the subject's thorax and abdomen. The subject's production rate was instead estimated based on the measured user's cycle time.

The results showed that, whereas the robot leadership entails a higher productivity of the team at the expense of an increase in the physiological stress, the human leadership induces lower stress but also determines a lower productivity.

¹ This research was approved on July 2019 by the Ethical Committee of the Department of Psychology of Università Cattolica del Sacro Cuore.

3 Real-Time GT-Based Method to Maximize the Human Performance and Mitigate the Cognitive Stress in HRC

In Sect. 2 the ‘open-loop’ effects of the robot interaction role on the human productivity and stress have been presented. The expression ‘open-loop’ indicates that the effects of the robot role are simply evaluated, but not exploited to vary the robot behavior accordingly (i.e. in closed-loop). The results of this exploratory study have highlighted two important aspects. The first is that there exists a trade-off between the maximization of human productivity and the minimization of the related cognitive stress. The second is that the robot interaction role might serve as a suitable control variable to effectively manage this trade-off. Based on these outcomes, we developed a novel closed-loop control strategy that makes the robot able to simultaneously stimulate the human productivity and mitigate the cognitive stress, by autonomously adapting online its role. Hence, this strategy allows to jointly optimize both human performance and well-being. The difficulties in developing such an optimization strategy are manifold:

- the dynamics of the human response (in terms of stress and performance) to a certain external stimulus (i.e. stressor), such as the variation of the robot behavior, is neither a priori known nor predictable;
- due to the intrinsic complexity of the stress regulatory mechanism, the stress response cannot be considered a fully observable phenomenon and thus cannot be properly modeled through mathematical equations;
- the human response to an external stressor is subject-specific;
- in the related literature, no guidelines are provided to determine the optimal stress-performance state that a certain human being can achieve;
- the relationship between the dynamics of stress and performance is not known and varies from individual to individual.

For these reasons, a model-based control strategy would have been unsuitable. To realize the optimization strategy, the human stress s_k and performance p_k are continuously monitored by the robot through the following statistics:

$$s_k = -\left(\frac{LF_k/HF_k}{RMSSD_k}\right) \quad p_k = -(\bar{t}_k + \sigma_k) \quad (1)$$

where \bar{t}_k and σ_k are the average human cycle time and related standard deviation, respectively. To express the existing trade-off between the minimization of the human stress and the maximization of his/her production performance, a GT-based approach has been developed. In other words, the game theory provides the theoretical foundations for modeling how the interaction between human and robot is influencing the stress and performance of the human co-worker. In particular, to realize the model of the above mentioned trade-off, we propose to interpret the overall interaction between the human and the robot as a repeated non-cooperative normal-form game between two players [10], i.e., the human (H) and the robot (R). Moreover, we assume the players H and R to be rational (self-interested), i.e. each one of them

aims at maximizing its own profit throughout the game. Hence, these two players allows us to account for the two competing aspects of whatever working environment: on the one hand, the importance of increasing the productivity, on the other hand the need of mitigating the worker stress. Thus, we consider that the goal of player R is to maximize the productivity and the one of player H to minimize the stress. Then, we exploit the concept of game actions [10] to express the potential attitudes of human and robot during the interaction. More specifically, we assume that each player might adapt to the other player (i.e. playing action D), if it meets the goal of the other player, or do not adapt (i.e. playing action ND), in the opposite case. Hence the GT-based model makes it possible to account for the admissible states that can be achieved during HRC. In the GT framework these states are denoted as the outcomes (of the game) which result from the combination of two simultaneous players' game actions. Besides, according to the principles of GT, each outcome is associated with a pair of values, known as utility functions (or payoffs), each of which expresses how much the related player is satisfied with that outcome. In the considered case, at each step, the payoffs of player H and R associated with the current game outcome are updated with the measured s_k and p_k , respectively. Thus, the proposed HRC game model can be exploited to evaluate the admissible compromises in terms of stress and performance, that might result from the interaction between the robot and the specific human co-worker. This formulation allows us to identify a particular outcome of the game, known as Nash Equilibrium (NE), that is a local equilibrium state of the game and represents a "no-regrets" situation for the players. Indeed, by definition, an outcome of the game is a NE if, once reached, no player has incentive to unilaterally change its action to increase its payoff. So, in the proposed formulation, the utility functions associated with the NE are considered the stress-performance equilibrium trade-off against which to locally evaluate the quality of the current stress-performance compromises that are resulting from the HRC and thus realize the optimization. Based on these evaluations, a Learning Automaton (LA) [4], which is the full-fledged decisional part of the proposed control strategy, is exploited to command the robot how to adjust its interaction role to optimize this trade-off. More specifically, at each step, we assume that the robot can be commanded to apply one of the following actions: α_F or α_L . The first indicates that the robot has the role of follower during cooperation and, as such, it has to synchronize with the production rhythm dictated by the human. The latter, instead, indicates that the robot becomes the leader of the cooperation and thus increases the production pace to stimulate the human. In other words, the GT model is used to evaluate how the robot action affects the human co-worker in terms of stress and performance. The LA, by rewarding or penalizing the robot action proportionally to its positive or negative effect on the human, is then able to determine the best next action (interaction role) that the robot has to do to optimize the stress-performance state of the human co-worker.

The proposed method was validated in a realistic collaborative assembly task involving the ABB YuMi cobot and 15 volunteers.² During the experimental campaign, the cycle time and the ECG signal of each subject were acquired as described in Sect. 2. By the way, in this case, each user was asked to perform a collaborative task consisting in assembling the components of a box that exemplifies an electrical circuit. Even in this case a between-subjects design was adopted. The subjects were evenly distributed by gender and age into 3 groups: NC, CP and CPS. Each group was associated with a different testing condition, i.e. a different robot behavior. However, the experimental protocol was the same for all groups. Volunteers of group NC (i.e. No Control) were leaders of the cooperation, namely the robot simply followed their production pace without optimizing productivity or stress. Volunteers of group CP (i.e. Control Productivity) tested a strategy where the robot always dictated the production rhythm to solely stimulate the productivity, regardless of the subject's stress. Eventually, group CPS (i.e. Control Productivity and Stress) experienced the full strategy just described where the robot adjusted its role, and thus the production pace, to jointly optimize the human well-being and productivity. In view of the outcomes of the study described in Sect. 2, strategy CP, which coincides with the robot being leader of the HRC, is considered the reference situation in terms of maximum productivity (throughput) that can be achieved by the worker. Strategy NC, which coincides with the human being leader of the HRC, is regarded as the reference situation in terms of minimum stress that the worker can experience during cooperation.

3.1 Results

The results of the experimental campaign and the related statistical analysis highlighted that, for what concerns the productivity (see Fig. 1b), the average throughput of group CP and CPS could not be considered statistically different. Thus the proposed strategy (CPS) is able to maximize the human productivity. Regarding the stress (see Fig. 1a), the average ECG-based stress statistic of Group NC and of Group CPS could not be considered statistically different. Hence, the proposed method is also able to minimize the human stress. To conclude, the novel control strategy, by suitably adjusting the robot role during cooperation, effectively enhances the productivity of the human-robot team, while significantly mitigating the work-related cognitive stress.

² This research was approved on November 19, 2020 by the Ethical Committee of Politecnico di Milano.

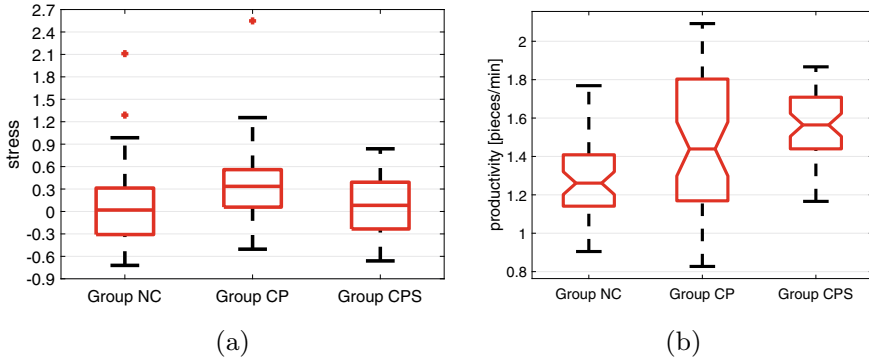


Fig. 1 a Relative ECG-based stress statistics obtained for volunteers of Group NC, CP and CPS. b Productivity rate obtained for volunteers of Group NC, CP and CPS

4 A Dynamic Task Allocation Strategy to Mitigate the Human Physical Fatigue During HRC

Besides the mitigation of the cognitive stress, it is becoming increasingly important to develop methods that also mitigate the physical workload undergone by the shop-floor worker. Indeed, in the last few years, the incidence rate of MSDs cases in the manufacturing industry has registered very high levels. In these frameworks, typically, the human operator is involved in several repetitive manual activities that might overload his/her musculoskeletal apparatus, undermining his/her health (risk of MSDs) and work-related performance.

In this section we present a novel dynamic task allocation strategy aiming at relieving the physical workload accumulated by the worker during cooperation. Indeed, given a certain task composed by a number of activities that human and robot have to do, the proposed method dynamically decides which task activities should be better allocated to the robot and which ones to the human to minimize his/her accumulated fatigue. This dynamic allocation method is based on the possibility of evaluating in real-time the muscle fatigue associated with each human movement. The proposed method assumes that, as it typically happens, the collaborative workstation is equipped with an external vision system that identifies the positions of a set of relevant points along the human silhouette (so-called ‘skeletal positions’) and thus tracks the human motions. To evaluate the fatigue associated with the human motions, we first developed a new musculoskeletal model of the human upper-body including the full musculoskeletal description of the two arms, the neck, and the most relevant muscles of thorax/back in the OpenSim³ digital human modeling software. In the developed model, we also placed a set of virtual markers in correspondence to the locations of the skeletal positions retrieved by the vision system. This makes it possible to replicate on the virtual human model the real human motions and evaluate

³ <https://simtk.org/projects/opensim>.

the related muscle effort. The latter is evaluated through the OpenSim Static Optimization (SO) technique which allows to estimate the muscle activations⁴ that fulfill positions, velocity and acceleration of a certain movement. So, the SO technique can be interpreted as an extension of the inverse dynamics method that, at each time step, transforms the net joint moments of the model into estimates of single muscle forces, by minimizing the sum of the squared muscle activations. However, the SO tool is not fast enough to be used online. To solve this problem, we artificially generated a huge, generic motion dataset (i.e. input) and we exploited the SO method offline to retrieve the related muscle activations (i.e. output). Then, we offline trained a Deep Neural Network (DNN) to learn the underlying mapping function between the above-mentioned input-output dataset. To reduce the complexity of this mapping, the upper-body muscles have been clustered in 6 groups according to the body joint they contributed to actuate. The groups thus obtained (from now on called joints) are: head, shoulders (left and right), elbows (left and right) and trunk. Hence, the DNN is used online to estimate the muscle activations associated with the human motions tracked during the execution of each activity. Then, given the average joints activations and the duration of each task activity, a joint-specific model denoted as Power Model [2] is used to estimate the fatigue experienced by the considered joint after the execution of the current activity. Indeed, the Power Model relates the muscle activation to the activity duration (which corresponds to the effort endurance time) and provides a curve that identifies, with increasing muscle activation, the maximum time before the considered joint experiences overloading. Then, the fatigue undergone by each joint is estimated as the product of the muscle activation and activity duration. An estimate of the maximum fatigue that can be sustained by the joint before overloading is then obtained as the product of the maximum activation and the activity duration. By computing the ratio between the estimated fatigue quantities, a relative indicator expressing the actual fatigue experienced by the joint with respect to the maximum fatigue that can be sustained for that activity is obtained. Given the above, in the proposed formulation, each i -th task activity is then associated with a vector, $\hat{\mathbf{f}}^i$, of 6 features that describes the fatigue which each joint is subject to at the end of that activity. An estimate of the actual fatigue accumulated by the worker, $\hat{\mathbf{f}}_{acc}$, due to the sequence of activities done so far is then obtained by applying the Exponentially Weighted Moving Average (EWMA) filter to the fatigue vectors associated with the last N activities performed by the human. To suggest the human the best activity to do to minimize his/her accumulated fatigue, we propose to evaluate how much the execution of the i -th activity would increase the norm of $\hat{\mathbf{f}}_{acc}$. This contribution is computed for each i -th activity as follows:

$$c^i(\hat{\mathbf{f}}^i, \hat{\mathbf{f}}_{acc}) = \|\hat{\mathbf{f}}^i\|_2 \left(\frac{\|\hat{\mathbf{f}}^i \cdot \hat{\mathbf{f}}_{acc}\|}{\|\hat{\mathbf{f}}^i\|_2 \|\hat{\mathbf{f}}_{acc}\|_2} \right) = \frac{\|\hat{\mathbf{f}}^i \cdot \hat{\mathbf{f}}_{acc}\|}{\|\hat{\mathbf{f}}_{acc}\|_2} \quad \forall i = 1, \dots, Nt \quad (2)$$

⁴The muscle activation is defined as the amount of force that a muscle actively produces with respect to the maximum capability of actively producing force.

where $\|\cdot\|_2$ indicates the vector L^2 norm, $|\cdot|$ the absolute value, \cdot the scalar product and Nt is the number of task activities. Then, the human is suggested to do the activity α_k^H that minimizes this contribution, whereas the robot is commanded to do the activity α_k^R that maximizes it (see Eq. (3)).

$$\alpha_k^H = \min_i \{c^1\{\hat{f}^1, \hat{f}_{acc}\}, \dots, c^{Nt}\{\hat{f}^{Nt}, \hat{f}_{acc}\}\} \quad \alpha_k^R = \max_i \{c^1\{\hat{f}^1, \hat{f}_{acc}\}, \dots, c^{Nt}\{\hat{f}^{Nt}, \hat{f}_{acc}\}\} \quad (3)$$

The proposed strategy was validated in a realistic collaborative task, composed by 4 packaging activities. The task involved the ABB YuMi cobot and 14 volunteers.⁵ During each activity the user was requested to perform manual handling of small and lightweight loads at high frequency. However the execution of the activities mainly entailed the activation of different body regions. For the experimental campaign a between-subject design was exploited: volunteers were evenly distributed by gender and age into two homogeneous groups, i.e. group S and group D. The latter experienced the proposed dynamic allocation strategy, while group S tested a static allocation of the task activities consisting in six consecutive repetitions of a fixed sequence of all the activities. The proposed dynamic allocation method requires an initial (online) calibration phase during which the user is requested to do two consecutive repetitions of the sequence of activities 1–2–3–4 (i.e. ‘baseline’ phase).

4.1 Results

The results of the experimental campaign and the related statistical analysis highlighted that the static allocation strategy entails an increase of the average accumulated fatigue by about 7% with respect to the fatigue accumulated during the baseline (see Fig. 2a). Conversely, the dynamic allocation strategy does not increase the average accumulated fatigue (w.r.t. the baseline). Thus, the proposed method is effective in mitigating the human accumulated fatigue. For what concerns the production performance, evaluated in terms of cycle time (see Fig. 2b), the results of the experimental campaign showed that the dynamic method entails a reduction of the average cycle time by more than 16% with respect to the static method. Hence, the proposed strategy is also effective in improving the user performance.

5 Conclusions

In this chapter we first presented an exploratory study aimed at analyzing how the robot interaction role influences the cognitive stress and production performance of the human co-worker. This study showed that the robot leadership entails a higher

⁵ This research was approved on November 19, 2020 by the Ethical Committee of Politecnico di Milano.

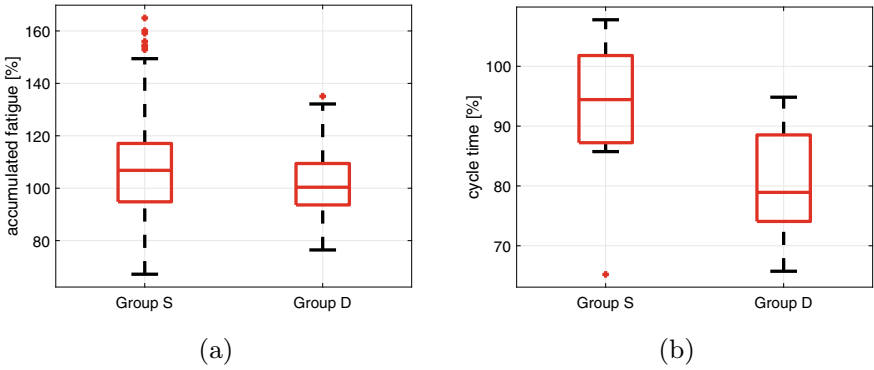


Fig. 2 **a** Relative fatigue accumulated by the volunteers of Group S and Group D. **b** Relative cycle time obtained for volunteers of Group S and Group D

human productivity, but also a higher human stress with respect to the case where the human takes the lead. This highlights the existing trade-off between the maximization of the human productivity and the mitigation of the human stress and opens the door for the development of an online closed-loop robot control strategy aiming at simultaneously optimizing both aspects. This strategy was then described in Sect. 3. The proposed method exploits game theory to model the above-mentioned trade-off and evaluate how the robot behavior (role) is influencing the stress and performance of the human co-worker. A control unit, represented by a learning automaton, was used to reward or penalize the robot behavior proportionally to its positive or negative effect on the human. Through that, the robot was enabled to autonomously determine the best action to do (which role to take on) to optimize the stress-performance state of the human co-worker. The proposed strategy turned out to effectively increase the human performance while significantly mitigating the stress, by solely adjusting the robot production rhythm according to the specific human co-worker. Eventually, we proposed an innovative dynamic task allocation strategy that, by relying on a novel human upper-body model, estimates online and non-invasively the muscle fatigue accumulated by the user during each activity and suggests the human the best next activity to do to minimize his/her accumulated fatigue. By optimally distributing the admissible task activities between human and robot, the method effectively reduced the physical workload accumulated by the worker and improved his/her performance.

References

1. Arai, T., Kato, R., Fujita, M.: Assessment of operator stress induced by robot collaboration in assembly. *CIRP Ann.* **59**(1), 5–8 (2010)

2. Frey Law, L.A., Avin, K.G.: Endurance time is joint-specific: a modelling and meta-analysis investigation. *Ergonomics* **53**(1), 109–129 (2010)
3. Matheson, E., Minto, R., Zampieri, E.G.G., Faccio, M., Rosati, G.: Human-robot collaboration in manufacturing applications: a review. *Robotics* **8**(4), 100 (2019)
4. Narendra, K.S., Thathachar, M.A.: Learning automata—a survey. *IEEE Trans. Syst. Man Cybern.* **4**, 323–334 (1974)
5. Nikolaidis, S., Lasota, P., Ramakrishnan, R., Shah, J.: Improved human-robot team performance through cross-training, an approach inspired by human team training practices. *Int. J. Robot. Res.* **34**(14), 1711–1730 (2015)
6. Paletta, L., Pszeida, M., Nauschnegg, B., Haspl, T., Marton, R.: Stress measurement in multi-tasking decision processes using executive functions analysis. In: *International Conference on Applied Human Factors and Ergonomics*, pp. 344–356. Springer (2019)
7. Peternel, L., Tsagarakis, N., Caldwell, D., Ajoudani, A.: Robot adaptation to human physical fatigue in human-robot co-manipulation. *Auton. Robot.* **42**(5), 1011–1021 (2018)
8. Shaffer, F., Ginsberg, J.: An overview of heart rate variability metrics and norms. *Front. Public Health* **5**, 258 (2017)
9. Villani, V., Sabattini, L., Secchi, C., Fantuzzi, C.: A framework for affect-based natural human-robot interaction. In: *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 1038–1044. IEEE (2018)
10. Weiss, G.: *Multiagent Systems* (ed.). The MIT Press, Cambridge (2013)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Dynamic Sediment Connectivity Modelling for Strategic River Basin Planning



Marco Tangi 

Abstract Sediment connectivity is a distributed property of river systems that emerges from the connected transfer of sediment between multiple sources and sinks. Its disruption, brought by anthropic disturbances, can have severe and unforeseen consequences on both fluvial ecosystems and human livelihood. Modeling network-scale sediment connectivity provides a foundational understanding of river processes and their response to new pressures and can be used to forecast future system evolutions. In this chapter, we present the basin-scale, dynamic sediment connectivity model D-CASCADE (Dynamic CATCHment Sediment Connectivity And DELivery), which quantifies spatiotemporal patterns of sediment delivery in river networks. D-CASCADE considers multiple factors affecting transport, including heterogeneities in hydrology and sediment supply, different grain sizes, channel morphological evolution, and reservoir presence and management. The model is designed to be flexible, data parsimonious, and computationally efficient. We also present two applications of D-CASCADE in real-world case studies for historic geomorphic evolution reconstruction and future dam impacts forecasting. D-CASCADE is intended for integrated, basin-scale water management efforts, to perform multiple screening of various decision portfolios for hydromorphological impact assessments.

Keywords River basin management · Sediment connectivity · Sediment processes modeling

1 Introduction

River sediment connectivity is defined as the connected transfer of solid material mediated by water from erosion areas in a river catchment to deposition areas. Human presence and activities have deeply affected natural river sediment connectivity,

M. Tangi (✉)

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via Ponzio 34/5, 20133 Milano, Italy
e-mail: marco.tangi@polimi.it

© The Author(s) 2023

C. G. Riva (ed.), *Special Topics in Information Technology*,

PoliMI SpringerBriefs, https://doi.org/10.1007/978-3-031-15374-7_3

whether directly, e.g., reservoirs and barriers construction, channels reshaping, sand and gravel mining, or indirectly, e.g., catchment land-use change, climate change, or water withdrawal for industrial, domestic or agricultural uses. Alterations in sediment delivery may have long-lasting and profound impacts on river systems, including the decline in fluvial biodiversity and floodplain fertility, delta land loss, and bank instability. Given the interconnected and distributed nature of sediment connectivity in rivers, multiple alterations may have cumulative effects, and these impacts may be displaced in space and time from their causes [11, 17].

Characterizing basin-scale sediment connectivity is paramount to improving our capacity to quantify possible future alterations [9, 10]. However, modeling such processes at the large temporal and spatial scales necessary for basin-scale decision-making is often challenging, due to their complexity and the lack of available datasets for initialization and validation. The recent increase in the availability of large-scale hydro-geomorphological datasets from remote sensing [3] has led to the development of new numerical and conceptual sediment connectivity models. These tools exchange part of the local-scale accuracy found in more traditional, physic-based morphodynamic models to reconstruct sediment delivery and transport patterns over the entire river network and for longer timeframes [4, 6, 13, 16, 19].

In this chapter, we introduce D-CASCADE (Dynamic CAtchment Sediment Connectivity And DELivery), a dynamic, network-based sediment connectivity model for quantifying spatiotemporal patterns of sediment supply and delivery across a basin [15]. The model uses a 1-D graph structure to represent the river system, and traces the position of individual sediment transport processes called *cascades* over time across this network.

By considering the movement of multiple cascades and the interactions between each other and the river channels, D-CASCADE can effectively characterize network sediment connectivity, its dynamic evolution over time, and its response to multiple heterogeneous drivers of change.

Add-on components can be included in the framework to model local processes that the 1D structure of D-CASCADE cannot capture. These processes may include 2D and 3D simulations of the morphological response of river channels to sediment delivery alterations, e.g., channel accretion or erosion or floodplain interactions, or the effects of human infrastructures like reservoirs on the fluvial system.

D-CASCADE is adaptable to different research purposes, output requirements, and data availability. In particular, the model is designed to be integrated with multi-objectives, optimization-based planning, and management frameworks, to derive sediment connectivity indicators and identify trade-offs and synergies between conflicting objectives.

The chapter is organized as follows. Section 2 describes the D-CASCADE model structure, core principles and components. Section 3 showcases two applications of D-CASCADE: on the Bega river system in Australia [15], and on the 3S river system in South East Asia. The two applications differ widely in terms of spatial and temporal scales, network hydromorphological features, data availability, and research objectives. Section 4 draws some concluding remarks and highlights future research opportunities.

2 Model Structure

In D-CASCADE, sediment transport is modeled by routing and tracing cascades as they move through the network over time. Cascades are identified by the reach where they were initially delivered or stored, and the total sediment volume transported [m³], subdivided into different grain size classes.

The structure, core components, and procedures of the D-CASCADE model are shown in Fig. 1. The model setup is divided into two phases: initialization and main D-CASCADE loop.

2.1 Initialization

The initialization phase contains all the operations necessary to apply D-CASCADE on a fluvial system. The model’s routing is performed on a river network characterized

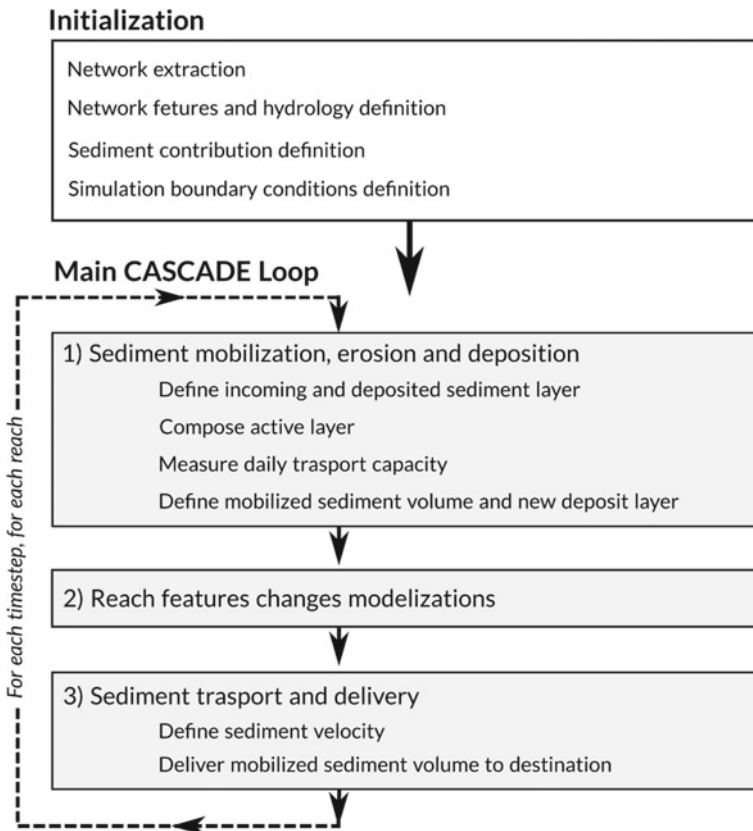


Fig. 1 D-CASCADE framework modelling steps. Figure from Tangi et al. [15]

as a direct, acyclic graph composed of nodes and reaches (Fig. 2a). The reach is the core modeling unit of D-CASCADE and represents a stretch of the river network with homogeneous or semi-homogeneous hydro-geomorphological properties. The level of detail in the network partitioning can be tailored according to data availability and computational effectiveness.

Each reach must be characterized by a unique set of parameters representing hydro-geomorphological features, e.g., channel active width, gradient, bed composition and roughness, and hydrological conditions such as discharge, water depth, and velocity.

Depending on the knowledge of the case study system and the data availability, a feature can be characterized as static, dynamic, or modeled. Static features do not change over time and are defined once for each reach. Dynamic features may change in each timestep, and must be specified as a time-varying data series. Modeled features are automatically updated in each timestep by specific add-on components, so they just require initialization for the first timestep.

Sediment volumes routed through the network may be derived from material initialized in the reaches as a sediment deposit or as external, user-defined sediment contributions supplied to individual reaches in specific timesteps and at specific rates, e.g., to reproduce sediment delivery from the hillslopes.

During initialization, the user defines the sediment classes to be considered in the analysis, according to the characteristic of the river system of the river and the level of detail desired.

2.2 *Main D-CASCADE Loop*

D-CASCADE model structure consists of two nested loops. The inner loop contains operations to determine cascade mobilization, deposition, and routing for each network reach, as well as all other operations modelled by add-on components. The outer loop repeats such operations for the entire simulation timeframe.

As shown in Fig. 1, the reach D-CASCADE loop is comprised of a series of operations repeated for each reach: (i) sediment mobilization, erosion and deposition, (ii) geomorphic features changes modelization, (iii) and sediment transport and delivery.

Figure 2 summarizes the steps necessary to define the mobilized sediment in each reach for each timestep. D-CASCADE traces the evolution over time of each reach deposit layer, conceptualized as a series of tiers stacked on top of each other, with each tier comprised of a single cascade. Incoming cascades, instead, can be comprised of material delivered by upstream reaches or external sediment inputs defined by the user (Fig. 2b). The active layer thickness in Fig. 2c must be defined by the user and may depend on the type of material composing the uppermost layer of the deposit.

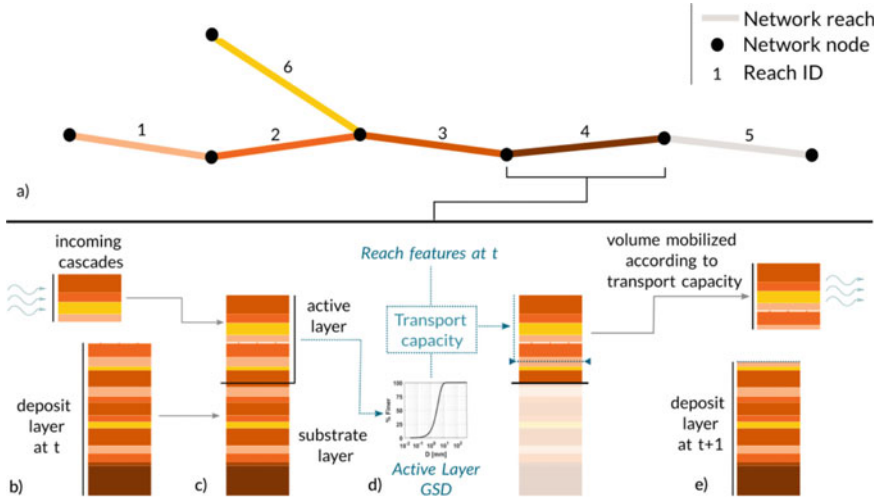


Fig. 2 An example of how the mobilized sediment volume is defined in a reach in a single model timestep. **a** shows the modeled network as a graph composed of reaches and nodes, **b** to **e** show the passages to define the mobilized sediment volume for reach 4. The colors of the tiers indicate the reach of provenance in the network. In **b** the model extracts the incoming cascades and deposit layer in the timestep, in **c** the deposit is divided into active and substrate layers, in **d** the model calculates the transport capacity for the sediment in the active layer, according to the layer Grain Size Distribution (GSD). Finally, in **e** the mobilized volume and new deposit layer are defined. Figure from Tangi et al. [15]

The volume mobilized in the reach in Fig. 2d depends on the sediment transport capacity, i.e., the amount of energy available in a reach for the transport of sediment of a particular grain class, determined by the hydromorphological characteristics of the reach and the composition of the sediment volume affected. The transport capacity is measured via empirical equations, selected according to the river’s hydromorphological features, the sediment classes considered, and the type of transport analyzed (total, suspended, or bedload transport).

Changes in geomorphic features can be modeled via selected add-on components to quantify changes in reach features between timesteps. These add-ons can include 2D or 3D morphodynamical models, simplified models calibrated on observed historic channel evolution, or conceptual models based on literature knowledge and field observations. Two-directional interactions between add-on components and D-CASCADE allow local changes to influence basin-scale sediment connectivity and vice-versa.

Finally, the destination of the mobilized volumes is determined according to the characteristic velocity of sediment transport, estimated via empirical equations [6]. According to the formula selected, sediments of different sizes may be transported at different velocities.

3 Case Studies

3.1 *Bega River System*

The D-CASCADE model was first applied to the Bega river network, in NSW, Australia (drainage area: 1930 km²), to reconstruct the historical geomorphological response of the river network to anthropic pressures brought by European settlements (ES) in terms of both sediment connectivity disruption and the consequent river changes [15].

The geomorphic evolution of the Bega river catchment is well-documented in literature [5, 7], and data are available on sediment budgets, hydrological variability and magnitude, timing, and location of historical drivers of change over the last two centuries. These data are used to define the model boundary conditions and provide validation for the model outputs. Uncertainty in the reconstruction of the hydrological record are accounted for by simulating multiple scenarios.

D-CASCADE simulations are set up considering a temporal horizon from the first ES to the present days (1850-2020). The river network features pre-ES are defined from literature reconstructions. The most significant drivers of change: deforestation, swamp drainage, and re-forestation, are introduced in the simulation in their historical chronological sequence. Add-on components are added to reproduce two of the most prominent morphological processes observed historically: channel width expansion from bank erosion and overbank flooding. Several indicators are identified to compare model outputs to the observed river dynamics.

D-CASCADE simulates scenarios of morphological evolution in the Bega river network over two centuries which match the historical observations with suitable accuracy. Observed changes are reconstructed in the model with the correct timing and location, including lowland channel expansion, valley fill incision, and sediment slug mobilization and delivery.

Finally, multiple forecasting simulations have been conducted to test the model performances for the estimation of the future morphological evolution of the system. Sediment connectivity patterns are modeled with D-CASCADE under different future hydrological and land-use change scenarios. Simulations suggest that vegetation removal, e.g., from wildfires or human interventions, may trigger a new phase of sediment mobilization in the lowland, most likely accompanied by new channel adjustments.

3.2 *3S River System*

The river system composed of the Se Kong, Se San, and Sre Pok rivers, often referred to as 3S, is a large tributary network (drainage area: 82,500 km²) of the lower Mekong river. In recent years, the 3S river basin has been interested in large-scale dam development projects aiming to exploit its hydropower potential, with approximately 40

dams are either planned or already built [12]. The construction of reservoirs, and the subsequent trapping of sediments in the impounded area, threatens to significantly reduce sediment delivery to the Mekong Delta, increasing the risk of delta subsidence and land loss [11].

D-CASCADE was applied to the 3S river system to assess the cumulative impact of multiple dams on the river sediment delivery and evaluate the benefits of coordinated reservoir sediment management techniques for sediment connectivity restoration.

As the 3S system is a data-scarce environment, we employed large-scale datasets, e.g., satellite imagery and digital elevation models to extract the river network and characterize the hydro-geomorphological features of its reaches. Different catchment sediment supply scenarios are tested using D-CASCADE and validated using literature estimates on outlet sediment delivery [8], to extract credible boundary conditions for future simulations with reservoirs.

Add-on components are developed and included in the simulations to integrate a dynamic representation of reservoir water and sediment management in D-CASCADE. These components simulate processes such as the alteration of morphological features in the flooded reaches inside the impoundment, the changes in water and sediment storage due to different management strategies and the effects of dam release on downstream hydrology.

We evaluated the impact of three different dam development portfolios in the lower 3S river (Fig. 3a): two with large reservoirs, either existing or planned (*LSS2Only* and *FullDam*) [14], and one with an alternative configuration comprised of three smaller reservoirs (*FullDam_Alt*) [1].

The results show a significant reduction in sediment transport for all dam development portfolios. Portfolios with large reservoirs (*LSS2Only* and *FullDam*) would reduce the average annual sediment delivery to the Mekong by 50%. The alternative configurations *FullDam_Alt*, due to the lower trapping efficiency of its reservoir, would lead to a less significant but still noticeable reduction (26% of the total load) (Fig. 3b).

Sediment management strategies such as drawdown sediment flushing can be applied to reduce the impact of reservoirs on sediment connectivity [9]. During drawdown flushing, the reservoir is emptied via low-level gates, and river-like conditions are established in the impoundment for a fixed number of days, during which the inflow is used to scour deposited sediments and transport them downstream [2, 18] (Fig. 4a). These operations have the added benefit of removing sediment from the impoundment of the reservoir, thus increasing its water storage capacity and decreasing maintenance costs [2]. However, they are only feasible in smaller reservoirs like the ones in the *FullDam_alt* portfolio.

To assess the potential benefits of these techniques for sediment connectivity restoration, we included drawdown sediment flushing operations in the management of the dams in the *FullDam_alt* portfolio via a specific add-on component. Sediment flushing is performed at the beginning of the wet season (July-August, as suggested from literature [1, 18, 19]), as reservoirs are typically at their lowest level after the

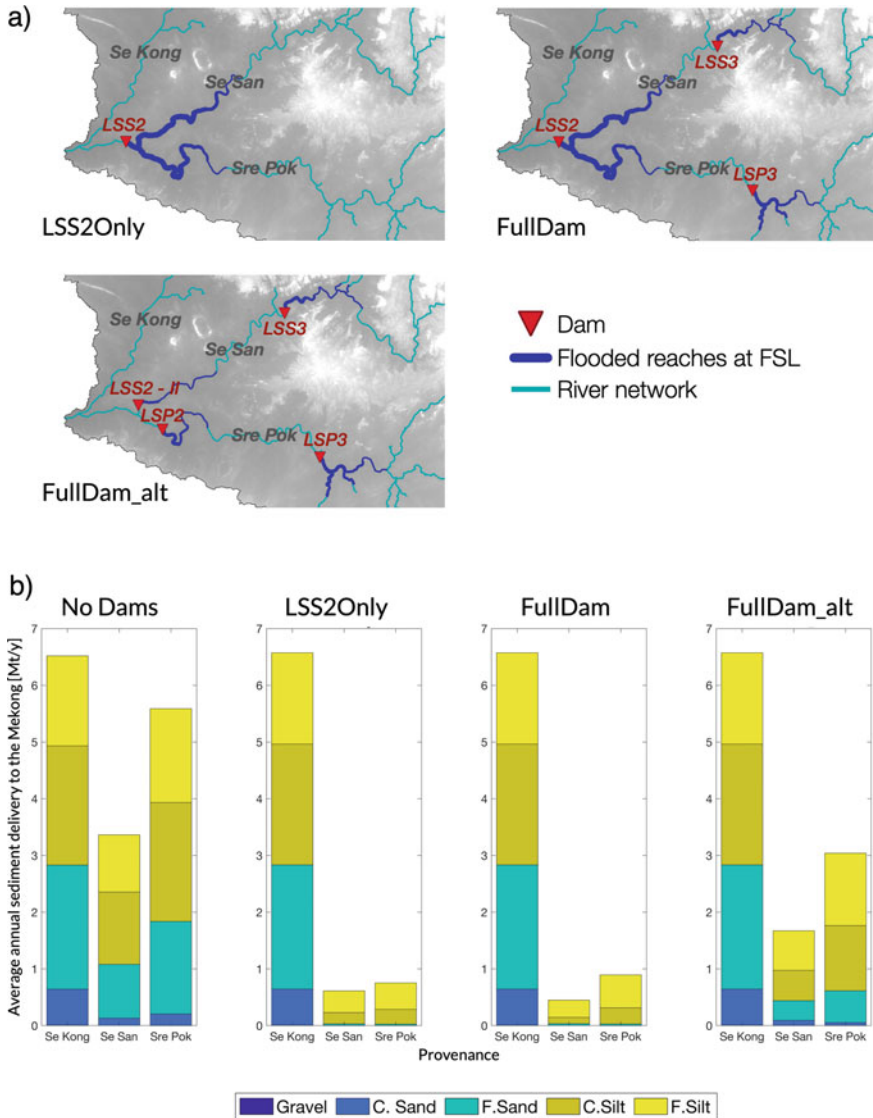


Fig. 3 Results for the application of D-CASCADE on the 3S system for the cumulative impact assessment of multiple dam development projects. **a** Dams location, impoundment area at full supply level for three dam development portfolios. **b** Average annual sediment yield to the Mekong river for the three dam development scenarios, partitioned by river provenance and grain size classes

dry season, and monsoon-induced floods increase both sediment scouring potential during flushing and refill velocity.

