MECHANICAL ENGINEERING | PHYSICS | PRESERVATION OF THE ARCHITECTURAL HERITAGE | STRUCTURAL, SEISMIC AND GEOTECHNICAL ENGINEERING | SCIENCE, TECHNOLOGY, AND POLICY FOR SUSTAINABLE CHANGE | URBAN PLANNING, DESIGN AND POLICY | AEROSPACE ENGINEERING | ARCHITECTURE, BUILT ENVIRONMENT AND CONSTRUCTION ENGINEERING | ARCHITECTURAL, URBAN AND INTERIOR DESIGN | BIOENGINEERING | DATA ANALYTICS AND DECISION SCIENCES | DESIGN | ELECTRICAL ENGINEERING | ENERGY AND NUCLEAR SCIENCE AND TECHNOLOGY | ENVIRONMENTAL AND INFRASTRUCTURE ENGINEERING | INDUSTRIAL CHEMISTRY AND CHEMICAL ENGINEERING | INFORMATION TECHNOLOGY | MANAGEMENT ENGINEERING | MATERIALS ENGINEERING | MATHEMATICAL MODELS AND METHODS IN ENGINEERING

DOCTORAL PROGRAM IN INFORMATION TECHNOLOGY

Chair:
Prof. Luigi Piroddi

Introduction
The Ph.D. programme in Information Technology (Ph.D. IT) covers research topics in four scientific areas, associated to different facets of the field of Information and Communication Technology, namely Computer Science and Engineering, Electronics, Systems and Control, and Telecommunications. This broad variety of research topics perfectly captures the core mission of the corresponding sections of the Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB). New research topics and cross-areas research fields are also covered, such as machine learning, big data, intelligent data analysis, Industry 4.0, internet of things, bioinformatics, quantum computing, ecology, environmental modelling, operations research, and transportation systems.
The Ph.D. IT programme is the largest at the Politecnico di Milano in terms of number of students. Every year, more than 90 new students join the programme, for an overall number of around 400.

Topics
Research at DEIB in the field of Information Technology is supported by 35 laboratories, and is organized in 4 main areas.

- Computer Science and Engineering (Vice-Chair: Prof. Cinzia Cappiello
  Information systems, Database management, Information design for the web, Methods and applications for interactive multimedia, Embedded systems design and design methodologies, Dependable systems, Cybersecurity, Autonomous robotics, Artificial intelligence, Computer vision and image analysis, Machine learning, Dependable evolvable pervasive software engineering, Compiler technology, Natural language processing and accessibility.
- Electronics (Vice-Chair: Prof. Ivan Rech):
  Circuits and systems, Single photon detectors and applications, Radiation detectors and low noise electronics, Electronic circuit design, Electron devices.
- Systems and Control (Vice-Chair: Prof. Simone Garatti):
  Control theory and its applications, Robotics and industrial automation, Dynamics of complex systems, Planning and management of environmental systems, Operations research and discrete optimization.
- Telecommunications (Vice-Chair: Prof. Massimo Tornatore):
  Networking, Applied electromagnetics, Optical communications, Quantum communications, Wireless and space communications, Remote sensing, Signal processing for multimedia and telecommunications.

Industrial collaborations
Due to its intrinsic technological nature, the Ph.D. IT programme features many industrial collaborations. More than 50% of the Ph.D. candidates are funded by companies or by international research projects involving industrial partners. Indeed, the Ph.D. School envisions the collaboration between university and companies as the ideal ground to convert invention and scientific research into technological innovation. Nevertheless, alongside applied research projects in collaboration with industrial partners, the programme is also able to preserve a strong characterization in fundamental research.
To monitor the activities and development of the Ph.D. programme, the Faculty Board cooperates with an industrial Advisory Board, composed by members of public and private companies, working in management, production, and applied research. The two boards jointly meet once a year to identify and suggest new emerging research areas and to foster the visibility of the Ph.D. IT programme in the industrial world.

Educational aspects
The teaching organization and the course subjects reflect the scientific interests of DEIB faculties. The curricula include a wide choice of courses (about 20 per year), and more than 30 courses for basic soft and hard skills offered by the Ph.D. School of the Politecnico di Milano. Access to external courses and summer schools is also encouraged. The challenge is to promote interdisciplinary research while offering advanced help to students to make the best choices according to the regulatory scheme of the programme. Students must undergo a yearly evaluation of the progress in their research and course work.
Internationalization
Every year, several courses are delivered by visiting professors from prestigious foreign universities. Moreover, the Ph.D. IT programme encourages joint curricula with foreign institutions. The programme has several Double Degree and Joint Degree agreements with institutions from countries in all continents. Every year the programme receives more than 150 applications from foreign countries and about 15% of our selected Ph.D. candidates have applied from outside Italy.

Conclusions
The core mission of the Ph.D. IT programme is to offer an excellent doctoral level curriculum, through high-quality courses, a truly interdisciplinary advanced education, cutting-edge research, and international and industrial collaborations.

PHD BOARD OF PROFESSORS	
Amigoni Francesco	Jabali Ola
Ardagna Danilo	Luini Lorenzo
Bertuccio Giuseppe	Maier Guido Alberto
Bestagini Paolo	Mainardi Luca
Bisoffi Andrea	Margara Alessandro
Bonfanti Andrea Giovanni	Mari Lorenzo
Boracchi Giacomo	Martinenghi Davide
Cappiello Cinzia, vice-coordinator of Computer Science and Engineering area	Matera Maristella
Carminati Marco	Monti-Guarnieri Andrea Virgilio
Corno Matteo	Mottola Luca
Dercole Fabio	Parolari Paola
Farina Marcello	Pelosi Gerardo
Ferrari Giorgio	Piroddi Luigi, Coordinator
Filippini Ilario	Rech Ivan, vice-coordinator of Electronics area
Fornaciari William	Restelli Marcello
Garatti Simone, vice-coordinator of Systems and Control area	Santambrogio Marco Domenico
Gatti Nicola	Silvano Cristina
Geraci Angelo	Tornatore Massimo, vice-coordinator of Telecommunications area
Gulinatti Angelo	Zanero Stefano

PHD ADVISORY BOARD	
Mario Caironi	IIT
Paolo Cederle	Everis
Cristina Cremonesi	The European Ambrosetti
Paolo Costa	European Space Agency
Giuseppe Desoli	STMicroelectronics
Alessandro Ferretti	Tre-Altamira
Giuseppe Fogliazza	MCE Srl
Bruno Garavelli	Xnext s.r.l.
Maurizio Griva	Reply SpA
Sabino Illuzzi	Prospera
Renato Lombardi	Huawei Technologies
Renato Marchi	KPMG
Giorgio Parladori	PoliHub (ex Research Program Director, SM Optics srl)
Francesco Prelz	INFN
Enrico Ragaini	ABB S.p.A.
Paolo Giuseppe Ravazzani	CNR
Dario Regazzoni	Amazon Web Services (AWS) Italy
Carlo Sandroni	RSE S.p.A.
Massimo Valla	TIM
Luisa Venturini	Vodafone Italy
Stefano Verzura	Huawei Technologies
Roberto Villa	IBM Italy
Stefano Verzura	Huawei Technologies
Roberto Villa	IBM Italy

Prizes and awards

In 2025 the following awards have been obtained by Ph.D. candidates:

- **Dimitris N. Chorafas Prize** – Bernardo Camajori Tedeschini, Marco Mussi
- **Prof. Florian Daniel PhD Thesis Award** – Bernardo Camajori Tedeschi
- **Springer Award** – Bernardo Camajori Tedeschini, Marco Mussi, Silvia Cascianelli, Diego Morra, Andrea Buffoli, Antonio Consolo, Davide Salvi, Qiaolun Zhang, Riccardo Poiani, Matteo Baldo, Andrea Costa, Riccardo Busetto, Araavind Sridhar
- **Leonardo Fibonacci Best Paper Award** - Giacomo Savazzi
- **I-RIM 3D 2025 Best Paper Award** - Isacco Zappa
- **J3C 2025 Young Author Award** - Christian Cella
- **IAISE Best Paper Award** - Arianna De Vecchi
- **DAFx Second Best Paper Award** - Oliviero Massi
- **14th IFAC Symposium on Robotics Best Paper Award (finalist)** - Giuseppe Fabio Preziosa
- **DEBS Best Poster Award** - Luca De Martini
- **TODAES Rookie Author of the Year Award** - Davide Baroffio
- **Innovation for Nuclear 2025** - Filippo Ghisio, Tommaso Ferri
- **Reconfigurable Architectures Workshop Best Poster Award** - Marco Venere
- **IEEE International Symposium on Circuits and Systems Best Track Manuscript Recognition** – Runner Up - Oliviero Massi
- **TREDI 2025 Best Poster Award** - Iurii Eremeev

EDUCATIONAL TECHNOLOGIES IN THE IMMERSIVE AGE: THE ROLE OF IMMERSION AND CONTROL FACTORS IN VR LEARNING

Paolo Boffi - Supervisor: Pier Luca Lanzi

Virtual Reality has become an increasingly accessible educational technology, offering immersive, interactive, and multisensory experiences that can support the understanding of complex topics, increase motivation, and enrich learner engagement. Yet, the actual educational value of immersion remains debated. Although immersive virtual reality is often assumed to be inherently superior to desktop or non-immersive solutions, the literature still reports contradictory results, and the mechanisms through which immersion affects learning are not fully understood. This dissertation addresses this gap by investigating how different levels of immersion, together with control factors and interaction design choices, influence learning in virtual environments.