According to the results, drawdown sediment flushing operations repeated every one or two years may increase network sediment yield by 4%–10% respectively

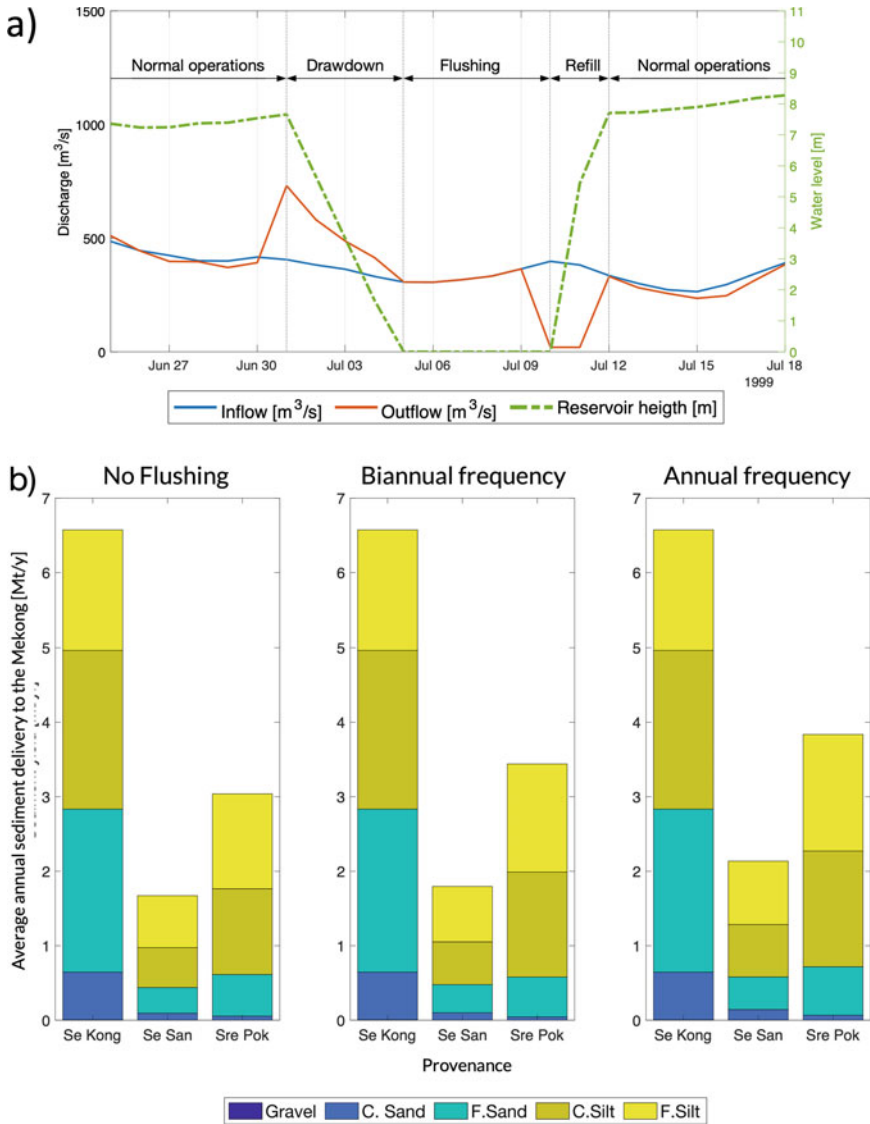


Fig. 4 a Visualization of a successful drawdown flushing operation as modelled in D-CASCADE. The figure refers to a 25-days flushing operation in the LSS3 dam. b Average annual sediment yield for the *FullDam_alt* portfolio, with no flushing, or annual and biannual flushing operations

(Fig. 4b). Flushing implementation also removes a fraction of the reservoirs sediment deposits, although the benefits are contained (7–11% reduction in the total sediment storage for the upstream dams).

4 Conclusion

In this chapter, we contribute the large-scale, dynamic sediment connectivity model D-CASCADE. D-CASCADE conceptualizes sediment transport as composed of individual sediment processes routed across a graph-like network. This modeling framework returns disaggregated, spatiotemporal information on sediment transport in each section of the river, together with data on sediment provenance, composition, and properties.

The model is flexible and data-parsimonious, making it a versatile tool for many applications with different scales, data availability and objectives. Add-on components may be added to D-CASCADE to include the representation of local-scale morphological dynamics inside the large-scale connectivity modelling framework.

The application of D-CASCADE on the Bega river case study showcases the model's potential for basin-scale geomorphological analysis to reconstruct historical sediment budget trajectories, increase our knowledge of the river sediment connectivity patterns, and forecast their future evolutions.

The study on the 3S system marks the first application of a dynamic, basin-scale sediment transport model to quantify the cumulative impact of a series of dams on river sediment delivery, and evaluate how sediment connectivity disruption could be mitigated by employing coordinated sediment management strategies.

Reservoir sediment management operations like drawdown sediment flushing come with noticeable costs, for example, due to losses in water storage and interruptions in hydropower production. Thus, these operations should be designed strategically and synchronized between multiple reservoirs to maximize benefits and reduce costs [19]. Future research will aim to integrate D-CASCADE in a multi-objective decision-making framework to extract optimal designs of water and reservoir sediment management strategies under competing objectives, e.g., hydropower generation, environmental conservation and sediment connectivity restoration.

Acknowledgements The author would like to thank Prof. Andrea Castelletti and Prof. Simone Bizzi, that supervised this research, and all the colleagues in the EiLab team at Politecnico di Milano for their help, support and insights.

References

1. Annandale, G.: Climate resilient mekong: Sediment pass-through at lower se san 2. Technical report, NHI, San Francisco (2013)
2. Atkinson, E.: The Feasibility of Flushing Sediment from Reservoirs. HR Wallingford Ltd (1996)
3. Bizzi, S., Piégay, H., Demarchi, L., Van de Bund, W., Weissteiner, C., Gob, F.: Lidar-based fluvial remote sensing to assess 50–100-year human-driven channel changes at a regional level: the case of the piedmont region, italy. *Earth Surf. Proc. Land.* **44**(2), 471–489 (2019)
4. Bizzi, S., Tangi, M., Schmitt, R.J., Pitlick, J., Piégay, H., Castelletti, A.F.: Sediment transport at the network scale and its link to channel morphology in the braided vjosa river system. *Earth Surf. Proc. Land.* **46**(14), 2946–2962 (2021). <https://doi.org/10.1002/esp.5225>

5. Brooks, A.P., Brierley, G.J.: Geomorphic responses of lower bega river to catchment disturbance, 1851–1926. *Geomorphology* **18**(3–4), 291–304 (1997)
6. Czuba, J.A., Foufoula-Georgiou, E.: A network-based framework for identifying potential synchronizations and amplifications of sediment delivery in river basins. *Water Resour. Res.* **50**(5), 3826–3851 (2014)
7. Fryirs, K., Brierley, G.J.: Variability in sediment delivery and storage along river courses in Bega catchment, NSW, Australia: implications for geomorphic river recovery. *Geomorphology* **38**(3), 237–265 (2001). [https://doi.org/10.1016/S0169-555X\(00\)00093-3](https://doi.org/10.1016/S0169-555X(00)00093-3)
8. Koehnken, L.: Discharge sediment monitoring project (dsmp) 2009–2013 summary and analysis of results. Final Report (2014)
9. Kondolf, G.M., Gao, Y., Annandale, G.W., Morris, G.L., Jiang, E., Zhang, J., Cao, Y., Carling, P., Fu, K., Guo, Q., et al.: Sustainable sediment management in reservoirs and regulated rivers: experiences from five continents. *Earth's Fut.* **2**(5), 256–280 (2014)
10. Kondolf, G.M., Rubin, Z.K., Minear, J.: Dams on the mekong: cumulative sediment starvation. *Water Resour. Res.* **50**(6), 5158–5169 (2014)
11. Kondolf, G., Schmitt, R., Carling, P., Goichot, M., Keskinen, M., Arias, M., Bizzi, S., Castelletti, A., Cochrane, T., Darby, S., et al.: Save the mekong delta from drowning. *Science* **376**(6593), 583–585 (2022)
12. MRC (Mekong River Commission): Hydropower project database. Basin Development Plan Programme (2014)
13. Schmitt, R.J., Bizzi, S., Castelletti, A.: Tracking multiple sediment cascades at the river network scale identifies controls and emerging patterns of sediment connectivity. *Water Resour. Res.* **52**(5), 3941–3965 (2016)
14. Schmitt, R., Bizzi, S., Castelletti, A., Kondolf, G.: Improved trade-offs of hydropower and sand connectivity by strategic dam planning in the mekong. *Nat. Sustain.* **1**(2), 96 (2018)
15. Tangi, M., Bizzi, S., Fryirs, K., Castelletti, A.: A dynamic, network scale sediment (dis) connectivity model to reconstruct historical sediment transfer and river reach sediment budgets. *Water Resour. Res.* **58**(2), e2021WR030784 (2022)
16. Tangi, M., Schmitt, R., Bizzi, S., Castelletti, A.: The cascade toolbox for analyzing river sediment connectivity and management. *Environ. Model. Softw.* **119**, 400–406 (2019)
17. Vörösmarty, C.J., Meybeck, M., Fekete, B., Sharma, K., Green, P., Syvitski, J.P.: Anthropogenic sediment retention: major global impact from registered river impoundments. *Global Planet. Change* **39**(1–2), 169–190 (2003)
18. White, R.: *Evacuation of Sediments from Reservoirs*. Thomas Telford Publishing (2001). <https://doi.org/10.1680/eosfr.29538>
19. Wild, T.B., Loucks, D.P., Annandale, G.W., Kaini, P.: Maintaining sediment flows through hydropower dams in the mekong river basin. *J. Water Resour. Plan. Manag.* **142**(1), 05015004 (2016)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Electronic

Gamma-Ray Spectroscopy and Imaging with SiPMs Readout of Scintillators: Front-End Electronics and Position Sensitivity Algorithms



Luca Buonanno 

Abstract This is an introductory article to the topics more widely discussed in the PhD thesis from the same author. Following a short introduction and the motivations for researching innovative gamma-ray detector systems, this article describes a novel 85 dB dynamic range per channel integrated circuit for SiPM charge signal readout, named GAMMA, and the custom FPGA-based readout system. Experimental results presented in this article, obtained using a planar array of NUV-HD SiPMs, encompass the single-photon sensitivity achieved by GAMMA ASIC and the 2.6% resolution at the ^{137}Cs peak emission energy of 662 keV, when using GAMMA ASIC to collect current signal from a detector array that is coupled to a LaBr_3 scintillation crystal. Pixellation of the detector matrix allows for coarse position of interaction sensitivity in the scintillation crystal using machine learning reconstruction algorithms.

Keywords SiPM · ASIC · LaBr_3 .

1 Introduction

Gamma radiation detectors find applications in multiple fields, such as nuclear physics experiments, nuclear medicine, molecular imaging, and homeland security [6, 7, 12, 14]. Often, the novel detector architectures, electronic systems, or data processing techniques introduced in one of these fields also offer potential improvements for the other applications. A substantial push in technological innovation is driven in this context by fundamental research.

This article focuses on presenting a detection module developed for Istituto Nazionale di Fisica Nucleare (INFN) research institute, aiming at spectroscopy and position sensitivity of gamma photon at medium and high energies, while superseding photomultiplier tube-based detection modules with silicon photomultipliers (SiPM). The use of SiPMs allows to fully exploit scintillation crystal technologies, e.g. enabling a wide dynamic range (from 1 to 10^5 photons per cm^2) and excellent res-

L. Buonanno (✉)
Politecnico di Milano, DEIB, Milano, Italy
e-mail: luca.buonanno@polimi.it

© The Author(s) 2023
C. G. Riva (ed.), *Special Topics in Information Technology*,
PoliMI SpringerBriefs, https://doi.org/10.1007/978-3-031-15374-7_4

olution (2.6% at 662 keV, the photopeak emission energy of ^{137}Cs) when using a coped lanthanum bromide crystal. Motivations and architecture of GAMMA detection module are respectively introduced in Sects. 2 and 3, while the GAMMA ASIC is described in detail in Sect. 4. A machine learning-based reconstruction algorithm and the experimental validation of the spectroscopic module are presented in Sect. 5.

2 GAMMA: A Novel Indirect Conversion Detector

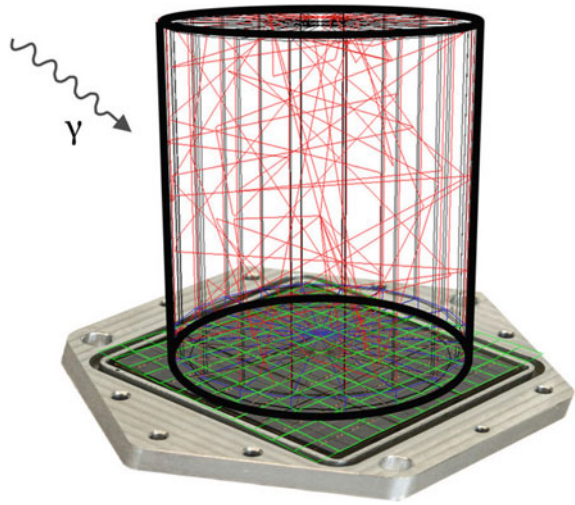
Gamma photons are commonly probed in physics experiments, aiming at collecting sets of data that are representative of their energy, the point of interaction with the detection system, or the time of interaction. Conversion of the physical quantity into a probable signal is required during the experiment, and the physical variable is often transformed in an electric signal.

In indirect conversion methods a scintillation material (which, following an interaction with a gamma photon, emits many low-energy photons, typically in the near infra-red, optical or near ultra-violet energy band) is coupled to a photosensor array. The amount of light produced is ideally proportional to the energy deposited by the incident radiation in the detector.

Scintillator-based gamma detectors fall in one of the following categories: detectors with monolithic scintillators, and detectors with pixelated scintillators. A monolithic scintillator consists of a continuous block of scintillator material. In Anger Cameras, a single monolithic scintillator crystal is coupled to an array of photodetectors. These cameras incrementally gained attention in the years: besides requiring smaller fabrication costs per unit area, they also allow to estimate the depth-of-interaction and to improve the spatial resolution beyond pixel size, being the resolution of this solution related instead to the statistical fluctuation occurring in the physical process of light sharing among pixels and to the specific reconstruction algorithms implemented. Secondly, the use of bulkier scintillators implies a higher stopping power, which is a key factor in accelerating the data collection of physics experiment in the range of MeV energies, often resulting in better spectroscopy performances with respect to pixelated crystals, given the higher probability of total energy absorption. Scientific literature on spectroscopy and imaging modules based on monolithic scintillators is extremely rich of applications, and often a test bench for novel technologies of photon sensing and reconstruction algorithms.

It is shown in Fig. 1 a 3D model of the Anger Camera that is introduced in this section, named GAMMA, coupling a $3''\times 3''$ crystal to a 12×12 array of SiPMs (simulation run in ANTS2 environment). The development of front-end readout, system electronics, back-end software, and data processing algorithms for this modern, compact gamma spectrometer was commissioned by the GAMMA experiment, supported by the INFN. Main investigation methods of GAMMA experiment involve the use of gamma spectroscopy: focus of the research activity is the study of shell

Fig. 1 3D model of a monolithic scintillator with SiPMs readout, and simulated paths of the scintillation light emissions in the cylindrical $3'' \times 3''$ crystal. The nine tiles of SiPM are arranged in a planar, 12×12 pixels array. Simulation run in ANTS2 environment [11]



structure and collective modes, with particular interest towards nuclei produced in exotic beams. Many experiments are based in ion beam facilities, and involve the use of arrays of detectors; the GAMMA experiment is often involved in the R&D phase of innovative detection modules, following with particular interest the latest discoveries in the fields of scintillator materials and readout electronics.

The lanthanum bromide crystals was selected for GAMMA module for its excellent conversion efficiency (allowing energy resolution $<3\%$ at the ^{137}Cs emission energy of 662 keV) and fast decay time (16 ns, which potentially enables a theoretical count rate capability higher than 10 MCPS). Lanthanum bromide crystals also have a satisfactory radiation tolerance to high-energy proton irradiation with fluences typical of those in the interplanetary space (LaBr_3 was studied for various space missions), and show sufficient radiation hardness when exposed to a high flux of MeV γ -rays.

The GAMMA module is depicted in Fig. 2. In the exploded view, we can identify the main components of the $13\text{ cm} \times 13\text{ cm} \times 15\text{ cm}$ module: the aluminum enclosure, the scintillation crystal coupled to the SiPMs matrix, and the readout electronics. The front-end collects the charge signals from SiPMs in the low-impedance input stage of the ASIC and provides the conditioned signal to the FPGA-controlled ADCs; the signal is digitized and sent to the central unit.

The lanthanum bromide crystal is coupled to a square array of 144 SiPMs, made up of 9 individual tiles; each tile is a 16 elements square array of $6\text{ mm} \times 6\text{ mm}$ SiPM detectors.

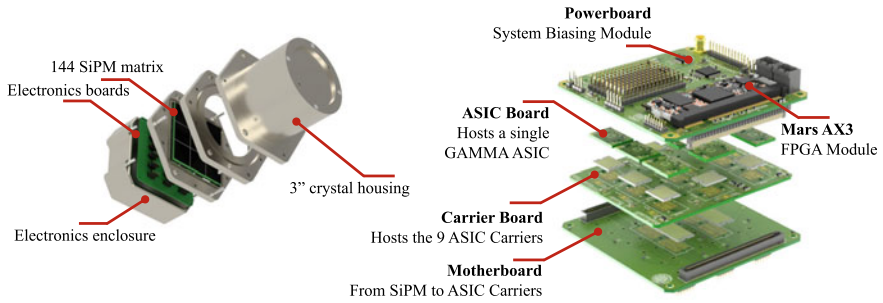


Fig. 2 (left) The $13\text{ cm} \times 13\text{ cm} \times 15\text{ cm}^3$ aluminium enclosure encompasses scintillation crystal, the 144 SiPM detectors, front-end electronics and the DAQ. (right) The SiPMs hosted by the Motherboard generate 144 current signals, which are monolithically collected by 9 GAMMA ASICs and digitized by the FPGA-based DAQ on the Powerboard

3 Architecture of GAMMA Module

When a gamma photon interacts with the scintillation crystal, photocarriers generated in the SiPMs induce current signal pulses that are processed by the electronics and converted to a digital data. Each SiPM is individually coupled to a low-impedance input node of the front-end GAMMA ASIC [13] via the Motherboard (Fig. 2), in order to convey signal current pulses from SiPM anodes to the analog filter stages while keeping a constant voltage across the detector. If a current pulse is detected from one of the 144 channels, the integration starts for all channels; the readout is therefore monolithic and self-synchronized to particles or gammas interactions in the crystal.

Multiple instances of a 16-channel custom electronic front-ends, the GAMMA ASICs, collect the charge from each of the SiPM anodes individually (for a total of 144 analog channels). Each ASIC channel features a gated integrator stage; each channel has three automatically set gain. The gain is individually selected in each channel by the Adaptive Gain Control (AGC) circuit, the central role of which is discussed in further detail in Sect. 4.2, integrating a support circuitry which predictatively estimates in the analog domain the charge accumulated during the integration period. The voltage dynamic of the gated integrator is therefore optimized on an event-per-event basis and the dynamic range of each channel is extended to 84 dB, which in turn allows to fully exploit the SiPM detectors dynamic range. Superior spectroscopic performance on a wide energy interval also allows to overcome in physics experiments the trade off between energy resolution and observed energy interval, cutting the experiment beamtime.

The multiplexed voltage outputs of the ASICs are digitized by 13-bit, 5 MSPS, differential input ADCs, one for each ASIC. The converted value is gathered by a Xilinx Artix-7 FPGA, which also serves as a bridge with the custom control application hosted on an external PC.

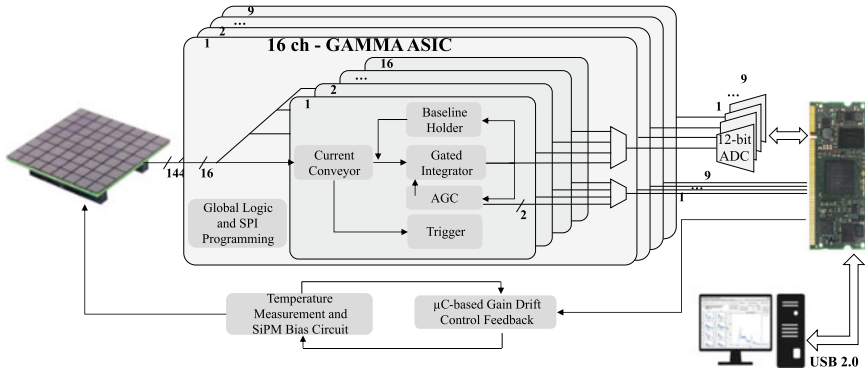


Fig. 3 Block scheme of GAMMA module. The front-end circuit is made up of 16-channel GAMMA ASICs, allowing in the module an individual, low-noise readout of each pixel. The SiPM bias voltage is controlled in closed loop, in order to compensate for gain drifts of the system due to temperature fluctuations

The module has an analog output proportional to the event energy that, together with the trigger signal coming from the ASICs, can be used to embed the detector in pre-existing facilities that use dedicated multichannel analyzers or digitizer synchronized with the arrival of a gamma photon.

An ARM Cortex-M4 microcontroller collects the temperature data from the SiPM's temperature sensors and acts on their bias to maintain constant the gain even in the presence of strong temperature variations [8]. The block scheme of the module is shown in Fig. 3.

4 GAMMA ASIC

The GAMMA experiment targets spectroscopy in an energy interval from 20 keV to 20 MeV, and count rates up to 50 kcps. In order to determine the amount of optical photons collected by each SiPM (and correspondingly the range of charge at the input of each ASIC channel), the light emission of the $3'' \times 3''$ crystal, collected by an array of 144 SiPMs of $6 \times 6 \text{ mm}^2$ area each, has been simulated by means of ANTS2 software [11] up to 20 MeV (Fig. 4).

The light distributions extracted by these simulations showed that, considering the minimum SiPM signal at the minimum gamma energy and the maximum signal at the maximum energy, the number of photons collected by each SiPM in the array ranges from few photons to about 10^4 photons, providing a maximum charge at the ASIC input of 3 nC. Maximization of scintillation light collection is a key factor to enhance resolution, but in order to have a circuit that allows both single-photon resolution and the collection of 10^4 photons, the dynamic range of each channel has to be larger than 80 dB.

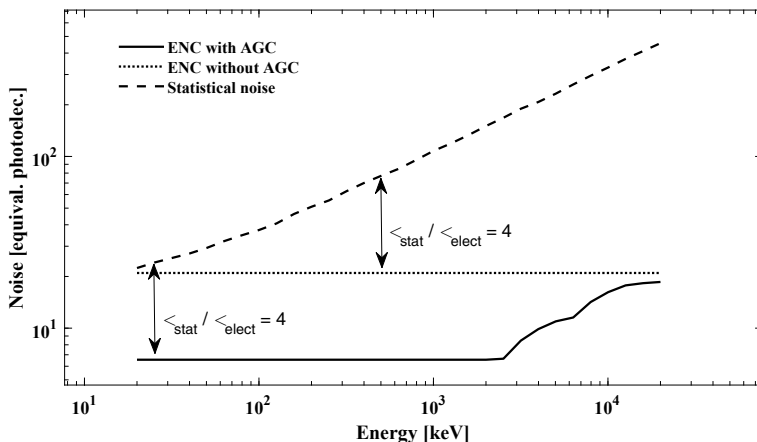


Fig. 4 Comparison of equivalent input noise contributions to energy resolution as a function of the incoming photon energy in the 100 keV–20 MeV range, obtained from numerical simulations in ANTS2 software. The adaptive gain control circuit allows for lower noise at low energies and, in turn, to extend the energy interval of the measurement. Fixed gain acquisitions are degraded at low energies by the electronics noise, in the energy interval where the electronics noise results comparable or even higher than the one associated to the collection of photons in the detector (statistical noise)

Many ASICs have already been developed for the readout of SiPMs, and a review can be found, for instance, in [5]. Commercially available ASICs are mainly developed for medical applications like PET, with limited DR. Considering our requirements (200 fC equivalent noise charge, 3 nC full scale range) we have decided to develop a new ASIC that is able to provide at the same time a low noise in the lower energy region and high DR capability.

4.1 GAMMA ASIC Top Level Architecture

The block diagram of the ASIC is shown in Fig. 5. The circuit can be connected to 16 SiPMs providing a simultaneous readout of the charge delivered by the detectors with independent AGC in each channel. The outputs of the channels (analog amplitude and digital code corresponding to the used gain for each integration) are multiplexed, converted to fully-differential signal, and routed towards the chip output pads.

4.2 Adaptive Gain Control circuit

In Fig. 4 the overall statistical noise and the total electronics noise (both extracted from simulations and numerical analysis, assuming that the gain is constant or it is set by the AGC circuit) are reported for the complete energy range. Both contributions

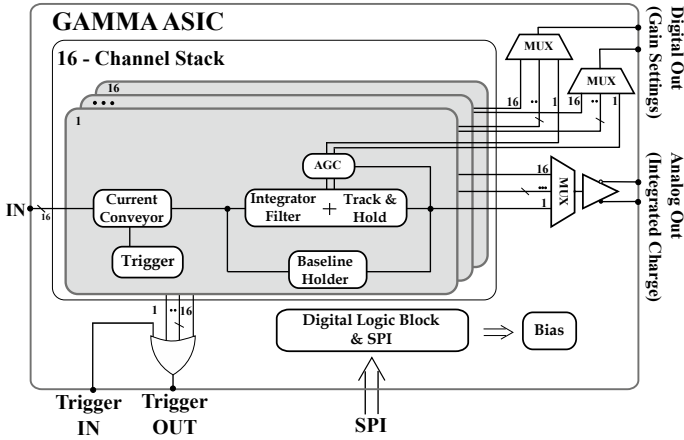


Fig. 5 GAMMA block diagram. The chip features 16 SiPM inputs and a multiplexed analog output. The Current Conveyor stages collect the signal charge, that is integrated in all filter stages when the input signal overcomes a programmable threshold in the Trigger circuit. The parallel outputs are multiplexed and sent to an external ADC by the fully differential output buffer. The AGC circuit adapts the gain of the filter to the signal amplitude

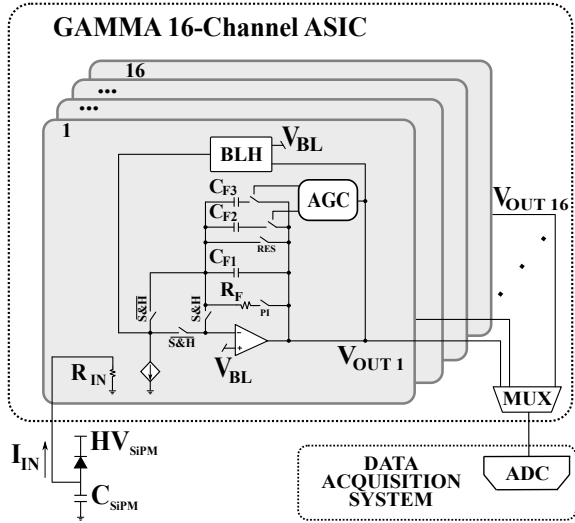
are extracted from ANTS2 simulations; if the AGC is active we assume that each ASIC channel has adopted the best gain (that is, the highest gain that does not induce saturation) in the measurement, out of three available settings.

If a constant gain is considered for the amplifier, the simulation showed that the electronics noise is dominant with respect to the statistical noise in the low energy range, affecting the overall resolution up to many hundreds keV; the relation $\sigma_{stat}/\sigma_{elect} \leq 4$ only holds down to 600 keV. The constant gain of the readout electronics implies in fact a trade-off for the readout electronics between low-noise and dynamic range capabilities.

We have introduced in the ASIC the AGC circuit, which automatically increases the integration capacitance of the gated integrator stage to reduce the gain for large signals. This technique, already implemented in other ASICs for x-ray detection [2] or astrophysics experiments [3], allows to implement a piece-wise linear characteristic and an overall gain compression.

Because the dynamic range of each individual energy range showing constant gain has to decrease in each interval in order to cope with design requirements (the statistical noise doesn't increase linearly with the input signal), the author found non beneficial to implement more than 3 possible integrator feedback capacitance settings. A transistor-level schematic description of the gated integrator stage is shown in Fig. 6.

Fig. 6 Implemented gated integrator filter with adaptive gain control block. The feedback resistance R_f is connected until the start of the integration phase. The S&H switch is open until the end of the integration phase, then it is closed (the two complementary switches are opened) and the integrator holds the charge to be read. At the end of the read phase, the feedback capacitance is discharged by switch RES



4.3 Current Buffer Input Stage

The input stage requirements strongly depend on the characteristics of the SiPM detectors. A typical circuit model for SiPM NUV-HD detector that was adopted in our simulation phase is described in [1, 5, 9]. Key parameters of such model are the SiPM detectors reset time, the micro-cell capacitance, and the ASIC input impedance, the values of which also define the minimum time constant of the input current signal [5]. We considered that in potential applications of our ASIC (e.g., in compact spectrometers for radiation monitoring, or in measurements with a much larger number of channels), it could be useful to merge several SiPMs (N_{SiPM}) to a single readout channel. In this case, the input capacitance scales linearly with the number of detectors connected. In particular, a large capacitive load at the input stage may impose severe challenges to the frequency response and even stability issues when a classical negative-feedback regulated cascode configuration, adopted in several commercial ASICs, is considered [10]. The current-buffer Temes architecture allowed for the merging, in the experimental measurements described in [4], of few tens of SiPMs for a total input capacitance up to 70 nF, while providing an input impedance sufficiently low for the integration to last few microseconds.

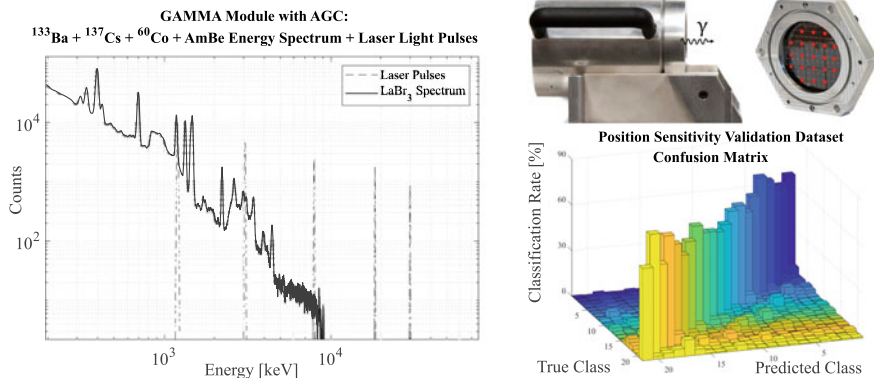


Fig. 7 (left) Superposition of the decay spectrum from ^{133}Ba , ^{137}Cs , ^{60}Co and AmBe radioactive sources and spectral lines generated by a calibrated laser-based test setup, measured with GAMMA module. (right) Experimental position sensitivity characterization setup and 3D plot of the validation dataset confusion matrix for the Decision Tree algorithm embedded in the FPGA. Each True Class row corresponds to a different position of the collimated ^{137}Cs source. Values on the diagonal correspond to the correct classification rates

5 Spectroscopy and Position Sensitivity Experimental Measurements

The spectroscopy capability of the system was tested using ^{133}Ba , ^{137}Cs , ^{60}Co and AmBe radioactive sources and using a custom laser-based setup, in order to simulate the scintillation light of LaBr₃ generated by the interaction with gamma photons at energies higher than the ones available.

The superposed spectra are shown in Fig. 7. The experimental measurements show a 2.7% energy resolution at 662 keV that monotonically decreases at higher energies, reaching the value of 1.0% at 9.0 MeV. The energy resolution in the measured energy interval is well fit by a curve proportional to $1/\sqrt{E}$. The non-linearity error of the measurement is 0.6%, and is limited by detector non-linearity at higher energies. Maximum measured count-rate is 80 kCps and the overall power consumption of the module is 6 W.

Position sensitivity was tested scanning the scintillation crystal with γ -rays emitted from a ^{137}Cs source, collimated on the orthogonal direction to the SiPM detectors plane in a 1 mm large pinhole. The scintillation crystal is irradiated from the top side on a square grid that has 1.5 cm pitch. Collected data are labelled using the scan positions, which are known, and are used as training datasets in classification algorithms.

Training a decision tree algorithm and embedding the classifier in the FPGA-based DAQ, from 80 to 90% (depending on the real beam position) of the events 2D reconstructed position are classified with an error smaller than 2 cm, computed from a validation dataset acquired using the same procedure of the training dataset (Fig. 7).

6 Conclusion

GAMMA module is a compact spectrometer designed for nuclear science experiments (13 cm × 13 cm × 15 cm, 6 W power consumption), which successfully exploits the codoped LaBr₃ resolution capability (2.7% at 662 keV, and 1% at 9.0 MeV) and SiPM linearity (0.6% error at 9 MeV). Pixellation of the detector matrix allows for coarse position of interaction sensitivity in the scintillation crystal.

The module that features nine GAMMA ASICs will soon be used in beamline facilities in order to validate its energy resolution performances using higher energy gamma-ray sources, and to validate the relativistic Doppler effect correction with machine learning algorithms.

References

1. Acerbi, F., Gundacker, S.: Understanding and simulating sipms. *Nucl. Instrum. Methods Phys. Res., Sect. A* **926**, 16–35 (2019)
2. Allahgholi, A., Becker, J., Bianco, L., Delfs, A., Dinapoli, R., Goettlicher, P., Graafsma, H., Greiffenberg, D., Hirsemann, H., Jack, S., et al.: Agipd, a high dynamic range fast detector for the european xfel. *J. Instrum.* **10**(01), C01023 (2015)
3. Bonvicini, V., Orzan, G., Zampa, G., Zampa, N.: Casis1. 0: a prototype vlsi front-end asic with ultra-large dynamic range and integrated adc for silicon calorimetry in space experiments. *Nuclear Instrum. Methods. Phys Res. Sect. A: Accel., Spectr., Detect. Assoc. Equipm.* **572**(1), 340–344 (2007)
4. Buonanno, L., Di Vita, D., Carminati, M., Fiorini, C.: A directional gamma-ray spectrometer with microcontroller-embedded machine learning. *IEEE J. Emerg. Select. Top. Circuits Syst.* **10**(4), 433–443 (2020)
5. Calo, P.P., Ciciriello, F., Petrignani, S., Marzocca, C.: Sipi readout electronics. *Nucl. Instrum. Methods Phys. Res., Sect. A* **926**, 57–68 (2019)
6. Camera, F., Ceruti, S., Lodetti, S., Frega, S., Miani, A., Giaz, A., Pellegrini, L., Riboldi, S., Bracco, A., Crespi, F., et al.: Characterization of large volume 3.5" × 8" labr3: Ce detectors for the hector+ array (2014)
7. Carminati, M., Di Vita, D., Buonanno, L., Montagnani, G.L., Fiorini, C.: A lightweight sipm-based gamma-ray spectrometer for environmental monitoring with drones. In: *International Conference on Applications in Electronics Pervading Industry, Environment and Society*, pp. 55–61. Springer (2020)
8. Di Vita, D., Buonanno, L., Montagnani, G., Minerva, A., Giannoni, A., Carminati, M., Camera, F., Fiorini, C.: A high dynamic range 144-sipm detection module for gamma spectroscopy and imaging with 3" labr 3. In: 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), pp. 1–3. IEEE (2019)
9. Gundacker, S., Heering, A.: The silicon photomultiplier: fundamentals and applications of a modern solid-state photon detector. *Phys. Med. Biol.* **65**(17), 17TR01 (2020)
10. de Medeiros Silva, M., Oliveira, L.B.: Regulated common-gate transimpedance amplifier designed to operate with a silicon photo-multiplier at the input. *IEEE Trans. Circuits Syst. I Regul. Pap.* **61**(3), 725–735 (2013)
11. Morozov, A., Solovov, V., Martins, R., Neves, F., Domingos, V., Chepel, V.: Ants2 package: simulation and experimental data processing for anger camera type detectors. *J. Instrum.* **11**(04), P04022 (2016)

12. Richter, C., Pausch, G., Barczyk, S., Priegnitz, M., Keitz, I., Thiele, J., Smeets, J., Vander Stappen, F., Bombelli, L., Fiorini, C., et al.: First clinical application of a prompt gamma based in vivo proton range verification system. *Radiother. Oncol.* **118**(2), 232–237 (2016)
13. Temes, G., Ki, W.: Fast cmos current amplifier and buffer stage. *Electron. Lett.* **23**(13), 696–697 (1987)
14. Zeiser, F., Tveten, G.M., Garrote, F.B., Guttormsen, M., Larsen, A.C., Ingeberg, V.W., Gørgen, A., Siem, S.: The γ -ray energy response of the oslo scintillator array oscar. *Nucl. Instrum. Methods Phys. Res., Sect. A* **985**, 164678 (2021)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Ultra-High Performance Digital Electronic Architectures for Events Management in Real Time Environments



Fabio Garzetti

Abstract The research spans several application areas, including biotechnology, medical imaging, and environmental monitoring. Precise and specialized processing techniques are often required for measurements of signal parameters with high efficiency, for example, in terms of resolution and count rate, such as time of occurrence of events. Digital solutions have thus received the most significant attention since they are the most effective at enabling flexible, application-oriented elaboration systems. The research is finalized to develop high-resolution time measurement systems in Field Programmable Gate Array (FPGA) devices. The goal is to optimize resolution, linearity, and processing speed at the state-of-the-art in modern, most dynamic branches of very high-efficiency time measurement. To maximize the number of events handled in real-time, the research is also concentrated on developing cutting-edge communication and data transfer methods for performing multiple measures in parallel.