The work is grounded in the Cognitive Affective Model of Immersive Learning (CAMIL), which identifies immersion, control factors, and representational fidelity as key technological dimensions shaping presence and agency and ultimately affecting learning outcomes. Starting from this framework, the dissertation first reviews the state of the art in educational virtual reality and shows that the impact of immersion is highly context-dependent, while many empirical studies still lack a strong theoretical grounding. CAMIL is therefore adopted as the conceptual backbone for examining not only whether immersive systems work, but under which conditions, for which tasks, and through which experiential mechanisms. Methodologically, the thesis adopts a horizontal approach across multiple educational contexts rather than focusing on a single topic. This choice allows the role of immersion to be examined in heterogeneous learning scenarios and supports broader reflections on how technological affordances should be matched to pedagogical goals. The research is organized around two main questions: how immersion affects learning outcomes across different domains, and how immersion and control factors jointly influence users' experience and performance in virtual environments. To answer these questions, the dissertation combines literature analysis, experimental evaluation, user-experience assessment, and interaction design.

Four case studies structure the empirical work. The first, Domus Romana, investigates cultural heritage education through a guided virtual reconstruction of an ancient Roman house, implemented in immersive and desktop versions and compared with a slideshow-based control condition. The study shows that both virtual versions are at least as effective as traditional teaching in transmitting basic historical knowledge, while the immersive version is perceived as more engaging and enjoyable. The second case study, Envisioning Corals, addresses environmental awareness and climate change through a virtual embodiment experience in which users inhabit coral-reef creatures and observe the effects of anthropogenic pollution over time. The results indicate that the immersive version elicits stronger embodiment and higher perceived agency, but these advantages do not translate into significantly better immediate factual learning than the desktop version. This finding is relevant because it shows that emotional engagement, embodiment, and learning outcomes do not always vary in parallel. The third case study, Physics Playground, explores physics learning through a virtual educational environment designed around fundamental concepts such as gravity, projectile motion, orbital dynamics, and escape velocity. The project evolves from an initial laboratory study to a more ecological evaluation carried out in high schools. A key contribution of this work is the integration of virtual reality with traditional instruction through video lectures and guided explanations. The results show that, when combined with a theoretical introduction, all conditions can support short-term factual learning. More importantly, the retention test reveals that the immersive condition better preserves knowledge over time than the desktop version, while a simpler but interactive web-based control condition also achieves strong long-term retention. These results suggest that the value of immersive systems emerges most clearly when learning is evaluated beyond the immediate post-test and when virtual reality is considered as part of a broader instructional strategy rather than as a standalone replacement for teaching. The fourth case study, BaggingHook, shifts the perspective from content delivery to interaction design. It investigates how interface-specific control techniques affect performance, usability, cognitive load, and user preference in immersive environments through the design of new target-selection strategies for dense 3D scenes. Although not focused on school content in a narrow sense, this project is important for educational technology because it addresses a fundamental precondition for effective immersive learning: the quality of interaction. The findings show that control strategies substantially shape users' sense of agency and the overall effectiveness of the experience. They also reveal that the most efficient technique is not necessarily the most enjoyable one, highlighting a trade-off that educational designers must consider when selecting interaction methods.

Across these studies, the dissertation shows that immersive virtual reality is effective in supporting factual knowledge acquisition, but not consistently superior to desktop virtual reality in immediate post-test performance. This pattern emerges across different domains, suggesting that a higher level of immersion does not automatically translate into better short-term learning outcomes. At the same time, the results indicate that immersion can play a more relevant role in long-term retention. In particular, learners in immersive conditions tend to preserve factual knowledge better over time than those in desktop settings, although interactive two-dimensional solutions can also be highly effective when they are well designed and cognitively focused. The dissertation therefore argues against deterministic assumptions about immersive media and supports a more nuanced interpretation centered on the interaction among immersion, control, motivation, and cognitive demand. Another important contribution concerns agency. The results show that agency cannot be reduced to the amount of movement performed in a virtual environment. Rather, it depends strongly on the quality, coherence, and naturalness of the control strategies available to the user. In some cases, immersive conditions produce higher agency with less physical movement, suggesting that efficient perceptual access and intuitive interactions may matter more than sheer bodily activity. This insight expands the discussion around CAMIL by showing that technological factors should not be considered in isolation, but as interdependent design variables.

Overall, this dissertation contributes to the understanding of when and how immersive technologies should be used for learning. Its results argue against a simplistic view in which more immersion necessarily means better education. Instead, they support a design-oriented perspective: immersive virtual reality can be highly valuable when it is aligned with the learning task, integrated with appropriate instructional support, and paired with effective control strategies. In this sense, the dissertation provides both empirical evidence and practical guidance for researchers, educators, and developers interested in building meaningful learning experiences in the immersive age.

TOWARD EXPRESSIVE AUTONOMY FOR NEURODIVERSE USERS: A DESIGN TRAJECTORY FROM CREATIVE INTERACTION TO ADAPTIVE END-USER DEVELOPMENT

Giacomo Caslini - Supervisor: Franca Garzotto

Thanks to the growing diffusion of smart objects and other connected devices embedded in everyday artifacts and physical spaces, many environments we live in are now augmented through an interleaving between the cyber and the physical worlds. Interactive Smart Spaces (ISSs) are environments in which IoT devices, smart objects, and software services cooperate and interact with users to support tasks and activities at varying levels of complexity. ISSs differentiate themselves from conventional Smart Spaces by their high degree of interactivity, which places the central focus on the active participation of human actors.

In this dissertation, I focus specifically on people with Neuro-Developmental Disorders (NDD). NDD is an umbrella term used to identify different disorders that can arise during the development period and characterize the concurring deficits in the cognitive, social, and motor spheres, and require interventions continuously tailored to individual needs. Intellectual Disabilities (ID), Attention Deficit and Hyperactivity Disorders (ADHD), and Autism Spectrum Disorders (ASD) are some of the most common examples of NDD. These deficits can deeply affect these individuals and their families since some of them may be unable to complete even simple daily tasks, making them strongly dependent on others for everyday functioning.

This thesis explores the interplay between two approaches to and with ISSs, generating two lines of research. On the one hand, it examines the appropriateness and effectiveness of external multisensory, interactive interventions. On the other hand, it investigates how to foster the emancipation and agency of users through the design and development of their own technologies.

The overarching goal of this dissertation is to investigate how ISSs can support individuals with NDD. These two lines of research translate into two corresponding research questions.

Previous work in this area suggests that multisensory interventions are effective in promoting well-being among people with NDD. First, building on this foundation, it is important to evaluate whether the new specific intervention presented in the thesis can positively contribute to the well-being of people with NDD, impacting specific dimensions, such as social communication and emotional regulation. Second, it is important to understand whether people with NDD can become active creators of interactive technologies by engaging in end-user development and making-oriented activities.

The first part of the thesis focuses on the development of an interactive activity for people with NDDs: MusicTraces. It is an open-ended music-and-painting activity developed for the Magic Room, a multisensory ISS. The project aims to support social interaction, emotional regulation, and well-being in adults with NDDs by allowing participants to make music and paint simultaneously in a flexible, customizable setup.

MusicTraces was evaluated with two user studies and a total of 18 adults with moderate-to-severe NDDs. Results indicate that MusicTraces positively supported social communication, emotional regulation, and well-being. In the comparative study, outcomes were comparable to or exceeded those of the control condition (traditional art therapy with music). Caregivers were frequently surprised by the high levels of engagement, even among participants with severe ID and/or motor impairments. Several participants interacted socially more than they typically do in daily life. The open-ended design of the activity also accommodated diverse self-regulation strategies and personal preferences, and sessions frequently exceeded participants' usual attention spans.

Fig.1 - Participants exploring the MakeNodes toolkit during a workshop.

The second part of the thesis introduces two separate research projects. MakeNodes is a tangible toolkit designed for people with ID. It consists of a set of sensing and actuating nodes that can be placed in the environment to augment physical objects and spaces with smart capabilities. The toolkit was evaluated in three workshops with a total of 12 adults with ID in a day-care center, showing a high level of personal satisfaction with the production and appreciation of the toolkit. Caregivers were also interested in integrating the toolkit into the day-care center activities. The evaluation also enabled a comparison of two interaction modalities and contributed a set of design guidelines for people with ID.

Building on the experience with MakeNodes, a second tool, Smartifier, is introduced. It uses two LLM-based chatbots to support users in designing and implementing Interactive Smart Spaces, both in terms of enabling smart objects and system behavior. Through conversational assistance that adapts to user input, Smartifier aims to provide richer and more flexible support while preserving a high degree of expressive freedom. The tool was evaluated in a preliminary study involving 17 neurotypical participants.

The results indicated a very high usability score (SUS) of 84.26, a recommendation score (NPS) of 64.71, and high correctness scores (over 91% for all tasks). A thematic analysis of conversation logs also contributed to a set of deeper insights and some design guidelines.

This thesis overall shows that Interactive Smart Spaces can meaningfully enhance well-being in people with NDDs through carefully designed multisensory activities, and that these individuals can also take an active role as designers of their own technological environments, given the appropriate support. The findings contribute both practical tools and methodological insights for the design of inclusive, user-developed ISSs. Future research will extend these interventions into new media, such as extended reality, increase the expressive capabilities of the toolkits, and assess their long-term impact in real-life contexts.

UNDERSTANDING LEARNING ENGAGEMENT IN VIDEO-BASED LEARNING: HOW TO MONITOR, ASSESS, AND ENHANCE

Giacomo Cassano - Supervisor: Nicoletta Di Blas

VIDEO-Based Learning (VBL) is becoming an increasingly popular approach in education due to its flexibility and accessibility. It allows students to engage with learning materials at their own pace and from any location, making education more inclusive and adaptable. However, while VBL offers these significant advantages, it also presents challenges. On one side, the lack of real-time interaction can make it difficult for instructors to monitor their students' engagement; on the other, VBL often leads to passive consumption of content, making it harder for the students to stay motivated. This research explores how technology can address these challenges, helping to better monitor and enhance engagement in VBL through both pedagogical and technological innovations.

Firstly, it builds on the theoretical background of the research by conducting a systematic literature review that identifies key indicators and facilitators of engagement in VBL. Then, with the knowledge obtained, two systems are developed and validated: **Evoli**, a *Video Annotation Tool (VAT)* that collects self-reported feedback from students and video *Learning Analytics (LA)* to make the teacher aware of students' engagement during video learning, and **SmartQuiz**, a *Large Language Model (LLM)-based framework* designed to generate high-quality in-video *Multiple Choice Questions (MCQs)*, recognized in the review as a key facilitator for engagement.

Evoli enhances instructor awareness by collecting self-reported feedback and LA on student interactions with video content. A study demonstrated the tool's good usability and its perceived usefulness by students and instructors. A second study highlights how a VAT like Evoli can independently influence student engagement, showing a positive impact on the emotional and agentive dimensions.

Finally, a study examined whether explicit feedback from students or Learning Analytics (LA) was a better predictor of performance, revealing that LA, especially video interaction data, is a more accurate predictor of achievement.

Recognizing the role of MCQs as a key facilitator of learning engagement in VBL, **SmartQuiz** tackles the time-consuming process of MCQ creation by utilizing LLMs to generate pedagogically sound questions aligned with best practices. Following the development of the system, an initial study assessed the quality of AI-generated MCQs, confirming their compliance with established guidelines. A subsequent study compared the impact of SmartQuiz-generated MCQs with those authored by instructors, revealing similar student learning outcomes.

These findings suggest that AI can generate high-quality MCQs, thereby reducing instructor workload without compromising quality. This thesis demonstrates how, by utilizing annotation tools enhanced with MCQs, it is possible to create a VBL learning experience that not only increases instructor awareness but also enriches and enhances student engagement.

Future research will investigate the impact of these tools on teaching and learning processes, examining how they influence student engagement, instructor effectiveness, and overall learning outcomes. Additionally, efforts will be made to further refine and enhance the technological capabilities of these tools, informed by the insights gained through ongoing evaluation and user feedback.

PROGRAMMABLE PHOTONIC CIRCUITS FOR ANALOG COMPUTING

Gabriele Cavicchioli - Supervisor: Andrea Melloni
Co-Supervisor: Francesco Morichetti - Tutor: Carlo Riva

In recent years, there has been a growing demand for high-speed computational units capable of performing mathematical operations with minimal latency. This demand arises from the widespread adoption of AI-driven tools, as well as from fields where low-latency computations are critical, such as complex telecommunication networks and avionics. Consequently, the performance requirements for computational units have become increasingly stringent, while conventional electronic circuits struggle to meet these demands. As a result, researchers have begun exploring novel processor architectures that do not necessarily rely on the traditional digital electronic paradigm. In this context, programmable photonic integrated circuits (PICs) have emerged as a promising solution for accelerating linear algebra operations, which constitute a significant portion of the computational workload in many applications. Several implementations of programmable PICs designed for matrix-vector multiplications have been proposed and demonstrated, with notable applications in signal processing and artificial intelligence. However, the broader adoption of these circuits for solving more complex problems remains limited due to the bottleneck introduced by electro-optical conversions.

This thesis tries to expand the number of mathematical problems which can be solved by PICs without electro-optical conversion and investigates the use of coherent programmable photonic integrated circuits in addressing linear inverse problems, such as solving linear systems of equations and matrix inversion, all-optically. Specifically, we present the design of a recursive programmable optical processor (POP) based on a Mach-Zehnder interferometer (MZI) mesh embedded within a feedback loop (Fig. 1), capable of tackling these challenges. The inclusion of a feedback loop enables the optical implementation of iterative solution schemes for linear systems, analogous to well-established methods such as the Jacobi iteration. In this framework, each completion of a round-trip by the light within the feedback loop corresponds to one step of the iterative solution. The time required for this step is equivalent to the light's round-trip duration, which is on the order of a fraction of a nanosecond, making this kind of POP far faster than a digital processor executing the same algorithm.

This thesis explores the capabilities of this architecture both theoretically and experimentally. A prototype was fabricated using commercial-grade silicon photonics technology, assembled, and tested. This prototype represents one of the first demonstrations of matrix inversion performed entirely in the optical domain, achieving an average accuracy of 84% for matrix inversion problems.

Furthermore, we demonstrate that the same architecture can function as a programmable multiple-input multiple-output (MIMO) infinite impulse response (IIR) filter, with a frequency response tailored for specific applications. This tunable spectral response can be leveraged in various applications, including dynamic signal equalization in optical networks or the optical-domain solution of systems of ordinary differential equations, as illustrated through numerical simulations.

Additionally, this thesis addresses the challenges associated with calibrating and accurately programming MZI meshes in the presence of thermal crosstalk effects. To mitigate these challenges, a data-driven control system (Fig. 2) is proposed. This system employs a deep learning model trained to pre-compensate the working points of MZI mesh phase shifters, thereby compensating for crosstalk effects or deviations from nominal circuit behavior. The deep learning model is based on a fully connected artificial neural network trained using a supervised learning approach. Emphasis was placed on analyzing the requirements of the training dataset, including the number of data points and the impact of noise present in the circuit. Through synthetic dataset simulations, it was shown that satisfactory performance can be achieved with approximately one million data points, which can be sampled from a real device in about five minutes. Moreover, the noise introduced by standard test and measurement equipment was found to have a negligible impact on system performance, eliminating the need for specialized low-noise electronic instrumentation to drive the PIC. The system also proved scalable to MZI meshes of varying sizes and architectures.

Finally, an experimental demonstration was conducted using a silicon photonics 3×3 triangular mesh. A training dataset was sampled from the chip, and the artificial neural network was trained and subsequently tested online, achieving a relative mean squared error of 2%. This result demonstrated the potential of such data-driven approaches in the control of programmable photonic circuits.

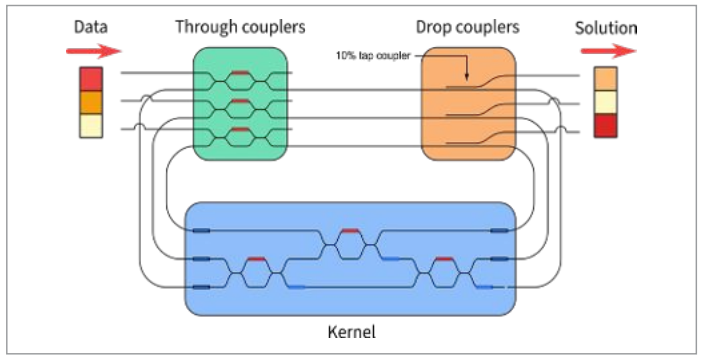

Fig.1 - Schematic representation of a recursive programmable optical processor (POP) based on Mach-Zehnder interferometers. The optical signals represent the data and the solution of a linear system of equations.

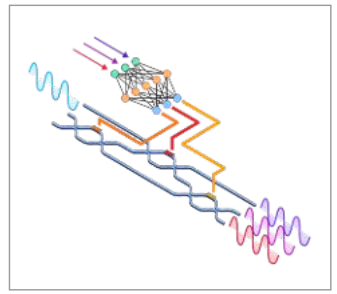

Fig.2 - Conceptual representation of the proposed Data-driven control system for feed-forward programmable photonic circuits

ADVANCED PERCEPTION METHODOLOGIES FOR OFF-HIGHWAY AUTONOMOUS VEHICLES

Marcello Cellina - Supervisor: Matteo Corno
Co-Supervisor: Sergio Matteo Savaresi

Autonomous Vehicles operating in Off-Highway scenarios provide demanding yet controlled settings to evaluate full-stack autonomy, from high-speed multi-vehicle interaction to navigation in unknown and unstructured terrain.
This thesis presents and validates novel algorithmic methodologies for three research problems concerning sensing methodologies for AVs: Vehicle Detection and Tracking, Navigable Area Estimation, and LiDAR-based Mapping and Localization, across two complementary vehicle platforms, namely the Dallara AV-21 for autonomous racing and the Autonomous LMV for off-road navigation.
These methods are implemented on the edge computers, operate in real time, and have been validated experimentally in track and field trials, achieving fully autonomous overtakes up to 275 km/h over different racetracks, and completing multi-kilometer autonomous navigation, exploration and return-to-home missions in off-road proving grounds.

Fig.1 -The two vehicles used in this thesis work

INCREASING DEXTERITY OF ROBOTIC MANIPULATION TASKS IN UNSTRUCTURED INDUSTRIAL ENVIRONMENTS

Matteo Colombo - Supervisor: Paolo Rocco

In the last decades, robotic arms have revolutionized the automation of industrial manufacturing processes, dramatically enhancing their productivity. Robots have been used in several applications, such as palletizing, welding, painting, and machine tending. Nevertheless, all these operations are usually carried out in well-known and organised workstations in a continuous and repetitive way. The robots are pre-programmed to follow a fixed set of instructions, which is repeated several times to complete the task at hand. This approach translates to a complete lack of dexterity and adaptability, and needs to be evolved in order to answer the new demands for flexible and quickly reconfigurable robotic systems brought forth by the latest industrial paradigm shifts. While the introduction of collaborative robots marked a significant step towards expanding automation capabilities, their effectiveness in dynamic and unstructured scenarios remains limited by insufficient dexterous skills, which are critical in tasks such as kitting, packaging, or random bin picking. These tasks are characterized by the presence of uncertainty when planning for object manipulation, which needs to be accounted for. Grasp poses are often random; different placement poses for the same object may be required during the same task, and the grasp target could be potentially occluded by other objects. Furthermore, as the robotic system should be flexible and readily reconfigurable, it should not excessively rely on external fixtures.