Keywords Time-to-Digital Conversion · TDC · FPGA · Data serialization · Synchronization

1 Introduction

One of the most significant aspects of an event is its time of occurrence, and the management of the temporal properties associated with the events is thus one of the cutting-edge issues in electronic processing.

The research focuses on the high-performance processing of temporal information, starting from its measurement and moving on to management and, most crucially, transmission.

The Time-to-Digital Converter (TDC), a device that makes it possible to measure events' times in digital form, is nowadays the primary method for doing that measure. The TDC detects events and provides a digital code indicating when events occur.

F. Garzetti (✉)

Dipartimento di Elettronica, Informazione e Bioingegneria – DEIB, Sezione di Elettronica, Politecnico di Milano, via Camillo Golgi 40, 20133 Milano, MI, Italy
e-mail: fabio.garzetti@polimi.it

© The Author(s) 2023

C. G. Riva (ed.), *Special Topics in Information Technology*,

PoliMI SpringerBriefs, https://doi.org/10.1007/978-3-031-15374-7_5

Because of the rising throughput (50–100 million measurements per second are becoming more frequent) and the rising number of parallel channels, the most recent generation of TDCs must deal with several operations beyond the measurement of the time an event occurs that translate into real architectural challenges [4].

System performance is greatly influenced by data routing, which is undoubtedly a foundational component of systems. To get beyond the restrictions of traditional timestamp management, a novel approach has been conceptualized and created in order to perform parallel-to-serial conversion while keeping the chronological order of the output measurements. The new approach is called BeltBus, and it gathers timestamps from the channels and sends them to the serial output like cargo on a conveyor belt. This technique is far from simple, even merely because a TDC component of the architecture is not synchronized with the system clock, making the transition between clock domains crucial [2]. The immediate result is vastly more straightforward processing in the phases after the measurement (for example, the histogram), avoiding the usage of multiplexers that necessarily become slower and slower as the number of channels increases and eliminating the single bus. The number of channels needed by an ever-increasing number of applications has reached a point where it is impossible to concentrate the acquisition and processing of the information transmitted by each of them in a single integrated circuit, even in an Application Specific Integrated Circuit (ASIC). The essential synchronization issue arises when the channels are divided across many processing circuits, which is made worse because time measurements are always differential measurements. Finding a solution to the synchronization issue in distributed systems has now taken on fundamental significance, in part since the jitter of the signals precludes the possibility of disseminating the same timing signal to every component of the system. Three distinct strategies were developed to correct the synchronization error, which was also looked at from the implementation perspective.

Finally, a TDC instrument that is entirely reconfigurable and based on an FPGA chip has been developed. The instrument underwent testing and validation with outside academic and industrial institutions, with excellent outcomes.

2 New TDC at the State-Of-Art in FPGA

To suit the requirements of a wide range of applications, numerous TDCs based on FPGA devices are available in various feature and performance combinations.

The proposed TDC [2] maximizes system performance, including parallelizing various tapped delay lines with high performance for multichannel purposes, an interpolation mechanism, and a sub-interpolation to improve resolution significantly below the tapped line bin propagation delay, and a calibration process to maintain linearity.

In doing so, a new fully configurable 16 channels TDC architecture with a dynamic range above 10 s, resolution around 300 fs, and single-shot precision just above 10 ps

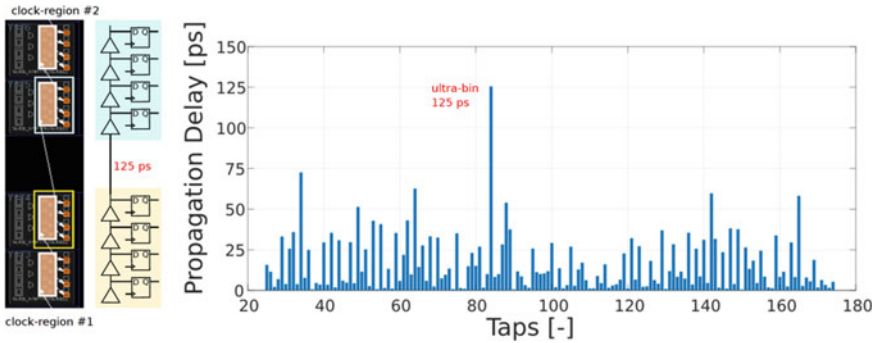


Fig. 1 Histogram of the propagation delays of the buffers making up the implemented tapped delay line. The diagram shows how the tapped delay line can span different clock areas, causing additional delays mainly due to crossing several clock regions. The most significant value of delay is named “ultra-bin”

has been created. Moreover, high linearity to 250 fs for the differential and 2.5 ps has been tested for the 16 independent channels without performance degradation [2].

The design was challenging. First, the regular usage of the FPGA device does not cover the propagation delays across the buffers of the delay lines fully equal. The buffer delays are consequently too dissimilar to the processing requirements. Furthermore, as the signal moves from one zone to the next, the FPGA partition in clock areas results in diversity in the delays. As seen in Fig. 1, the delay difference affects both resolution and linearity.

For instance, sub-interpolation and calibration, which are necessary to address this issue, have received much attention. Sub-interpolation is used to correct the resolution, and calibration compensates for the linearity. This ground-breaking architecture was created to have a high dynamic range, high resolution, high accuracy, high linearity, and multichannel operability; all the information can be found in [2].

Special attention has been spent on the feasibility of migrating the new TDC firmware among different FPGA devices. Individual users are frequently unwilling to give up their outdated devices, especially for interventions that can be difficult and expensive. It is obvious how not straightforward it is to pursue the objective of making firmware transfer between various devices extremely simple. There are clues of common sense, not set criteria or design rules, for doing that. Understanding and utilizing the inherent FPGA structure is necessary to implement a TDC with such excellent performance. It is impossible to transfer the firmware rapidly due to the significant differences between devices made by various manufacturers. There will be a new work because the TDC core will change. Additionally, the challenge increases if the primitives point to global logic functions rather than specific hardware, allowing the synthesizer flexibility over what will be implemented. This is true, for instance, of Intel (Altera) FPGAs.

3 High-Performance Data-Serialization

A physical event is first detected and converted into an electrical signal by detectors and discriminators as part of the time-resolved measurement process. Then, the TDC performs digital coding of the event's instant of occurrence to provide a timestamp. Timestamps play a crucial role in a TDC system because they must get to the elaboration modules to provide different types of information, like a histogram or a count of the acquired measures. They arrive in a parallel manner from different channels of the TDC core. In a world where speed and measurement rate are becoming critical in many applications, multichannel architectures are essential because they enable the simultaneous detection of more physical events [1]. In order to meet this need, time-resolved setups are adding more channels. While adopting a multichannel structure has many benefits, it also adds some complexity to managing the resources that are accessible. Typically, a parallel bus sends timestamps from the TDC-Core to the elaboration modules.

Additional restrictions on the incoming timestamps are also necessary if more elaboration is provided because they cannot be delivered randomly and must follow the temporal order. With a small number of channels, ordering may be straightforward, but as the number increases, ordering becomes more challenging. As a result, performance, ease of implementation, and reusability must be sacrificed due to parallel architecture.

To get beyond the restrictions of traditional timestamp management, a unique module that can perform a parallel-to-serial conversion while preserving the output measurements' chronological sequence has been designed and created. The new solution is called BeltBus, and it gathers timestamps from the channels and sends them to a serial output (like a conveyor belt with packages, as Fig. 2 makes evident).

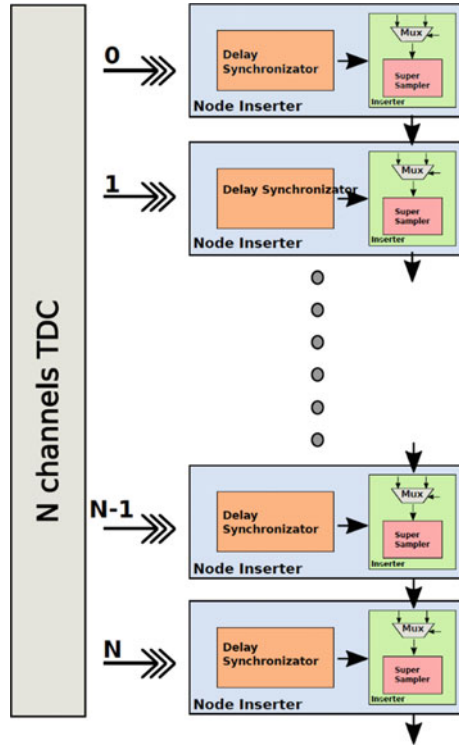
A series of temporally ordered measurements are produced via the parallel to serial conversion made possible by the BeltBus architecture and synchronization mechanism. The trade-off between multichannel structures and elaboration complexity can be resolved thanks to this elaboration structure. The timestamps can reach external modules via the BeltBus' single serial output, where they can be used to perform various functions without the need for a pre-processing phase. Furthermore, a sizable amount of routing resources is conserved compared to parallel outputs.

Finally, because it enables combining the outputs of numerous TDCs, this structure is particularly suited for multi-board applications.

However, this structure does have some restrictions. First, execution frequency is reduced because serialization is limited by bus frequency. Second, throughput is highly constrained because there are no separated buses, and all the measures must flow in the same bus.

The Cocotb co-simulation, which connects Python test benches to GHDL, an open-source Linux VHDL code simulator, was used to assess the performance of the BeltBus module. Create a folder containing the test bench, MakeFile, and top module for simulation, open it in the terminal and type "make" to run the Cocotb co-simulation. As soon as the simulator starts, Cocotb quickly turns over control to

Fig. 2 Implemented serialization structure in the BeltBus. The timestamps from the parallel channels propagate along with the nodes and reach the serial output. Internal resources available in each serial node are put in evidence



GHDL from the test benches. An info message is produced on the monitor after each instance of an overflow, giving the total number of overflows. The simulation time and the number of coarse overflows that should occur can be chosen in the test.

4 Multi-TDCs Synchronization Techniques

Due to the growing complexity of the events being seen, many applications would benefit from having many measuring channels of TDCs connected to maintain a high-resolution measurement [3]. Making a network of independent time-meters, or TDCs, in which every device can communicate with one another, would be an intriguing idea. Due to the differences in the clock frequencies of various network components and the differences introduced by the FPGA-based TDC, synchronizing numerous TDCs is challenging and complex.

The critical point is that using various devices as implementation hosts is forced by the increasing number of channels that more and more applications demand. This problem makes it hard to compare and use timestamps from many TDCs together; the research has aimed to develop methods for treating timestamps from various TDCs as though the same machine produced them.

The research has introduced a new method of synchronization that makes timestamps from several TDCs appear to have been generated by the same device.

Many operating parameters are unavoidably distributed when using systems with various physical properties. The first crucial parameter is the clock period, T_{CLK} . Despite their apparent similarity, two different boards may have clock frequency incompatibilities. Since the interpolation process yields a resolution that is proportional to the clock period T_{CLK} [2], errors in clock frequencies immediately result in resolution errors. Gain errors are inconsistencies in the resolution LSB.

Moreover, the TDCs turning on at various periods results in distinct dispersion. An offset issue arises when different TDCs turn on at somewhat different times. In a network with two or more TDCs, each source of error must be considered to provide synchronization and enable communication.

Mismatches in the values of the various TDC clocks' period are the root cause of the Gain Error. The gain error increases directly to the timestamp value because it results from a mismatch between the LSBs.

The Offset Error results mainly from discrepancies in the time instant when the various TDCs are turned on. Let us consider two TDCs with the same clock period that turns on at different times. Because of this circumstance, the timestamps of the two devices in question would differ in time.

In a network of TDCs, gain and offset errors regularly occur. These errors pile up analytically, creating an offset between them.

In the quantitative analysis, the global variance (2CH of the single-channel measure) includes the contributions provided by the fluctuation of the measure on the generic TDC channel, the variance due to non-ideal features of the frontend, and the intrinsic jitter of the signal. The TDC REF-measure channel's variation and the variance of the frontend to which the REF is connected are usually added together to form the variance σ^2_{REF} .

The intrinsic jitter of the signal governs the accuracy of the synchronization algorithms and is a component influenced by the accuracy and non-idealities of the instrument utilized. In order to lower the precision loss brought on by the periodic REF signal jitter, a high order means filtering is actuated. As a result, the REF signal's contribution to the compensation is minimized by making it insignificant compared to channel one.

A synchronization signal is generated and distributed among the various devices as part of the method designed to correct the faults discussed before and realign the timestamps. Several algorithms have been examined based on this signal.

A representation of the synchronization process based on a standard reference signal is reproduced in Fig. 3.

To assure synchronization, the easiest way would be to spread a high-frequency clock among all FPGAs and use it as a common clock. This approach is incredibly straightforward, but it has the drawback of degrading timing signals and making signal integrity challenging to maintain. Along with these problems, the spread of the clock signal across all FPGAs would amplify its jitter, which would be impossible to minimize. Due to these problems, a low-frequency signal that does not affect signal integrity or transfer rate was chosen for the network's numerous components.

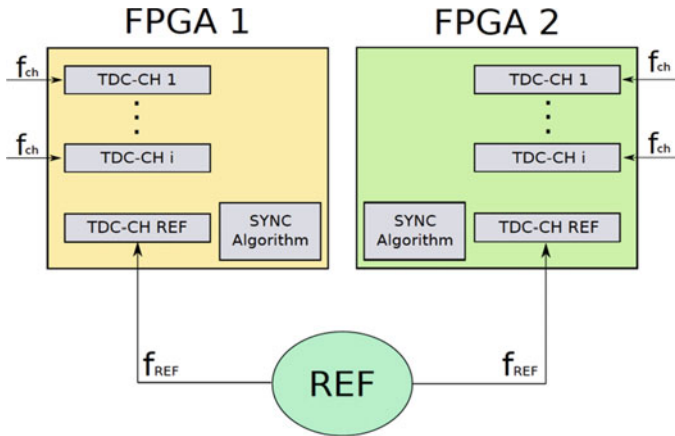


Fig. 3 Several TDCs utilize a REF signal put on two FPGAs. It is important to note that each TDC has an i -th number of data generation channels and a separate channel for the REF signal

The only measurements that matter for multichannel applications are relative ones.

The first method can be referred to as Reference Based Gain Error-based Synchronization. The reference based gain error-based synchronization method is the most logical because it compares measurements from various TDCs. By utilizing one of the TDCs as a reference and figuring out a correction factor for all the others, the objective is to make all timestamps homogeneous. It is necessary to normalize the timestamps from the generic TDC_i in relation to a calculated compensation factor using the TDC_1 as a benchmark. This implies modifying LSB_i regarding LSB_1 's value as a starting point.

The second method can be referred to as Self Gain Error-based Synchronization. It does not use a reference device because it is based on TDC's self-compensation. This approach aims to always refer to the TDC's synchronization signal period.

The third method can be referred to as Offset Error-based Synchronization. Another method to synchronize the network is a continuous offset error compensation. An initial zeroing would be sufficient to correct the offset error after all the TDCs have been turned on. By measuring the first offset between the first timestamp and the REF signals, each TDC can be zeroed at a rate equal to the REF signal frequency. In addition, to offset mistakes, also gain errors exist. However, they can be ignored provided the zeroing procedure is carried out repeatedly and more quickly than the difference in the two frequencies.

Because the first two strategies employ the same compensation method in two different ways, the results are comparable since they are based on the same engine. Both enable straightforward offset error correction by doing an initial zeroing and thorough gain error correction. These compensating methods' main drawback is that they involve multiplications that are pretty challenging to implement in FPGA designs without using dedicated hardware.

In contrast to the first two approaches, the third one improves by increasing the REF frequency, compromising the signal's integrity. The third compensation method is far easier to execute in an FPGA design than the others because it simply calls for additions and subtractions, which are straightforward to carry out in an FPGA.

5 Felix Instrument

The research activity's findings have undergone comprehensive validation in numerous experimental settings, verifying the anticipated performances and functional requirements. It is significant to present one of the instruments developed for the experimental activities rather than reporting every single experiment. This would demand an unnecessarily comprehensive treatment due to their number and complexity.

The Felix board is one of the instruments developed for practical applications in several settings and is an example of technology transfer of the research outputs.

The Felix board (Fig. 4) is programmable and implements the best-suited TDC architecture on FPGA. In addition to the cutting-edge measurement performance, the most crucial features are portability, ease of use, and compactness up to the plug-and-play operation.

The instrument's cost-to-performance ratio is quite good. It features two input channels (START and STOP) plus a SYNC input, for example, if a laser synchronization is required. Thanks to its configuration, it is suitable for testing detectors (e.g., CDL, SiPM, SPAD, PMT); it can perform time-correlation measurements and

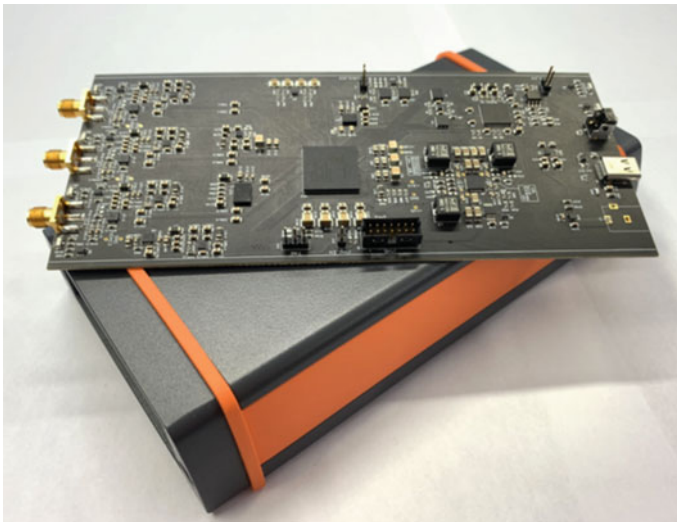


Fig. 4 Substrate for instrument hardware. The FPGA device can be seen in the center of the PCB

complete rapid characterization of measurement systems. The device was created with various firmware and software modules, including a sophisticated Graphical User Interface (GUI) and a software Application Programming Interface (API) that support the user in all conceivable application scenarios.

At acquisition rates up to 100 Msps, the single-shot precision settles below 10 ps r.m.s.. The resolution (LSB) reaches 250 fs with a short FSR. Dead time is less than 5 ns, whereas integral and differential nonlinearity (DNL, INL) is measured at less than 85 fs and 5.6 ps, respectively. The frontend circuits translate a temporal external analog signal into digital pulses using a comparator with a fully configurable threshold.

It is challenging to substantiate the assertion that the proposed design achieves the state-of-the-art in terms of both small size and high performance compared to existing alternatives. In actuality, there are no programmable devices that simultaneously maximize both of the features requested by modern uses. An application of the instrument in a configuration for coincidence measurements is shown in Fig. 5.

The instrument's core consists of an FPGA that implements the firmware that implements the TDC architecture and allows software connecting with the outside world, such as a PC. The instrument's connection gates include CH1, CH2, SYNC

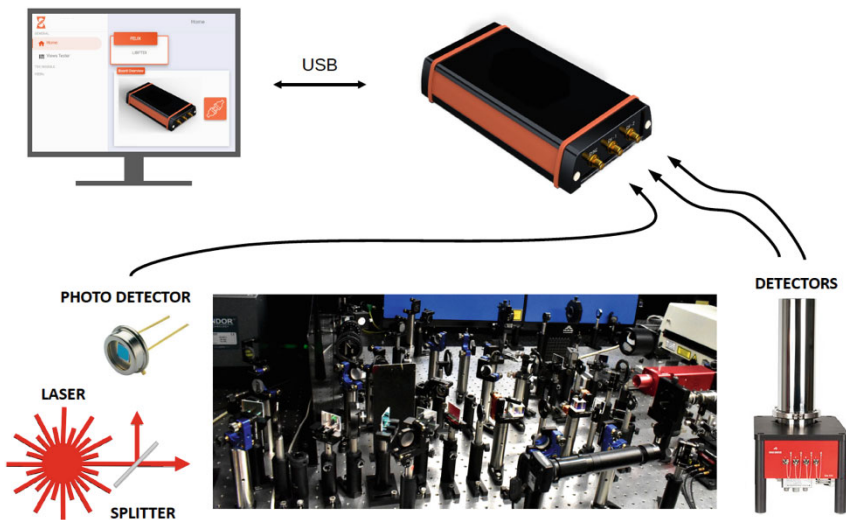


Fig. 5 Photo of the device captured when it was being used to measure coincidence. The beam from a laser source is delivered in parallel to the SYNC input of the instrument that is being presented by a photodetector, as well as to the setup where the measure for verifying the coincidence between the two likely emissions resulting from excitation is carried out. The instrument calculates the timestamps of the laser source (cause) and the timestamps from two detectors that collect the sample's emissions (effects). Here, the device decides if two detected events caused by the same laser pulse coincide. It may be possible to conduct more research, such as calculating the statistical frequency between causes and effects on a histogram. The PC's only use is to show processing results

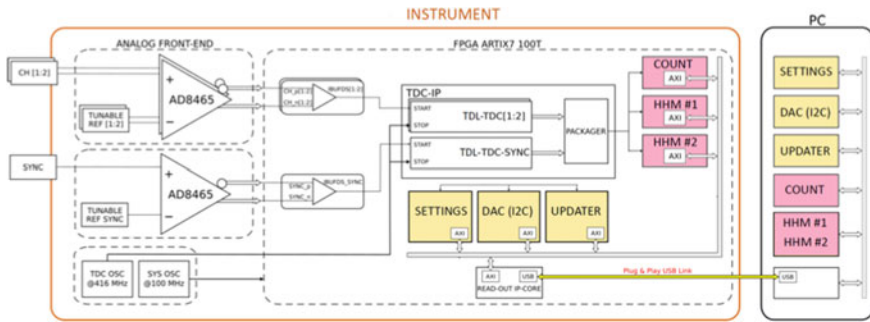


Fig. 6 A schematic diagram shows the parts of the instrument’s architecture comprising hardware, firmware, and software. The instrument’s firmware and software are highlighted in yellow, through which the user easily programs to best suit his operational requirements

channels, and a USB connector with communication and a power supply connections. The FPGA device is a 28 nm Xilinx Artix-7 on which different processing modules are structured in the form of IP-Cores and the primary core that implements the TDC. The software is made up of plug-ins, one for each type of firmware processing. As a result, the user can entirely modify the instrument by adding their own IP-Cores and accompanying software plug-ins. A synoptic diagram of the instrument’s architecture is shown in Fig. 6, emphasizing the hardware, firmware, and software elements and the connections between them.

The instrument has been tested in the lab and in many specific application scenarios, fully legitimating its functions and features.

6 Conclusions and Future Developments

An innovative, fully configurable TDC architecture for implementation on FPGA has been designed, engineered, and experimentally validated. Features are above most of the available ASIC and FPGA comparable solutions at state of the art. Providing dynamic-range above 10 s, resolution below 300 ps, single-shot precision just above 10 ps, high linearity limited to 250 fs differential and 2.5 ps integral, and up to 16 independent channels without any performance degradation.

To face modern architectural challenges of these systems, i.e., throughput beyond tens of millions of measurements per second and number of channels in parallel vertiginously growing, two primary topics have been investigated and innovative solutions proposed.

For addressing the keystone of highly efficient data routing, an innovative technique named BeltBus has been developed, which performs parallel-to-serial conversion while maintaining the chronological order of the output measurements to overcome the limitations of standard timestamp management.

The ever-increasing number of parallel channels will make it mandatory to partition them among different processing circuits that must be synchronized. A strategy for periodic correction of the synchronicity error in distributed systems has been devised, and consequently, three different operative methods have been derived.

The evolutionary trend of the systems is not to improve the resolution, already at levels beyond the need, but to increase the number of channels. Therefore, it is natural that the research's evolution will be to provide the techniques presented with the ability to manage hundreds of channels in parallel.

References

1. Eda, H., et al.: Multichannel time-resolved optical tomographic imaging system. *Rev. Sci. Instrum.* **70**(9), 3595–3602 (1999)
2. Garzetti, F., Corna, N., Lusardi, N., Geraci, A.: Time-to-digital converter IP-core for FPGA at state of the Art. *IEEE Access* **9**, 8515–85528 (2021)
3. Pan, W., et al.: High resolution distributed time-to-digital converter (TDC) in a White Rabbit network. *Nucl. Instrum. Methods Phys. Res. Sect. A* **738**, 13–19 (2014)
4. Tancock, S., Rarity, J., Dahnoun, N.: The wave-union method on DSP blocks: improving FPGA-based TDC resolutions by 3x with a 1.5x area increase. *IEEE Trans. Instrum. Measur.* **71**, 2001811 (2022)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Computer Science and Engineering

Temporal Logic and Model Checking for Operator Precedence Languages: Theory and Applications



Michele Chiari

Abstract Temporal logic is an established tool for writing requirement specifications for computer systems, thanks to its balance between expressive power and efficiency of verification algorithms. Linear Temporal Logic (LTL), one of the most commonly used, allows for naturally expressing *safety* and *liveness* requirements on a linear timeline, but incurs into some limitations when utilized to express requirements of procedural programs. In fact, such programs exhibit a typically context-free behavior, which LTL formulas cannot represent. Precedence Oriented Temporal Logic (POTL), a temporal logic based on Operator Precedence Languages (OPLs), a subclass of Deterministic Context-Free Languages. With POTL, we can express requirements involving Hoare-style pre/post-conditions, stack inspection, and others, also in the presence of exception-like constructs. We prove that POTL is as expressive as First-Order Logic (FOL) on its algebraic structure, and devise and implement an explicit-state satisfiability and model-checking algorithm for it, obtaining some promising experimental results.

Keywords Linear temporal logic · Operator precedence languages

1 Introduction

Ensuring the correctness of software artefacts is one of the most important issues in software engineering. Testing is one of the most frequently used quality assurance activities aimed at discovering software defects. However,

Program testing can be used to show the presence of bugs, but never to show their absence!
E.W. Dijkstra [12]

A way of actually *ensuring* the correctness of a program is to write a proof for it, either manually or with the help of a proof assistant [4, 21]. Although founders

M. Chiari (✉)
Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via Ponzio 34/5, 20133 Milano, Italy
e-mail: michele.chiari@polimi.it

of computer science such as Floyd [14], Hoare [15] and others developed logical frameworks that support such proofs, the practice of proving program correctness is mostly limited to academia and few other contexts. In fact, writing program proofs requires mathematical skills that are uncommon in ordinary software developers, as well as considerable amounts of time.

1.1 Automated Verification

These difficulties made the case for trying to automate program proof making, with the ideal aim to produce tools that prove or disprove a program's correctness with little to no user intervention. Unfortunately, computability theory tells us that most program properties, albeit trivial, are undecidable [16]. This, however, did not prevent the field of *program verification* from arising.

We can identify two main lines of research on program analysis: *Abstract Interpretation* [11] and *Model Checking* [10]. While abstract interpretation bypasses decidability issues by means of abstractions of the properties to be verified, model checking does so by resorting to less powerful, but decidable, models of computation. In fact, if we limit the program's memory to be finite, the whole system can be modeled as a finite *Transition System (TS)*, which admits verification algorithms for many interesting properties.

The general idea behind model checking is the following: we use an operational formalism to create a model of the system to be checked, we formulate its requirements in an appropriate mathematical formalism, and then we automatically check whether the model satisfies the requirements. If the requirements are not satisfied, then we get a *counterexample*, i.e. the description of a behavior of the system model that violates the requirements. Of course, we must employ a combination of formalisms for model and requirements that admits a (possibly efficient) model-checking algorithm. While systems are often modeled as finite TS's, which are quite similar to Finite-State Automata (FSAs) and State Charts, the choice for representing requirements usually falls on *temporal logics* such as LTL [22], Computation Tree Logic (CTL), and CTL* [9]. These logics allow for naturally expressing constraints about the evolution of the system's behavior over time, and admit relatively efficient model-checking algorithms.

The trade-off made with model checking to avoid undecidability brings in two main drawbacks [10]:

1. the computational complexity of model checking algorithms makes them intractable in theory-although not in practice-due to the *state-space* explosion problem;
2. the use of finite transition systems (or even FSA) limits the accuracy with which we can model certain systems, and the use of decidable logics (e.g., LTL or CTL) to express requirements limits the properties we can verify.

Most of the recent research on model checking addresses one or both such issues. My work, in particular, tries to address number 2.

1.2 Model-Checking Computer Programs

Requirements written in LTL say something about a program's execution traces, seen as a linear timeline of discrete events. We can specify *safety* properties, which state that certain unwanted behaviors will never occur, with the *globally* operator: $\Box \varphi$ states that φ will always be true from the moment in which $\Box \varphi$ is evaluated. E.g., $\Box \mathbf{exc}$ states that the program will never throw an exception. The *eventually* operator $\Diamond \varphi$ states that φ will happen sometime in the future. Together with \Box , it can be used to express *liveness* requirements: $\Box \Diamond (\mathbf{call} \wedge p_A)$ states that procedure p_A is called infinitely many times. The *until* operator $\psi \mathcal{U} \varphi$ is true at an instant i if there is a future time instant j where φ holds, and ψ holds in all instants between i and j . E.g., with $\neg p_A \mathcal{U} p_B$ we say that procedure p_A cannot be called before p_B is executed.

While LTL can express many useful properties, the fact that it considers a linear timeline can be a limitation when reasoning about programs. E.g., suppose we want to say that procedure p_A always returns normally (i.e., it does terminate and not due to an exception). We could try with $\Box (\mathbf{call} \wedge p_A \implies \Diamond (\mathbf{ret} \wedge p_A))$, but this would be true in an execution trace such as $\{\mathbf{call}, p_A\}\{\mathbf{call}, p_A\}\{\mathbf{ret}, p_A\}\{\mathbf{exc}\}$, where p_A calls itself recursively, but only the second instance terminates normally.

The reason why we cannot express this property in LTL is that it is equivalent to the FOL definable fragment of *Regular Languages*, while the behavior of procedures is context-free, as it is driven by a stack. Thus, a solution would be to devise a temporal logic that can express context-free properties. This was already attempted with logics based on *nested words*, such as CaRet [3] and Nested Words Temporal Logic (NWTL) [2]. These logics add a binary nesting relation connecting procedure calls with their returns, so that properties like the one above can be expressed.

Nested words, however, have some limitations. In particular, they cannot easily model exceptions, which are quite widespread in modern programming languages. The main contribution of my work fills this gap by presenting a temporal logic able to reason on procedural programs with exceptions. This logic, called POTL [7], is based on OPL, a subclass of deterministic context-free languages introduced by Floyd [13]. We studied POTL in terms of expressiveness and introduced and implemented a satisfiability and model checking procedure for it. The fact that we implemented a model checker for POTL is particularly relevant, considering that other context-free logics (e.g., CaRet and NWTL) had been previously studied only theoretically.

2 Background on Operator Precedence Languages

Here we give an intuitive overview of OPL that only requires basic familiarity with formal language theory; a more extensive treatment is given in [20] and in [5].

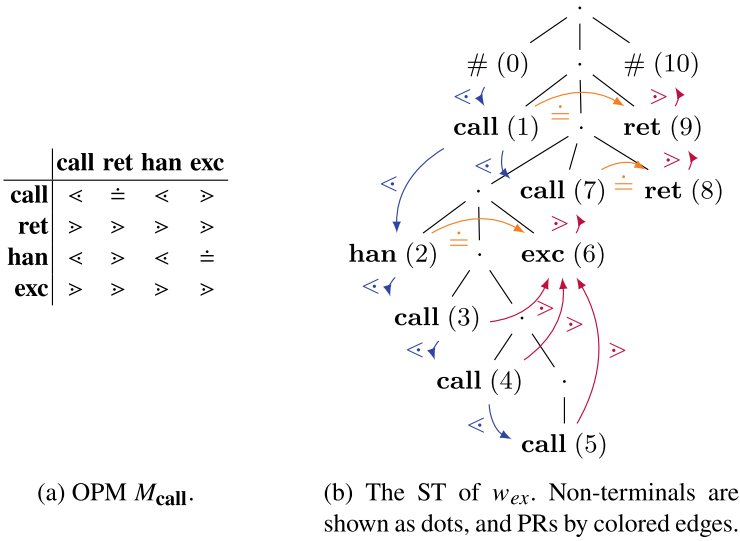
OPL have been inspired by precedence relations among operators in arithmetic expressions. They are generated by grammars in operator form, i.e. whose rules' right-hand sides (rhs's) have no consecutive non-terminals. Their parsers are guided in recognizing and reducing grammar rhs by three binary Precedence Relations (PRs) among terminal symbols. Given two terminals a, b , for any non-terminals A, B, C and mixed terminal/non-terminal strings α, β, γ , we say a yields precedence to b ($a < b$) if there exists a rule $A \rightarrow \alpha a C \beta$, s.t. a string $B b \gamma$ or $b \gamma$ derives from C in any number of passes; a is equal in precedence to b ($a \doteq b$) if there exists a rule $A \rightarrow \alpha a C b \beta$ or $A \rightarrow \alpha a b \beta$; and a takes precedence over b ($a > b$) if there is a rule $A \rightarrow \alpha C b \beta$, s.t. $\gamma a B$ or γa derives from C . In practice, $a < b$ if b is the beginning of a rhs; $a \doteq b$ if they belong to the same rhs; $a > b$ if a is the end of a rhs. If at most one PR holds between any terminal pair, once all PR are collected into an Operator Precedence Matrix (OPM), the Syntax Tree (ST) of any word on the same alphabet is fully determined. By convention, we delimit all words by #, s.t. $\# < a$ and $a > \#$ for any terminal a .

In the following, we consider execution traces that only record when a function is called (**call**), returns (**ret**), installs an exception handler (**han**) or throws an exception (**exc**). We use the example trace

$$w_{ex} = \text{call han call call call exc call ret ret}$$

resulting from a program where a procedure is called (**call**) and installs a handler (**han**). Then, three procedures are called recursively (**call call call**), and the last one throws an exception (**exc**). Finally, another function is called that terminates immediately (**call ret**), and the procedure called at the beginning also returns.

Figure 1a shows OPM M_{call} , which we use to extract the context-free structure from execution traces, while Fig. 1c shows the steps of the parsing algorithm guided by M_{call} on the example word w_{ex} . To illustrate it, we refer to rows of Fig. 1c. First, we add the delimiter # at w_{ex} 's boundaries and write all PR between consecutive characters: the result is row 0. Then, we select all innermost patterns of the form $a < c_1 \doteq \dots \doteq c_\ell > b$. In row 0, the only such pattern is the underscored **call** enclosed within the pair ($<$, $>$). This means that the ST we are going to build, if it exists, must contain an internal node with the terminal character **call** as its only child. We mark this fact by replacing the pattern $<\text{call}>$ with a dummy non-terminal character, say N -i.e., we reduce **call** to N . The result is row 1. Next, we apply the same labeling to row 1 by simply ignoring symbol N and find a new candidate for reduction, the pattern $<\text{call } N>$. The reduction of row 2 is similar, so we come to row 3. This time the terminals to be reduced are two, with an \doteq and an N in between. This means that they embrace a subtree of the ST whose root is the node represented by N . By executing the reduction leading from row 3 to 4 we produce a new N immediately



0 # < call < han < call < call < call > exc > call ≐ ret > ret > #
 1 # < call < han < call < call N > exc > call ≐ ret > ret > #
 2 # < call < han < call N > exc > call ≐ ret > ret > #
 3 # < call < han ≐ N exc > call ≐ ret > ret > #
 4 # < call < N call ≐ ret > ret > #
 5 # < call ≐ N ret > #
 6 # ≐ N #

(c) The sequence of bottom-up reductions during the parsing of w_{ex} .

Fig. 1 Demonstration of the Operator-Precedence parsing algorithm

to the left of a **call** which is matched by an equal-in-precedence **ret**. We repeat the procedure until we obtain row 6, where by convention we state the \doteq relation between the two #'s. The resulting ST is shown in Fig. 1b, with PR highlighted.

The way PR determine the ST of a string is formalized by *chains*:

Definition 1 A *simple chain* ${}^{c_0}[c_1c_2 \dots c_\ell]^{c_{\ell+1}}$ is a string $c_0c_1c_2 \dots c_\ell c_{\ell+1}$, such that: $c_0, c_{\ell+1} \in \Sigma \cup \{\#\}$, $c_i \in \Sigma$ for every $i = 1, 2, \dots, \ell$ ($\ell \geq 1$), and $c_0 < c_1 \doteq c_2 \dots c_{\ell-1} \doteq c_\ell > c_{\ell+1}$.

A *composed chain* is a string $c_0s_0c_1s_1c_2 \dots c_\ell s_\ell c_{\ell+1}$, where ${}^{c_0}[c_1c_2 \dots c_\ell]^{c_{\ell+1}}$ is a simple chain, and $s_i \in \Sigma^*$ is either the empty string or is such that ${}^{c_i}[s_i]^{c_{i+1}}$ is a chain (simple or composed), for every $i = 0, 1, \dots, \ell$ ($\ell \geq 1$). Such a composed chain will be written as ${}^{c_0}[s_0c_1s_1c_2 \dots c_\ell s_\ell]^{c_{\ell+1}}$.

In a chain, simple or composed, c_0 (resp. $c_{\ell+1}$) is called its *left* (resp. *right*) *context*; all terminals between them are called its *body*.