This thesis addresses the challenge of enhancing robotic manipulation dexterity in unstructured industrial environments using simple, industry-ready hardware such as parallel grippers. We propose a set of model-based and learning-based control strategies that enable flexible object reorientation, regrasping, and retrieval tasks without external fixtures or specialized multi-fingered hands. We start by focusing on the problem of arbitrarily reconfiguring an object from a random pose to a desired one. To address it, we first introduce a dual-arm manipulation planning framework that performs fixtureless object reorientation through optimized handover sequences. We leverage object data and gripper parameters to generate offline a compact regrasp graph and sets of feasible placement poses. The graph representation, combined with an online optimal handover planner, allows efficient planning and execution of long-horizon manipulation sequences, which are composed of a concatenated sequence of motion primitives. Then, we present a multimodal approach that integrates vision and tactile feedback to plan and supervise in-hand object reorientation using a dual-arm robot, merging open-loop trajectory planning with tactile-based supervision. Vision-based data are used by a path planner to generate a feasible sliding path on grasped surface, along which the second arm can push the object held by the first one, in order to change its in-hand configuration without opening the gripper. This sliding motion is supervised by a tactile classifier that learns from RGB and point cloud data if the current sliding motion is feasible. If it is not, the operation is halted and a replanning step is executed. Finally, a deep reinforcement learning framework is proposed for object retrieval in cluttered environments, where complementary pushing and grasping policies are learned to coordinate exploration and target extraction in a push-to-grasp approach. This enables a robotic arm to manipulate the clutter in its workspace to reveal and grasp a target object. The push and grasp actions are generated by three deep neural networks, concurrently trained through a Deep Q-Network approach in simulated scenes. Two kinds of push actions are learned: find and retrieve. find is rewarded for revealing the object and for removing occlusions on top of it, while retrieve is rewarded for increasing the grasp score on the target, assigned by the grasp network after each executed push. Action selection is coordinated by a vision module that segments the scenes and generates a region of interest where the push actions are most likely to be effective once the object is revealed.

All methods have been validated in simulation and real-world experiments on collaborative robotic platforms to show their applicability in realistic scenarios.

LEARNING-BASED OPTIMAL MANAGEMENT AND CONTROL FOR SUSTAINABLE ENERGY SYSTEMS WITH STORAGE CAPABILITIES

Andres Felipe Cordoba Pacheco - Supervisor: Fredy Ruiz

Tutor: Andrea Maria Zanchettin

The growing decentralization of energy systems and the integration of distributed energy resources have introduced new challenges and opportunities for microgrid management. This thesis addresses these complexities by introducing a comprehensive hierarchical control framework that combines day-ahead dispatch, intra-day control, multi-objective optimization, and data-driven approaches, with a focus on energy storage systems, flexible loads, grid stability, and optimal performance both technically and economically.

This approach is based on a multi-objective day-ahead dispatch strategy for microgrids that balances operational costs, including the battery aging cost, and employs both centralized and distributed Alternating Direction Method of Multipliers methods. The strategy optimizes energy dispatch while incorporating battery degradation, offering valuable insights into the long-term economic viability of battery energy storage systems in volatile market conditions.

Building on this, an intra-day hierarchical control is proposed incorporating a Scenario-Based MPC approach together with a deterministic MPC for intra-day management, where it effectively manages the uncertainties from non-controllable distributed energy resources improving cost efficiency and power control.

In order to design the low-level controllers, data-driven methods were employed to enhance system flexibility and performance. Using the adaptability and effectiveness in improving the behavior of the system with the Set-Membership Data-Driven approach, the methodology developed was extended from Single Input Single Output (SISO) to Multiple Input Multiple Output (MIMO) systems. Furthermore, Data-Driven Control strategies were successfully validated in a real-world application of Battery Energy Storage Systems (BESS). Additionally, flexible loads, specifically thermoelectric refrigerators, were controlled using these strategies to provide frequency containment services.

The findings of this thesis emphasize the transformative potential of data-based approaches in enhancing the efficiency, reliability, and flexibility of microgrid operations, contributing significantly to the future of microgrids composed of distributed energy systems.

CHANGE DETECTION FOR PHASE ESTIMATION IN LARGE INSAR DATASETS

Giovanni Costa - Supervisor: Andrea Virgilio Monti Guarnieri

Synthetic Aperture Radar (SAR) and SAR Interferometry (InSAR) have fundamentally transformed the capacity to monitor surface deformation and infrastructure stability with millimetric precision. However, a persistent bottleneck in current processing pipelines remains the inherent vulnerability to temporal surface changes.
When a target undergoes a physical or dielectric transformation, the resulting signal decorrelation often leads to its exclusion from the analysis or to erroneous displacement estimations. This research addresses this critical gap by developing a comprehensive "change-conscious" InSAR framework, shifting the paradigm from treating change as a noise source to utilizing it as a vital diagnostic tool.
The core of this work lies in the development of upstream change detection (CD) algorithms that operate prior to or during the deformation analysis, providing a resilient foundation for time-series interpretation. While traditional approaches are often divided into amplitude-based (ACD) or coherence-based (CCD) methods, they frequently fail in real-world scenarios due to an over-reliance on rigid statistical models or the need for "steady-state" calibration datasets, which are rarely available in dynamic environments.

To overcome these limitations, the research first establishes a rigorous interferometric formalization of change phenomena. By analyzing the unique signatures within the coherence matrix (as illustrated in Fig. 1), a clear taxonomy is defined between:
Definitive Change (DC): Abrupt and permanent topographic perturbations, such as new construction or landslides, which result in a distinct block structure in the coherence matrix (Fig. 1a).
Temporary Change (TC): Transient phenomena, such as seasonal vegetation or meteorological events, that affect the signal temporarily without permanently altering the target's scattering properties, manifesting as localized decorrelation "stripes" (Fig. 1b).
Building on this theoretical foundation, the thesis introduces three major algorithmic contributions. First, the Permutational Change Detection (PCD) is proposed as a pioneering step toward fully non-parametric CCD. By leveraging Permutation Tests, this algorithm specializes in the detection of Definitive Changes without requiring the strong Gaussian assumptions that limit traditional detectors, making it robust across diverse land-cover types.
Second, the Band-limited Uncalibrated Detector (BUD) represents a significant advancement in non-parametric detection. It is designed to infer changes, regardless of their nature, by directly testing the multi-temporal evolution of observed coherence samples, without the need for calibration datasets, a major hurdle in operational SAR monitoring. Furthermore, BUD transitions the output from traditional binary maps to quantitative information layers, allowing for a nuanced, multi-temporal analysis of surface dynamics. A prime example is the Object Counter (Fig. 2), which identifies and counts the number of stable temporal "objects" (i.e., distinct coherent periods) within a target's coherence matrix. This metric provides a spatial proxy for the target's temporal reliability, effectively mapping the complexity of the change phenomena across the scene.
Finally, DANI-NET (Deep Analysis for Non-stable InSAR targets Network) represents the first effective Deep Learning framework capable of simultaneously inferring both the nature (DC vs. TC) and the spatio-temporal location of a change.
While the SAR community has traditionally viewed amplitude and coherence as orthogonal sources, DANI-NET unifies them into a single, cohesive multidimensional feature space. To solve the lack of labeled SAR data, a fully synthetic-based training strategy was designed. This dataset is a physically-grounded model of how changes manifest in the complex SAR signal, ensuring that DANI-NET provides a transparent, interpretable, and explainable method.
The robustness of PCD, BUD, and DANI-NET has been rigorously demonstrated through comparative analyses against state-of-the-art methods. Beyond synthetic validation, the frameworks were applied to real-world datasets ranging from high-rise urban infrastructures to rural areas prone to natural hazards. The results confirm that a change-conscious approach significantly enhances the reliability of InSAR products, offering the SAR community a physically-consistent toolkit for monitoring applications.

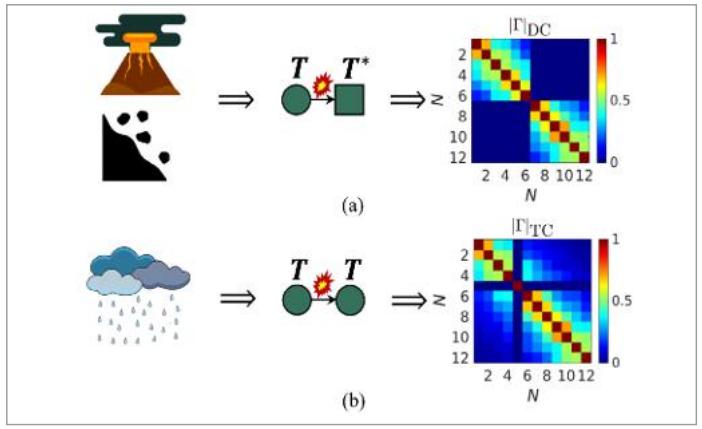

Fig.1 - Comparison of coherence matrix signatures for different change phenomena: (a) Definitive Change (DC), illustrated by abrupt events such as volcanic activity or landslides, resulting in a persistent block-diagonal decorrelation pattern in the coherence matrix ; (b) Temporary Change (TC), such as rainfall or transient meteorological events, characterized by a localized decorrelation "stripe" in $|\Gamma|_{TC}$ that does not permanently alter the target's long-term stability

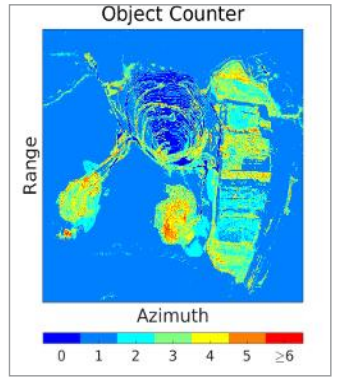

Fig.2 - Example of a quantitative information layer generated by the BUD algorithm. The Object Counter map displays the number of distinct coherent temporal periods identified for each pixel. Lower values indicate stable targets, while higher values pinpoint highly dynamic areas, such as active mining sectors, requiring a "change-conscious" InSAR framework

ETHICS IN DATA SCIENCE: EXPLORATION OF FAIRNESS IN DATA-DRIVEN DECISION SYSTEMS

Chiara Criscuolo - Supervisor: Davide Martinenghi

This dissertation provides a comprehensive investigation of fairness in data-driven decision-making systems, a topic of growing importance as machine learning (ML) models increasingly support or automate decisions in socially sensitive domains such as finance, healthcare, hiring, and recommendation platforms. While these systems promise efficiency and accuracy, they may also introduce or amplify biases embedded in historical data, system design, or organizational processes. Fairness, however, is neither universally defined nor straightforward to operationalize: it involves a complex interplay of statistical, causal, and contextual factors that influence how decisions are produced and interpreted. This thesis addresses these challenges by proposing conceptual foundations, methodological tools, and practical frameworks to support the identification, measurement, mitigation, and communication of fairness throughout the lifecycle of ML systems.

The first research question concerns how fairness can be rigorously defined and measured. The thesis develops a unified taxonomy of fairness metrics that spans both statistical and causal definitions. Statistical metrics—such as group fairness, equal opportunity, equalized odds, and conditional use accuracy—compare predictions across subgroups and highlight disparities in model behavior. Causal metrics provide deeper insight by examining the underlying generative mechanisms and distinguishing between fair and unfair causal pathways. This taxonomy clarifies assumptions, identifies incompatibilities between metrics, and shows why no single definition is universally appropriate. To support practitioners in navigating these alternatives, the thesis introduces the *Fairness Decision Tree*, a structured, question-driven framework that guides users in selecting the most suitable metric by considering domain constraints, reliance on historical decisions, tolerance for different types of errors, and the presence of causal dependencies. Since fairness assessments are often poorly understood by non-expert stakeholders, the thesis also presents *FairnessFriend*, an interpretability-oriented chatbot that explains fairness metrics, their implications, and the reasoning behind metric selection. By integrating human-centered design principles with automated explanations, *FairnessFriend* helps bridge the gap between technical fairness analysis and stakeholder comprehension.

The second research question focuses on whether and how unfairness can be mitigated in practice. The thesis compiles a detailed catalog of mitigation techniques, grouped into pre-processing, in-processing, and post-processing approaches. Pre-processing methods modify the training data to reduce bias, in-processing methods adjust the learning algorithm to enforce fairness constraints, and post-processing methods correct predictions after training. While many techniques exist, their effectiveness depends heavily on model choice, dataset characteristics, and the fairness definition being pursued. To address this complexity, the thesis introduces *FAIR-CARE* (Fig. 1), a scalable and extensible evaluation pipeline for benchmarking mitigation strategies. *FAIR-CARE* systematically applies mitigation techniques across multiple datasets, ML models, fairness metrics, and performance indicators, enabling a rigorous comparison of their trade-offs. Applied to seven real-world healthcare datasets—where fairness concerns are particularly critical due to the high-stakes nature of clinical decisions—the pipeline reveals that no mitigation technique performs uniformly well. Instead, effectiveness varies with data distribution (balanced vs. skewed), protected attribute, and type of error prioritized. The experiments also highlight that improvements in fairness often come at the expense of performance metrics such as accuracy or precision, and that post-processing techniques, while flexible, may not consistently achieve the same gains as targeted pre- or in-processing approaches. To support broader accessibility, the thesis introduces *LLM4Fairness*, a conversational agent powered by large language models that explains mitigation results in plain language and helps users understand fairness-performance trade-offs.

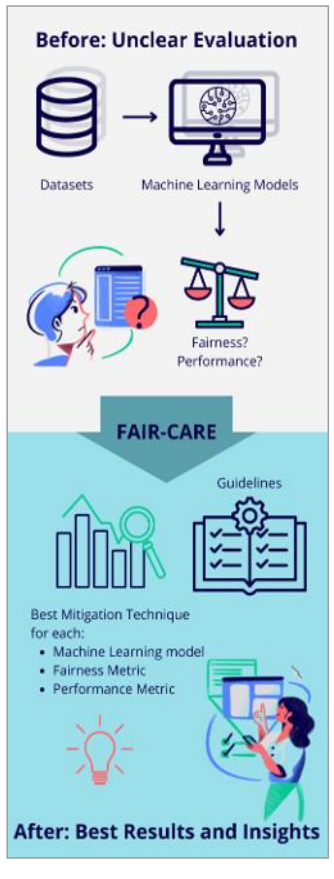

Fig.1 - FAIR-CARE objectives

The third research question extends the study of fairness to other requirements and decision-making contexts. The thesis explores the relationship between fairness and data quality, showing that data imperfections such as missing values, inconsistencies, or sampling imbalances not only hinder model accuracy but also distort fairness assessments. It proposes methodological guidelines to jointly evaluate data quality interventions and fairness outcomes, emphasizing that fairness cannot be treated in isolation from the broader data lifecycle. The thesis further examines fairness in ranking systems by analyzing intersectionality, a paradigm that acknowledges how multiple, overlapping protected attributes can create compounded forms of disadvantage. It offers a structured tutorial of intersectional fairness methods and highlights challenges such as data sparsity and conflicting constraints. Finally, the thesis introduces *SOCRATE*, a novel framework designed to ensure fairness in recommender systems with limited item availability. Unlike traditional fairness approaches that focus on static recommendations, *SOCRATE* compensates users over time, ensuring that temporary scarcity does not lead to systematic disadvantage across sessions.

Taken together, these contributions form an integrated approach to fairness in data-driven systems. By combining theoretical clarity, methodological rigor, and user-centered tools, the thesis supports the development of ML systems that are not only accurate and efficient but also transparent, equitable, and aligned with emerging regulatory frameworks such as the EU AI Act. The proposed tools—*FairnessFriend, FAIR-CARE, and LLM4Fairness*—translate complex fairness concepts into accessible explanations for diverse stakeholders, promoting informed decision-making and enhancing trust in automated systems. The thesis ultimately demonstrates that fairness is not a single, static requirement but a dynamic property shaped by context, design choices, and system constraints, and that addressing it requires a holistic perspective across the entire ML pipeline.

MODELLING AND CONTROL OF INNOVATIVE OXY-TURBINE POWER GENERATION SYSTEMS

Matteo De Pascali - Supervisor: Francesco Casella

Until renewable energy sources will not be able to cover the entire energy demand to achieve a fully green energy sector, balancing electrical power systems will be a challenging task.
To fill in the gaps left by the intermittent renewable sources, innovative, low- or zero-emissions thermal power generation units can be employed in the next decades to replace outdated and pollutant power plants.Among these new solutions, the ones resorting to carbon capture utilization and storage (CCUS) technologies are thoroughly studied by researchers and have attracted industry interest for their high efficiency and decreasing investment costs.
The objective of this work is the study of innovative oxy-turbine power generation systems and in particular of the Solid-Oxide Semi-Closed CO2 (SOS-CO2) cycle and the assessment of its flexibility to operate in the energy scenarios described above. This concept unit resorts to the oxy-combustion CCUS technology, a technique to easily obtain a pure stream of CO2 to be stored after the fuel combustion. By operating the combustion with pure O2 instead of air, which contains N2, the only products of the combustion process are CO2 and H2O, that can be easily separated through water condensation. The SOS-CO2 cycle, thus, features an oxy-combustion Brayton cycle connected to a Solid-Oxide Fuel Cell (SOFC) module, which efficiently oxidizes about 70% of the fuel before it is sent to the oxy-combustor.
As its name suggests, the SOS-CO2 cycle operates with peculiar working fluids, i.e., a pure CO2 stream recycled from the turbine exhaust and a CO2 and O2 mixture, which replaces air in the cathode channels of the SOFC.
The efficiency and the flexibility of such an innovative system should thus be assessed before the power plant commissioning. To perform such analysis, the object-oriented and equation-based Modelica language is used to code a library of components to assemble the SOS-CO2 cycle model. The models of the components, which are employed to perform plant-wide studies, are described, with particular attention on the strategies employed to complete the plant steady-state initialization successfully.
Since the phenomena occurring in the power plant are described by highly nonlinear equations, the initialization of the model simulation is a difficult and critical task, prone to convergence failures if not properly addressed. The modeling techniques employed aim at simplifying and decoupling the solution of the huge original nonlinear system of equations of the initialization problem and provide ways to consistently perform the initialization of this class of power plants. Moreover, specific utility blocks for object-oriented and equation-based models are presented to directly initialize the plant in off-design conditions and employ the model for linearization, open-loop and closed-loop simulations.
Once the power plant model is ready for simulation, the characterization of off-design policies to control efficiently the SOS-CO2 cycle during partial load operations is addressed. Taking as inspiration the conventional off-design strategies for thermal power generation systems, four strategies were defined, each one with the aim of maximizing the efficiency of the power plant while respecting the technological limitations of its components.
The first two strategies limit the fuel cell usage, while the last ones shift remarkably the power production towards this component.
The partial load trajectories formulated are employed as reference signals for a suitably tuned control system. First, a novel algorithm to support the formulation of reduced-order models for thermal power generation systems is presented, together with the procedure adopted for the synthesis of a model-based controller.
Second, a preliminary analysis on the coupling between the input and output variables of the plant reveals that the power block of the SOS-CO2 cycle should be controlled with a centralized control system to handle the input-output coupling, while quantities related to the technological limitations of the plant (e.g. maximum SOFC inlet temperatures, minimum pressure of the cycle...) can be controlled with decentralized PI controllers. Finally, following the indications provided by the developed algorithm, a reduced-order model of the power block is built to capture the main dynamic characterization of the considered plant portion. The reduced-order model is employed to build the control architecture and simulations are performed to track fast load variation ramps and simulate a rapid intervention of the plant to balance the electrical grid.
As final topic of the dissertation, optimization tasks are considered. Since Modelica tools do not offer open-source solutions to code optimization problems involving Modelica models, a compiler from Modelica to Pyomo language is coded to automatically translate simulation problems into optimization ones. The compiler takes care of scaling properly the generated models and performs code manipulations to substitute difficult to handle expressions like logarithms. As test case, the fuel cell model included in the SOS-CO2 cycle is converted into a Pyomo model and static and dynamic optimization routines are performed. First, parameter identification problems are addressed and finally open-loop dynamic optimal control problems are solved while enforcing constraints on the variation of critical quantities like the temperatures of the fuel cell solid layers.