If we delimit chain bodies in w_{ex} by square brackets, we obtain the following:

$$\#[\text{call}[[\text{han}[\text{call}[\text{call}[\text{call}]]]\text{exc}]\text{call ret}]\text{ret}]\#$$

Some simple chains in w_{ex} are $\text{call}[\text{call}]^{\text{exc}}$ and $\text{call}[\text{han exc}]^{\text{call}}$; some composed chains are $\text{call}[\text{call}[\text{call}]]^{\text{exc}}$ and $\text{call}[[\text{han}[\text{call}[\text{call}[\text{call}]]]\text{exc}]\text{call ret}]^{\text{call}}$.

In the ST, each chain body corresponds to a sub-tree. Thus, the structure of chains in a given string is isomorphic to its ST.

OPL also have a defining class of pushdown automata, Operator Precedence Automaton(OPA) [19]. We use them for model checking POTL, but we omit their definition for lack of space.

3 OP-Words and POTL

Here we present the temporal logic POTL. Before, however, we need to define Operator Precedence (OP) words, the algebraic structure we use for modeling execution traces.

3.1 Operator-Precedence Words

POTL is a linear-time temporal logic, which extends LTL: its algebraic structure is an extension of LTL's. The semantics of LTL [22] is defined on a set of word positions,¹ respectively $U = \{0, 1, \dots, n\}$, with $n \in \mathbb{N}$ for the finite-word semantics, and $U = \mathbb{N}$ for the infinite-word one, equipped with a total ordering $<$ and monadic relations, called *atomic propositions* (AP).

OP words augment this linear order with an additional binary relation, which we call the *chain* relation and denote as χ . The χ relation is isomorphic with the context-free structure of a word, and we use it to encode the structure we want to embed in our execution traces. When modelling procedural programs, for example, the χ relation models the program's stack: function calls are linked to their respective returns or to exceptions that terminate them. More formally,

Definition 2 (*OP word*) An *OP word* on a finite set of atomic propositions AP is the tuple $\langle U, <, M_{AP}, P \rangle$, where U and $<$ are as above; M_{AP} is an OPM on $\mathcal{P}(AP)$, and $P: AP \rightarrow \mathcal{P}(U)$ is a function associating each atomic proposition with the set of positions where it holds, with $0, (n+1) \in P(\#)$.

Here we consider only finite-word languages, while the extensions needed to deal with ω -languages are covered in [5].

¹ Here we consider a discrete timeline, hence $U \subseteq \mathbb{N}$. However, LTL can be defined on any Dedekind-complete set.

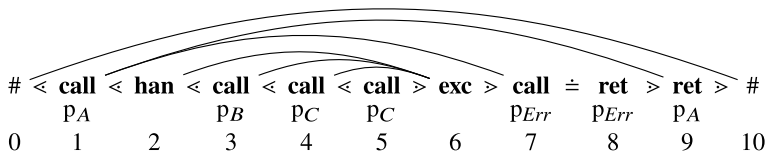


Fig. 2 w_{ex} as an OP word. Chains are highlighted by arrows joining their contexts; structural labels are in bold, and other atomic propositions are shown below them. p_j means a **call** or a **ret** is related to procedure p_j . First, procedure p_A is called (pos. 1), and it installs an exception handler in pos. 2. Then, three nested procedures are called, and the innermost one (p_C) throws an exception, which is caught by the handler. Function p_{Err} is called and, finally, p_A returns

Definition 3 (*Chain relation*) The *chain relation* $\chi(i, j)$ holds between two positions $i, j \in U$ if and only if $i < j - 1$, and i and j are respectively the left and right contexts of the same chain according to M_{AP} and the labeling induced by P .

In the following, given two positions i, j and a PR π , we write $i \pi j$ to say $a \pi b$, where $a = \{p \mid i \in P(p)\}$, and $b = \{p \mid j \in P(p)\}$. For notational convenience, we partition AP into structural labels, written in bold face, which define a word's structure, and normal labels, in round face, defining predicates holding in a position. Thus, an OPM M can be defined on structural labels only, and M_{AP} is obtained by inverse homomorphism of M on subsets of AP containing exactly one of them.

Figure 2 shows w_{ex} as an OP word, with additional propositions denoting function names. The χ relation is represented by edges. Notice how it describes the structure of the execution trace: each call is connected to the event that terminates it (either a **ret** or an **exc**), and to **calls** to nested functions (e.g., $\chi(1, 7)$).

3.2 Precedence-Oriented Temporal Logic

Here we describe the most important operators in POTL [7], leaving the remaining ones—namely, *hierarchical* operators—for [5]. Given a finite set of atomic propositions AP , the syntax of POTL follows:

$$\varphi = a \mid \neg\varphi \mid \varphi \vee \varphi \mid \bigcirc^t \varphi \mid \ominus^t \varphi \mid \chi_F^t \varphi \mid \chi_P^t \varphi \mid \varphi \mathcal{U}_\chi^t \varphi \mid \varphi \mathcal{S}_\chi^t \varphi$$

where $a \in AP$, and $t \in \{d, u\}$.

The truth of POTL formulas is defined w.r.t. a single word position. Let w be an OP word, and $a \in AP$. Then, for any position $i \in U$ of w , we have $(w, i) \models a$ if $a \in P(i)$. Propositional operators \wedge , \vee and \neg have the usual semantics. Next, while giving the formal semantics of POTL operators, we illustrate it by showing how it can be used to express properties on program execution traces, such as Fig. 2.

Next/back operators. The *downward* next and back operators \bigcirc^d and \ominus^d are like their LTL counterparts, except they are true only if the next (resp. current) position

is at a lower or equal ST level than the current (resp. preceding) one. The *upward* next and back, \bigcirc^u and \ominus^u , are symmetric.

Formally, $(w, i) \models \bigcirc^d \varphi$ if and only if $(w, i + 1) \models \varphi$ and $i < (i + 1)$ or $i \doteq (i + 1)$, and $(w, i) \models \ominus^d \varphi$ if and only if $(w, i - 1) \models \varphi$, and $(i - 1) < i$ or $(i - 1) \doteq i$. Substitute $<$ with $>$ to obtain the semantics for \bigcirc^u and \ominus^u .

For instance, \bigcirc^d **call** means that the next position is an inner call (it holds in pos. 2, 3, 4 of Fig. 2), \ominus^d **call** to say that the previous position is a **call**, and the current is the first of the body of a function (pos. 2, 4, 5), or the **ret** of an empty one (pos. 8).

The *chain* next and back operators χ_F^t and χ_P^t evaluate their operand respectively on future and past positions in the chain relation with the current one. The *downward* (resp. *upward*) variant only considers chains whose right context goes down (resp. up) in the ST.

Formally, $(w, i) \models \chi_F^d \varphi$ if and only if there exists a position $j > i$ such that $\chi(i, j)$, $i < j$ or $i \doteq j$, and $(w, j) \models \varphi$. $(w, i) \models \chi_P^d \varphi$ if and only if there exists a position $j < i$ such that $\chi(j, i)$, $j < i$ or $j \doteq i$, and $(w, j) \models \varphi$. Replace $<$ with $>$ for the upward versions.

E.g., in pos. 1 of Fig. 2, $\chi_F^d p_{Err}$ holds because $\chi(1, 7)$, meaning that p_A calls p_{Err} at least once. Also, $\chi_F^u \mathbf{exc}$ is true in **call** positions whose procedure is terminated by an exception thrown by an inner procedure (e.g. pos. 3 and 4). $\chi_P^u \mathbf{call}$ is true in **exc** statements that terminate at least one procedure other than the one raising it, such as the one in pos. 6. $\chi_F^d \mathbf{ret}$ and $\chi_F^u \mathbf{ret}$ hold in **calls** to non-empty procedures that terminate normally, and not due to an uncaught exception (e.g., pos. 1).

Until/Since operators. The *summary* until $\psi \mathcal{U}_\chi^t \theta$ (resp. since $\psi \mathcal{S}_\chi^t \theta$) operator is obtained by inductively applying the \bigcirc^t and χ_F^t (resp. \ominus^t and χ_P^t) operators. It holds in a position if either θ holds, or ψ holds with $\bigcirc^t(\psi \mathcal{U}_\chi^t \theta)$ (resp. $\ominus^t(\psi \mathcal{S}_\chi^t \theta)$) or $\chi_F^t(\psi \mathcal{U}_\chi^t \theta)$ (resp. $\chi_P^t(\psi \mathcal{S}_\chi^t \theta)$). It is an until operator on paths that move not only between consecutive positions, but also between contexts of a chain, skipping its body. With $M_{\mathbf{call}}$, this means skipping function bodies. The downward variants move between positions at the same level in the ST (i.e., in the same simple chain body), or down in the nested chain structure. The upward ones move at the same or to higher ST levels.

For example, formula $\top \mathcal{U}_\chi^u \mathbf{exc}$ is true in positions contained in the frame of a function that is terminated by an exception. It is true in pos. 3 of Fig. 2 because of path 3-6, and false in pos. 1, because no path can enter chain $\chi(1, 9)$. Formula $\top \mathcal{U}_\chi^d \mathbf{exc}$ is true in call positions whose function frame contains **exc**s, but that are not directly terminated by one of them, such as the one in pos. 1 (with path 1-2-6). Moreover, $\mathbf{call} \mathcal{U}_\chi^d (\mathbf{ret} \wedge p_{Err})$ holds in pos. 1 because of path 1-7-8, $(\mathbf{call} \vee \mathbf{exc}) \mathcal{S}_\chi^u p_B$ in pos. 7 because of path 3-6-7, and $(\mathbf{call} \vee \mathbf{exc}) \mathcal{U}_\chi^u \mathbf{ret}$ in 3 because of path 3-6-7-8.

3.3 Theoretical Results

We proved that POTL is equivalent to FOL on OP words, which is a strong guarantee for its expressive power, as state-of-the-art temporal logics, namely LTL [18] and NWTL [2], are equivalent to FOL on their respective word structures.

Theorem 1 ([5, Theorems 9.9 and 9.21] and [8]) *POTL = FOL with one free variable on finite and infinite OP words.*

Moreover, we developed a procedure for building an automaton (precisely, an OPA) that accepts exactly the execution traces that satisfy a given POTL formula. Thanks to OPA forming a Boolean algebra, we derived a satisfiability and model checking procedure, which allows us to state

Theorem 2 ([5, Theorems 10.8, 10.13]) *POTL model checking and satisfiability are EXPTIME-complete.*

This places POTL at the same level of computational complexity as less expressive logics such as NWTL [2].

4 Verifying Programs with Exceptions

We show how we can use POTL to verify programs with exceptions through a case study. The properties we are interested in are related to *exception safety*, which has been introduced to help developers deal with pitfalls caused by exceptions in C++ [1].

Let $\Box\psi$ be the LTL *globally* operator, which states the truth of its operand in all future word positions. POTL can express Hoare-style pre/post-conditions with formulas such as $\Box(\mathbf{call} \wedge \rho \implies \chi_F^d(\mathbf{ret} \wedge \theta))$, where ρ is the pre-condition, and θ is the post-condition. We can lift this formula to exceptions with the following shortcut:

$$\mathit{CallThr}(\psi) := \bigcirc^u(\mathbf{exc} \wedge \psi) \vee \chi_F^u(\mathbf{exc} \wedge \psi),$$

which, evaluated in a **call**, states that the procedure currently started is terminated by an **exc** in which ψ holds. So, $\Box(\mathbf{call} \wedge \rho \wedge \mathit{CallThr}(\top) \implies \mathit{CallThr}(\theta))$ means that if precondition ρ holds when a procedure is called, then postcondition θ must hold if that procedure is terminated by an exception. In object-oriented programming languages, if $\rho \equiv \theta$ is a class invariant asserting that a class instance's state is valid, this formula expresses *weak (or basic) exception safety*, and *strong exception safety* if ρ and θ express particular states of the class instance. The *no-throw guarantee* can be stated with $\Box(\mathbf{call} \wedge p_A \implies \neg \mathit{CallThr}(\top))$, meaning procedure p_A is never interrupted by an exception.

We take our case study from a famous tutorial [24] on how to make exception-safe generic containers in C++. It presents two implementations of a generic stack data structure, parametric on the element type T . The first one is not exception-safe: if the constructor of T throws an exception during a pop action, the topmost element is removed, but it is not returned, and it is lost. This violates the strong exception safety requirement that each operation is rolled back if an exception is thrown, as the stack is effectively popped even though the pop action failed due to the exception. The second version of the data structure instead satisfies such requirement.

While exception safety is, in general, undecidable, we can prove the stronger requirement that each modification to the data structure is only committed once no more exceptions can be thrown. We can check such requirement with the following formula:

$$\begin{aligned} & \Box(\mathbf{exc} \implies \\ & \neg((\Theta^u \text{ modified} \vee \chi_p^u \text{ modified}) \wedge \chi_p^u(\text{Stack} :: \text{push} \vee \text{Stack} :: \text{pop}))) \end{aligned} \quad (1)$$

Additionally, we can prove that both implementations are *exception neutral*, i.e. `Stack` methods do not block exceptions thrown by the underlying element type `T`'s methods and constructors. This can be done by checking the following formula:

$$\Box(\mathbf{exc} \wedge \Theta^u \top \wedge \chi_p^d(\mathbf{han} \wedge \chi_p^d \text{Stack}) \implies \chi_p^d \chi_p^d \chi_F^u \mathbf{exc}). \quad (2)$$

We implemented an explicit-state model checker for POTL, called POMC [6], which accepts in input programs written in MiniProc, a simple procedural programming language with exceptions, compiles them to an OPA, and checks them against POTL formulas. To check the two `Stack` implementations against the above requirements, we modelled them in MiniProc, and checked them against (1) and (2). POMC successfully found a counterexample to (1) for the first implementation in 3 s, and proved the safety of the second one in 12 s. Exception-neutrality was proved for both implementations by checking (2) in resp. 13 and 18 s.

More case studies can be found in [5], including a more complex one on a Java implementation of the QuickSort algorithm [23].

5 Conclusions and Future Work

We have presented POTL, a temporal logic based on OPL, proved its expressive completeness, and implemented a model checker for it. Its evaluation on several case studies shows promising results for its applicability to model-checking of procedural programs.

One natural further step could be the investigation of more efficient model checking algorithms, based on symbolic or bounded model-checking techniques [10]. Moreover, POTL model checking could be applied to programs written in real-world programming languages by means of predicate abstraction mechanisms [17].

References

1. Abrahams, D.: Exception-safety in generic components. In: *Generic Programming*. LNCS, vol. 1766, pp. 69–79. Springer (1998). https://doi.org/10.1007/3-540-39953-4_6
2. Alur, R., Arenas, M., Barceló, P., Etesami, K., Immerman, N., Libkin, L.: First-Order and Temporal Logics for Nested Words. *LMCS* **4**(4) (2008). [https://doi.org/10.2168/LMCS-4\(4:11\)2008](https://doi.org/10.2168/LMCS-4(4:11)2008)
3. Alur, R., Etesami, K., Madhusudan, P.: A temporal logic of nested calls and returns. In: *TACAS '04*. LNCS, vol. 2988, pp. 467–481. Springer (2004). https://doi.org/10.1007/978-3-540-24730-2_35
4. Bertot, Y., Castéran, P.: *Interactive Theorem Proving and Program Development - Coq'Art: The Calculus of Inductive Constructions*. Texts in Theoretical Computer Science. An EATCS Series, Springer (2004). <https://doi.org/10.1007/978-3-662-07964-5>
5. Chiari, M.: *Temporal Logic and Model Checking for Operator Precedence Languages: Theory and Applications*. Ph.D. thesis, Politecnico di Milano (2022)
6. Chiari, M., Bergamaschi, D., Pontiggia, F.: Precedence oriented model checker (POMC) (2021). <https://github.com/michiari/POMC>
7. Chiari, M., Mandrioli, D., Pradella, M.: Model-checking structured context-free languages. In: *CAV '21*. LNCS, vol. 12760, pp. 387–410. Springer (2021). https://doi.org/10.1007/978-3-030-81688-9_18
8. Chiari, M., Mandrioli, D., Pradella, M.: A First-order Complete Temporal Logic for Structured Context-free Languages. *LMCS* (2022). <https://arxiv.org/abs/2105.10740>, to appear
9. Clarke, E.M., Emerson, E.A., Sistla, A.P.: Automatic verification of finite-state concurrent systems using temporal logic specifications. *ACM Trans. Program. Lang. Syst.* **8**(2), 244–263 (1986). <https://doi.org/10.1145/5397.5399>
10. Clarke, E.M., Henzinger, T.A., Veith, H., Bloem, R. (eds.): *Handbook of Model Checking*. Springer (2018). <https://doi.org/10.1007/978-3-319-10575-8>
11. Cousot, P., Cousot, R.: Abstract interpretation: a unified lattice model for static analysis of programs by construction or approximation of fixpoints. In: *POPL'77*, pp. 238–252. ACM (1977). [10.1145/512950.512973](https://doi.org/10.1145/512950.512973)
12. Dijkstra, E.W.: Notes on structured programming (1970). <http://www.cs.utexas.edu/users/EWD/ewd02xx/EWD249.PDF>, EWD249
13. Floyd, R.W.: Syntactic analysis and operator precedence. *J. ACM* **10**(3), 316–333 (1963). <https://doi.org/10.1145/321172.321179>
14. Floyd, R.W.: Assigning meaning to programs. In: *Proceedings of the Mathematical Aspects of Computer Science: American Mathematical Society Symposia*, vol. 19, pp. 19–31. AMS (1967)
15. Hoare, C.A.R.: An axiomatic basis for computer programming. *Commun. ACM* **12**(10), 576–580 (1969). <https://doi.org/10.1145/363235.363259>
16. Hopcroft, J., Ullman, J.: *Introduction to Automata and Formal Languages*. Addison-Wesley, Reading (1979)
17. Jhala, R., Podelski, A., Rybalchenko, A.: Predicate abstraction for program verification. In: *Handbook of Model Checking*, pp. 447–491. Springer (2018). https://doi.org/10.1007/978-3-319-10575-8_15
18. Kamp, H.: *Tense logic and the theory of linear order*. Ph.D. thesis, University of California, Los Angeles (1968)
19. Lonati, V., Mandrioli, D., Panella, F., Pradella, M.: Operator precedence languages: Their automata-theoretic and logic characterization. *SIAM J. Comput.* **44**(4), 1026–1088 (2015). <https://doi.org/10.1137/140978818>
20. Mandrioli, D., Pradella, M.: Generalizing input-driven languages: theoretical and practical benefits. *Comput. Sci. Rev.* **27**, 61–87 (2018). <https://doi.org/10.1016/j.cosrev.2017.12.001>
21. Nipkow, T., Paulson, L.C., Wenzel, M.: Isabelle/HOL – A Proof Assistant for Higher-Order Logic. LNCS, vol. 2283. Springer (2002). <https://doi.org/10.1007/3-540-45949-9>

22. Pnueli, A.: The temporal logic of programs. In: FOCS '77, pp. 46–57. IEEE Computer Society (1977). <https://doi.org/10.1109/SFCS.1977.32>
23. Pontiggia, F., Chiari, M., Pradella, M.: Verification of programs with exceptions through operator-precedence automata. In: SEFM '21. LNCS, vol. 13085, pp. 293–311. Springer (2021). https://doi.org/10.1007/978-3-030-92124-8_17
24. Sutter, H.: Exception-safe generic containers. C++ Report (1997). https://ptgmedia.pearsoncmg.com/imprint_downloads/informit/aw/meyerscddemo/DEMO/MAGAZINE/SU_FRAME.HTM

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Deep and Wide Tiny Machine Learning



Simone Disabato 

Abstract In the last decades, on the one hand, Deep Learning (DL) has become state of the art in several domains, e.g., image classification, object detection, and natural language processing. On the other hand, pervasive technologies—Internet of Things (IoT) units, embedded systems, and Micro-Controller Units (MCUs)—ask for intelligent processing mechanisms as close as possible to data generation. Nevertheless, memory, computational, and energy requirements characterizing DL models are three or more orders of magnitude larger than the corresponding memory, computation, and energy capabilities of pervasive devices. This work aims at introducing a methodology to address this issue and enable pervasive intelligent processing. In particular, by defining Tiny Machine Learning (TML) solutions, i.e., machine and deep learning models that take into account the constraints on memory, computation, and energy of the target pervasive device. The proposed methodology addresses the problem at three different levels. In the first approach, the methodology devices inference-based Deep TML solutions by approximation techniques, i.e., the TML model runs on the pervasive device but was trained elsewhere. Then, the methodology introduces on-device learning for TML. Finally, the third approach develops Wide Deep TML solutions that split and distribute the DL processing over connected heterogeneous pervasive devices.

Keywords Tiny machine learning · Deep learning · Internet of Things · Micro-controllers · Distributed deep learning · On-device learning

1 Introduction

Machine Learning (ML) and Deep Learning (DL) techniques have widely spread across the most diverse areas in the last decade, achieving state of the art in several fields. Convolutional Neural Networks (CNNs) models—e.g., the ResNet [29] or the Inception [51]—and their recent extensions with Vision Transformers [24, 55],

S. Disabato (✉)
Politecnico di Milano, Milan, Italy
e-mail: simone.disabato@polimi.it

© The Author(s) 2023
C. G. Riva (ed.), *Special Topics in Information Technology*,
PoliMI SpringerBriefs, https://doi.org/10.1007/978-3-031-15374-7_7

provide a classification of input images. Similar architectures—e.g., the Yolo [47], can also identify the position of the detected objects. Other examples of DL applications are natural language processing [14, 62], anomaly detection [52], event-based cameras [10], and autonomous driving [11, 12].

All DL techniques share some common characteristics. At first, the huge number of parameters. To give an idea, the ResNet CNN has 11 to 60 million parameters, whereas the Inception-V3 24 to 43 million. The image and language modeling architectures based on transformers have one or more order of magnitude more parameters, e.g., the ViT architectures [24] have 86 to 632 million parameters, BERT [14] 110 to 345 million, and GPT-3 [8] 175 billion. It is straightforward that only the memory required by the parameters ranges from dozens of megabytes to various gigabytes. Secondly, the DL training—i.e., the optimization problem to find the parameters' values—is time-consuming (and extremely energy-hungry), requiring days or weeks on high-performance computers with clusters of GPUs or TPUs. For instance, [48] roughly calculates 4.5 GPU-years to complete the optimal training of their DL models, whereas [5] estimates 16 GPU-weeks to train the Inception-V3 CNN. Finally, although the inference—i.e., the processing of one possibly unseen input by a trained DL model—is significantly simpler than training, the number of multiply-add operations required is huge. As an example, the ResNet CNN requires 5 to 11 million multiplications to classify one image.

Following the terrific spread of pervasive devices—Embedded Systems, Internet of Things (IoT), and MicroController units (MCU)—the *intelligence for pervasive units* emerged as a novel, challenging, and breakthrough research direction and application of ML and DL. On the one hand, pervasive devices are nowadays used in a broad range of applications, such as autonomous driving, smart cities, and smart industries. On the other hand, moving the (machine and deep learning-based) intelligent processing—e.g., fault detection or change detection algorithms [22]—as close as possible to data generation is crucial to support real-time applications, prolong the system lifetime, and increase the Quality-of-Service [2, 50, 53].

In contrast to these benefits of intelligence for pervasive devices, there are the constraints on memory, computation, and energy characterizing such devices. On IoT units, the memory capacity ranges from Kilobytes to Megabytes [21], and the power consumption should be in the order of milliWatts [44]. Instead, MCUs rarely have 1 Megabyte of memory and require their application to run in the microWatts area [38]. For instance, the most powerful MCU is the STM32H743ZI MCU (of STMicroelectronics), which has a 480 MHz Cortex M7 processor and 1024 KB of RAM that is further split into five blocks with different speeds and purposes. The majority of MCUs have instead 96 to 256 KB of RAM.

This work addresses the challenges mentioned above by designing deep learning-based intelligent techniques for pervasive devices, i.e., DL models whose requirements in terms of memory footprint, computational load, and energy consumption do not overcome the corresponding technological constraints of the IoT or microcontroller unit where they are executed. In the related literature, Tiny Machine Learning (TML) [16, 27, 38, 49] is a newborn research direction with the same goals. In particular, TML takes into account all the three main approaches developed in the related

literature to reduce DL model's complexity (see Sect. 2) and provides a unified and comprehensive approach to the problem [2, 20]. More in detail, almost all the TML works focus on allowing the ML or DL inference on IoT units or MCUs, with very few works tackling and proposing on-device learning [9, 19, 20]. On-device learning is a demanding and tough task, but it allows to adapt the TML model over time from the incoming novel data and in response to concept drift, i.e., statistical variations of the data generation process. Concept drift is widespread in real-world scenarios—typical causes are sensor faults, aging or seasonality effects, and user behavior changes—and, if not correctly addressed, might result in catastrophic impacts on the TML model [22].

In the aforementioned scenario, this work proposes a methodology to design and deploy *Deep and Wide Tiny Machine Learning* solutions for the considered tiny device(s), i.e., IoT units, embedded systems, or MCUs. Specifically, the methodology deal with the problem from two different perspectives. The first one, presented in Sect. 3, details how to design inference-based Deep TML solutions, i.e., TML solutions based on Deep Learning. Section 4 extends such an approach by introducing on-device learning mechanisms. Section 5 discusses the second approach, namely Wide (Deep) TML, that splits the DL computation over a set of possibly heterogeneous tiny devices.

2 Related Literature

The related literature addresses the problem of reducing the DL models' complexity (with or without the goal of matching the technological constraints of some target devices) in three main ways, detailed in the following paragraphs. Please refer to [15] for a detailed version of this analysis.

DL Hardware Design. The design of dedicated hardware solutions provides the best power consumptions and inference times, at the expense of a complex design phase and reduced flexibility. Examples of newly designed hardware are Tensor Processing Units (TPUs) and Neural Processing Units or Accelerators [4, 34].

Approximating DL. The two main approximation techniques for DL models are task dropping (or pruning) and precision scaling [2, 40, 45]. Pruning techniques are organized into structured algorithms that prune the structure of DL models (e.g., the layers) and unstructured ones that can also trim the parameters within a level without maintaining the structure [23, 37, 43, 59]. Precision scaling mechanisms aim instead at changing the classical 32-bit floating-point representation of parameters with 8-bit data types, novel or fixed-point data-types, as well as ternary or binary ones [33, 64].

Other approaches encompass the definition of optimized architectures or layers—e.g., the depth-wise convolutions in MobileNet [31], or the groups of convolutions in ShuffleNet [61]—, the definition of sparse architectures [39, 42], and the introduction of Early-Exits [7, 17].

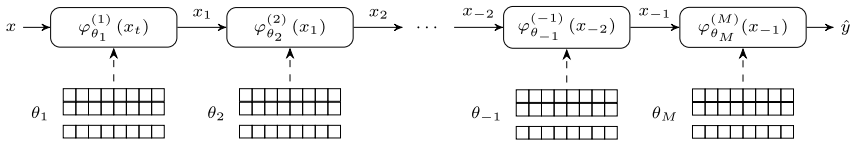
Distributing DL. In this approach, the DL computation is distributed over a set of heterogeneous tiny devices, without introducing any approximation but adding as a new layer of complexity the data transmission [32, 54, 63].

Machine Learning in Presence of Concept Drift. The literature about ML or DL in presence of concept drift organizes the works into two main classes [22, 28]: passive solutions [25, 36] always adapt the model independently from the evidence of a concept drift; whereas active approaches [18, 57, 60] employ techniques to detect changes and then adapt the model.

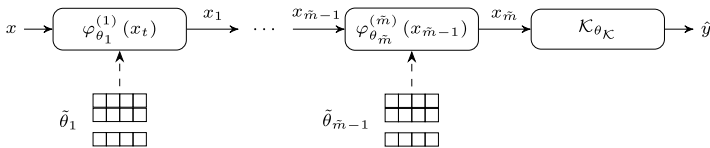
3 Deep Tiny Machine Learning

Let \mathcal{D} be a DL model that provides an output \hat{y} for an input x , hence $\hat{y} = \mathcal{D}(x)$. Its processing pipeline is a sequence of M layers $\varphi_{\theta_\ell}^{(\ell)}$ s, for each $\ell \in \{1, \dots, M\}$, with parameters θ_ℓ , where each layer can be for instance a convolution, a pooling operator, or a non-linearity. Such pipeline is then formalized as $y = \mathcal{D}(x) = \varphi_{\theta_M}^{(M)}(x_{M-1})$, with $x_\ell = \varphi_{\theta_\ell}^{(\ell)}(x_{\ell-1})$, and where x_0 degenerates to the input x .

Let Ω be a transformation function that takes as inputs a DL model \mathcal{D} and a set of approximation parameters, and provides as output the corresponding approximated DL model $\hat{\mathcal{D}}$. The definition of such a function, as well as the set of possible approximation parameters, strictly depend on the considered approximation techniques. For the purpose of this presentation, two approximation techniques are employed and shown in Fig. 1:



(a) A generic DL model \mathcal{D} architecture with M layers and 32-bit floating-point parameters θ_ℓ , where the shortcut $-x$ in layers numbering corresponds to $M - x$.



(b) The corresponding approximated DL model $\hat{\mathcal{D}}$ with $\tilde{m} \leq M$ after task dropping, a final layer \mathcal{K} , and parameters $\tilde{\theta}_\ell$ after parameter scaling (with parameter q).

Fig. 1 A sketch of the proposed mechanisms to reduce the memory footprint of a Deep Learning model \mathcal{D}

- *task dropping* (or structured pruning), that is defined as the skipping of some tasks to save memory and computation [45]. In our case, task dropping, parameterized by $\tilde{m} \leq M$, consists in keeping the first \tilde{m} layers of the original DL model \mathcal{D} , i.e., $\hat{\mathcal{D}} = \Omega(\mathcal{D}, \{\tilde{m}, \mathcal{K}\}) = \mathcal{K}_{\theta_{\mathcal{K}}}(\varphi_{\theta_{\tilde{m}}}^{(\tilde{m})}(x_{\tilde{m}-1}))$, where \mathcal{K} is an additional layer (e.g., a classifier) that may be added on top of the approximated pipeline to maintain the original DL model scope. Other task dropping techniques can be considered as well, such as considering non-consecutive layers in the approximated pipeline, unstructured pruning techniques, or layer-specific mechanisms (e.g., pruning filters within convolutional layers);
- *precision scaling* that consists in changing the representation of the parameters θ_ℓ of a DL model \mathcal{D} . The transformation is parameterized by q and is defined as $\hat{\mathcal{D}} = \Omega(\mathcal{D}, \{q\}) = \varphi_{\tilde{\theta}_M}^{(M)}(\dots(\varphi_{\tilde{\theta}_1}^{(1)}(x)))$, where $\tilde{\theta}_\ell$ represents the approximation of parameters θ_ℓ . In the simplest version, q represents the number of bits at which quantize the parameters (e.g., 1, 2, or 8 bits). Still, custom data types or different quantization schemes per each DL model layer could be considered as well.

Finally, let \bar{m} and \bar{c} be the memory and computational constraints of a target tiny device, respectively. The definition of energy constraints is more complex since every DL inference affects the energy budget, whereas the memory and the computation remain fixed. As a consequence, energy constraints are not employed in the following with the device assumed plugged into an energy source.

The methodology. In this scenario, a methodology \mathcal{M} to automatically design Deep Tiny Machine Learning solutions, i.e., DL-based TML models satisfying the technological constraints of the tiny device executing them, is defined as the function $\mathcal{M}(\mathcal{D}, \Omega, \mathcal{T}, \bar{m}, \bar{c})$, as in [2] and with \mathcal{T} being a validation set. Such a methodology operates in two steps. The first step identifies the set \mathcal{F}_{all} of all the feasible solutions. Formally, $\mathcal{F}_{all} = \{\hat{\mathcal{D}} \text{ s.t. } m_{\hat{\mathcal{D}}} \leq \bar{m} \text{ and } c_{\hat{\mathcal{D}}} \leq \bar{c}\}$, where $m_{\hat{\mathcal{D}}}$ and $c_{\hat{\mathcal{D}}}$ represent the memory and the computational requirement of the approximation $\hat{\mathcal{D}}$, respectively. The set is generated by applying the transformation Ω for all the considered approximations and the range of their parameters (in our case $\tilde{m} = \{1, \dots, M\}$, a set of possible layers \mathcal{K} s, and a set of precision scaling mechanisms q s). A trade-off between the granularity of explored approximations and the corresponding exploration time is hence expected.

The methodology's second step takes as input the set of feasible solutions \mathcal{F}_{all} and the validation set \mathcal{T} and provides as output a Pareto Front \mathcal{F} of those solutions. More in detail, the front is defined by the triples $\langle \hat{\alpha}_i, m_{\hat{\mathcal{D}}_i}, c_{\hat{\mathcal{D}}_i} \rangle$, computed after training (and evaluating it with the metric $\hat{\alpha}$) only layer \mathcal{K} (in a transfer-learning fashion [58]) by relying on disjoint subsets of \mathcal{T} .

Selecting the target solution: a discussion. Since the methodology provides as output a Pareto front, the choice of the optimal solution for each application scenario is left

to the methodology’s user. As an example, real-time scenarios suggest choosing the solution with a smaller computational complexity (probably paying in terms of accuracy). In contrast, in scenarios where accuracy is crucial, this reasoning no longer holds.

A second aspect to consider is the generalization of the chosen solution besides the validation set \mathcal{T} (as it happens in autoML tools [26, 56]). Although not investigated for the proposed methodology, we expect that the introduced technological constraints will mitigate the problem, not allowing the creation of over-complex (and prone to overfitting) DL models.

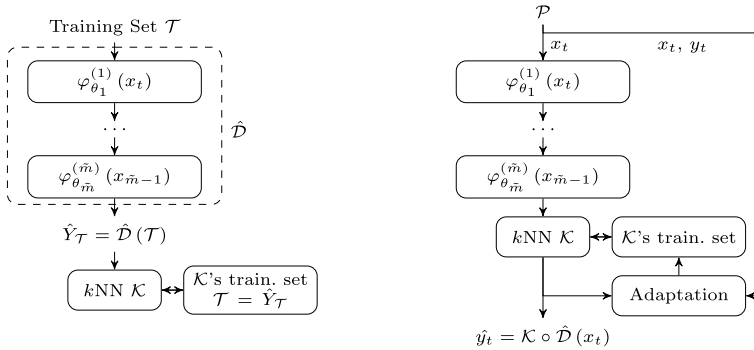
4 On-Device Tiny Machine Learning

The need for adaptation. The methodology briefly introduced in Sect. 3 proposes a set of solutions whose inference is feasible on the desired tiny device. Consequently, a deployed solutions can process novel unseen inputs x and provide corresponding outputs \hat{y} , but it will remain fixed over time. The latter is a major limitation in real scenarios for two main reasons. At first, the data needed to train (more precisely, fine-tune) the DL solution might be extremely limited or available incrementally. Hence, it is desirable to deploy a first (approximated) DL model on the tiny device and then consider some incremental learning mechanisms to improve the model on newly arrived data [19]. The second reason is the intrinsic non-stationarity of the environment in which the DL model operates due to seasonality, faults in the sensors, or aging effects, to make a few examples. A model that does not adapt in response to an environmental change—formally, a concept drift that can be viewed as a statistical variation of the data-generation process’ behavior—is likely to become useless quickly [22].

The proposed solution, presented in [19, 20], exploits the freedom the methodology discussed in Sect. 3 allows in the choice of the additional layer \mathcal{K} to be added on top of the approximated DL model $\hat{\mathcal{D}}$. Although a straightforward approach to the problem suggests to rely on classical DL-based classifiers (e.g., fully-connected layers) or machine learning parametric and supervised classifiers (e.g., Support Vector Machines [13] or Random Forests [30]), here the choice is to rely on the non-parametric k -Nearest Neighbor [3].

The k NN algorithm predicts according to the majority voting of the k samples of the input x within its knowledge (or training) set \mathcal{T} , where k is typically set as the ceil of the square root of the cardinality of \mathcal{T} [1]. The main advantages of choosing the k NN are the absence of a training phase (providing a training set is enough) and the cost-free adaptation procedure, where it is sufficient to modify the training set \mathcal{T} to update the k NN classifier as well. The drawbacks are two: the training set has to be kept entirely into memory and the prediction time increases with the cardinality of the training set.

Figure 2 shows the resulting architecture $\mathcal{K} \circ \hat{\mathcal{D}}$ of a DL model that can learn on a tiny device, with an approximated DL model $\hat{\mathcal{D}}$ followed by the k NN classifier



(a) Training: definition of the classifier’s knowledge (or training set). (b) Testing with \mathcal{P} being the data generation process.

Fig. 2 The architecture of a DL model $\mathcal{K} \circ \hat{\mathcal{D}}$ that can learn on device

\mathcal{K} . More in detail, Fig. 2a shows the *training* of such architecture, that consists in providing a training set \mathcal{T} to the k NN, whereas Fig. 2b shows the architecture during the operational life, with an adaptation module that, in a test-then-train scenario [18, 22], receives as input the classification $\hat{y} = \mathcal{K} \circ \hat{\mathcal{D}}(x)$ of an input x and then the true label y .

Adaptation module. The adaptation module can adapt the model over time according to different policies (see Sect. 2 to understand nomenclature):

- *Incremental Approach* [19]. This approach adds every supervised sample to the k NN’s training set. It may be useful when a few data are available at the training stage.
- *Passive Update* [20]. By definition, a passive approach adapts the model at every incoming (supervised) sample. In particular, in this scenario, only the misclassified samples are added to the k NN’s training set since they are assumed to bring more information about the underlying problem.
- *Active Update* [20]. Active approaches rely on an explicit mechanism to detect a concept drift, namely, Change Detection Tests (CDTs). Examples of CDTs can be found in [6, 41, 46, 60]. Once CDTs detect a change, the adaptation takes place in two steps, assuming that the algorithm kept a (history) window of the last seen samples. The first one aims at identifying the samples within the history window representing the new knowledge after the concept drift. Then, those samples are used to update the model (here, they entirely replace the k NN’s training set).
- *Hybrid Update* [20]. The hybrid approach fuses a passive and active update to combine the benefits and mitigate the drawbacks of each approach. More in detail, the active update aims at detecting the concept drift and thus removing obsolete knowledge from the model. In contrast, the passive update continuously updates the model (also in stationary conditions). In such a way, the passive updates can alleviate the case where the active approach does not detect a concept drift, whereas

the active approach should keep under control the memory, which in this passive approach indefinitely grows.