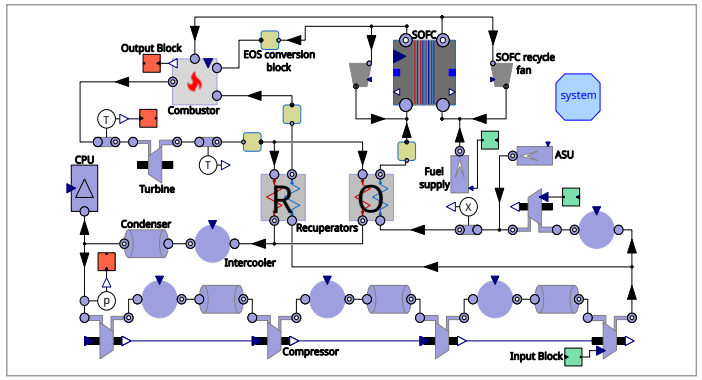

Fig.1 - Diagram of the SOS-CO2 cycle in Modelica with initialization blocks pairs connected to some of the input (green blocks) and outputs (orange blocks) of the system.

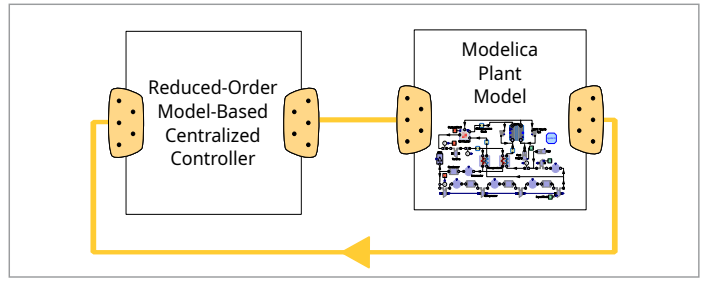

Fig.2 -The plant and control diagram.

OPTIMIZATION FRAMEWORKS FOR TWIN-IN-THE-LOOP SYSTEMS IN CONTROL AND ESTIMATION

Giacomo Delcaro - Supervisor: Sergio Matteo Savaresi

Simulation models have moved from offline design tools to active components capable of supporting estimation, control, and online adaptation. This PhD investigates how high-fidelity digital twins can be embedded directly inside closed-loop algorithms, introducing a unified Twin-in-the-Loop (TiL) perspective. In TiL schemes, the virtual plant runs in parallel with the real system and becomes part of the decision-making loop: it can predict unmeasured states, generate nominal control actions, or provide informative virtual experiments for parameter tuning. The central challenge is that such twins are typically accurate but analytically intractable black boxes, which makes classical observer and controller design tools difficult to apply. The thesis therefore focuses on data-efficient optimization methods that make TiL architectures practical.
The first contribution concerns TiL observers for vehicle dynamics estimation. A digital twin is used as predictor inside the state observer, while measured outputs correct the virtual states. The observer calibration process is then cast as a large-scale zeroth-order optimization problem. The thesis studies dimensionality-reduction strategies to handle large numbers of tuning parameters and then develops a high-dimensional Bayesian optimization approach based on gradient steps and line searches that improves robustness and reduces calibration effort. These methods allow a unified state estimator to exploit rich nonlinear vehicle models while remaining deployable in real time.
The second contribution concerns TiL control. Here a nominal controller is closed on the digital twin, while an additional compensator corrects the mismatch between simulated and real dynamics. This idea is studied for lateral vehicle control and active suspensions, showing that high-fidelity virtual models can be used not only for simulation but also as real-time control assets. The work also clarifies how different sim-to-real perturbation directions affect compensator design and proposes robust tuning strategies when the nominal twin is not perfectly favorable.
The final part of the thesis introduces an Adaptive Twin-in-the-Loop framework for controller tuning. Several randomized digital twins run in parallel with the real plant and are queried together by a novel multi-source Bayesian optimization algorithm. A symmetric multi-output Gaussian-process model learns correlations between real and virtual data, while a dedicated acquisition strategy selects the most informative real and virtual experiments. Benchmark studies, vehicle dynamics applications, and experimental motor tests show faster convergence and improved data efficiency than standard Bayesian optimization. Overall, the thesis turns digital models from passive validation models into active closed-loop components and provides practical tools for future adaptive cyber-physical systems.

Fig.1 - Real plant and digital twin in a closed-loop learning and optimization interaction

ON DATA & DECISION-MAKING: PERSPECTIVES FROM HEALTHCARE APPLICATIONS CONSIDERING REPRESENTATIVENESS AND ANALYTICAL SCOPE

Martina Doneda - Supervisor: Giuliana Carello
Co-Supervisor: Ettore Lanzarone

This thesis consists of a commented collection of nine papers, organized in three parts, which explore several possible interactions between data science and operations research (OR) from the perspective of decision-making applied to the healthcare sector, populating a theoretical framework of analysis.
The rationale behind this research work stems from the fact that the volume of data produced worldwide has grown exponentially – less than half a zettabyte between the invention of writing and 2006, 180 zettabytes between 2006 and 2011. However, not all data are useful: a substantial portion is irrelevant or biased, making it difficult to identify what information is actually valuable. As decision-making processes increasingly rely on data-driven approaches, a need emerges for structured methods that can transform data into actionable insights.
OR is the science of decision-making by definition, and data science is the branch of knowledge that enables the use of data: their integration is at the same time natural and increasingly important in complex decision environments. Moreover, the two disciplines are closely related, since many data methods, in particular machine learning (ML) rely on optimization techniques that are central to OR. Healthcare is a particularly relevant domain to test hybrid approaches. Healthcare systems involve many interacting actors and processes, and their management require systematic decision-support tools. The sector is also economically and environmentally significant: in the European Union healthcare expenditure accounts for more than 10% of its GDP (as of 2022), and it is estimated that healthcare activities account for about 5% of global greenhouse gases emissions worldwide.
The relationship between data and decisions can be described along two dimensions: data representativeness and analytical scope.
Representativeness, or the capability of representing a system determines the usefulness of available data for decision-making. Clinical data are the most common source of information and are widely used to predict outcomes, estimate mortality risk, or forecast disease progression. Operational data (such as timestamps or procedure codes) can instead help identify inefficiencies and bottlenecks in healthcare processes. Advances in language processing models can also allow for the analysis of human-produced clinical notes. However, most healthcare data are collected for clinical, administrative, or legal purposes rather than operational planning. As a result, they may not adequately represent the system from a managerial perspective, leaving decision-makers with incomplete or biased information.
Analytical scope can be explicit, such as in reporting, statistical analysis, and forecasting, or implicit, to produce actionable solutions through optimization or prescriptive modeling methods. To simplify, an explicit scope describes or predicts system behavior but do not directly produce decisions, while an implicit scope is used to generate solutions that decision-makers can implement.
The combination of these two dimensions creates a conceptual framework for positioning different analytical approaches, depicted in Figure 1.

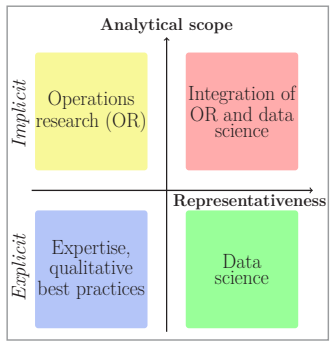

Fig.1 - Theoretical framework

OR methods can operate efficiently with limited or imperfect data, and they focus on delivering solutions obtained through optimization techniques. Conversely, data science methods typically rely on large, representative datasets and are more often used for descriptive or predictive analyses. A third area emerges at the intersection of OR and data science, where predictive models and optimization techniques are integrated to support decision-making. Contexts with limited data and explicit analytical goals often rely primarily on expert judgement rather than quantitative methods, and are therefore out of the scope of this thesis.
The nine papers contained herein explore the remaining three relevant quadrants, illustrating how different techniques can be applied across this above-describe framework in the healthcare context.
In the first part of the thesis, the focus is on OR approaches working with limited data: the first study develops a scheduling model for pediatric surgeries that generates both an initial plan and a set of backup schedules to handle disruptions caused by emergencies or patient no-shows. The second one proposes a tool for planning networks of primary care facilities, designed as a proof of concept despite the limited availability of both demand and offer data. A third study evaluates a new organizational model for home blood donation, using historical records to simulate future scenarios and assess feasibility.
In the second part, other works that focus on data-driven analyses with explicit scope are presented. In one paper, text mining techniques applied to Italian newspaper articles are used to analyze public perceptions of blood donation and support more effective communication strategies. The second paper develops a ML model based on electrocardiogram data to estimate five-year mortality risk in individuals whose tests were originally considered normal, demonstrating the potential of routine clinical data for mass screening. A third paper uses ML to assist surgical planning for arteriovenous fistulae in dialysis patients, providing fast patient-specific predictions performing better than physical models.
Lastly, the integration of OR and ML for decision-making is explored. One paper addresses surgical scheduling with uncertain operating times by combining optimization with probabilistic predictions. The last two papers included in the thesis propose a simulation-based framework to evaluate the potential value of predictive models before actually training them. In the first one, absenteeism predictions are simulated to determine how accurate a ML model would need to be to improve nurse rostering decisions. In the second one, policies for patient admission scheduling under uncertain hospital length-of-stay are evaluated by simulating prediction errors and analyzing their operational impact.
Overall, the thesis presents a roadmap for the evaluation of the appropriateness of data-driven decision systems, as a function of both the representativeness of the available data and the desired analytical scope. OR and data science offer complementary tools for healthcare decision-making: in particular, ML should not be viewed as a universal solution; instead, its potential benefits must be assessed in relation to the representativeness of the available data and the nature of the decisions to be supported. Lastly, simulation-based approaches that approximate the behavior of predictive models should be viewed as a practical way to evaluate whether investing in data-driven solutions is worthwhile.