Limitations. The main limitation of the proposed approach is the definition of supervised on-device mechanisms. Introducing semi-supervised or unsupervised adaptation mechanisms will broaden the range of applications since supervised information is rarely or sometimes available in many real scenarios.

5 Wide (Deep) Tiny Machine Learning

Sections 3 and 4 enable the inference and the on-device training of DL models on tiny devices by relying on approximation mechanisms to reduce the memory and computational requirements of the DL model until the technological constraints in terms of memory and computation of the tiny device are met. This approach comes at the expense of a drop in the DL model evaluation metric. A different approach to the problem aims at distributing the computation over a set of possibly heterogeneous tiny devices. In this way, the original DL model is split into smaller tasks—e.g., each DL layer can be a task—that are feasible for the tiny devices executing them, but since no approximation is introduced, the DL model is expected to behave in the same way as if it is not distributed. Tiny devices are referred to as IoT units or units in the following.

The methodology. A methodology working in this scenario requires to define:

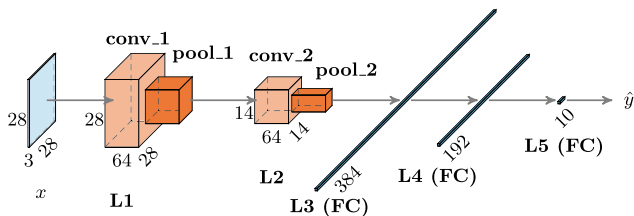
- *The IoT System model.* A sample IoT system comprises C data-acquisition units $\{s_1, \dots, s_C\}$, C target units $\{f_1, \dots, f_C\}$, and a set $\mathbb{N}_N = \{1, \dots, N\}$ of N IoT computing units. Each unit has a memory capacity \bar{m}_i and available computation \bar{c}_i , for each $i \in \mathbb{N}_N$. See Fig. 3b and c as an example.

Such a system is modeled as a graph with nodes $V = \{\mathbb{N}_N \cup \{s_1, \dots, s_C, f\}\}$ and arcs connecting units within the range of the transmission technology. In this formalization, the distance d_{i_1, i_2} , for each $i_1, i_2 \in V$, is the shortest communication path between units i_1 and i_2 within the graph.

- *The split-policy.* The simplest way to split a DL model is layer-wise. Other policies, e.g., also splitting within the DL layers (e.g., the convolutional filters as in [63]), could be considered as well.

In the most general formulation, the methodology assumes to distribute C DL models, i.e., one per each source. The u -th DL model, for each $u \in \mathbb{N}_C = \{1, \dots, C\}$, has M_u layers, each of which has: memory demand $m_{u,j}$; computation $c_{u,j}$; and activations of size $K_{u,j}$, for each $j \in \{1, \dots, M_u\}$.

- *The latency.* The transmission latency $t_{u,i,k,j}^{(t)}$ (between units $i, k \in \mathbb{N}_N$) models the time required to transmit the activations $K_{u,j}$ of DL model u 's layer j and may include possible delays (e.g., due to transmission failures or queues). The processing latency $t_{u,i,j}^{(p)}$ models the time the unit i requires to carry out the computation $c_{u,j}$ of j -th layer of DL model u .



(a) The architecture of a 5-layer DL model, where x is a 3D input (e.g., a RGB image).

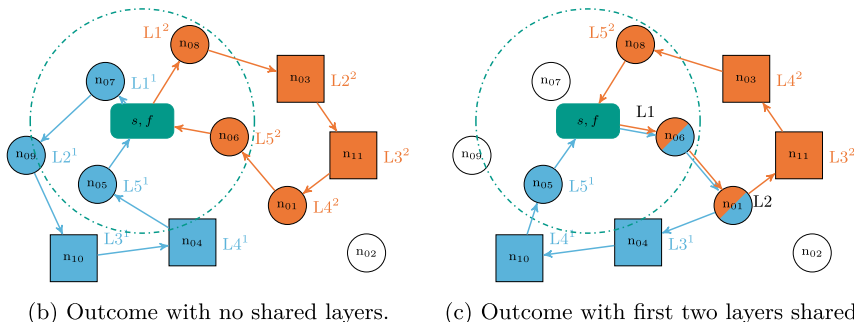


Fig. 3 Distributing two 5-layer DL models over an IoT system with or without shared layers. The squares are Raspberry Pi 3B+ and the circles STM32H7 MCUs. The transmission range is shown by the green dash-dotted line

In these settings, the methodology [21] is a quadratic binary optimization problem on variables $\alpha_{u,i,j}$ (that are ones if unit i runs the layer j of DL model u) that minimizes the *decision-making latency* [2] (i.e., from source to target):

$$\sum_{u=1}^C \sum_{i=1}^N \sum_{k=1}^N \sum_{j=1}^{M-1} \alpha_{u,i,j} \cdot \alpha_{u,k,j+1} \cdot t_{u,i,k,j}^{(t)} + t_s^{(t)} + t_f^{(t)} + \sum_{u=1}^C \sum_{i=1}^N \sum_{j=1}^M \alpha_{u,i,j} \cdot t_{u,i,j}^{(p)},$$

where $M = \max\{M_u\}$; the first term models the transmission times among IoT units; $t_s^{(t)}$ and $t_f^{(t)}$ “hidden” the transmission times from sources and to targets; and the last term models the processing time. Please refer to [21] for details.

Such optimization problem is subject to (at least) three set of constraints, two ensuring the memory and computational technological constraints, and one guaranteeing that each DL layer is assigned exactly once:

$$\forall i \in \mathbb{N}_N, \sum_{u=1}^C \sum_{j=1}^M \alpha_{u,i,j} \cdot m_{u,j} \leq \bar{m}_i \quad \text{and} \quad \forall i \in \mathbb{N}_N, \sum_{u=1}^C \sum_{j=1}^M \alpha_{u,i,j} \cdot c_{u,j} \leq \bar{c}_i,$$

$$\forall u \in \mathbb{N}_C, \forall j \in \mathbb{N}_M, \sum_{i=1}^N \alpha_{u,i,j} = 1 \text{ [if } j \leq M_u, 0 \text{ otherwise].}$$

Other constraints can be introduced to model specific situations as explained in [21]. For instance, Fig. 3 shows an application of the methodology where two 5-layer DL models (Fig. 3a) has to be placed in the IoT system shown in Fig. 3b and c. In particular, to ensure that the first two layers are shared between the two DL models in Fig. 3c, additional constraints force the first two layers of the DL models to the same IoT unit and avoid that those layers are counted twice in the memory or computational budget [21].

Limitations. The proposed methodology is a quadratic binary optimization problem that is NP-complete since it can be converted to a binary linear program, which is one of Karp's 21 NP-complete problems [35].

The second main limitation is that the optimization problem does not consider energy constraints for the reasons stated in Sect. 3.

6 Conclusions

This work addressed the problem of designing DL models in a TML scenario, i.e., DL models whose requirements in terms of memory, computation, and energy match with the corresponding technological constraints introduced by the IoT unit, or the micro-controller on which they are deployed. To achieve this goal, two main approaches have been proposed: the first one relies on approximation techniques to reduce the memory and computational requirements, with the possibility of defining solutions that can learn directly on the device; the second one does not introduce any approximation in the DL model, but splits it into smaller tasks and distributes them over a set of possibly heterogeneous devices.

Future work encompasses the definition of semi-supervised on-device mechanisms, the introduction of energy constraints, and a different approach for the methodology in Sect. 5 that is not NP-complete.

References

1. Alippi, C., Boracchi, G., Roveri, M.: Just-in-time classifiers for recurrent concepts. *IEEE Trans. Neural Netw. Learn. Syst.* **24**(4), 620–634 (2013)
2. Alippi, C., Disabato, S., Roveri, M.: Moving convolutional neural networks to embedded systems: the AlexNet and VGG-16 case. In: 17th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN), pp. 212–223. IEEE, Porto (2018)
3. Altman, N.S.: An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **46**(3), 175–185 (1992)
4. Ando, K., Ueyoshi, K., Orimo, K., Yonekawa, H., Sato, S., Nakahara, H., Takamaeda-Yamazaki, S., Ikebe, M., Asai, T., Kuroda, T., et al.: Brein memory: A single-chip binary/ternary reconfigurable in-memory deep neural network accelerator achieving 1.4 tops at 0.6 w. *IEEE J. Solid-State Circuits* **53**(4), 983–994 (2017)

5. Badre, P., Bandiwadekar, S., Chandanshive, P., Chaudhari, A., Jadhav, M.S.: Automatically identifying animals using deep learning. *Int. J. Recent Innov. Trends Comput. Commun* **6**(4), 194–197 (2018)
6. Basseville, M., Nikiforov, I.V., et al.: *Detection of Abrupt Changes: Theory and Application*, vol. 104. Prentice Hall Englewood Cliffs (1993)
7. Bolukbasi, T., Wang, J., Dekel, O., Saligrama, V.: Adaptive neural networks for efficient inference. In: *Proceedings of the 34th International Conference on Machine Learning*, vol. 70. pp. 527–536 (2017)
8. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al.: Language models are few-shot learners. *Adv. Neural. Inf. Process. Syst.* **33**, 1877–1901 (2020)
9. Cai, H., Gan, C., Zhu, L., Han, S.: Tinytl: reduce memory, not parameters for efficient on-device learning. *Adv. Neural Inf. Process. Syst.* **33** (2020)
10. Cannici, M., Ciccone, M., Romanoni, A., Matteucci, M.: Attention mechanisms for object recognition with event-based cameras. In: *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1127–1136. IEEE (2019)
11. Chen, C., Seff, A., Kornhauser, A., Xiao, J.: Deepdriving: Learning affordance for direct perception in autonomous driving. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2722–2730 (2015)
12. Chen, D., Krähenbühl, P.: Learning from all vehicles. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17222–17231 (2022)
13. Cortes, C., Vapnik, V.: Support-vector networks. *Mach. Learn.* **20**(3), 273–297 (1995)
14. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, vol. 1 (Long and Short Papers), pp. 4171–4186 (2019)
15. Disabato, S.: *Deep and Wide Tiny Machine Learning* (2022)
16. Disabato, S., Canonaco, G., Flikkema, P.G., Roveri, M., Alippi, C.: Birdsong detection at the edge with deep learning. In: *2021 IEEE International Conference on Smart Computing (SMARTCOMP)*, pp. 9–16. IEEE (2021)
17. Disabato, S., Roveri, M.: Reducing the computation load of convolutional neural networks through gate classification. In: *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE (2018)
18. Disabato, S., Roveri, M.: Learning convolutional neural networks in presence of concept drift. In: *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE (2019)
19. Disabato, S., Roveri, M.: Incremental on-device tiny machine learning. In: *Proceedings of the 2nd International Workshop on Challenges in Artificial Intelligence and Machine Learning for Internet of Things*, pp. 7–13 (2020)
20. Disabato, S., Roveri, M.: Tiny machine learning for concept drift (2021). arXiv preprint [arXiv:2107.14759](https://arxiv.org/abs/2107.14759)
21. Disabato, S., Roveri, M., Alippi, C.: Distributed deep convolutional neural networks for the internet-of-things. *IEEE Trans. Comput.* (2021)
22. Ditzler, G., Roveri, M., Alippi, C., Polikar, R.: Learning in nonstationary environments: a survey. *IEEE Comput. Intell. Mag.* **10**(4), 12–25 (2015)
23. Dong, X., Chen, S., Pan, S.J.: Learning to prune deep neural networks via layer-wise optimal brain surgeon. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 4860–4874 (2017)
24. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16 × 16 words: transformers for image recognition at scale. In: *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3–7, 2021* (2021)
25. Elwell, R., Polikar, R.: Incremental learning of concept drift in nonstationary environments. *IEEE Trans. Neural Netw.* **22**(10), 1517–1531 (2011)

26. Fabris, F., Freitas, A.A.: Analysing the overfit of the auto-sklearn automated machine learning tool. In: International Conference on Machine Learning, Optimization, and Data Science, pp. 508–520. Springer (2019)
27. Fedorov, I., Stamenovic, M., Jensen, C., Yang, L.C., Mandell, A., Gan, Y., Mattina, M., Whatmough, P.N.: Tynilstms: efficient neural speech enhancement for hearing aids (2020). arXiv preprint [arXiv:2005.11138](https://arxiv.org/abs/2005.11138)
28. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., Bouchachia, A.: A survey on concept drift adaptation. *ACM Comput. Surv. (CSUR)* **46**(4), 44 (2014)
29. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
30. Ho, T.K.: Random decision forests. In: Proceedings of 3rd International Conference on Document Analysis and Recognition, vol. 1, pp. 278–282. IEEE (1995)
31. Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H.: Mobilenets: efficient convolutional neural networks for mobile vision applications (2017). arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861)
32. Hu, D., Krishnamachari, B.: Fast and accurate streaming cnn inference via communication compression on the edge. In: 2020 IEEE/ACM Fifth International Conference on Internet-of-Things Design and Implementation (IoTDI), pp. 157–163. IEEE (2020)
33. Hubara, I., Courbariaux, M., Soudry, D., El-Yaniv, R., Bengio, Y.: Quantized neural networks: training neural networks with low precision weights and activations. *J. Mach. Learn. Res.* **18**(1), 6869–6898 (2017)
34. Jouppi, N.P., Young, C., Patil, N., Patterson, D.: A domain-specific architecture for deep neural networks. *Commun. ACM* **61**(9), 50–59 (2018)
35. Karp, R.M.: Reducibility among combinatorial problems. In: Complexity of computer computations, pp. 85–103. Springer (1972)
36. Krawczyk, B., Woźniak, M.: One-class classifiers with incremental learning and forgetting for data streams with concept drift. *Soft. Comput.* **19**(12), 3387–3400 (2015)
37. Li, H., Kadav, A., Durdanovic, I., Samet, H., Graf, H.P.: Pruning filters for efficient convnets. In: 5th International Conference on Learning Representations, ICLR 2017. Toulon, France (2017)
38. Lin, J., Chen, W.M., Lin, Y., Cohn, J., Gan, C., Han, S.: Mxnet: Tiny deep learning on iot devices. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) *Advances in Neural Information Processing Systems*, vol. 33, pp. 11711–11722. Curran Associates, Inc (2020)
39. Liu, B., Wang, M., Foroosh, H., Tappen, M., Pensky, M.: Sparse convolutional neural networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 806–814 (2015)
40. Liu, Z., Sun, M., Zhou, T., Huang, G., Darrell, T.: Rethinking the value of network pruning. In: 7th International Conference on Learning Representations, ICLR 2019. New Orleans, LA, USA (2019)
41. Lorden, G., et al.: Procedures for reacting to a change in distribution. *Ann. Math. Stat.* **42**(6), 1897–1908 (1971)
42. Lu, Y., Lu, G., Zhang, B., Xu, Y., Li, J.: Super sparse convolutional neural networks. *Proc. AAAI Conf. Artif. Intell.* **33**(01), 4440–4447 (2019)
43. Luo, J.H., Wu, J., Lin, W.: Thinet: a filter level pruning method for deep neural network compression. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 5058–5066 (2017)
44. Martinez, B., Monton, M., Vilajosana, I., Prades, J.D.: The power of models: modeling power consumption for iot devices. *IEEE Sens. J.* **15**(10), 5777–5789 (2015)
45. Mittal, S.: A survey of techniques for approximate computing. *ACM Comput. Surv. (CSUR)* **48**(4), 1–33 (2016)
46. Page, E.S.: Continuous inspection schemes. *Biometrika* **41**(1/2), 100–115 (1954)
47. Redmon, J., Divvala, S., Girshick, R., Farhadi, A.: You only look once: unified, real-time object detection. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788. CPVR '16, IEEE (2016)

48. Richards, L.E., Nguyen, A., Capps, R., Forsyth, S., Matuszek, C., Raff, E.: Adversarial transfer attacks with unknown data and class overlap. In: Proceedings of the 14th ACM Workshop on Artificial Intelligence and Security, pp. 13–24 (2021)
49. Rusci, M., Fariselli, M., Capotondi, A., Benini, L.: Leveraging automated mixed-low-precision quantization for tiny edge microcontrollers. In: IoT Streams for Data-Driven Predictive Maintenance and IoT, Edge, and Mobile for Embedded Machine Learning, pp. 296–308. Springer (2020)
50. Sanchez-Iborra, R., Skarmeta, A.F.: Tinymml-enabled frugal smart objects: challenges and opportunities. *IEEE Circuits Syst. Mag.* **20**(3), 4–18 (2020)
51. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)
52. Tack, J., Mo, S., Jeong, J., Shin, J.: Csi: Novelty detection via contrastive learning on distributionally shifted instances. In: Advances in Neural Information Processing Systems, vol. 33, pp. 11839–11852 (2020)
53. Tang, J., Sun, D., Liu, S., Gaudiot, J.L.: Enabling deep learning on iot devices. *Computer* **50**(10), 92–96 (2017)
54. Teerapittayanon, S., McDanel, B., Kung, H.T.: Distributed deep neural networks over the cloud, the edge and end devices. In: 2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS), pp. 328–339. IEEE (2017)
55. Wang, Y., Huang, R., Song, S., Huang, Z., Huang, G.: Not all images are worth 16×16 words: dynamic transformers for efficient image recognition. *Adv. Neural Inf. Proc. Syst.* **34** (2021)
56. Wong, C., Houlsby, N., Lu, Y., Gesmundo, A.: Transfer learning with neural automl. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems, pp. 8366–8375 (2018)
57. Yang, Z., Al-Dahidi, S., Baraldi, P., Zio, E., Montelatici, L.: A novel concept drift detection method for incremental learning in nonstationary environments. *IEEE Trans. Neural Netw. Learn. Syst.* **31**(1), 309–320 (2019)
58. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? In: Advances in Neural Information Processing Systems, pp. 3320–3328 (2014)
59. Yu, R., Li, A., Chen, C.F., Lai, J.H., Morariu, V.I., Han, X., Gao, M., Lin, C.Y., Davis, L.S.: Nisp: pruning networks using neuron importance score propagation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9194–9203 (2018)
60. Zambon, D., Alippi, C., Livi, L.: Concept drift and anomaly detection in graph streams. *IEEE Trans. Neural Netw. Learn. Syst.* **29**(11), 5592–5605 (2018)
61. Zhang, X., Zhou, X., Lin, M., Sun, J.: Shufflenet: an extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6848–6856 (2018)
62. Zhang, Z., Han, X., Liu, Z., Jiang, X., Sun, M., Liu, Q.: Ernie: Enhanced language representation with informative entities. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pp. 1441–1451 (2019)
63. Zhao, Z., Barijough, K.M., Gerstlauer, A.: Deepthings: distributed adaptive deep learning inference on resource-constrained iot edge clusters. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **37**(11), 2348–2359 (2018)
64. Zhu, F., Gong, R., Yu, F., Liu, X., Wang, Y., Li, Z., Yang, X., Yan, J.: Towards unified int8 training for convolutional neural network. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1969–1979 (2020)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Learning in the Presence of Multiple Agents



Giorgia Ramponi

Abstract Reinforcement Learning (RL) has emerged as a powerful tool to solve sequential decision-making problems, where a learning agent interacts with an unknown environment in order to maximize its rewards. Although most RL real-world applications involve multiple agents, the Multi-Agent Reinforcement Learning (MARL) framework is still poorly understood from a theoretical point of view. In this manuscript, we take a step toward solving this problem, providing theoretically sound algorithms for three RL sub-problems with multiple agents: Inverse Reinforcement Learning (IRL), online learning in MARL, and policy optimization in MARL. We start by considering the IRL problem, providing novel algorithms in two different settings: the first considers how to recover and cluster the intentions of a set of agents given demonstrations of near-optimal behavior; the second aims at inferring the reward function optimized by an agent while observing its actual learning process. Then, we consider online learning in MARL. We showed how the presence of other agents can increase the hardness of the problem while proposing statistically efficient algorithms in two settings: Non-cooperative Configurable Markov Decision Processes and Turn-based Markov Games. As the third sub-problem, we study MARL from an optimization viewpoint, showing the difficulties that arise from multiple function optimization problems and providing a novel algorithm for this scenario.

Keywords Multi-agent learning · Reinforcement learning · Inverse reinforcement learning

1 Introduction

Learning is one of the most fascinating open problems of our days. The first thing we can do is to observe the world and try to infer how this process happens in nature. For example, think about how humans learn to perform a task: humans adapt their

G. Ramponi (✉)
Politecnico di Milano, Milano, Italy
e-mail: giorgia.ramponi@polimi.it

© The Author(s) 2023
C. G. Riva (ed.), *Special Topics in Information Technology*,
PoliMI SpringerBriefs, https://doi.org/10.1007/978-3-031-15374-7_8

behavior to maximize a signal from the environment. Imagine a child who has to learn to ride a bicycle. The child starts by sitting on the seat, and nothing happens. Then she puts her feet on the pedals but rides too slowly, causing the bicycle to lose its balance, and she falls. She thus learned from her experience that she must pedal faster to avoid falling again. This concept is mathematically modeled by *Reinforcement Learning* [20]. However, we need to consider that humans, and animals, do not live alone: we are “social beings”, i.e., we act in a social system in which multiple entities interact with each other. Therefore, the actions of each individual can modify the learning process of all the other entities involved. For example, if we decide to buy a stock on the stock exchange, the result of our action will affect not only us but also the entire stock market. It is easy to imagine that these interactions can be very complex, and we can hardly understand how our decisions can affect the world around us. One of the sciences that mathematically model these interactions is called *Game Theory* [8]. From these considerations, we can conclude that in order to create a system that is capable of acting autonomously, we must study how to build an autonomous learning agent (*Reinforcement Learning*) and model how this is influenced by the other entities that surround it (*Game Theory*). *Multi-agent Reinforcement Learning* (MARL) [4, 21] is a bridge between these two worlds. The MARL framework studies the problem of learning by interacting with an unknown system, considering that it is composed of more than one entity.

In this contribution, we provide a summary of the Ph.D. dissertation entitled “Challenges and Opportunities in Multi-Agent Reinforcement Learning” [13], focused on studying the aspects of learning in multi-agent environments. In Sect. 2 we provide the preliminaries and background on RL, MARL, and IRL. Then, we start outline the contribution of the work. For each contribution, we introduce the setting, provide some insights on the proposed algorithm, and discuss the results. In Sect. 3 we analyze the problem of IRL in a multi-agent environment from two viewpoints: first, we consider the problem of learning the intentions of another agent which is learning a new task; second, we propose an algorithm to deal with IRL about Multiple Intentions, i.e., the problem of recovering the reward functions from a set of experts. In Sect. 4 we address the problem of online learning in Stochastic Games. Specifically, we propose an algorithm to deal with the online learning problem in the Configurable Markov Decision Process, where the two entities involved are the configurator of the environment and the RL agent. Then, we introduce a new lower bound on the online learning problem in Stochastic Games, proposing an algorithm that nearly matches this lower bound. Finally, in Sect. 5 we summarize the results and discuss some future research directions. We do not discuss the fourth part “Policy Optimization in Multi-Agent Reinforcement Learning” due to lack of space. In this part, the author reported the result of the paper [17], where they introduced a new optimization algorithm for Continuous Stochastic Games, proving convergence results to equilibrium solutions in general games.

2 Preliminaries

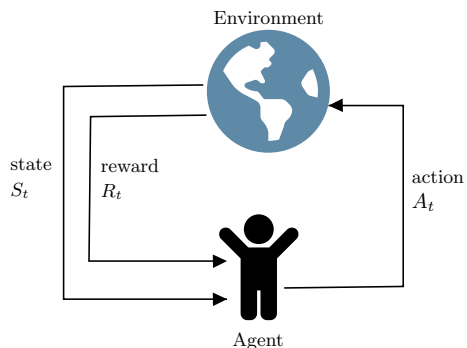
In this section, we provide the necessary background on Reinforcement Learning, Multi-Agent Reinforcement Learning and Inverse Reinforcement Learning.

2.1 (Multi-agent) Reinforcement Learning

Reinforcement Learning (RL) is a framework to learn by trial-and-error in a sequential-decision way: the agent performs an action and receives feedback from the environment. The RL problem involves a learning *agent* (or learner) which interacts with an *environment* during a sequence of discrete-time steps. This interaction is described by three components: the *state* s , the *action* a and the *reward* r (see Fig. 1). The *state* describes the actual configuration of the environment perceived by the agent, which can be a subset of the environment state's characteristics. The *action* consists of the decision taken by the RL agent. The environment responds to every performed action changing the state and giving the agent a *reward*, where the *reward* is numeric feedback of the agent's performances. The interactions are formally described by a Markov Decision Process (MDP) [3, 12] $\mathcal{M} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma, \mu, H)$, where \mathcal{S} and \mathcal{A} are respectively the set of states and actions, \mathcal{P} is the probability of transitioning from a state to another taking an action, \mathcal{R} describes the immediate reward obtained in a state taking an action, γ is the discount factor, μ the initial states distribution and H is the horizon (the number of sequential interactions between the agent and the environment). An agent's behavior is described by a policy π which represents the probability of choosing an action a in a state s at a particular timestep h . The performance index of an RL agent's policy is the expected discounted sum of the rewards collected during the interaction with the environment:

$$J(\pi) := \mathbb{E} \left[\sum_{t=0}^{H-1} \gamma^t \mathcal{R}(s_h, a_h) \right], \quad (1)$$

Fig. 1 The agent-environment interaction in a Markov decision process



where the expectation is taken with respect to $s_0 \sim \mu$, $s_{h+1} \sim \mathcal{P}(\cdot|s_h, a_h)$, $a_h \sim \pi(\cdot|s_h)$. When multiple agents are involved the MDP framework is extended to the Markov Game (or Stochastic game) framework. A Markov Game $\mathcal{MG} = (\mathcal{S}_1, \dots, \mathcal{S}_n, \mathcal{A}_1, \dots, \mathcal{A}_n, \mathcal{P}, \mathcal{R}_1, \dots, \mathcal{R}_n, \gamma_1, \dots, \gamma_n, \mu, H)$ is an MDP where the transition distribution over the next state depends on the actions of all the agents and each agent i has its own reward function \mathcal{R}_i . In this case the performances of all the systems are described by equilibrium concepts. The most famous equilibrium is the Nash Equilibrium (NE), where a joint policy $\pi^* = \{\pi_i^*\}_{i=1}^n$ is an NE if:

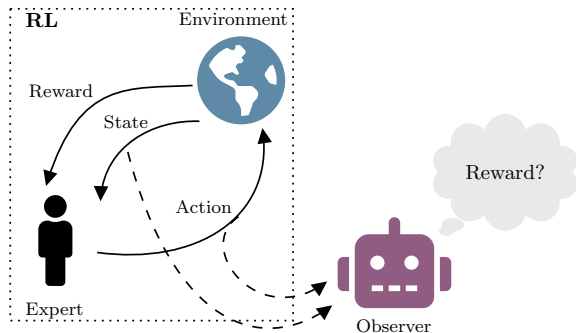
$$J_i(\pi^*) \geq J_i(\pi_i, \pi_{-i}^*) \quad \forall \pi_i \in \Pi_i, \quad i \in [n]. \tag{2}$$

The idea behind the NE is that each agent cannot improve its performance, when the other agents' policies are fixed. Also other equilibrium concepts are relevant as Stackelberg Equilibrium and Coarse Correlated Equilibrium.

2.2 Inverse Reinforcement Learning

As we wrote in the previous section, RL is a framework to learn how to perform a task described by a reward function. However, in some cases, it is extremely difficult to design a suitable reward function. On the other hand, for many tasks, we already have experts (for example, humans) who know how to accomplish the same task. The Imitation Learning (IL) [10] paradigm aims to exploit the expert information to clone the experts' behavior or formalize the expert's intentions. The IL framework is divided into two main subareas: Behavioral Cloning (BC) and Inverse Reinforcement Learning (IRL). BC, as the name says, aims to clone the expert's behavior in order to use it as a policy. IRL, instead, can recover the expert's reward function to understand its intentions and use this reward function to learn an optimal policy in any environment, even different from the one in which the expert acts. The IRL problem is composed of two agents: an expert who shows how to perform a task and an observer who watches the expert's demonstrations and learns from them the reward function (see Fig. 2). The framework used to model the problem is known

Fig. 2 The expert-observer interaction in the Inverse Reinforcement Learning framework



as Markov Decision Process without Rewards (MDP $\setminus\mathcal{R}$). An MDP $\setminus\mathcal{R}$ is defined as a tuple $\mathcal{M}\setminus\mathcal{R} = (\mathcal{S}, \mathcal{A}, \mathcal{P}, \gamma, \mu, H)$ which is the same as an MDP, but without the notion of a reward function. The expert plays a policy π^E which is (nearly) optimal for some unknown reward function \mathcal{R} , and we are given a dataset $D = \{\tau_1, \dots, \tau_m\}$ of trajectories from π^E . The IRL problem consists in recovering the reward function \mathcal{R} that the agent is optimizing.

3 Inverse Reinforcement Learning in Multi-agent Environments

The IRL [9] framework aims at recovering the reward function of an optimal agent. In the classical setting, an expert, i.e., an agent that has already learned a task, makes available a dataset of its interactions with the environment. From this, the IRL algorithm recovers the reward function that the expert is optimizing. When there are multiple agents in the environment, the IRL framework changes its objective too. For example, there could be multiple experts who show their possible different behaviors, leading to an increase in available data, but a necessity to cluster them by their intentions. Or, an agent can be interested in learning the other agents' reward functions, to use it to compute their strategy or to cooperate; however, it has to discover it without waiting for the other agent's convergence to an optimal policy. In this section we briefly describe two algorithms we designed to recover the reward function when there are multiple experts (Sect. 3.1) and when we can observe a learning agent (Sect. 3.2).

3.1 Multiple-intentions Inverse Reinforcement Learning

The first multi-agent framework that we studied is the Multiple-Intentions IRL (MI-IRL) [2] which involves an observer who has access to the demonstrations performed by multiple experts. The observer has to recover the reward functions and use them to cluster the observed agents. Solving this problem is helpful for reasons of explainability since it could be used to understand the similarity between apparently different agents. Moreover, as an immediate benefit, grouping experts who show other behaviors but share the same intent allows for enlarging the set of demonstrations available for the reward recovery process. This has a significant impact on several realistic scenarios, where the only information available is the demonstration dataset, and no further interactions with the environment are allowed.

In [15] we propose a novel batch model-free IRL algorithm, named Σ -Gradient Inverse Reinforcement Learning (Σ -GIRL), and then we extend it to the multiple-intention setting. Σ -GIRL, similarly to [11, GIRL], searches for a reward function that makes the estimated *policy gradient* [5] vanish, i.e., a reward function that is a stationary point of the expected return. However, differently from GIRL, Σ -GIRL

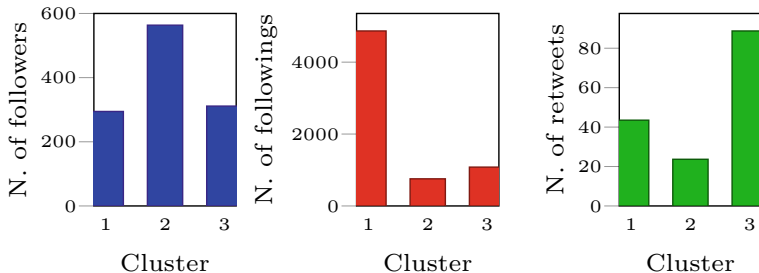


Fig. 3 Twitter clustering statistics. Average number of followers (left), followings (center) and retweets (right) for each cluster

explicitly considers the uncertainty in the gradient estimation process, looking for the reward function that maximizes the likelihood of the estimated policy gradients, under the constraint that such reward is a stationary point of the expected return. Then, we embed Σ -GIRL into the multiple-intention framework by proposing a *clustering algorithm* that, by exploiting the likelihood model of Σ -GIRL, groups the experts according to their intentions. The optimization of the multiple-intention objective is performed in an *expectation-maximization* (EM) fashion, in which the (soft) agent-cluster assignments and the reward functions are obtained through an alternating optimization process. In Fig. 3 we present the result of Σ -GIRL in recovering and clustering the intents of a group of Twitter users. The algorithm divided the 14 Twitter accounts into three clusters. The first cluster is interested in retweeting posts with high popularity. As we can observe from Fig. 3, this cluster represents a *normal* Twitter user: he/she follows many users and has a lower number of followers. In the second cluster, the agents do not retweet often, and the reason could be they have not used the social network much, as they have few retweets and follow a small number of people. In the last cluster (to which a bot, a company, and two HR managers belong) the agents tend to retweet all popular tweets.

3.2 Inverse Reinforcement Learning from a Learning Agent

The second setting that we take into account is the Inverse Reinforcement Learning from a Learner (IRL_L), proposed by [6]. The standard IRL setting assumes that an observer receives the interactions between another agent, the expert, which already knows how to perform the task, and the environment. However, in some cases, the observer can observe the learning process of this other agent, and so it can try to infer the agent’s reward function beforehand. In this setting, the observer recovers the reward function from a learning agent and not from an expert. In [6] the authors assume that the learner is learning under an entropy-regularized framework,

Table 1 Reward weights for the autonomous simulate driving scenario

	Recovered weights	Real weights
Time	0.0401	0.0017
Jerk	0.0174	0.0003
Slow	0.0001	0.0000
Crash	0.9424	0.9980

motivated by the assumption that the learner is showing a sequence of constantly improving policies. However, many Reinforcement Learning (RL) algorithms [5] do not satisfy this assumption, and human learning is also characterized by mistakes that may lead to non-monotonic learning process. In our work [14], we proposed an algorithm for this relatively new setting, IRLfL, called Learning Observing a Gradient not-Expert Learner (LOGEL), which is not affected by the violation of the constantly improving assumption. Given that many successful RL algorithms are gradient-based [5] and there is some evidence that the human learning process is similar to a gradient-based method [19], we assume that the learner is following the gradient direction of her expected discounted return. The algorithm learns the reward function that minimizes the distance between the actual policy parameters of the learner and the policy parameters that should be obtained if she were following the policy gradient using that reward function. In [14], we provide a finite-sample analysis that bounds the correctness of the recovered weights. Table 1 reports some results on a simulated autonomous driving task. It is easy to see that, the proposed algorithm succeeds in recovering the correct reward from the driver’s trajectories.

4 Online Learning in Multi-agent Reinforcement Learning

In this section we consider the problem of online learning in MARL. In this case, we are not only interested in finding the optimal policies but also in measuring the performance of our algorithm during the learning process. The performance is measured using the *regret*, i.e., comparing the performance of the agent’s actual policy with the optimal one. This problem is important in practice where we cannot learn from a simulator or using offline data, but we actually learn interacting with the system. In our work we consider the challenging problem of minimize the regret in a multi-agent game when we can control only one of the agents. We consider two different multi-agent framework. In the first one, called Configurable Markov Decision Process, the two entities are the configurator and the agent. The configurator can partially control the environment, i.e., the transition probability distribution, and the agent is the classical RL agent. The second is a Turn-based Markov Game, where the two agents involved optimize two (possibly different) reward functions.

4.1 *Online Learning in Non-cooperative Configurable Markov Decision Processes*

In this section, we briefly expose the result published in [16] where we solved an online learning problem in the Configurable Markov Decision process framework. A Configurable Markov Decision process involves two entities, the configurator, and the agent. This framework is motivated by real-world scenarios in which an external supervisor (configurator) can *partially* modify the environment. Previous to [16], the Configurable Markov Decision Processes [7, Conf-MDPs] consists of simultaneously optimizing a set of environmental parameters and the agent’s policy to reach the maximum expected return. In many scenarios, however, the configurator does not know the agent’s reward and its intention differs from that of the agent. In [16] the authors introduce the Non-Cooperative Configurable Markov Decision Processes (NConf-MDP), a new framework that handles the possibility of having different reward functions for the agent and the configurator. An NConf-MDP allows modeling a more extensive set of scenarios, including all the cases in which the agent and configurator display a non-cooperative behavior, modeled by the individual reward functions. Obviously, this setting cannot be solved with a straightforward application of the algorithms designed for Conf-MDPs that focus on the case in which both entities share the same interests. In this novel setting, the authors consider an online learning problem, where the configurator has to select a configuration, within a finite set of possible configurations, in order to maximize its own return. This framework can be seen as a *leader-follower* game, in which the *follower* (the agent) is self-ish and optimizes its own reward function, and the *leader* (the configurator) has to decide the best configuration w.r.t. the best response of the agent. Clearly, to adapt its decisions, the configurator has to receive some form of feedback related to the agent’s behavior. The authors analyze two settings based on whether the configurator observes only the agent’s actions or also a noisy version of the agent’s reward function. For the two settings, they propose algorithms based on the Optimism in the Face of Uncertainty [1] principle, which yield bounded regret.

4.2 *Online Learning in General-Sum Stochastic Games*

In this setting, we consider the learning problem in two-player general-sum Markov Games when we can control only one agent which is playing against an arbitrary opponent to minimize the regret. Previous works only consider the zero-sum setting, in which the two agents are completely adversarial. However, in some cases, the two agents may have different reward functions without having conflicting objectives. This class of games is called general-sum Markov Games. In our work [18], we show that the regret minimization problem is significantly harder than in standard Markov Decision Processes and zero-sum Markov Games. We derive a lower bound on the expected regret of any “good” learning strategy. The lower

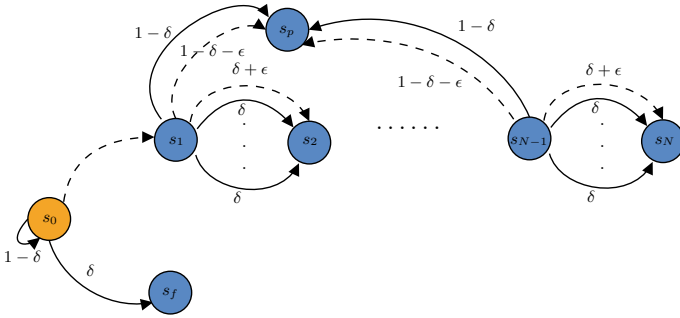


Fig. 4 The Turn-based Stochastic Game for the lower bound. The states belonging to the non-controllable agent are in orange, the ones belonging to the controllable one in blue. The dashed lines corresponds to the transition probabilities taking the optimal action a^* , and the others taking any other action

bound shows the constant dependencies with the number of deterministic policies that the agent can play, which is not present in zero-sum Markov Games and Markov Decision Processes. To have an intuition, we can look at Fig. 4. The controllable agent controls the blue states, while the non-controllable one controls the orange ones. The idea behind the proof is the following: if the controllable agent does not play the optimal policy, the uncontrollable one will always choose the action which leads to the down path, preventing the agent from exploring the environment. After this result, we propose a novel optimistic algorithm that nearly matches the proposed lower bound.