SILICON LOW-GAIN AVALANCHE DETECTORS FOR X-RAY SPECTROSCOPY

Iurii Eremeev - Supervisor: Giuseppe Bertuccio

Co-Supervisor: Filippo Mele

Semiconductor radiation detectors (SRD) are nowadays among the most developed technologies for X-ray imaging and spectroscopy, widely employed in various research and industrial systems, including equipment for medical industry, material structure and composition analysis, security screening, food control systems etc. The evolution of cutting-edge SRD technologies, which achieve state-of-the-art energy, spatial, and timing resolution, has been driven intensively by the most demanding fields of SRD application in scientific experiments, such as Astrophysics and High-Energy Physics (HEP), including the celebrated Large Hadron Collider (LHC) at CERN. Various advanced SRD systems are born to satisfy the needs of LHC in the precise tracking of minimum ionizing particles (MIP), such as silicon drift detectors (SDD), which are now widely used in high-energy resolution X-ray spectroscopy. A recent example of this technology transfer from HEP experiments to Photon Science applications is the Low-Gain Avalanche Diodes (LGAD), which combine an outstanding timing resolution of less than 30 ps with a high spatial resolution of tens of μm. LGADs are based on a conventional Diode SRD structure with a highly doped gain layer implemented inside the low-doped region of the structure, inducing a large peak of the electric field (above 200 kV/cm), which ensures impact ionization of the structural atoms inside the gain region. This LGAD multiplication gain increases the signal amplitude, thereby reducing relative noise contributions of the preamplifier and subsequent stages of the electronic chain, which improves the system performance. Recently, the increase of the signal amplitude in LGAD detectors has attracted significant research interest for hybrid soft X-ray spectrometers operating from 200 eV to 2 keV, whose energy resolution is expected to improve when LGAD replaces a standard Diode-based sensor. However, the LGAD multiplication gain introduces an excess noise component due to the non-deterministic nature of impact ionization. This excess noise must be precisely determined at various operating conditions to optimize the performance of any LGAD-based radiation detector. Thus, this thesis targets an analysis of different noise components in LGAD-based X-ray spectrometers, including the noise of the electronic chain and excess noise due to multiplication gain, and defines their ultimate performance limits.

The performance characterization of LGAD-based X-ray detectors required the use of various LGAD structures, which were provided by the Brookhaven National Laboratory (BNL), including the very low signal charge gain LGAD (1.2<Ms<1.9), the moderate-gain LGAD (5.5<Ms<12.6), and the high-gain LGADs (10.2<Ms<19.3), so covering a wide gain range to explore the potential advantage of LGAD in X-ray spectroscopy. To analyze the performance of the LGAD sensors, first, the LGADs' current noise power spectral densities (NPSD) were measured under various operating conditions, utilizing a custom low-noise wide-bandwidth transimpedance amplifier (TIA), which demonstrates performance comparable to state-of-the-art current noise measurement systems. The LGAD NPSD comprises both low-frequency and white noise components, which vary significantly with operating parameters – including the device reverse bias voltage, current, and temperature – due to the change in the LGAD multiplication gain.

The dependence of the LGAD multiplication gain on operating parameters has been studied both theoretically and experimentally, revealing a significant increase with device bias voltage and a decrease with temperature, consistent with the multiplication gain model developed for LGAD devices. A novel activation energy method has been proposed in this work for determining the LGAD dark current multiplication gain ($M_{DC} = I_{mult} / I_{inj}$). This method enables the experimental extraction of the multiplied LGAD current component (I_{mult}) by subtracting the not multiplied current component ($I_{n/mult}$) from the measured LGAD current (I_{LGAD}). The value of $I_{n/mult}$ is determined at low LGAD bias voltages – before the sensor structure depletion and the onset of multiplication gain – at which the current activation energy measured with LGAD is consistent with that of an equivalent Diode structure without internal multiplication gain and with the literature data for conventional silicon sensors. The I_{inj} was approximated using the measured Diode current (I_{Diode}) and subtracting the not injected Diode current component ($I_{n/inj}$), which does not traverse the region of the Diode structure geometrically equivalent to the gain region in LGAD. The value of $I_{n/inj}$ was determined by matching the M_{DC} to the Ms measured at low bias voltages, assuming negligible gain suppression. Obtained M_{DC} enabled a correct reconstruction of the dependency of the measured LGAD excess noise factors (ζ), originating from the charge multiplication statistics, on M_{DC}, which is aligned with the theoretical predictions for the avalanche structures with short multiplying regions.

The last part of the work consists of a comprehensive performance analysis of the LGAD-based X-ray detection systems using ^{241}Am and ^{55}Fe sources, separating the noise contributions from the electronic setup and from the sensor itself, including both the LGAD dark current noise and the excess noise of the spectral lines, associated with the statistical distribution of the multiplication gain and a spatial gain inhomogeneity. It was found that the Ms increase decreases the white series and 1/f+dielectric system noise contributions. However, the increase in Ms broadens the spectral lines due to the distribution of the multiplication gain and the increase in the LGAD dark current noise contribution. Consequently, the LGAD sensor improves the energy resolution of the X-ray spectrometers compared to the Diode based counterparts only when the relative electronic noise contributions – including the white series, 1/f+dielectric, and the system parallel noise component – are higher than the noise components related to the multiplication gain distribution inside the sensor and its dark current noise. As far as the performance of the X-ray spectrometer based on LGAD is concerned, this work achieves a maximum energy resolution of 4% on the 13.9~keV line (FWHM = 560 eV) using the low-gain LGAD sensor operating at Ms = 1.5 coupled to state-of-the-art CMOS preamplifier (SIRIO).

This thesis proposes a robust methodological approach to measure various noise components of LGAD-based radiation detection systems using LGAD dark current noise measurements and high-energy X - γ ray spectroscopy, which can be used to optimize different LGAD-based radiation detection systems, including those developed for soft X-ray spectromicroscopy and 4D-trackers in HEP experiments.

NOVEL AVALANCHE PHOTODIODES FOR QUANTUM AND SENSING INFRARED APPLICATIONS

Lorenzo Finazzi - Supervisor: Alberto Tosi

Co-Supervisor: Giovanni Isella - Tutor: Giorgio Ferrari

The Complementary Metal-Oxide-Semiconductor (CMOS) process remains the global benchmark for cost-effective and scalable semiconductor fabrication. While silicon and germanium are the cornerstones of this technology, their intrinsic material properties limit photon detection to wavelengths below 1.8 µm. This spectral constraint poses a significant challenge for emerging applications in environmental sensing and infrared spectroscopy, which demand efficient detection beyond 2 µm. This doctoral dissertation addresses this gap by exploring the development of Germanium-Tin (GeSn) alloys as a disruptive, CMOS-compatible solution for Short-Wave and Mid-Infrared (SWIR/MIR) detection. Through Technology Computer-Aided Design (TCAD) modeling and experimental characterization, this research validates the use of GeSn in avalanche photodiodes (APDs), successfully demonstrating a spectral cutoff extension to approximately 2950 nm and providing a path toward monolithic infrared integration. In parallel with the development of novel group-IV alloys, the research investigates the optimization of state-of-the-art III-V semiconductor devices. Specifically, the work focuses on enhancing InGaAs/InP Single-Photon Avalanche Diodes (SPADs) through advanced back-end microfabrication techniques. Significant improvements were achieved in photon collection efficiency and optical crosstalk reduction within SPAD arrays through the integration of microlenses and optical isolation. Furthermore, the study explores the reliability of these sensors in harsh environments, evaluating radiation-induced damage from proton irradiation and demonstrating the efficacy of laser annealing as a technique for crystal lattice recovery. These insights are critical for the deployment of single-photon detectors in space-related missions and high-energy physics applications.

The final phase of this research, conducted during a visiting period at the Research Laboratory of Electronics at the Massachusetts Institute of Technology (MIT), integrates these detection capabilities with integrated photonics. By leveraging Time-Correlated Single Photon Counting (TCSPC), the study characterizes the dispersion of silicon nitride waveguides and the fast pulse dynamics of on-chip silicon light sources. Collectively, the findings presented in this thesis bridge the gap between fundamental material science and system-level photonic integration, offering versatile tools for fields ranging from biomedical imaging to quantum information processing.

IMPACT CHARACTERIZATION OF SMART RADIO ACCESS NETWORK DEVICES IN A NETWORK-WIDE CONTEXT

Paolo Fiore - Supervisor: Ilario Filippini

This thesis investigates Smart Radio Devices (SRD) as a transformative technology for addressing current radio access network (RAN) challenges through a comprehensive system-level analysis. Rather than treating communication and sensing as separate problems, we examine how emerging technologies like Reconfigurable Intelligent Surfaces (RIS) and Network-Controlled Repeaters (NCR) can reshape network deployment strategies while simultaneously enabling new services.