5 Conclusions

The work [13] addressed different problems in MARL, going from IRL to online learning and policy optimization. The MARL framework provides a useful way to model multi-agent decision-making problems such as smart grids, autonomous driving, financial markets, drone delivery, and robotic control problems. The main purpose of the work is to show the flexibility of the MARL framework to model real-world problems compared to the single-agent one, and, on the other hand, how it leads to novel challenges: the learning objective changes, the environment is no more stationary and standard algorithms from single-agent RL cannot be applied. We consider three sub-problems, inspired by the single-agent literature: IRL in Multi-Agent Systems, Online Learning in MARL, and Optimization in MARL. Although we provide novel algorithms for various problems which arise in these contexts, many problems remain unsolved, and also new open questions, practical and theoretical, come out.

Inverse Reinforcement Learning in Multi-Agent systems From the MI-IRL setting, the Σ -GIRL algorithm presents some limitations: we need to specify the number

of existing clusters as hyperparameter, and the algorithm can converge to stationary points which are not local maximizers. In the IRL from a Learner setting, the main challenge arises from the assumption that we know when the agent changes its policy and so a natural extension could be the automatic detection of the policy change. Moreover, if we want to play simultaneously and/or the other agents are not rational, it is necessary to build different algorithms to recover the reward function.

Online Learning in Markov Games There are many future directions in the online learning problem in Stochastic Games. From our work in Non-cooperative Configurable MDPs, a direct follow-up will be extending the approach to deal with continuous state and action spaces using, for example, function approximation. From our findings in general-sum Markov Games, it is clear that the algorithm design and the resulting performance guarantees heavily depend on any knowledge about the opponents, either known a priori or obtainable during the learning process. An interesting future direction is to assume to have the possibility to observe other agents' interactions with the environment or having some previous knowledge about the other agents (as having access to a finite set of opponents or considering a larger set of opponents' classes with some regularity assumptions).

Acknowledgements The author acknowledges Marcello Restelli for the support during this work.

References

1. Auer, P., Cesa-Bianchi, N., Fischer, P.: Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.* **47**(2–3), 235–256 (2002)
2. Babes, M., Marivate, V.N., Subramanian, K., Littman, M.L.: Apprenticeship learning about multiple intentions. In: *ICML* (2011)
3. Bellman, R.: A markovian decision process. *J. Math. Mech.* 679–684 (1957)
4. Buşoniu, L., Babuška, R., De Schutter, B.: Multi-agent reinforcement learning: an overview. *Innovat. Multi-agent Syst. Appl.* **1**, 183–221 (2010)
5. Deisenroth, M.P., Neumann, G., Peters, J., et al.: A survey on policy search for robotics. *Found. Trends Robot.* **2**(1–2), 388–403 (2013)
6. Jacq, A., Geist, M., Paiva, A., Pietquin, O.: Learning from a learner. In: *International Conference on Machine Learning*, pp. 2990–2999. PMLR (2019)
7. Metelli, A.M., Mutti, M., Restelli, M.: Configurable Markov decision processes. In: *International Conference on Machine Learning*, pp. 3491–3500. PMLR (2018)
8. Morgenstern, O., Von Neumann, J.: *Theory of Games and Economic Behavior*. Princeton University Press (1953)
9. Ng, A.Y., Russell, S.J., et al.: Algorithms for inverse reinforcement learning. In: *Icml*. vol. 1, p. 2 (2000)
10. Osa, T., Pajarinen, J., Neumann, G., Bagnell, J.A., Abbeel, P., Peters, J.: An algorithmic perspective on imitation learning. *CoRR* (2018). [arXiv:abs/1811.06711](https://arxiv.org/abs/1811.06711)
11. Pirotta, M., Restelli, M.: Inverse reinforcement learning through policy gradient minimization. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30 (2016)
12. Puterman, M.L.: *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley (2014)

13. Ramponi, G.: Challenges and opportunities in multi-agent reinforcement learning (2021)
14. Ramponi, G., Drappo, G., Restelli, M.: Inverse reinforcement learning from a gradient-based learner **33**, 2458–2468 (2020). <https://proceedings.neurips.cc/paper/2020/file/19aa6c6fb4ba9fcf39e893ff1fd5b5bd-Paper.pdf>
15. Ramponi, G., Likmeta, A., Metelli, A.M., Tirinzoni, A., Restelli, M.: Truly batch model-free inverse reinforcement learning about multiple intentions. In: International Conference on Artificial Intelligence and Statistics, pp. 2359–2369. PMLR (2020)
16. Ramponi, G., Metelli, A.M., Concetti, A., Restelli, M.: Online learning in non-cooperative configurable Markov decision process. In: AAAI Workshop on Reinforcement Learning in Games (2021)
17. Ramponi, G., Restelli, M.: Newton optimization on helmholtz decomposition for continuous games. In: Thirty-Fifth AAAI Conference on Artificial Intelligence (2021)
18. Ramponi, G., Restelli, M.: Learning in markov games: can we exploit a general-sum opponent? In: The 38th Conference on Uncertainty in Artificial Intelligence (2022)
19. Shteingart, H., Loewenstein, Y.: Reinforcement learning and human behavior. *Curr. Opin. Neurobiol.* **25**, 93–98 (2014)
20. Sutton, R.S., Barto, A.G., et al.: Introduction to Reinforcement Learning, vol. 135. MIT Press Cambridge (1998)
21. Zhang, K., Yang, Z., Başar, T.: Multi-agent reinforcement learning: a selective overview of theories and algorithms (2019). arXiv preprint [arXiv:1911.10635](https://arxiv.org/abs/1911.10635)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Post-cloud Computing: Addressing Resource Management in the Resource Continuum



Michele Zanella

Abstract The exponential growth of interconnected IoT devices, highlights the infrastructure limitations of Cloud-based computing approaches. In this context, novel solutions (i.e., Fog and Edge computing) aim to exploit a continuum resource space composed of nearby and mobile devices as a single heterogeneous and distributed system to move part of the computation closer to data sources. In this regard, the heterogeneous nature of these devices (performance, features, capabilities...) requires proper programming models and run-time management layers. This chapter proposes an overview of recent modeling premises and quantitative results in a resource management perspective through the BarMan framework, which combines a task-based programming model, a run-time resource manager, and the BeeR task distribution software to deploy use-case applications-modules across the boards of a real Fog cluster.

Keywords Fog computing · Resource management · Task-based programming model

1 The Resource Continuum

The advent of the Internet of Things (IoT) era enables the possibility to interconnect together millions of devices, which are able to collect an enormous quantity of data processed somewhere in the Cloud for different purposes [9]. According to Gartner [6], in 2016 there were 6,4 billion edge devices around the world, but recently, IDC [8] predicted up to 60 billion connected entities by 2025, generating about 80 ZB of raw data. This enormous quantity of data highlights the limits of the Cloud approach: network saturation, the unsustainability of larger datacenters in terms of size and energy consumption, unreliability, and latency of internet connections which are not suitable for real-time or emergency use-cases. In this context, new post-Cloud trends

M. Zanella (✉)

Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Via Ponzio 34/5, 20133 Milano, Italy
e-mail: michele.zanella@polimi.it

© The Author(s) 2023

C. G. Riva (ed.), *Special Topics in Information Technology*,

PoliMI SpringerBriefs, https://doi.org/10.1007/978-3-031-15374-7_9

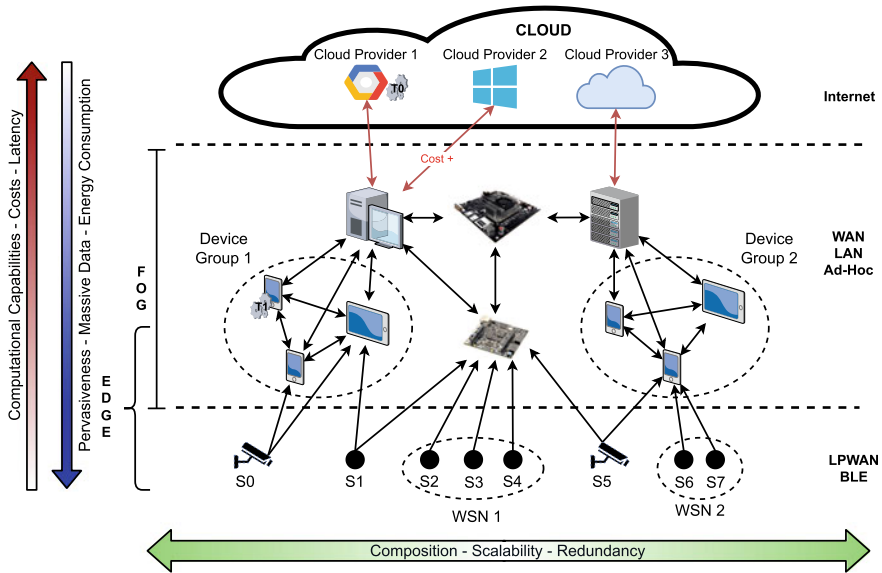


Fig. 1 The resource continuum reference architecture

aim to shift part of the processing near to the data, exploiting the increased capabilities of edge devices, which unfortunately are still limited by battery energy budget and thermal issues. The most promising approach is called Fog computing [4, 11] and it includes, in the latest definition, a layer between Cloud and edge that provides computational and storage services, including also aspects related to networking and management. This approach follows a *resource continuum* concept since the system connects devices from the Cloud to the edge seamlessly, as shown by Fig. 1.

1.1 A Comprehensive Architecture

In order to understand the underlying problems, the first step is to define and model the infrastructure. In this regard, the infrastructure can be organized following a multi-tier approach [11], as highlighted in Fig. 1: the *computing nodes* expose different capabilities (i.e., processing, storage, and network connectivity) between the various level. Secondly, the resources are modeled following a bi-dimensional space. The vertical dimension spans different paradigms: in this way, moving from lower to higher levels, it is possible to gain performance and energy budget but also increases costs and latency; on the contrary, from Cloud to edge leads to more pervasiveness and low-power consumption. Instead, through the horizontal dimension, it is possible to perform scaling and balancing among sibling nodes, or implementing fault-tolerance policies by (a) switching between different providers at the Cloud

level, (b) composing dynamic worker groups exploiting Fog devices, (c) arranging sensors and Edge devices together.

As a consequence, such architecture is characterized by an high level of *computing resource heterogeneity*, which can be divided into three categories:

- *inter-level*: very different capabilities depending on the level;
- *intra-level*: different resources at the same level;
- *intra-node*: heterogeneous resources available on the same device/node (e.g., general-purpose single or multi-core CPUs, many-core GPU, HW accelerators).

Dealing with such a dynamic, modular and heterogeneous system requires to address some research challenges, which are the key objectives of our work: (a) the possibility to *control and manage* the entire system; (b) the development of a *unified programming model* and a *transparent task distribution* to fully exploit the connected devices; (c) the need for a *run-time resource manager (RM)* with *proper task mapping and resource allocation strategies*; (d) the availability of *real-world use-case applications* as well as *accessible and physical hardware test-bed* to perform evaluations.

2 The BarMan Framework

The *BarMan* framework in Fig. 2 integrates different frameworks in a single suite and extends their capabilities to enable a run-time managed execution of multi-tasking applications on a heterogeneous and distributed (embedded) computing system. In the following paragraphs, we highlight the main features of each module: in particular, it comprises The BarbequeRTRM resource manager for embedded, mobile, and HPC systems, the *libmango* programming model for modular applications, and the *BeeR* framework in order to distribute them among the devices. Eventually, as shown in Fig. 2, these modules can interface with various platforms and use-cases.

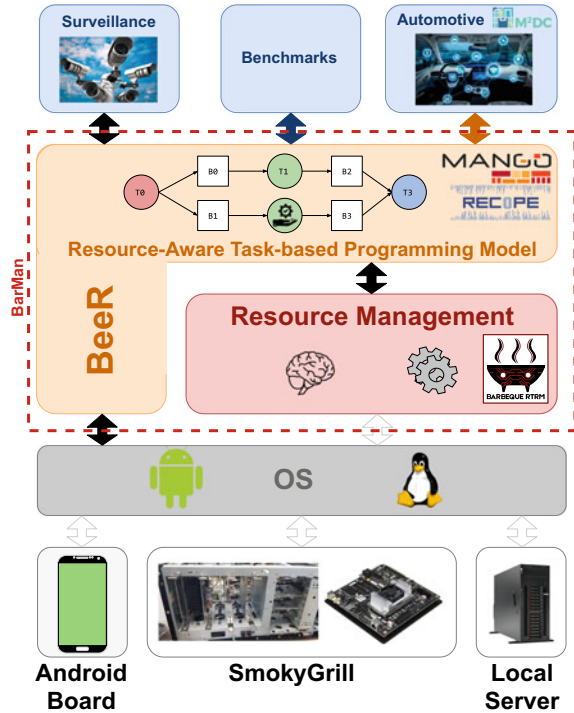
2.1 The BarbequeRTRM Resource Manager

The *Barbeque Run-Time Resource Manager* (BarbequeRTRM) [3]¹ is a modular, open-source, and extensible run-time RM able to manage the allocation of computing resources to concurrent applications while taking into account both applications' QoS requirements and dynamic resource availability. Supporting several different types of platforms (e.g., embedded, HPC, mobile...), it has been extended to support multi-device cooperation.

The main feature is the so-called *Adaptive Execution Model*, an Android-style execution flow in which each application and resource manager interact. This allows the applications to be controlled and reconfigured by the manager in order to respect

¹ <https://bitbucket.org/boisp>.

Fig. 2 Overview of the BarMan framework, applications and integrated devices



performance/power constraints. The proper management is performed by various allocation policies that can be easily plugged to deal with specific optimization metrics (e.g., latencies, bandwidth, power consumption...).

2.2 *libmango*: A Task-Based Programming Model

The `libmango` Programming Library [1]² has been initially developed inside the MANGO European Project [5] to develop multi-tasking applications on HPC platforms. This way, applications can be decomposed into small chunks of code (called tasks or kernels) that are offloaded and run on different nodes. Our solution outperforms state-of-the-art solutions [1], and differently from libraries like OpenCL, allows the developer only to provide a task-graph of the application (i.e., a Directed Acyclic Graph describing tasks, memory buffers, and their inter-dependencies) without implementing the task-to-device mapping logic, which is delegated to the management software for the optimization at run-time.

² https://bitbucket.org/mango_developers/mangolib.

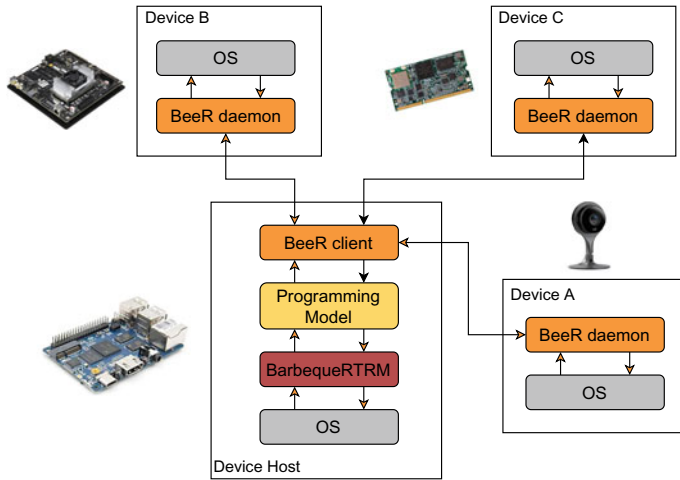


Fig. 3 BarMan framework’s modules deployment

2.3 Transparent Tasks Distribution: BeeR

The programming library has been extended to support dynamic content offloading by the open-source BeeR. It is composed of two components: a *client library*, which includes API to extend the `libmango`, and a *daemon server* deployed on the devices and is in charge of handling the execution of incoming tasks, as well as providing minimal support to retrieve device status. Figure 3 shows the interaction between the BeeR and the other BarMan modules. In the typical flow, when the user requests an application, the BarbequeRTRM provides an appropriate mapping plan based on the selected optimization metrics and the current state of the entire system (load, energy budget, availability...). Then, the BeeR client enables the communication between the remote daemon instances and defines the set of assigned resources for each task. On the remote side, the BeeR daemon performs the resource and buffer reservation and manages the task execution.³

2.4 Use-Case Evaluation Scenarios

One of the main demands in the research community is the need to have real-world use-case applications and an accessible cluster to perform experiments. Thus, in order to prove and evaluate the functionality of the entire framework, we developed two application scenarios tested on a real self-build Fog cluster: the *SmokyGrill*. The

³ <https://bitbucket.org/dark-corner/beer-framework>.

latter is composed of different interconnected embedded boards (i.e., Jetson TX2, Odroid H2, Freescale) that aim to replicate a typical Fog setup with both low-end and high-end devices.

Video Surveillance The first application is related to the video surveillance scenario, and its goal is to classify and track moving objects in a specific area. It is inspired by the simulated use-case by Gupta et al. [7], then implemented with some modifications and optimization using the `libmango` API. Given the application, we develop the *Latency Versus Accuracy* (LAVA) allocation policy to optimize three metrics (detection accuracy, detection latency, and tracking latency) by minimizing the execution and communication (wireless or wired) latencies of its tasks. Finally, the evaluation consists of (a) a kernel execution characterization; (b) network and framework overheads measurement; (c) policy execution analysis with a set of scenarios aiming to cover different devices' availability and configurations. Among the other results that can be found in [12], the most promising outcome shows the benefit of performing a distributed execution on proper available devices instead of considering the original monolithic version by improving the execution time up to 66% depending on the situation scenario considered (Fig. 4).

Large-scale Emergency system In this second use-case, we apply our architectural model to a large-scale emergency scenario [13]. The current emergency systems are outdated and can not satisfy the time-sensitive need for trustworthy emergency services when natural disasters happen. To overcome these limitations, we propose a semantic-based trustworthy information-centric Fog system that provides emergency services, and it is based on three components: (a) *edge devices*: to collect and pre-process the information at the Edge level; (b) the *information-centric (IC) Fog network*: to exploit the Fog paradigm by computing semantic information for secure emergency analysis and management; (c) a *Cloud emergency center (CEC)*.

The objectives of this application are: (1) proposing a novel emergency communication network to aggregate and analyze emergencies with semantic information

Fig. 4 Comparison of execution time

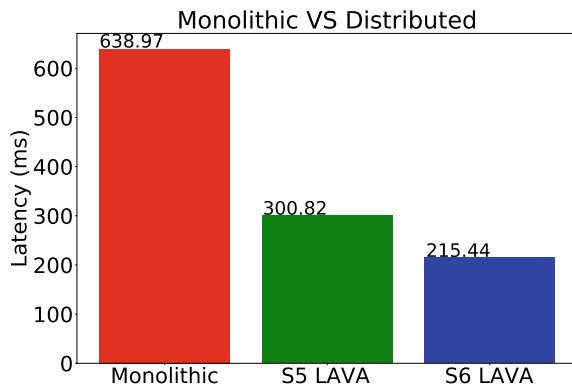
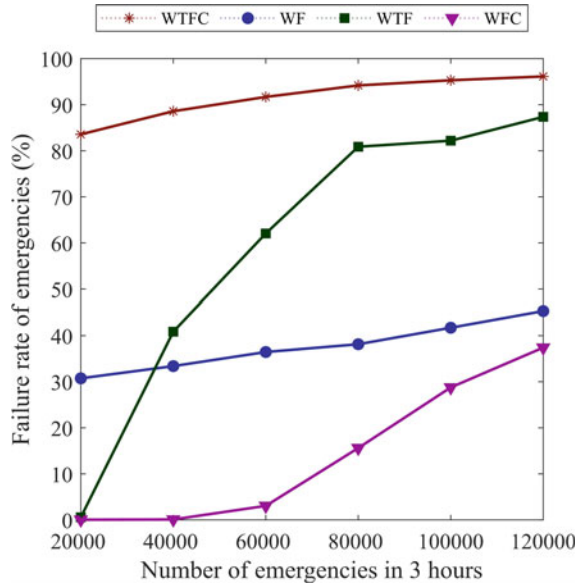


Fig. 5 Failure rate versus number of emergencies



in the network layer; (2) designing a semantic-based trustworthy routing scheme, able to filter untrusted content by improving the quality of emergency services; (3) evaluating the system through a real testbed and a modified Cloudsim simulation software for large-scale scenarios in order to provide valuable data to deploy the system on existing IIoVT devices.

After retrieving data from the testbed, we filled the simulation software to analyze how the system handles a 3-hour earthquake disaster varying the number of emergencies and the layers’ availability. In this regard, we evaluate four scenarios: (a) without Fog networks and Cloud (*WTFC*), (b) with only Fog networks (*WF*), (c) with working Fog networks and Cloud (*WFC*); (d) a traditional system without Fog networks (*WTF*). As shown in Fig. 5, using the resource continuity (i.e., *WFC* scenario), the failure rate decreases between 87% and 37% w.r.t. other system’s configurations.

3 Exploiting Mobile Devices

Integrating mobile devices into the resource continuum opens wide opportunities in terms of scalability and pervasiveness. However, since they are battery-supplied, an efficient and fine-grained power management is crucial to meet users’ satisfaction and maintain devices’ availability.

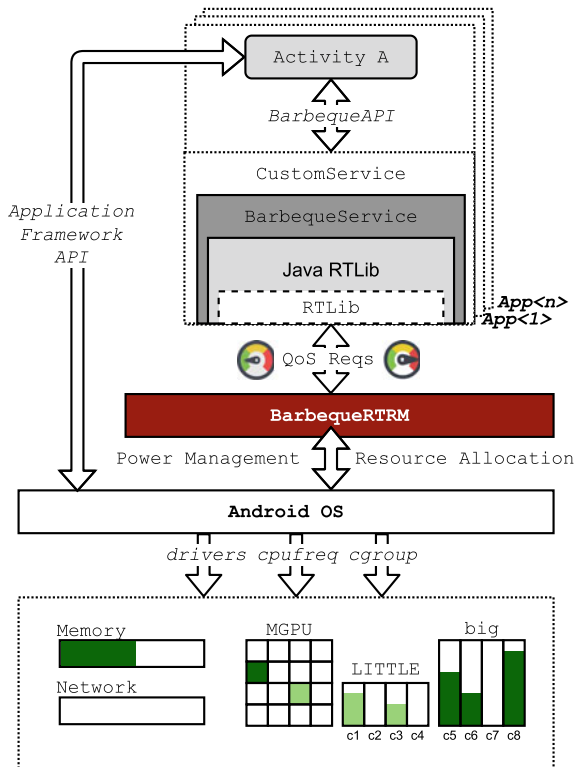
Moreover, nearby devices can be exploited to decrease power consumption by delegating part of the computation, thus leading to privacy and over-usage issues [10]. In this regard, an integrated resource manager can assure isolation and energy

budgeting for the different running applications or incoming tasks. Thus, we develop an Android version of the BarbequeRTRM, which supports mobile applications in terms of QoS and resource demands.

3.1 Run-Time Adaptive Application Execution

Figure 6 shows the API design, which bonds the native layer of Android OS where the BarbequeRTRM runs and the Java application framework layer through a custom service-based wrapper. In this way, applications can be integrated with the BarbequeRTRMAEM, enabling the possibility to (1) set the application performance/power saving goal (i.e., the throughput) or (2) require an explicit constraint on the resource allocation. Given this information, the RM can enforce power/resource management through the underlying Linux frameworks (e.g., *cpufreq*, *cgroup*) to set CPU operating points, and reserve CPU time quota or cores basing on the specific optimization policy plugged. In this regard, one of our experimental evaluations are devoted to providing a proof-of-concept of the prototype. We profile the execution of a mobile benchmark suite gathering performance (fps) and power consumption measures in

Fig. 6 Overview of the Android BarbequeRTRM software stack



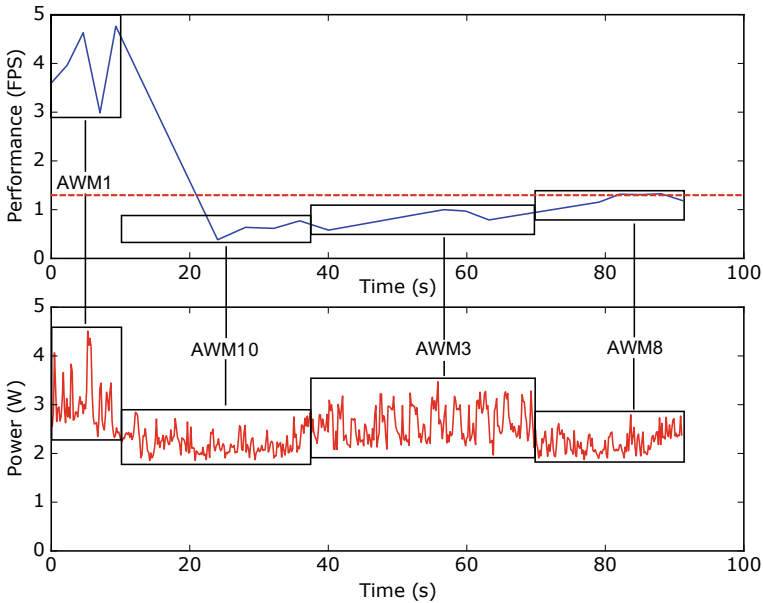


Fig. 7 Benchmark’s performance and system’s power consumption

order to define a set of pre-defined resource assignment configurations called *Application Working Modes* (AWM); each AWM contains information about the amount of CPU, the frequency setting, the average system power consumption, and the performance estimation. Then, a simple policy chooses different AWMs in a constrained range based on the maximum throughput required by the user. Figure 7 shows an evaluation scenario with the *Image Effect* benchmark, where the desired throughput (dotted-red line) is reached and the power consumption reduced by a subsequent changing of AWMs made by the policy.

4 Conclusions and Future Directions

Besides the different results highlighted in this brief research summary, there are still open issues that need to be addressed. In particular, regarding the task distribution problem, the main challenges are related to security and privacy, which require actuating hard isolation and lightweight cryptography techniques.

Moreover, from a resource management standpoint (1) machine learning technique can be used to predict incoming applications and their performance goal required by specific application or process, thus improving the current mobile policy; (2) the LAVA allocation policy will also be extended also with energy-related aspects.

Finally, other use-case applications can be integrated, like the automotive one developed during the M2DC EU project [2]; as well as deploying in real large-scale experiments the two presented use-cases to gather useful measurements and insights to improve simulation models and management strategies.

References

1. Agosta, G., Fornaciari, W., Massari, G., Pupykina, A., Reghenzani, F., Zanella, M.: Managing heterogeneous resources in hpc systems. In: Proceedings of PARMA-DITAM 2018. PARMA-DITAM'18, Association for Computing Machinery, pp. 7–12. New York, NY, USA (2018). <https://doi.org/10.1145/3183767.3183769>
2. Agosta, G., Brandolese, C., Fornaciari, W., Mainardi, N., Pelosi, G., Reghenzani, F., Zanella, M., Des Courchamps, G., Ducrot, V., Juilly, K., et al.: Accelerating automotive analytics: the m2dc appliance approach. In: International Conference on Embedded Computer Systems, pp. 455–469. Springer (2019)
3. Bellasi, P., Massari, G., Fornaciari, W.: Effective runtime resource management using linux control groups with the barbequeRTRM framework. *ACM Trans. Embed. Comput. Syst.* **14**(2), 39:1–39:17 (2015)
4. Bonomi, F., Milito, R., Zhu, J., Addepalli, S.: Fog computing and its role in the internet of things. In: Proceedings of the First Edition of the MCC Workshop on Mobile Cloud Computing, MCC'12, pp. 13–16. ACM, New York, NY, USA (2012)
5. Flich, J., et al.: Exploring manycore architectures for next-generation hpc systems through the mango approach. *Microprocess. Microsyst.* **61** (2018)
6. Gartner: Top 10 strategic predictions for 2015 and beyond: digital business is driving big change (2015). <https://www.gartner.com/newsroom/id/3165317>. Accessed 30 Jun 2022
7. Gupta, H., Vahid Dastjerdi, A., Ghosh, S.K., Buyya, R.: ifogsim: a toolkit for modeling and simulation of resource management techniques in the internet of things, edge and fog computing environments. *Softw. Pract. Exp.* **47**(9), 1275–1296 (2017)
8. (IDC), I.D.C.: Iot growth demands rethink of long-term storage strategies. <https://www.idc.com/getdoc.jsp?containerId=prAP46737220>. Accessed 30 Jun 2022
9. Massari, G., Zanella, M., Fornaciari, W.: Towards distributed mobile computing. In: 2016 Mobile System Technologies Workshop (MST), pp. 29–35 (2016). <https://doi.org/10.1109/MST.2016.13>
10. Zanella, M., Massari, G., Fornaciari, W.: Enabling run-time managed distributed mobile computing. In: Proceedings of PARMA-DITAM 2018. PARMA-DITAM'18, pp. 39–44. ACM, New York, NY, USA (2018). <https://doi.org/10.1145/3183767.3183778>, <http://doi.acm.org/10.1145/3183767.3183778>
11. Zanella, M., Massari, G., Galimberti, A., Fornaciari, W.: Back to the future: resource management in post-cloud solutions. In: Proceedings of the Workshop on INTElligent Embedded Systems Architectures and Applications. INTESA'18, pp. 33–38 (2018). <https://doi.org/10.1145/3285017.3285028>
12. Zanella, M., Sciamanna, F., Fornaciari, W.: Barman: a run-time management framework in the resource continuum. *Sustain. Comput. Inf. Syst.* **35**, 100663 (2022)
13. Zhang, Q., Wu, J., Zanella, M., Yang, W., Bashir, A.K., Fornaciari, W.: Sema-iiiov: Eemergent semantic-based trustworthy information-centric fog system and testbed for intelligent internet of vehicles. *IEEE Consum. Electron. Mag.* (2021)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Telecommunications

Intelligent Networked Music Performance Experiences



Luca Comanducci 

Abstract A Networked Music Performance (NMP) is defined as what happens when geographically displaced musicians interact together while connected via network. The first NMP experiments begun in the 1970s, however, only recently the development of network communication technologies has created the necessary infrastructure needed to successfully create an NMP. Moreover, the widespread adoption of network-based interactions during the COVID-19 pandemic has generated a renewed interest towards distant music-based interaction. In this chapter we present the Intelligent networked Music PERforMANce experiENCEs (IMPERMANENCE) as a comprehensive NMP framework that aims at creating a compelling performance experience for the musicians. In order to do this, we first develop the neTworkEd Music PERfoRmANCe rEsearch (TEMPERANCE) framework in order to understand which are the main needs of the participants in a NMP. Informed by these results we then develop IMPERMANENCE accordingly.

Keywords Networked music performance · Deep learning · Spatial audio · Microphone array

1 Introduction

During the last decades, the increasing speed of network communication technologies has grown steadily, creating a fertile environment for the development and consumption of online media-content. Moreover, the recent COVID-19 pandemic made online communications, such as live-streaming or online meetings, a part of

L. Comanducci (✉)
Politecnico di Milano, Milan, Italy

Dipartimento di Elettronica, Informazione e Bioingegneria, Via Ponzio 34/5, 20133 Milano, Italy
e-mail: luca.comanducci@polimi.it

© The Author(s) 2023
C. G. Riva (ed.), *Special Topics in Information Technology*,
PoliMI SpringerBriefs, https://doi.org/10.1007/978-3-031-15374-7_10

daily lives for a considerable amount of people. In this context, also the online fruition of music content has grown, both for what concerns listening, either streaming or remote-concert attendance, or performing, e.g. rehearsals or remote live-performances. We define a performance where musicians are connected via network as a Networked Music Performance (NMP). The research in NMPs has begun as early as the 1970s [7], but only recently it has grown enough to be actually a viable possibility for geographically displaced musicians. In order to create a NMP experience that feels compelling to the musicians, several aspects need to be considered, which can be broadly separated into temporal factors, related to the synchronicity between the musicians, and spatial factors, related to the audiovisual perception. Several NMP-dedicated software solutions have been created over the years [25, 31], such as LOLA [24], UltraGrid [27] or JackTrip [9]. Basically, all these software frameworks are low-latency audio/video streaming protocols. While the minimization of the latency is extremely important in creating a satisfying NMP, it does not suffice in creating an experience that feels compelling from the point of view of the musicians.

In this chapter, we present a general-purpose framework for NMP denoted Intelligent networked Music PERforMANce experiENCEs (IMPERMANENCE). In modeling this framework, we do not aim at creating a NMP that hopelessly tries to mimic the hypothetical real environment, instead we try to model the NMP as a medium on its own. In order to do this we build the components of IMPERMANENCE by first analyzing which are the needs of the musicians, by developing a research framework, denoted neTworkEd Music PERfoRmANCe rEsearch (TEMPERANCE). Following the guidelines established through TEMPERANCE, we organize a series of experiments aimed at understanding the impact of temporal and spatial factors on the perceived Quality of Experience (QoE) of the musicians. Informed by the results obtained through these experiments we develop IMPERMANENCE accordingly. We take advantage of both signal processing, well suited to treat audio signals, and deep learning techniques. The latter allow us to overcome limitations posed by classical signal processing techniques, related to the number of sensors needed and to the decrease in the performance quality in adverse scenarios, i.e. where reverberation and noise are present.

The research described in this chapter summarizes the results obtained in my Ph.D. thesis [12] and the scientific publications in [3, 4, 11, 13–22, 30], obtained during my Ph.D. The rest of this chapter is organized as follows. In Sect. 2 we present the preliminary analysis aimed at understanding which are the main aspects of a NMP that need to be tackled by the IMPERMANENCE framework. Specifically, we first present the research framework TEMPERANCE and two experimental studies performed with musicians. In Sect. 3 we present the general structure of the IMPERMANENCE framework, while in Sects. 4 and 5, we present how we plan to treat the temporal and spatial factors, respectively. In Sect. 6 we present preliminary results related to a technique that could provide audio compression. Finally, in Sect. 7 we draw some conclusions.

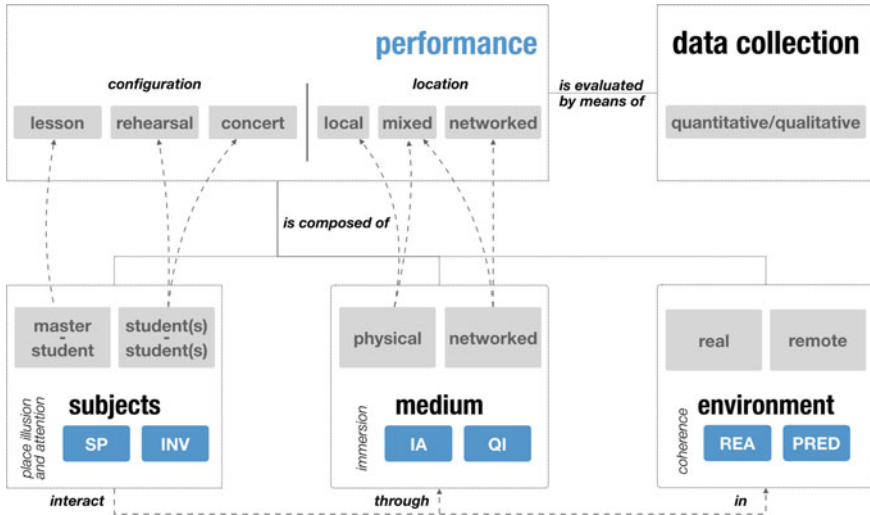


Fig. 1 The TEMPERANCE framework for NMP research [21]

2 The neTworkEd Music PERfoRmANCe rEsearch (TEMPERANCE) Framework

The TEMPERANCE framework [21] is devoted to the design and realization of NMP experiments and scenarios. While the framework is general enough to be applicable to a wide variety of music genres and types of remote performances, it was developed focusing on the remote teaching of chamber music, during the INTERactive environment for MUSIC learning and practising (INTERMUSIC)¹ project.

The general structure of the TEMPERANCE framework is depicted in Fig. 1. A **performance** may be defined as what occurs when two or more **subjects** interact together through a **medium** in a certain **environment**. The performance is the entity that stands at the highest conceptual level and can assume two main configurations, namely a *performed music composition*, or a *taught lesson*. It can either be a *rehearsal* or a *concert*, both involving the inclusion of at least two musicians and possibly additional people. Depending on the location, when the subjects are in the same room we have a *local performance*, when they are geographically displaced we have a *networked performance*, that is the focus of this chapter; finally in the case where more than one subject is located in one of the remote rooms we talk of a *mixed performance*. The *medium* is distinguished as being either *physical*, i.e. air propagation, or *networked*, i.e. internet connection. Given a musician, we

¹ <http://intermusicproject.eu>.

define the environment where she/he is playing as *real environment*, while the representation of the remote one, where the other musician/s are playing as the *remote environment*. We define a series of presence-based [23] constructs characterizing the NMP experience, general enough to be able to be applied in scenarios different from the chamber music ones. For a more thorough explanation, we refer the interested reader to [21]. In order to analyze/evaluate the performance it necessary to conduct a data collection step, which then allows to evaluate in terms of both objective and subjective measures. The former serve to the purpose of numerically evaluating the performance [31], while the latter, mainly consist of questionnaires investigating the proposed presence constructs and focus on the general QoE of the NMP.

We now present two studies, both performed at the Conservatorio di Musica “Giuseppe Verdi” di Milano using conservatory students aged from 14 to 29 years. The purpose of the studies it to both verify the hypotheses behind the creation of the TEMPERANCE framework and to understand which are the main aspects that need to be treated in order to create a satisfying NMP through IMPERMANENCE. More specifically, the two studies aim at separately exploring the impact of latency and of the audiovisual setup on the performance. In both cases the musicians were placed into two separated rooms connected via LOLA [24]. A server was placed between the two rooms in order to act as a network emulator simulating real-world-like conditions.

2.1 Study I: Latency Perception in NMP

The first study [18, 20] concerns the impact that the network latency has on the performance of the musicians. Ten volunteers participated, divided into five duets. Each of them performed under 6 different latency conditions, ranging from 28 ms to 134 ms (2-way). The order under which the conditions were performed was random for each couple. The stimuli (i.e. the music) were based on Béla Bartók *Mikrokosmos* [1] piano pieces, based on exercises exploring *rhythm-melody-expression* relationships [21]. The performance of the musicians were analyzed both in terms of objective measurements such as tempo trend, i.e. how the musicians vary the tempo during the performance, and asymmetry, i.e. the disalignment between the performance of the musicians, and in terms of subjective measurements, based on a questionnaire exploring the presence-based constructs. The results show that no single general trend is understandable from the way that musicians cope with different levels of latency, in the sense that each musician behaves differently. It is interesting to analyze how, even if confronted with high delay times, musicians are able to somehow adapt themselves or at least to adopt different type of strategies. These findings motivated us to avoid a latency-minimization approach, already present in the aforementioned NMP software solutions, and instead to focus on providing tools that are able to help the musicians in coping with the latency, that is, the adaptive metronome approach presented in the IMPERMANENCE framework.

2.2 Study II: Audiovisual Immersion in NMP

The second study concerns the audio-visual immersion in NMP environments. We considered 8 musicians, corresponding to 4 duets, playing Béla Bartók's "44 Duos for 2 Violins" [2], that are pedagogical pieces addressed to train motor responses to aural problems and rhythmic and structural features. During this study, the duets had to play using two different setups corresponding to different degrees of immersion. The first setup consisted of a 24-inch screen, positioned in front of the performer at a distance of approximately 1.76 m from the back of the chair, while sound was rendered through a pair of loudspeakers positioned at the sides of the monitor. The second setup was created in order to be more resemblant of a face-to-face situation. A 50-inch screen was positioned at the side in order to project the remote performer as if it was laterally positioned with respect to the local musicians. Sound was rendered through a pair of open headphones (Sennheiser HD-650), augmented with a custom-made head-tracker [22] whose data were fed to a Pure Data (PD) patch that provided binaural audio in real-time through a series of Head Related Transfer Functions (HRTFs). The impact of the two setups on the musicians' performances were analyzed through a questionnaire [22]. The obtained results provided important design implications. The simple frontal screen setup was perceived as plausible by the musicians, enabling us to avoid to specifically treat the visual perception in the IMPERMANENCE framework. On the other hand, the auditory perception was felt as more puzzling. While the 3D audio was perceived as useful by the musicians, the presence of the headphones was sometimes felt as obtrusive, motivating us to pursue research in 3D audio rendering based on loudspeaker setups.

3 The Intelligent Networked Music PERforMANce experiENCes (IMPERMANENCE) Framework

By exploiting the information extracted from the two experiments presented in Sects. 2.1 and 2.2 through the TEMPERANCE framework, we are able to propose a unified framework for NMPs whose characteristics are focused on satisfying the needs of the musicians, by tackling both temporal and spatial factors. In Fig. 2 we present the conceptual map summarizing IMPERMANENCE.

In order to mitigate the impact that latency has on the musicians, we propose to use a technique based on the adoption of an adaptive metronome [3, 4]. A metronome is a device that produces a tick at regular intervals, commonly used by musicians in order to keep the correct tempo while practicing. The metronome becomes adaptive, by combining it with a beat tracker, that is a device able to retrieve the tempo of a generic musical audio signal, in order to modify its tempo depending on the musicians' performance. Differently from most NMP frameworks proposed in the literature, we do not try to minimize the latency, but instead help the musicians in coping with it. This is due to the findings presented in Sect. 2.1 that demonstrate that the musicians

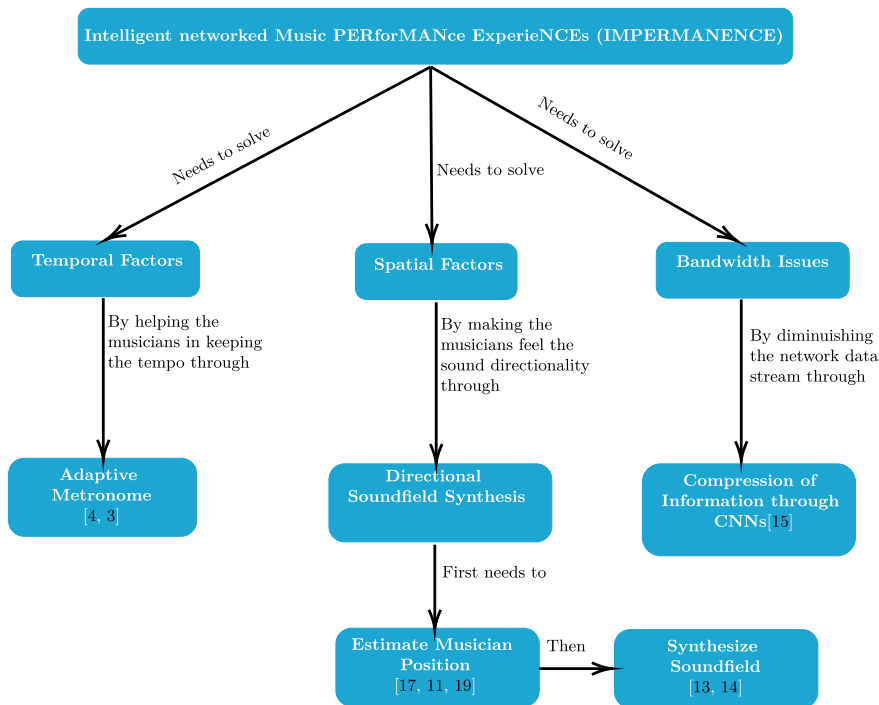


Fig. 2 The IMPERMANENCE framework for the implementation of Networked Music Performances

are able to devise strategies to contrast the effect of the latency. Treating spatial factors means to tackle both the auditory and visual perception. Following the findings obtained through Study II, presented in Sect. 2.2, where it was shown that a screen could suffice for what concerns the visual perception, we decided to focus on creating an interesting environment from the auditory point of view. For this reason, we present a technique for the rendering of the soundfield based on irregular loudspeaker setups. Since the IMPERMANENCE framework would imply that additional information needs to be transmitted via network (e.g. multichannel signals, metronome signal), we also preliminary explore the possibility of compressing the audio via Convolutional Neural Networks (CNNs).

4 Latency Compensation Through Adaptive Metronomes

The latency present in all network-based communications is of one of the main challenges that musicians face when performing a NMP. As already pointed out, several low-latency solution exists, however their major drawback is that they often need

high-end hardware, often not available to the musicians. In IMPERMANENCE, we propose to treat the latency, not by trying to minimize it, but instead, by trying to help the musicians to cope with it. A viable solution consists in using metronomes, light-weight devices, commonly used by musicians. The use of metronomes in NMP contexts was already explored in the literature [28]. In the IMPERMANENCE framework, we want to modify the metronome paradigm in order to make it akin to a Virtual Conductor (VC) [8], that is a software that gives tempo indications to the musicians, much like a conductor in a real orchestra does. In order to do this, we propose to use *adaptive metronomes*, that is, metronomes that are able to track the tempo of the musicians during the performance, through an off-the-shelf beat tracking algorithm [26], and to modify their tempo accordingly. We propose [3, 4] three adaptive metronomes techniques, differing by the complexity with which the tempo information is processed, namely: Single Beat Tracking with Master/Slave Approach (SBT), Crossed Beat Tracking (CBT) and Unique Metronome with Virtual Conductor (UMVC).

The SBT technique was proposed in order to get a first exploration of the viability of the adaptive metronome solution. It works by making the assumption that musicians behave following the *master-slave* latency compensation technique [10]. The musicians take two distinct roles: the *leader* determines the tempo of the performance and the *follower* tries to be synchronized with the leader's tempo. The SBT technique uses a single adaptive metronome that tracks the tempo of the leader and provides it to the follower, hopefully helping him/her in keeping a more steady tempo. SBT was tested in the Conservatorio Giuseppe Verdi di Milano [4] with real world musicians, showing promising results.

We then explored techniques that track the tempo of the two musicians at the same time. In the CBT technique, both musicians are tracked by two adaptive metronomes and each of them listens to a metronome whose tempo is based on the performance of the other musician. The UMVC technique instead uses two beat trackers to track both musicians, but provides only one common reference metronome signal to both of them, acting more similarly to a conductor. The details of the proposed algorithms are contained in [3], CBT and UMVC were tested only with amateur musicians and while results are still preliminary they encourage us to further developments of the techniques.

5 Spatial Audio Reproduction

The spatial perception of the sound emitted by the musical instruments, plays an important role in defining the quality of the NMP experience. However, as found out through study II, presented in Sect. 2.2, the physical devices used to perform reproduction should not be experienced as obtrusive by the musicians. For these reasons, in IMPERMANENCE we propose to reproduce the audio through an irregular array loudspeaker setup, that is an array where the spacing between the loudspeakers is not constant. This is motivated by the fact that while classic soundfield synthesis techniques are based on extensive loudspeaker arrays, NMP rooms are often cramped

with instruments and are hardly able to accommodate such setups. An irregular loudspeaker setup is more easily deployable in such scenario. However, irregular setups are challenging from the signal processing point of view, since the non-uniform sampling of the soundfield causes distortions during reproduction. A function modeling the correct driving signals for such loudspeaker configuration would be highly non linear and complex to model analytically. To overcome these limitations we propose to use techniques derived from deep learning, which enable the automatic extraction of highly non linear functions. In order to be able to reproduce the desired soundfield, we must know beforehand the position of the musicians in the room. This allows us both to reproduce correctly the perceived location (i.e. directionality) or to move them inside the remote room. Similarly to soundfield synthesis techniques, source localization ones are often based on extensive microphone array setups; we propose two techniques for the localization of the musicians based on minimal setups.

5.1 Source Localization Using Distributed Microphones Based on Ray Space Transform and Deep Learning

The first proposed localization technique is able to perform 2D localization of the musician, through the deployment of a small number of, arbitrarily deployed, microphones. If the microphones are synchronized, it is possible to compute the Generalized Cross Correlation (GCC) [29] between a reference microphone and the other ones. The highest peak of the GCC corresponds to the time-delay of the sound emitted by the source and measured at the microphones, which can then be exploited to perform localization. Unfortunately, noise and reverberation create several spurious peaks in the GCC, making the localization task harder. In the technique proposed in [17], we take advantage of CNNs and of the Ray Space Transform (RST) [6], a compact representation of the soundfield acquired by multiple Uniform Linear Arrays (ULAs) of microphones in terms of acoustic rays. Through a CNN we map the noisy GCC obtained at the real sparsely deployed microphones and map it to the simulated RST computed in anechoic conditions using ULAs surrounding the whole room. As demonstrated in [17] this procedure allows an accurate source localization even when dealing with highly challenging environments.

5.2 Source Localization Through Frequency-Sliding Generalized Cross-Correlation

The second proposed localization technique is based on an extension of the GCC framework, that entails the computation of the Frequency-sliding Generalized Cross-correlation (FS-GCC) following a sliding window approach. This enables us to obtain a set of sub-bands GCCs which can be stacked together into a single matrix, where

each row of the matrix corresponds to the GCC relative to the corresponding frequency band. The usefulness of this approach is given by the fact that in the anechoic (i.e. without noise and reverberation) scenario, the FS-GCC matrix is rank one. This allows us to exploit low-rank approximations and to separate the noisy components from the desired ones. Results presented in [11] show that the proposed approach is able to obtain better results than the GCC over real measurements. The proposed technique does not require any deep learning technique, and so it is more suitable to scenarios where extensive computational power may not be available. However, we experimented with the possibility of applying deep learning to the FS-GCC method in [19] by using a CNN to denoise the noisy input FS-GCC, showing better performances.

5.3 Soundfield Synthesis Through Irregular Loudspeaker Arrays

In order to reproduce the soundfield, we implement a technique that leverages on deep learning and on the Plane Wave Decomposition of the soundfield. Operatively, we consider the method already proposed in [5], where a Model-based acoustic Rendering (MR) technique is proposed. The MR technique reproduces the soundfield accurately when the considered setup is regular, but suffers from reproduction errors when dealing with irregular setups. The proposed technique, detailed in [13] considers a CNN that takes as input the driving signals obtained through the MR technique and outputs a compensated version of them. Since no ground truth is available for the compensated driving signals, we incorporate into the network architecture a soundfield estimation part, that simply convolves the obtained driving signals with the corresponding point-to-point Green's function, obtaining the estimated reproduced soundfield. The difference between estimated and ground-truth soundfield in terms of modulus and phase is used to compute the loss and train the CNN. We recently proposed an extension of this technique [14] following the same approach, with the difference that the driving signals are not obtained by compensating the ones obtained through a pre-existing method, but by extracting them through a CNN directly from discrete points of the environment where the soundfield is computed.

6 Speech Reconstruction from CNN Embeddings

Transmitting all the information needed in a NMP may create bottlenecks for what concerns the data transmission, especially when dealing with multidimensional signals such as the ones needed in a spatial audio framework. In [15] we explore a technique that could enable the compression of audio data. We do this by first studying a preliminary problem that has also relevant implications from an Explainable Arti-

ficial Intelligence (XAI) perspective. CNN models dealing with audio, usually take as input time-frequency representations of the audio signal, which is consecutively compressed by the layers of the model in order to extract high-level features. In [15] we consider pre-trained CNNs as feature extractors and build specular decoder networks in order to reconstruct the input time-frequency representation from the output of the intermediate layers. Results show that it is easy to reconstruct the input from convolutional layers, while it is considerably more complex when dealing with fully connected ones. This motivates us to further investigations, since by building proper networks, it would be possible to use their intermediate layers in order to perform compression on the multidimensional input audio signals.

7 Conclusion

In this chapter we presented the main results concerning the development of IMPERMANENCE as a comprehensive framework for the realization of NMPs. The biggest innovation of IMPERMANENCE is that it does not simply aim at reducing the issues already treated in the literature, e.g. low latency audio/video streaming, but it aims at creating an all around environment that feels coherent and whole to the musicians. In this sense, we treat the NMP as a Functional eXtended Reality (FXR) experience, where the objective is not the simplistic realism, but the creation of an environment that it is based on the needs of the musicians. In order to do this we base the design choices of IMPERMANENCE on the results obtained through experiments performed using the research framework TEMPERANCE, considering real musicians. We believe that the proposed method will enable NMP to be treated not as a mere approximation of the physical performance, but as a type of performance on its own.

References

1. Bartók, B.: *Mikrokosmos, 153 Progressive Piano Pieces. New Definitive Edition.* Boosey & Hawkes Music Publishers Limited, London (1987)
2. Bartók, B.: *44 duos for 2 violins, vol. 1–2.* Universal Edition (1992)
3. Battello, R., Comanducci, L., Antonacci, F., Cospito, G., Sarti, A.: Experimenting with adaptive metronomes in networked music performances. *J. Audio Eng. Soc.* **69**(10), 737–747 (2021)
4. Battello, R., Comanducci, L., Antonacci, F., Sarti, A., Delle Monache, S., Cospito, G., Pietrocola, E., Berbenni, F.: An adaptive metronome technique for mitigating the impact of latency in networked music performances. In: *2020 27th Conference of Open Innovations Association (FRUCT)*, pp. 10–17. IEEE (2020)
5. Bianchi, L., Antonacci, F., Sarti, A., Tubaro, S.: Model-based acoustic rendering based on plane wave decomposition. *Appl. Acoust.* **104**, 127–134 (2016)
6. Bianchi, L., Antonacci, F., Sarti, A., Tubaro, S.: The ray space transform: a new framework for wave field processing. *IEEE Trans. Signal Process.* **64**(21), 5696–5706 (2016)

7. Bischoff, J., Gold, R., Horton, J.: Music for an interactive network of microcomputers. *Comput. Music J.* 24–29 (1978)
8. Bos, P., Reidsma, D., Ruttkay, Z., Nijholt, A.: Interacting with a virtual conductor. In: *International Conference on Entertainment Computing*, pp. 25–30. Springer (2006)
9. Cáceres, J.P., Chafe, C.: Jacktrip: under the hood of an engine for network audio. *J. New Music Res.* **39**(3), 183–187 (2010)
10. Carôt, A., Werner, C.: Network music performance—problems, approaches and perspectives. In: *Proceedings of the “Music in the Global Village”-Conference, Budapest, Hungary*, vol. 162, pp. 23–10 (2007)
11. Cobos, M., Antonacci, F., Comanducci, L., Sarti, A.: Frequency-sliding generalized cross-correlation: a sub-band time delay estimation approach. *IEEE/ACM Trans. Audio Speech Lang. Process.* **28**, 1270–1281 (2020)
12. Comanducci, L.: *Intelligent networked music PERforMANce Experiences (IMPERMANENCE)*. Ph.D. thesis, Politecnico di Milano (2022)
13. Comanducci, L., Antonacci, F., Sarti, A.: Synthesis of soundfields through irregular loudspeaker arrays based on convolutional neural networks (2022). [arXiv:2205.12872](https://arxiv.org/abs/2205.12872)
14. Comanducci, L., Antonacci, F., Sarti, A.: A deep learning-based pressure matching approach to soundfield synthesis. In: *17th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE (2022)
15. Comanducci, L., Bestagini, P., Tagliasacchi, M., Sarti, A., Tubaro, S.: Reconstructing speech from cnn embeddings. *IEEE Signal Process. Lett.* **28**, 952–956 (2021)
16. Comanducci, L., Borra, F., Bestagini, P., Antonacci, F., Sarti, A., Tubaro, S.: Ray space transform interpolation with convolutional autoencoder. In: *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 261–265. IEEE (2018)
17. Comanducci, L., Borra, F., Bestagini, P., Antonacci, F., Tubaro, S., Sarti, A.: Source localization using distributed microphones in reverberant environments based on deep learning and ray space transform. *IEEE/ACM Trans Audio Speech Lang. Process.* **28**, 2238–2251 (2020)
18. Comanducci, L., Buccoli, M., Zanoni, M., Sarti, A., Delle Monache, S., Cospito, G., Pietrocola, E., Berbenni, F.: Investigating networked music performances in pedagogical scenarios for the intermusic project. In: *Proceedings of the 23rd Conference of Open Innovations Association (FRUCT)*, pp. 119–127. IEEE (2018)
19. Comanducci, L., Cobos, M., Antonacci, F., Sarti, A.: Time difference of arrival estimation from frequency-sliding generalized cross-correlations using convolutional neural networks. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Process (ICASSP)*, pp. 4945–4949. IEEE (2020)
20. Delle Monache, S., Buccoli, M., Comanducci, L., Sarti, A., Cospito, G., Pietrocola, E., Berbenni, F.: Time is not on my side: network latency, presence and performance in remote music interaction. In: *Machine Sounds, Sound machines-Proceedings of the XXII CIM Colloquium on Music Informatics*, pp. 152–158 (2018)
21. Delle Monache, S., Comanducci, L., Buccoli, M., Zanoni, M., Sarti, A., Pietrocola, E., Berbenni, F., Cospito, G.: A presence-and performance-driven framework to investigate interactive networked music learning scenarios. *Wirel. Commun. Mobile Comput.* **2019** (2019)
22. Delle Monache, S., Comanducci, L., Cospito, G., Sarti, A., Avanzini, F.: Synchronous online ensemble performance. In: Johnson, C., King, A. (eds.) *Music, Technology, Innovation: Industry and Educational Perspectives*. Routledge (2022)
23. Draper, J.V., Kaber, D.B., Usher, J.M.: Telepresence. *Hum. Factors* **40**(3), 354–375 (1998)
24. Drioli, C., Allocchio, C., Buso, N.: Networked performances and natural interaction via lola: Low latency high quality a/v streaming system. In: *International Conference on Information Technologies for Performing Arts, Media Access, and Entertainment*, pp. 240–250. Springer (2013)
25. Gabrielli, L., Squartini, S.: *Wireless Networked Music Performance*. Springer (2016)
26. Giorgi, B.D., Zanoni, M., Böck, S., Sarti, A.: Multipath beat tracking. *J. Audio Eng. Soc.* **64**(7/8), 493–502 (2016)

27. Holub, P., Matela, J., Pulec, M., Šrom, M.: Ultragrid: Low-latency high-quality video transmissions on commodity hardware. In: Proceedings of the 20th ACM International Conference on Multimedia, pp. 1457–1460. ACM (2012)
28. Hupke, R., Beyer, L., Nophut, M., Preihs, S., Peissig, J.: Effect of a global metronome on ensemble accuracy in networked music performance. In: Audio Engineering Society Convention 147. Audio Engineering Society (2019)
29. Knapp, C., Carter, G.: The generalized correlation method for estimation of time delay. *IEEE Trans. Acoust. Speech Signal Process.* **24**(4), 320–327 (1976). <https://doi.org/10.1109/TASSP.1976.1162830>
30. Pezzoli, M., Comanducci, L., Waltz, J., Agnello, A., Bondi, L., Canclini, A., Sarti, A.: A dante powered modular microphone array system. In: Audio Engineering Society Convention 145. Audio Engineering Society (2018)
31. Rottondi, C., Chafe, C., Allocchio, C., Sarti, A.: An overview on networked music performance technologies. *IEEE Access* **4**, 8823–8843 (2016)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



Fast and Robust Estimation of Atmospheric Phase Screens Using C-Band Spaceborne SAR and GNSS Calibration



Marco Manzoni 

Abstract Over the last few years, a growing interest has been observed in the field of Interferometric Synthetic Aperture Radar (InSAR) meteorology. The atmosphere has always been seen as a disturbance in interpreting interferograms (the output product of InSAR processing). A space-borne radar, however, can sense the refractive index of the medium it travels. The refractive index, in turn, is sensitive to pressure, temperature, and humidity of the air. Therefore, SAR data contains information about the atmosphere's status and can be exploited by Numerical Weather Prediction Models (NWPM) as additional information to improve weather forecasts. This chapter investigates a fast and robust method for generating the so-called Atmospheric Phase Screens (APS) from InSAR data. The method exploits both Permanent Scatterers (PS) and Distributed Scatterers (DS) in an optimal way leading to wide and dense APS maps. When operating at large scales, it is also mandatory to calibrate the data using a network of Global Navigation Satellite System (GNSS) receivers. The calibration can remove the so-called Orbital Phase Screens (OPS) that otherwise severely corrupt the atmospheric measurements. Results using real data acquired by the European Sentinel-1 mission show the potential of InSAR meteorology to provide valuable data to improve weather forecasts.

Keywords SAR · Atmospheric phase screen · GNSS calibration

1 Introduction

A precise weather forecast is crucial for several applications, such as agricultural planning, prevention of natural disasters, and prediction of solar energy production. Numerical Weather Prediction Models (NWPM) rely on abundant, continuous, and accurate input data to generate weather forecasts. These data can be provided by several different sensors such as radiosonde, ground-based, or space-borne radars and

M. Manzoni (✉)

Department of Electronics, Informatics and Bioengineering, Politecnico di Milano,
20133 Milan, Italy

e-mail: marco.manzoni@polimi.it

© The Author(s) 2023

C. G. Riva (ed.), *Special Topics in Information Technology*,

PoliMI SpringerBriefs, https://doi.org/10.1007/978-3-031-15374-7_11

Table 1 Requirements for integrated water vapor products into NWPM

Requirement	Threshold	Breakthrough	Goal
Temporal resolution	6 h	60 min	15 min
Spatial resolution	20 km	5 km	0.5 km

in-situ weather stations. Unluckily, the number of measures in sub-Saharan Africa is insufficient to provide an accurate weather forecast. In such a vast continent is indeed very complicated to install and maintain in-situ weather stations. The solution within the TWIGA¹ H2020 project (Transforming Water, weather, and climate Information through in situ observations for Geo-Services in Africa) comes from the exploitation of remotely sensed data and in particular from space-borne Synthetic Aperture Radar (SAR) data. While in passive optical systems, the sensor measures the reflection of the sunlight on an object, in a SAR system, an antenna brings the illumination to the scene, and the echo is recorded [2].

The most important feature of an active and coherent remote sensing system like a SAR is the capability to exploit the phase information of the echo signal [4]. The mentioned phase contains information about the optical path between the sensing platform and the targets on the ground. The optical path includes the geometrical distance between sensor and target and the refractive index of the medium (i.e. atmospheric conditions). The latter is the measure of interest for NWPM. The refractive index is related to the water vapor content in the atmosphere, thus it can be a valuable product for the ingestion [5].

A meteo-hydrological product must respect some requirements in terms of horizontal and temporal resolution to be helpful in the ingestion process. The Observing Systems Capability Analysis and Review (OSCAR) tool [1], developed by the World Meteorological Organization (WMO) sets the minimum horizontal resolution for Integrated Water Vapor (IWV) maps in high-resolution NWPMs at 20km with a goal, over which further improvements are not necessary, of 0.5 km. Concerning the temporal resolution, the minimum requirement is 6 h with a goal at every 15 min. The requirements are summarized in the Table 1.

A SAR system is generally able to respect the spatial resolution required by a NWPM. The field of view of a space-borne SAR is usually several thousands of km² with a few meters of resolution. Even if spatial filtering is employed, the final resolution will still be better than the goal one (below 500 m). It is not the same for the temporal resolution that, in the case of SAR systems, is much coarser than the threshold one. A temporal revisit of 6 days is the best we can achieve thanks to the Sentinel-1 constellation, and it is nowhere near the threshold of 6 h defined by OSCAR.

A Global Navigation Satellite System (GNSS) network is the opposite: it can provide continuous measurement in time; however, it is just a point-wise measurement in

¹ TWIGA has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 776691.

space. Nevertheless, the two can collaborate, and in particular, the GNSS can provide valuable additional information to calibrate SAR Atmospheric Phase Screen (APS) maps. In the following section, we will review the effects of the atmosphere on the SAR measurements, and then we will proceed in detailing the processing workflow for APS maps estimation.

2 The Essence of SAR and InSAR

The complex radar image (also called Single Look Complex, SLC) contains information both about the intensity of the reflection and about the sensor-target travel time [7]. The former is encoded in the image amplitude, while the latter in the image phase. The phase of a single image is, however, not exploitable since it is the result of the coherent superposition of many elementary scatterers inside the resolution cell. Information can be extracted through a process called interferometry. An interferogram is formed by coherently combining two SAR images:

$$I = P_1 P_2^* \propto \exp\{-j(\phi_1 - \phi_2)\} = \exp\{-j\psi\} \quad (1)$$

where P_1 is the first image, P_2 is the second image, ϕ_1 and ϕ_2 are their respective phases, ψ is called interferometric phase and * indicates the complex conjugate. The interferometric phase can be modeled as the sum of several terms:

$$\psi = \psi_{\text{flat}} + \psi_{\text{topo}} + \psi_{\text{defo}} + \psi_{\text{atmo}} + \psi_{\text{orbit}} + \psi_{\text{scat}} + \psi_{\text{noise}} \quad (2)$$

where ψ_{flat} is a reference phase, ψ_{topo} is a component linearly related with the topography of the scene, ψ_{defo} is proportional to the deformation of the scene between the two acquisitions, ψ_{atmo} is the APS we are interest into, ψ_{orbit} is the orbital phase screen induced by errors in the knowledge of the satellites orbits during acquisitions, ψ_{scat} is due to the difference of scattering mechanism between the two acquisitions and ψ_{noise} is the always-present noise in the measure. Each component may or may not be useful for a particular purpose. If, as in our case, the atmospheric conditions are the objective of the research, all the other terms except for ψ_{atmo} are considered as noise and must be removed from the interferometric phase.

3 Tropospheric Effects on SAR Data

When the radar signal travels through the atmosphere, it is delayed by a quantity called propagation delay. The delay depends on the refractive index of the medium itself. If the refractive index is not unitary, the optical distance R between the target and the satellite can be approximated as the sum of the geometrical distance R_g

and the equivalent excess path generated by the non-unitary refractive index R_a . A common value for R_a in the vertical (zenith) direction is between 2.2 and 2.7 m [10].

The excess path delay can be further divided into a spatially and temporally smooth component called dry (or hydrostatic) delay and a spatially turbulent one called wet delay. The latter is the one that is directly proportional to the water-vapor density in the medium. The dry delay is usually in the order of a couple of meters, while the wet one is no more than 0.3 m [10].

The wet delay is the main responsible for the so-called atmospheric artifacts in any interferometric SAR processing such as Digital Elevation Model (DEM) estimation or deformation monitoring. The fluctuations of this term are significant both in space and time, and we can refer to them as atmospheric turbulence. The reduced spatial scale and high temporal variability make it difficult to estimate the wet delay using external measures in an accurate and high-resolution fashion. Nevertheless, this contribution aims to estimate this delay that still contains valuable information about the water vapor content in the troposphere. In the following section, the complete processing scheme is described in detail.

4 Atmospheric Phase Screen Estimation

The complete workflow for the estimation of the APS maps is depicted in Fig. 1. Each step will be described in this section.

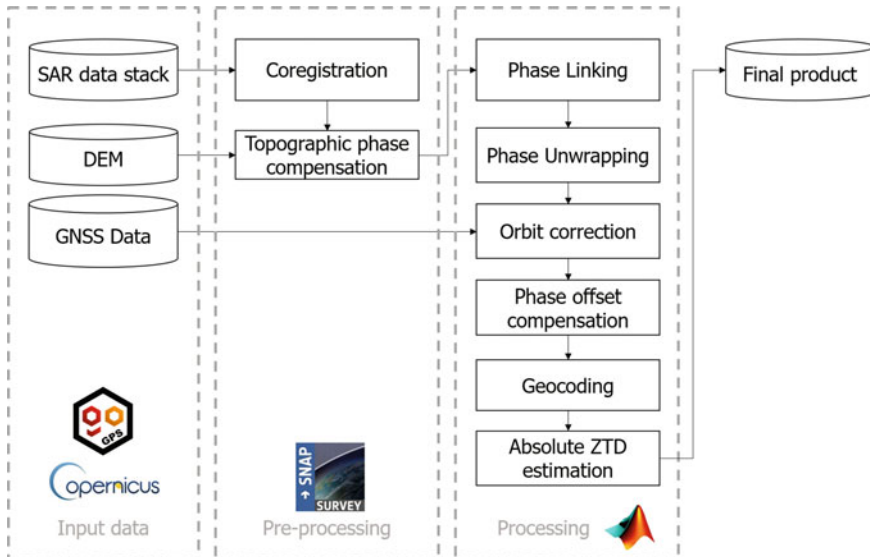


Fig. 1 Scheme of the complete workflow for the estimation of atmospheric phase screens

4.1 Data Pre-processing

We start from a set of freely available images from the European SAR constellation Sentinel-1. The temporal baseline between two consecutive images is in the order of 6/12 days. This particular temporal baseline is guaranteed by the revisit frequency of the constellation. First of all the images must be properly coregistered [9] and the topographic component of the interferometric phase must be compensated using a DEM. Both steps can be executed using the open source software SNAP provided by ESA.

Also the GNSS data must be processed at this stage to obtain an atmospheric product called Zenith Total Delay (ZTD) that is the instantaneous atmospheric delay as sensed by a GNSS station. The processing of GNSS data is done using the free and open source software GoGPS [8]. The details about the processing of GNSS data for the extraction of atmospheric products are out of the scope of this contribution.

4.2 Phase Linking for APS Estimation

Once the images have been coregistered and compensated for topography, the phase estimation can take place. The workflow implements the so-called Phase Linking [3] algorithm. With N images in the stack, we can form $N(N - 1)/2$ different interferograms. Phase Linking tries to justify all these interferograms with $N - 1$ equivalent (i.e., best-fitting) interferograms.

We can put the algorithm in a mathematical framework by saying that, if the temporal covariance matrix of the data contains all the possible interferograms between couples of images, Phase Linking finds the best Rank-1 approximation of the covariance matrix. The advantage of this method is that, by exploiting the whole covariance matrix, it automatically reaches the best possible phase estimation (in terms of variance) both in the presence of a PS and a DS.

In Fig. 2 we depicted three covariance matrices, and in yellow, we highlighted the portion of the matrix exploited by different phase estimation algorithms.

The Phase Linking exploits all the possible interferograms leading to optimal estimation of the interferometric phases. It is not the case for the AR(1) algorithm that foresees interferogram formation at a short temporal baseline followed by integration. The disadvantage is that during integration, the noise is accumulated; thus, the solution is not optimal. Another technique called DInSAR foresees the interferogram formation at varying temporal baselines. This way of estimating the InSAR phase is particularly noisy for decorrelating targets at long baselines.

The Phase Linking algorithm is a generic estimator for interferometric phases. Some practical precautions are needed to apply this algorithm for atmospheric phase screen retrieval:

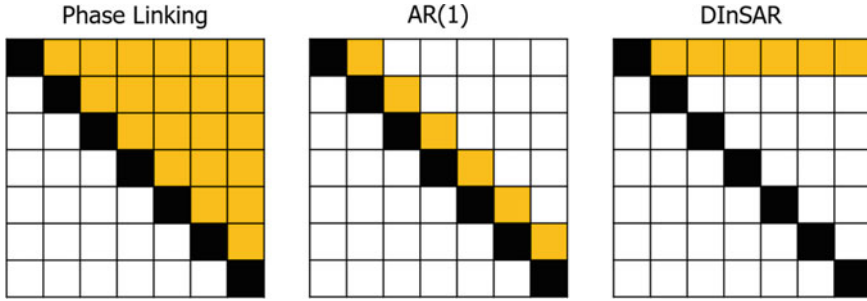


Fig. 2 Each algorithm exploits a different portion of the coherence matrix. Phase Linking, for example, utilizes all the available interferograms, the AR(1) just the interferograms at the shortest temporal baseline, while the standard DInSAR method exploits all the interferograms at varying temporal baseline

1. **The total temporal extent of the stack must be short.** The first consequence of reducing the number of images in the stack is to reduce the effect of terrain deformations. In the model of the interferometric phase (Eq. 2), a term due to the presence of ground deformation is present. With an average subsidence rate of 10–20 mm/year, if the stack is kept short, the impact of deformation on the interferometric phase is minimal. Considering 8 images with a total temporal extent of about 50 days (in the case of Sentinel-1), the error will be less than 2 mm. This bias in the estimate is tolerable for our purposes. Notice that these considerations are valid just in the presence of a small-medium subsidence rate, not for extreme deformations or earthquakes where the subsidence can be in the order of several tens of cm.

The stack temporal extent also needs to consider the average “life” of a distributed scatterer. As already pointed out, we exploit all possible interferograms with N images with Phase Linking. If the coherence of the interferograms with a very long temporal baseline is very low, they will bring noise into the final estimate.

2. **The number of independent samples used to estimate the covariance matrix must be carefully selected.** The interferometric phase’s quality is tightly related to the number of independent samples used to estimate the covariance matrix. We found that with roughly 600 independent samples, the standard deviation of the phase always remains below 1 mm. Sentinel-1 shows a nominal resolution of $5 \times 20 \text{ m}^2$ in IW acquisition mode in range and azimuth, respectively. The images are over-sampled at about $2.5 \times 15 \text{ m}^2$. To obtain a number of independent looks equal to 600, we need roughly 3000 pixels. We choose a 149×25 window spanning $375 \times 375 \text{ m}^2$. This last figure is the resolution of our APS maps which is well below the OSCAR requirements.

Once the interferometric phases have been derived, they are unwrapped: phase unwrapping is a well-known technique and will not be discussed in this contribution.

4.3 GNSS Calibration

In the interferometric phase model described in Sect. 2, the term ψ_{orbit} represents the so-called Orbital Phase Screen (OPS). This component arises due to the inaccurate knowledge of the satellite's orbit during the acquisition of each image. The OPS manifests as a phase plane (i.e., a trend) in the interferogram. The most common solution in literature is simply fitting and removing a plane in the interferometric phase. However, this process could corrupt the estimated APS's low-frequency components.

In this contribution, we developed a novel technique based on a network of GNSS stations. The workflow extract from SAR data the value of the APS over each station. The ZTD provided by each station is differentiated in time to form the differential ZTD (dZTD) that is then projected in the slant range using the SAR looks angle, obtaining a GNSS-derived equivalent APS. We recall now that the SAR interferogram contains both the APS and the OPS. If we remove from the SAR interferogram the GNSS-based APS, what is left is just the orbital plane.

Only then it is possible to fit a plane and estimate the OPS without the risk of filtering the signal of interest. Moreover, if the plane is calculated in radar coordinates, the parameters of the plane equation are related to physical quantities of the SAR acquisition, namely the normal baseline error and the derivative of the parallel baseline error. Once the parameters have been estimated, the forward problem can be computed for each pixel in the interferogram grid leading to a set of calibrated APS.

4.4 Transforming the Maps from Differential to Absolute

The last step before the ingestion into NWPM in the so-called *absolutization* process [6]. All the maps are differential with respect to a common but unknown reference atmosphere. The simplest and most effective solution is to provide an a-priori map to be summed to each interferogram. One of the sources of the prior could be a NWPM itself. It is sufficient to sum a single absolute ZTD map to all the phase-linked interferograms to obtain a stack of absolute ZTDs (see Eq. 1). We remark again that a NWPM or a GNSS network can provide this prior. In this latter case, the absolute ZTD must be interpolated over the whole SAR grid.

5 Results with Sentinel-1 Data

In this section we provide some results using real data acquired with the constellation Sentinel-1 (S1) over South Africa. The data, composed by five S1 frames covers an area of $250 \times 850 \text{ km}^2$. The footprint is depicted in Fig. 3. The area shows also a

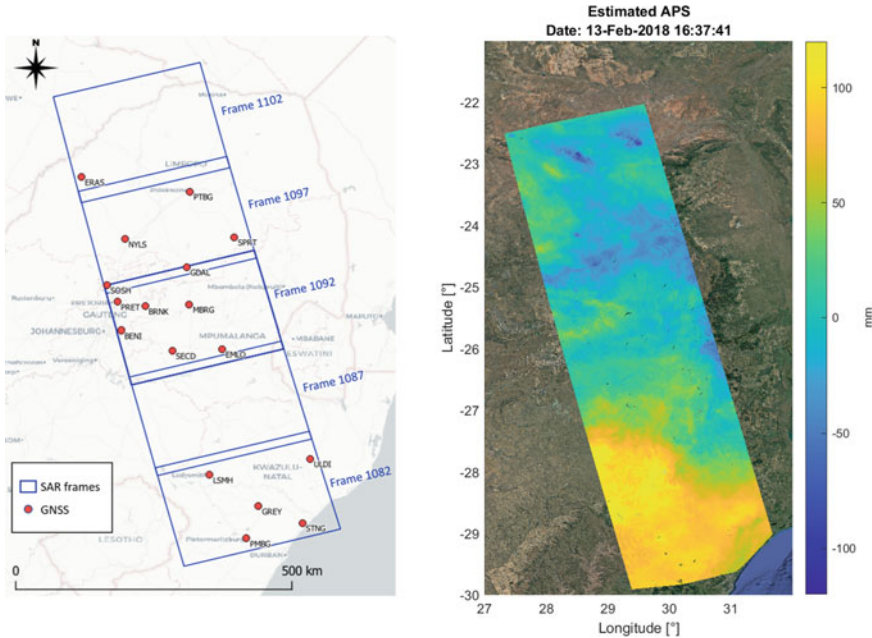


Fig. 3 The five Sentinel-1 frames used lead to a $250 \times 850 \text{ km}^2$ interferogram

very poor mean coherence meaning that the scene is very unstable due to temporal decorrelation making it hard to obtain a reliable estimate with traditional InSAR phase estimators.

In Fig. 3 we depicted also a single APS maps. Five images form the stack, but for brevity, we decided to depict just one of them. The visible holes are areas where the phase estimation is unreliable such as water bodies or dense forests with severe temporal decorrelation. Spatial variograms and spatial spectra have been computed to characterize the maps derived statistically.

In Fig. 4a the spatial variogram is depicted. In black dashed lines, all the variograms for the five images are represented, in blue the average variogram, while in red and green the $2/3$ power law and $5/3$ power laws. From the figure, it is evident that the data follow on average the theoretical model proving its ability to capture the APS dynamics. The same fit can be found for radially average spectra. In Fig. 4b the results are shown: the derived spectra fit well the theoretical model provided for the turbulent part of the APS. The model is a $1/f^\alpha$ power law where $\alpha = 5/3$ for radially averaged power spectra [4].

We also validated the maps with an external source. In our case, we used freely available data from the Generic Atmospheric Correction Online Service (GACOS). It can provide absolute ZTD and differential APS derived by NWP to correct atmospheric artifacts in InSAR products. In our study we did not use these datasets to *correct* the APS, but to *validate* it. In Fig. 5 one comparison is depicted where we

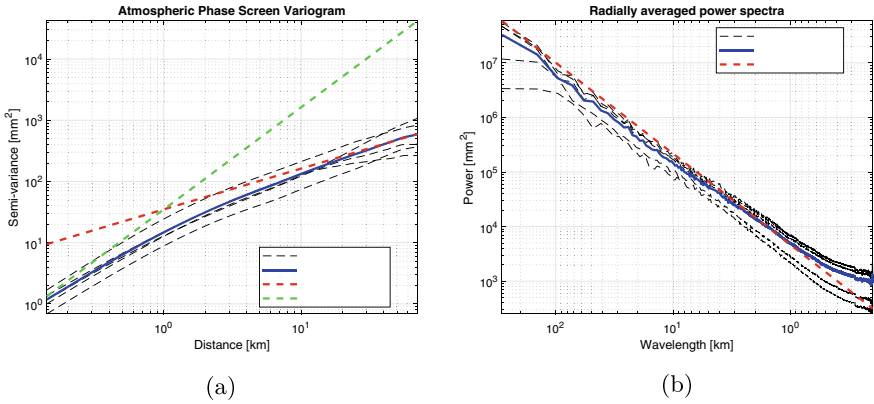


Fig. 4 **a** Variogram of the APS in the South African case study. The data follows the theoretical model. **b** Average power spectra of the data. Also in this case the theoretical model is respected

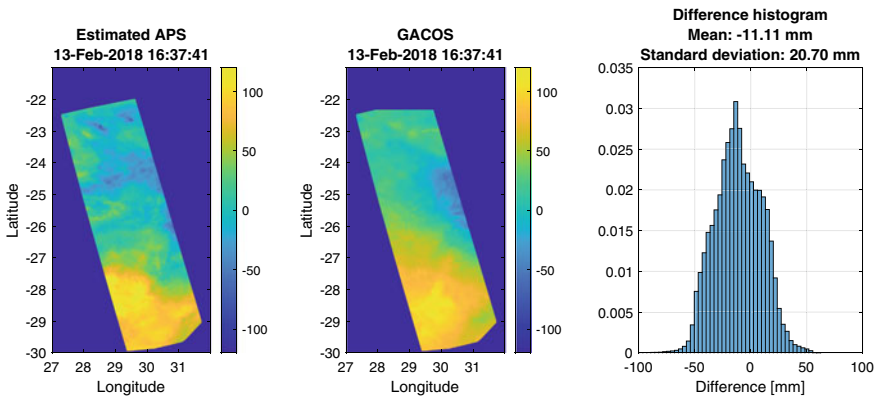


Fig. 5 Comparison between GACOS and one of the estimated APS maps. The bias between the two is roughly 1 cm, with a standard deviation around 2 cm

can see from the histogram that the error is slightly more than 1 cm in the mean and with a standard deviation of roughly 2 cm.

6 Conclusion

In this contribution, we presented a novel way to estimate Atmospheric Phase Screens (APS) from a stack of SAR data. The method exploits both Permanent and Distributed Scatterers (PS,DS): this feature allows the generation of wide and dense maps that can be ingested into NWPM to improve weather forecast. The algorithm has been extensively described in this chapter, its working principle is detailed, and results

using real data acquired by the European constellation Sentinel-1 are presented. The validation has been carried out using a statistical and theoretical analysis of the variograms and spectra of the maps and by comparing them with a NWPM (GACOS). In both cases, the accordance is very high, showing the capability of the proposed method to capture most of the dynamics of the atmosphere.

References

1. Requirements defined for Integrated Water Vapour, shorturl.at/kouUV
2. Cumming, I.G., Wong, F.H.C.: Digital processing of synthetic aperture radar data: algorithms and implementation. Artech House remote sensing library, Artech House, Boston (2005). (oCLC: ocm56805179)
3. Guarnieri, A.M., Tebaldini, S.: On the exploitation of target statistics for SAR interferometry applications. *IEEE Trans. Geosci. Remote Sens.* **46**(11), 3436–3443 (2008)
4. Hanssen, R.: Radar Interferometry Data Interpretation and Error Analysis, vol. 2 (2001). <https://doi.org/10.1007/0-306-47633-9>
5. Manzoni, M., Monti-Guarnieri, A.V., Realini, E., Venuti, G.: Joint exploitation of SAR and GNSS for atmospheric phase screens retrieval aimed at numerical weather prediction model ingestion. *Remote Sens.* **12**(4), 654 (2020)
6. Meroni, A.N., Montrasio, M., Venuti, G., Barindelli, S., Mascitelli, A., Manzoni, M., Monti-Guarnieri, A.V., Gatti, A., Lagasio, M., Parodi, A., Realini, E., Tagliaferro, G.: On the definition of the strategy to obtain absolute InSAR Zenith Total Delay maps for meteorological applications. *Front. Earth Sci.* **8**, 359 (2020)
7. Moreira, A., Prats-Iraola, P., Younis, M., Krieger, G., Hajnsek, I., Papathanassiou, K.P.: A tutorial on synthetic aperture radar. *IEEE Geosci. Remote Sens. Mag.* **1**(1), 6–43 (2013). <https://doi.org/10.1109/MGRS.2013.2248301>
8. Realini, E., Reguzzoni, M.: goGPS: open source software for enhancing the accuracy of low-cost receivers by single-frequency relative kinematic positioning. *Meas. Sci. Technol.* **24**(11), 115010 (2013)
9. Sansosti, E., Berardino, P., Manunta, M., Serafino, F., Fornaro, G.: Geometrical SAR image registration. *IEEE Trans. Geosci. Remote Sens.* **44**(10), 2861–2870 (2006). <https://doi.org/10.1109/TGRS.2006.875787>
10. Ulaby, F.T., Long, D.G.: Microwave Radar and Radiometric Remote Sensing. The University of Michigan Press, Ann Arbor (2014)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.



How Deep Learning Can Help Solving Geophysical Inverse Problems



Francesco Picetti 

Abstract This brief summarizes some of the main results I obtained during my Ph.D. studies at Politecnico di Milano, under the supervision of Professor Stefano Tubaro. The thesis provides contributions to understanding the advantages, and limitations, of data-driven deep learning approaches to geophysical inverse problems, with a special focus on Convolutional Neural Networks (CNNs). Exploration Geophysics aims at estimating accurate physical properties of the Earth subsurface from seismic data acquired close to the surface. Seismic data show a great variety of statistically relevant and independent patterns. I devise Deep Learning methods to solve several geophysical tasks by learning such patterns. First, I devise generative networks as a post-processing operator for refining reflectivity images. When trained on pure image datasets, these networks suffer from the lack of physical knowledge. Then, I show a different approach named Deep Priors, which are CNNs that precondition the inverse problem. In particular, I develop a scheme to interpolate seismic data. Finally, I leverage the features extraction ability of CNNs for buried landmine detection on Ground Penetrating Radar (GPR) acquisitions. While the presented methods are effective compared to the state of the art, improvements can be achieved by integrating pure data-driven algorithms within general inverse problems theory through a-priori information derived from domain knowledge.

Keywords Inverse problems · Deep learning · Deep priors · Geophysics

1 Introduction

During my doctoral studies, I mainly focused on reflection seismology problems. As the name suggests, reflection seismology processes seismic waves reflected by the subsurface features like faults, sediment layers, and rock bodies. Intrinsicly, reflection seismology is a collection of inverse problems [4, 18]. Additionally, given

F. Picetti (✉)

Dipartimento di Elettronica, Politecnico di Milano, Informazione e Bioingegneria via Ponzio 34/5, 20133 Milano, Italy
e-mail: francesco.picetti@polimi.it

© The Author(s) 2023

C. G. Riva (ed.), *Special Topics in Information Technology*,

PoliMI SpringerBriefs, https://doi.org/10.1007/978-3-031-15374-7_12

the dimensionality of the datasets, the costs related to acquisitions and processing, and the scarcity of ground truth, its tasks are often hard to solve and thus very challenging.

Advanced problems are ill-posed, which means that the solution is not unique, and are ill-conditioned, which means that small errors in the data result in dramatic changes in the solution [19]. These aspects and computational issues make the inverse problem theory long to be considered fulfilled.

For excellent physical reasons, seismic spectra imply that single-experiment data are insufficient for imaging purposes. This observation explains why seismic surveys are multi experiment [18]. The redundant data and the nonlinear nature of the modeling functions motivated the study of supervised deep learning imaging techniques. Although it is strongly based on statistical methods, machine learning has the particular focus on *making decisions* in an automated manner. In other words, it intrinsically aims at merging data processing with human interpretation. In exploration geophysics, solutions to inverse problems must fulfill some physical constraints, as they must be interpreted. In this sense, geophysicists are studying how to embed ML-generated features into traditional inversion schemes. Vice versa, physical constraints have been applied to machine learning scheme for lithofacies segmentation [11], data processing [14] and inversion problems [5]. Additionally, in [6] human interpretation methods are integrated into a CNN scheme.

Generative models were also proposed to precondition the inversion, under the assumption that CNNs can extract high-level features from data. This is the case of the so-called *Deep Prior* paradigm, initially developed for super-resolution, denoising, and inpainting of natural images [21].

In the light of these considerations, my research focused on understanding the advantages, and drawbacks, of machine learning methods for addressing geophysical inverse problems. To this end, the present brief is organized as follows. In Sect. 2, I first introduce a seismic imaging technique, the Reverse Time Migration (RTM), whose least-squares solution I address by means of generative networks [15]. Such architectures are a powerful tool for enhancing seismic images, but they lack physical information. Thus, I propose to change paradigm. In Sect. 3 I develop a Deep Prior interpolation scheme for seismic data [12]. In Sect. 4 I deploy convolutional autoencoders to detect buried landmine on radar data [3]; a CNN learns the secure soil features, thus failing to reconstruct signatures from possible threats. Finally, I draw my conclusions and devise a pathway for future research.

2 RTM Image Enhancement Through GANs

Reverse Time Migration (RTM) is a wave-equation-based imaging technique that represents the state-of-the-art industrial technology for depth imaging [23]. The goal of RTM is to image the reflectivity of the Earth subsurface from reflection data [2]:

$$\mathbf{d} = F(\mathbf{m}) + \mathbf{n}, \quad (1)$$

where \mathbf{d} are the observed scattered data, \mathbf{m} is the reflectivity model, F is a forward modeling operator and \mathbf{n} is an additive noise.

Usually, the reflectivity image is computed as:

$$\mathbf{m} = \mathbf{B}'\mathbf{d}, \quad (2)$$

where \mathbf{B} is the Born linear approximation of F that leverages finite difference Green's functions, and $'$ represents the adjoint operator (i.e., the conjugate transpose).

To overcome limitations such as spatial aliasing, limited aperture, and non-uniform illumination, Least-Squares Reverse Time Migration (LS-RTM) was proposed to invert the forward modeling operator through iterative algorithms [17] in a least-squares sense:

$$\hat{\mathbf{m}} = \arg \min_{\mathbf{m}} \|\mathbf{d} - \mathbf{B}\mathbf{m}\|_2^2 + \lambda R(\mathbf{m}), \quad (3)$$

where R is a proper regularizing operator that imposes a-priori information or desired features on the resulting solution (e.g., it can enforce solution sparsity, smoothness, etc., depending on the result on the desired goal).

2.1 Generative Adversarial Networks

We can split the least-squares problem into two different problems: a standard imaging operator followed by a post-processing operator. In particular, we can invert the model-data relationship (1) as

$$F^{-1} = \mathcal{G} \circ B', \quad (4)$$

where \mathcal{G} is a mapping operator (i.e., a post-processing machine) that transforms the migrated $\mathbf{m} = \mathbf{B}'\mathbf{d}$ into the desired image $\hat{\mathbf{m}}$.

I propose to use a convolutional neural network (CNN) as a general-purpose post-processing tool after RTM to produce images as if more sophisticated algorithms generated them. Specifically, I devise a Generative Adversarial Network (GAN), which is a class of CNNs that have been proposed in [8] to learn deep representations using a small training dataset. To achieve the goal of a realistic generation, they are trained singularly.

In supervised training, the generator \mathcal{G} minimizes a distance between the generated and reference images. In GANs, the generator is flanked by a fellow *discriminator* \mathcal{D} , which is a binary classifier trained to discriminate whether its input is a pristine image or generated by its mate. At the same time, \mathcal{G} is trained to obtain the desired output from a given input and fool \mathcal{D} . Therefore, \mathcal{D} can be seen as a regularizer driving \mathcal{G} to output realistic images. More formally, we can write the adversarial loss as the sum of two components. \mathbf{m} and $\bar{\mathbf{m}}$ being the input and the ground truth vectors, respectively, the first term, referred to as the *generator loss*, is often defined as the ℓ_2 -norm of the error introduced by the generator:

$$\mathcal{L}_G = \|\mathcal{G}(\mathbf{m}) - \bar{\mathbf{m}}\|_2^2, \quad (5)$$

it forces the generated model to be similar to the target. As second term, we define the *discriminator loss* starting from a query image $\hat{\mathbf{m}}$ as

$$\mathcal{L}_D = \log \mathcal{D}(\hat{\mathbf{m}}, \bar{\mathbf{m}}) + \log (1 - \mathcal{D}(\hat{\mathbf{m}}, \mathcal{G}(\mathbf{m}))), \quad (6)$$

which measures how likely the generator is able to fool the discriminator in terms of binary cross-entropy. Finally, these two terms are added together to form the GAN loss:

$$\mathcal{L} = \mathcal{L}_G + \lambda \mathcal{L}_D, \quad (7)$$

where the parameter λ balances the two contributions.

2.2 Experimental Results

Let us consider a fast track project: we desire high quality migrated images, but we have no resources to perform RTM over the entire data. We could perform a fine migration over a subset of the available dataset, and a faster migration (e.g., with subsampled data) over the whole dataset. Then, we train a GAN to refine the coarsely migrated images, and finally, we can deploy it over the entire coarse dataset.

Input images \mathbf{m} have been generated by migrating 10 equispaced sources and 80 equispaced receivers covering the whole acquisition surface. The target images $\bar{\mathbf{m}}$ have been migrated from 200 sources and 800 receivers.

The training was performed for 300 epochs; nonetheless, the net converged after 50 epochs approximately. The parameter λ was set to 0.1. Once the training is completed, the time required to create an output image of the network was about 2 min, almost entirely dedicated to migration with the coarse geometry, while a fine geometry would require 40 min.

It is interesting that results are pretty promising on the test image, showing that the proposed solution does not simply overfit the training set but can generalize. Nevertheless, a few details are lost on the test set. It is worth noticing a channel around the center of the target patch shown in Fig. 1, which is missing in the output and barely present in the coarse input. To increase the quality of the generated images, we would like to add physical information to the CNN training.

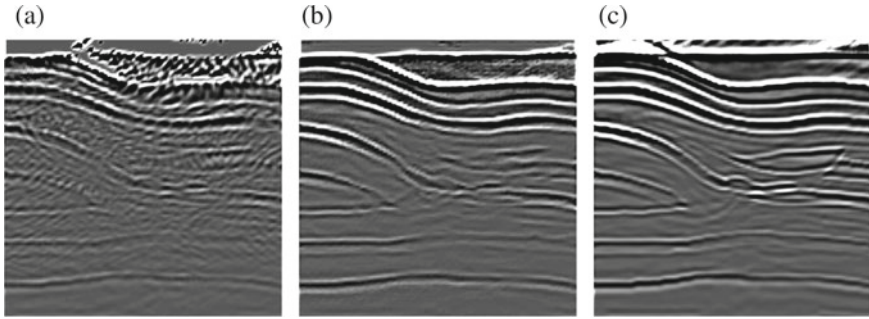


Fig. 1 Input \mathbf{m} (a), output $\hat{\mathbf{m}}$ (b) and desired $\bar{\mathbf{m}}$ (c) patches from the test set

3 Seismic Data Interpolation Through Deep Priors

Let us consider an ideal densely and regularly sampled three-dimensional cube of seismic data. This data cube represents the true model we aim at estimating through interpolation. Without loss of generality, we can represent ideal data as a vector $\bar{\mathbf{m}}$. We can think of the observed seismic data \mathbf{d} as generated by a linear sampling operator \mathbf{S} applied to $\bar{\mathbf{m}}$. Therefore, we can define the interpolation problem as the inversion of

$$\mathbf{d} = \mathbf{S}\bar{\mathbf{m}}. \quad (8)$$

Within the Deep Prior paradigm [21], the standard inverse problem cost function

$$J(\mathbf{m}) = \|\mathbf{S}\mathbf{m} - \mathbf{d}\|_2^2 \quad (9)$$

is recast as finding the set of parameters θ which minimizes

$$J(\theta) = \|\mathbf{S}f_\theta(\mathbf{z}) - \mathbf{d}\|_2^2, \quad (10)$$

where \mathbf{z} is a noise realization and θ the parameters of the CNN f_θ .

Once the optimum set of parameters θ^* has been computed, the inverted model is simply the output of such optimized network:

$$\hat{\mathbf{m}} = f_{\theta^*}(\mathbf{z}). \quad (11)$$

Figure 2 helps visualizing the deep prior scheme adopted for interpolation. The convolutional autoencoder f_θ is optimized by computing the loss function over the actual known traces only; no ground truth is required. As a byproduct, it also generates missing traces.

In the inversion process, the CNN implicitly assumes the role of prior information that exploits correlations in the data to learn their inner structure. Therefore, choosing a specific CNN architecture is critical for a suitable and well-performing solution.

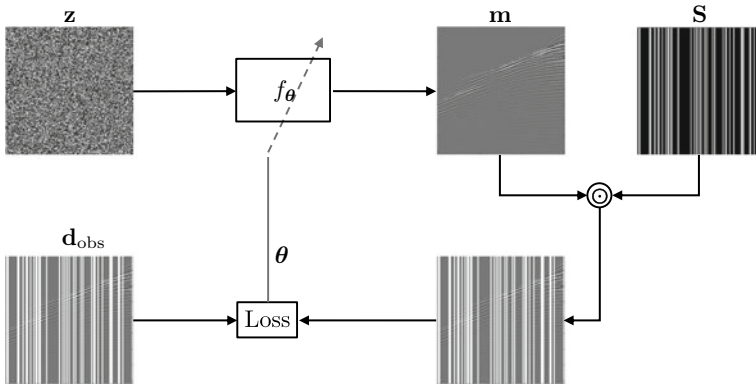


Fig. 2 Seismic data interpolation in the Deep Prior paradigm

3.1 Experimental Results

In this example, the desired shot gather is obtained via FD acoustic propagation over a 3D extension of the Marmousi model, depicted in the top-right pane of Fig. 3. It comprises 175 crosslines and 100 inlines; the sampling interval is 10m. The wavelet is a Ricker centered in 15 Hz; the recording time is 3.5 s. The top-left pane shows the input decimated data, obtained by randomly deleting 90% of the traces. Interpolation is performed with a modified version of the Multi-Resolution U-Net architecture [10], employing 2D kernels (thus processing slices of the volume) and 3D kernels. The interpolation results are shown in the bottom left and right panes, respectively. The 3D convolutions are the right choice to extract relevant information from the decimated volume; however, the computational cost is much higher than the 2D case.

3.2 Tackle Aliasing with Event Dips

Deep Priors have shown effective reconstruction performance when resolving irregular sampling. However, such techniques do not optimally perform when facing aliased data. A viable way to improve the performance is to drive the solution according to the events slopes.

Specifically, I first estimate a low-pass version of the data considering as

$$J(\theta) = \|\mathbf{S}f_{\theta}(\mathbf{z}) - \mathbf{L}\mathbf{d}\|_2^2, \quad (12)$$

\mathbf{L} being a second order Butterworth low-pass filter applied trace-by-trace, thus removing alias effects that could badly impact on standard Deep Prior inversion. Then, I estimate the slope angles ϕ through the structure tensor algorithm [22], to build

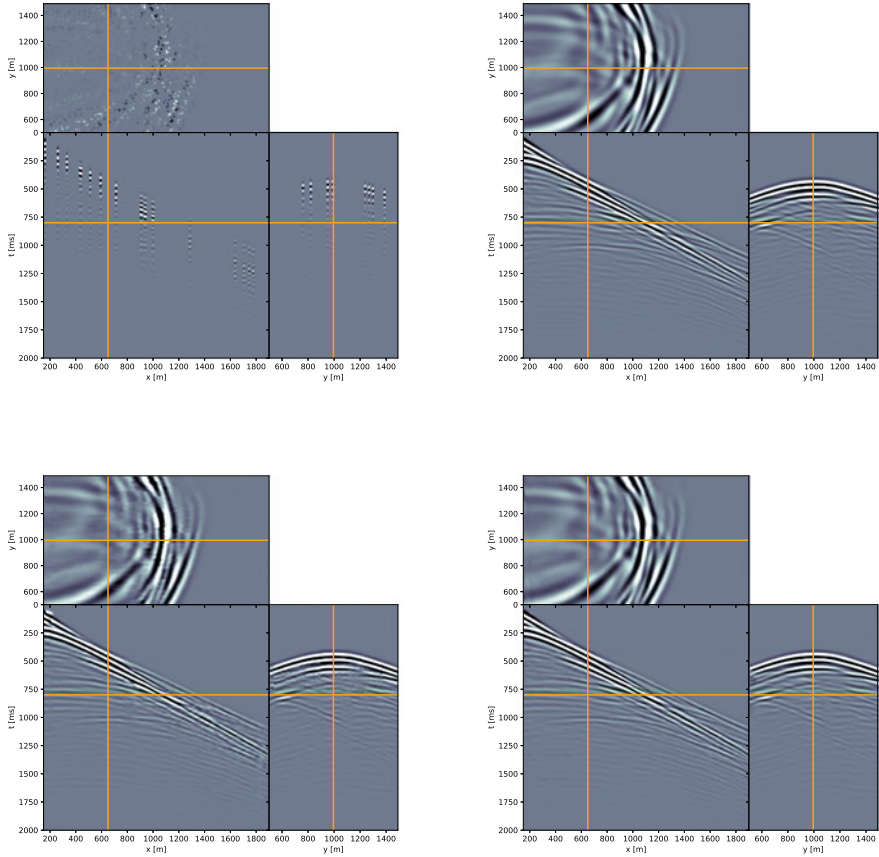


Fig. 3 Deep Priors interpolation of pre-stack seismic data. Top: input coarse data and desired dense data; bottom: interpolated data through 2D and 3D kernels, respectively

the directional Laplacian operator $\vec{\nabla}_\phi$ [9]. The latter is employed to regularize the inversion of the full-band seismic data:

$$J(\theta) = \|\mathbf{S}f_\theta(\mathbf{z}) - \mathbf{d}\|_2^2 + \varepsilon \mathbf{w}^2 \left\| \vec{\nabla}_\phi f_\theta(\mathbf{z}) \right\|_2^2, \quad (13)$$

where \mathbf{w} is the vector that collects the anisotropy of the estimated gradient square tensor for each data sample [22], and it can be interpreted as a confidence measure of the estimated slopes. Notice that, as the optimization goes on, the interpolated data can be used to produce a better estimate of the data slopes, updating $\vec{\nabla}_\phi$ at runtime.

I compare the interpolation results with and without the proposed regularizer on a synthetic dataset of 4 linear events with different dips. Three events are mildly steep, while the fourth event has been designed to produce aliasing effects. This example is simple yet particularly challenging due to the slope of the events. The

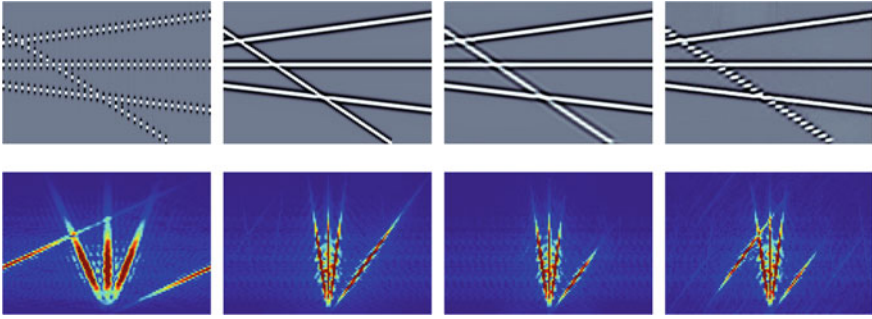


Fig. 4 Interpolation results on synthetic linear events: input decimated data, target complete data, output of Deep Prior optimization with and without the proposed regularization. Above, the data; below, the corresponding FK spectra

desired data are depicted in the second column of Fig. 4. I build a regular binary mask for simulating the decimation by keeping one trace over 3, as depicted in the first column. The network is optimized for 3000 iterations to fit the low-passed data (12). Then, the slopes are estimated, and the regularized optimization is performed for 7000 iterations with $\varepsilon = 5$.

The results of Deep Prior optimization when minimizing the regularized (13) and the standard data fitting (10) are reported in the third and fourth columns, respectively. Standard Deep Prior optimization could not resolve the aliasing of the steeper event, while the regularized results are more homogeneous. This effect is evident also in the spectra depicted in the fourth column, where we can notice a residual aliasing pattern; it has been mitigated by regularization.

4 CNN Landmine Detection

Landmines have been massively deployed in a vast amount of countries all over the world. According to the United Nations, the 99.6% of mines and unexploded ordnance must be safely removed from an area to consider it landmine-free. The majority of modern landmines are mainly composed of plastic materials, significantly reducing the efficacy of metal detectors. A suitable alternative is the Ground Penetrating Radar (GPR), as it is capable of detecting even small dielectric changes in the subsurface.

The 2D image $\mathbf{V}(t, x)$ whose coordinates are the reflection time t and the *inline* axis x is called B-scan. By concatenating Y consecutive B-scans, we sample the soil also in the *crossline* axis y , thus obtaining the volume $\mathbf{V}(t, x, y)$. If the y -th B-scan has been acquired over a buried target, we associate the binary label $l(y) = 1$; if it has been acquired over a target-free area, we label it with $l(y) = 0$. The buried

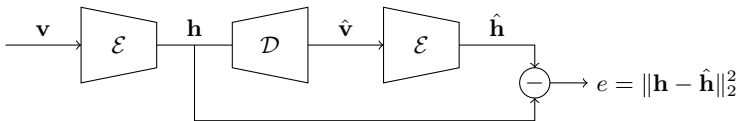


Fig. 5 The proposed anomaly detection scheme

object detection problem we aim at solving consists in taking the volume $\mathbf{V}(t, x, y)$ as input, and producing a label $\hat{l}(y)$ for each B-scan.

Instead of a naive supervised binary classification scheme, let us train a convolutional autoencoder on patches of mine-free B-scans by minimizing the MSE between the input volume \mathbf{v}_i and the autoencoder output $\hat{\mathbf{v}}_i$. Once the training has converged, we can state that the CNN has learned how to reconstruct background soil information correctly. When a volume of soil \mathbf{V} is to be analyzed, it is split into sub-blocks \mathbf{v}_i and the anomaly metrics depicted in Fig. 5 is computed as:

$$e_i = \|\mathcal{E}(\mathbf{v}_i) - \mathcal{E}(\mathcal{D}(\mathcal{E}(\mathbf{v}_i)))\|_2^2. \quad (14)$$

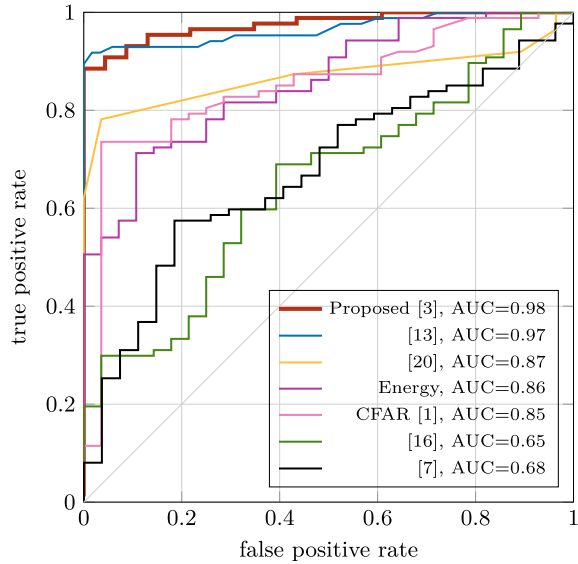
Then, all e_i values are merged into a volumetric anomaly mask \mathbf{E} of the same size of \mathbf{V} . Finally, the label $\hat{l}(y)$ is computed by hard-thresholding the maximum value of $\mathbf{E}(t, x, y)$ with a global value Γ .

4.1 Experimental Results

To compare against other methods proposed in the literature, I consider a simple detector that uses the maximum value of the B-scan energy, a standard constant false alarm rate (CFAR) technique [1], a model-based method [7], a computer vision approach [20] and two CNN-based methods [13, 16]. I evaluated my technique employing Receiver Operating Characteristic (ROC) curve, which represents the probability of correct detection (i.e., correctly finding a threat) and probability of false detection (i.e., detecting objects in clear areas) by spanning all possible values of the threshold Γ . Each point of a ROC curve is obtained by comparing the ground truth $l(y)$ with $\hat{l}(y)$ binarized with respect to threshold Γ . In a real scenario, traces generated by a single landmine can be observed in multiple B-scans. We consider a misdetection every time we miss a single B-scan. Therefore all presented results can then be considered a lower bound on the performance of the proposed method.

As depicted in Fig. 6, the proposed anomaly detector achieves the best Area Under the Curve (AUC) score. It is worth noticing that, along with the method proposed by my colleagues in [13], the proposed autoencoder is able to achieve a $\text{TPR}_{\text{FPR}=0}$ near 0.9, which is highly desirable in a demining system.

Fig. 6 ROC curves of the proposed landmine anomaly detector compared with state-of-the-art solutions



5 Conclusions

Convolutional Neural Networks (CNNs) are an uprising tool to leverage data patterns, especially when acquisition campaigns produce redundant datasets as in Exploration Geophysics. In many physical applications, however, the data features extracted by convolutional layers are not sufficient to describe the complex phenomena that we aim to invert. This consideration suggests that a-priori domain information should also be inserted in machine learning schemes.

In this brief, I presented three different machine learning paradigms applied to imaging, interpolation, and interpretation tasks. First, I study the Generative Adversarial Networks (GANs) as a tool to refine Reverse Time Migration (RTM) images. This supervised network produces very good results but it lacks of physical information on the geological features it processes. Then, I described how Deep Priors could be relevant to seismic data interpolation. While producing state-of-the-art results, this method cannot solve the intrinsic spectral incompleteness of seismic traces, especially in the presence of aliased data, for which I proposed a slope-informed regularization term. Finally, I show how learning patterns can be employed in an anomaly detection scheme for spotting landmines in Ground Penetrating Radar (GPR) data. The learning ability of a CNN is leveraged to discriminate safe background soil images and buried threats signatures. Such a scheme is based on a single class training, thus relaxing the requirements, and costs, of collecting datasets.

In the last years, a considerable number of architectures have been proposed for improving human-level tasks. A general trend is to go for deeper, heavier, and more

complex ensemble models. It is worth noticing that this race does not guarantee better results. Therefore, extra care should be spent studying new problem-setting approaches rather than problem-solving ones.

References

1. Bahadirlar, Y., Sezgin, M.: CA-CFAR detection against K-distributed clutter in GPR. In: *SPIE Detection and Remediation Technologies for Mines and Minelike Targets (2007)*
2. Baysal, E., Kosloff, D.D., Sherwood, J.W.: Reverse time migration. *Geophysics* **48**(11), 1514–1524 (1983)
3. Bestagini, P., Lombardi, F., Lualdi, M., Picetti, F., Tubaro, S.: Landmine detection using autoencoders on multipolarization gpr volumetric data. *IEEE Trans. Geosci. Remote Sens. (TGRS)*, pp. 1–14 (2020). <https://doi.org/10.1109/TGRS.2020.2984951>
4. Claerbout, J.: *Basic Earth Imaging*. Stanford University (2008)
5. Colombo, D., Turkoglu, E., Li, W., Sandoval-Curiel, E., Rovetta, D.: Physics-driven deep-learning inversion with application to transient electromagnetics. *Geophysics* **86**(3), E209–E224 (2021). <https://doi.org/10.1190/geo2020-0760.1>
6. Di, H., Li, C., Smith, S., Li, Z., Abubakar, A.: Imposing interpretational constraints on a seismic interpretation convolutional neural network. *Geophysics* **86**(3), IM63–IM71 (2021). <https://doi.org/10.1190/geo2020-0449.1>
7. Dou, Q., Wei, L., Magee, D.R., Cohn, A.G.: Real-time hyperbola recognition and fitting in GPR data. *IEEE Trans. Geosci. Remote Sens. (TGRS)* **55**(1), 51–62 (2017)
8. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) *Advances in Neural Information Processing Systems*, vol. 27, pp. 2672–2680. Curran Associates, Inc. (2014)
9. Hale, D.: Local dip filtering with directional laplacians. *CWP Report 657* (2007)
10. Ibtchaz, N., Rahman, M.S.: Multiresunet: rethinking the u-net architecture for multimodal biomedical image segmentation. *Neural Netw.* **121**, 74–87 (2020)
11. Jiang, C., Zhang, D., Chen, S.: Lithology identification from well-log curves via neural networks with additional geologic constraint. *Geophysics* **86**(5), IM85–IM100 (2021). <https://doi.org/10.1190/geo2020-0676.1>
12. Kong, F., Picetti, F., Lipari, V., Bestagini, P., Tang, X., Tubaro, S.: Deep prior based unsupervised reconstruction of irregularly sampled seismic data. *IEEE Geosci. Remote Sens. Lett. (GRSL)* (2020). <https://doi.org/10.1109/LGRS.2020.3044455>
13. Lameri, S., Lombardi, F., Bestagini, P., Lualdi, M., Tubaro, S.: Landmine detection from GPR data using convolutional neural networks. In: *European Signal Processing Conference (EUSIPCO)* (2017)
14. Pham, N., Li, W.: Physics-constrained deep learning for ground-roll attenuation. In: *First International Meeting for Applied Geoscience & Energy Expanded Abstracts*, pp. 1335–1339 (2021). <https://doi.org/10.1190/segam2021-3583447.1>
15. Picetti, F., Lipari, V., Bestagini, P., Tubaro, S.: Seismic image processing through the generative adversarial network. *Interpretation* **7**(3), SF15–SF26 (2019). <https://doi.org/10.1190/INT-2018-0232.1>
16. Sakaguchi, R.T., Morton, K.D., Collins, L.M., Torrione, P.A.: Recognizing subsurface target responses in ground penetrating radar data using convolutional neural networks. In: *SPIE Detection and Sensing of Mines, Explosive Objects, and Obscured Targets* (2015)

17. Schuster, G.T.: Least-squares cross-well migration. In: SEG Technical Program Expanded Abstracts 1993, pp. 110–113. Society of Exploration Geophysicists (1993)
18. Symes, W.W.: The seismic reflection inverse problem. *Inverse Probl.* **25**, 123008 (2009). <https://doi.org/10.1088/0266-5611/25/12/123008>
19. Tarantola, A.: *Inverse Problems Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier (1987)
20. Torrione, P.A., Morton, K.D., Sakaguchi, R., Collins, L.M.: Histograms of oriented gradients for landmine detection in ground-penetrating radar data. *IEEE Trans. Geosci. Remote Sens.* **52**(3), 1539–1550 (2014)
21. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2018)
22. van Vliet, L.J., Verbeek, P.W.: Estimators for orientation and anisotropy in digitized images. In: *Advanced School for Computing and Imaging Annual Conference (ASCI)* (1995)
23. Zhang, Y., Sun, J.: Practical issues of reverse time migration: true amplitude gathers, noise removal and harmonic-source encoding. In: *Beijing International Geophysical Conference and Exposition 2009: Beijing 2009 International Geophysical Conference and Exposition, Beijing, China, 24–27 Apr 2009*, pp. 204–204. Society of Exploration Geophysicists (2009)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