Mobile data traffic forecasts indicate a three-fold increase from 2023 to 2028, driven by ultra-high-definition video streaming, cloud gaming, AR/VR, and massive IoT. Traditional sub-6 GHz frequencies are rapidly saturating, creating urgent need for new spectrum resources.

The 3GPP standardization of Frequency Range 2 (FR2) within the millimeter-Wave (mmW) spectrum makes way for extreme data rates through vast bandwidth allocations.

Despite compelling advantages, mmW deployment introduces fundamental propagation challenges. Path loss increases quadratically with frequency, and signals exhibit poor penetration through building materials and limited diffraction around obstacles. When line-of-sight (LoS) paths are obstructed, signal strength quickly degrades, rendering links unstable.

To address deployment challenges while maintaining cost-effectiveness, 3GPP standardized Integrated Access and Backhaul (IAB) architecture in 5G NR Release 16. IAB introduces a hierarchical network structure where a single IAB Donor maintains wired connections while IAB Nodes provide access to users and relay traffic through wireless backhaul, reducing deployment costs by up to 85% compared to fiber-based solutions.

While IAB addresses backhaul challenges, intrinsic propagation limitations require additional solutions. RISs consist of planar arrays of passive elements that dynamically adjust the phase of incident electromagnetic waves to redirect signals toward desired directions, creating virtual LoS paths even when the direct ones are blocked. The quasi-passive operation consumes minimal power, and manufacturing costs are significantly lower than traditional active infrastructure, enabling massive deployment scenarios.

NCRs represent an evolution of traditional wireless repeaters with beamforming capabilities and intelligent network control. Unlike passive RIS, NCRs employ active amplification to boost signal strength, overcoming severe path loss though introducing noise amplification and potential self-interference.

Building upon 5G, 6G networks are expected to introduce Integrated Sensing and Communication (ISAC) as a centerpiece technology. Positioning and localization services promise cm-level accuracy with respect to satellite-centered solutions in challenging scenarios such as dense urban environments.

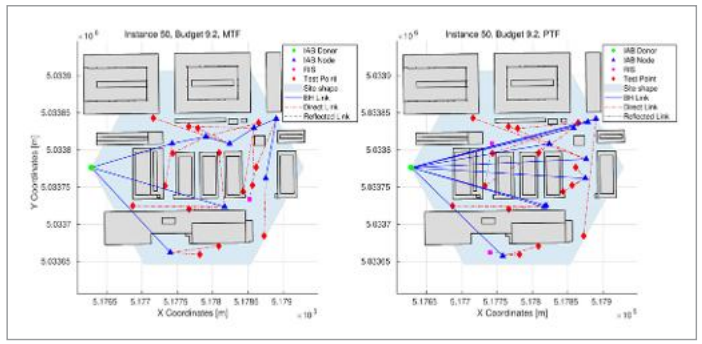

Fig.1 - Differences in network topology after planning for mean throughput (multi-hop, left) vs peak throughput (star-like, right) in a complex blockage model scenario

Our first contribution formulates Mixed-Integer Linear Programming (MILP) optimization models jointly optimizing IAB Node and RIS placement for redundant dual access, balancing link length minimization and angular diversity maximization. We incorporate novel and traditional modeling elements including angular diversity constraints, field-of-view limitations, and interactions between placement, association, and resource allocation. Monte Carlo simulations test resilience against random drops of opaque obstacles, demonstrating that RIS-enhanced deployments achieve comparable resilience to pure IAB densification at reduced costs.

Our second contribution introduces enhanced blockage modeling with three-category characterization: static obstacles (e.g. buildings), nomadic obstacles (mobile entities via Poisson point processes with a knife-edge diffraction model), and self-blockage (device orientation statistics). We develop peak-throughput optimization formulations differing from traditional mean-throughput approaches, recognizing burst-oriented traffic patterns dominated by video streaming and optimizing instantaneous capacity rather than average performance. This naturally favors star-like network configurations better suited to modern traffic characteristics, as shown in Fig.1.

Our third contribution introduces NCRs alongside RIS, incorporating 3GPP-compliant Modulation and Coding Schemes (MCS) and bidirectional traffic flows. Extensive parametric analysis investigates the impacts across diverse operating conditions. Results establish definitive conclusions: IAB infrastructure dominates capacity optimization, while RIS and NCRs provide marginal benefits except under extreme blockage density or artificially low costs. This fundamentally challenges assumptions about SRD applicability for pure communication enhancement.

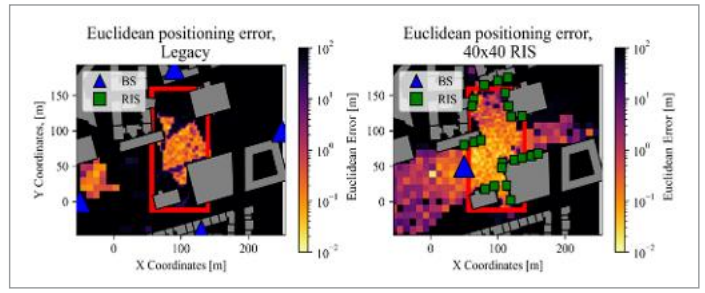

Fig.2 - Minimum achieved positioning error for the legacy case (left) with base stations only vs a RIS-enhanced approach (right): RIS visibly improve positioning accuracy, especially in urban non-LoS conditions

Our fourth contribution reveals an alternative use case for RIS in the sensing domain. We develop an opportunistic Time-Difference-of-Arrival (TDoA) positioning framework leveraging communication-oriented RIS deployments without requiring dedicated sensing infrastructure. Our approach integrates beamforming and data clustering for anchor selection robustness. Cramér-Rao Bound analysis establishes theoretical accuracy limits, while realistic ray-tracing simulations demonstrate sub-meter positioning accuracy where conventional approaches fail due to non-LoS conditions, as shown in Figure 2.

INTEGRATION OF QUANTUM COMMUNICATIONS IN PASSIVE OPTICAL NETWORKS

Alessandro Gagliano - Supervisor: Paolo Martelli

Tutor: Massimo Tornatore

The transition of telecommunication networks toward quantum-secure communications is one of the most urgent challenges in modern information technology. Classical cryptographic systems, which currently protect data communications, are increasingly threatened by the progress of quantum computing. In this context, Quantum Key Distribution (QKD) offers a radically different approach, enabling secret-key exchange with security guaranteed by the laws of quantum mechanics rather than by computational complexity. However, its large-scale adoption requires QKD to be integrated into the optical infrastructures already deployed, where Passive Optical Networks (PONs) represent the main architecture in the access segment.

This thesis investigates the coexistence of quantum and classical signals in shared optical access networks, with the aim of enabling the practical deployment of quantum communications in legacy PONs. A major obstacle to this integration is Spontaneous Raman Scattering (SpRS), the main noise source generated by high-power classical channels and one of the most critical impairments for quasi-single-photon quantum signals. The work provides an extensive experimental characterization of Raman noise in standard single-mode fibres over a broad spectral range and validates the results with commercial quantum and classical devices. Building on these measurements, a new analytical model is developed to estimate Raman noise generation in realistic PON architectures, showing how the impact of noise strongly depends on optical splitter (OS) placement, propagation direction and network configuration.

Starting from this physical and system-level analysis, the thesis defines design rules for the integration of both Discrete-Variable (DV) and Continuous-Variable (CV) QKD in different access network architectures, including single-fibre, dual-feeder and dual-fibre PONs. The results show that single-fibre PONs are severely constrained by cumulative SpRS noise, while architectures with additional fibre availability can significantly improve Secret Key Rate and network reach. In particular, the dual-feeder solution emerges as a promising compromise between performance and deployment feasibility, since it removes the dominant backscattered Raman contribution generated in the feeder section. A comparative analysis between DV-QKD and CV-QKD further shows that there is no universally optimal solution: CV-QKD is more effective in low-user or high-noise scenarios, whereas DV-QKD offers longer secure distances and better scalability. These findings indicate that future Quantum Passive Optical Networks should be designed with flexible deployment strategies, where the most suitable QKD technology is selected according to the physical infrastructure and service requirements.

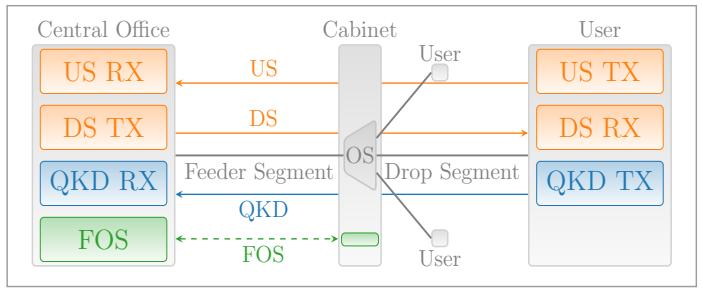

Fig.1 - Highly-secure passive optical network with data, quantum and sensing channels coexisting in the same optical infrastructure. US: upstream; DS: downstream; TX: transmitter; RX: receiver

Beyond data confidentiality, the thesis extends the concept of protection to the physical integrity of the network by integrating QKD with Fibre Optic Sensing (FOS). The proposed layered architecture, combining a DV-QKD system with a Michelson interferometric sensor, is analysed both theoretically and experimentally, demonstrating that quantum, sensing and classical channels can coexist within the same access network. The finite-key performance of QKD is also investigated, highlighting the importance of asynchronous key generation and request mechanisms for practical large-scale deployment. A preliminary experimental demonstration, focused on business-user scenarios with dedicated spare fibres, confirms the feasibility of this integrated approach with currently available technologies.

Overall, the thesis provides a comprehensive framework for bringing quantum communications into passive optical access networks. By combining analytical modelling, experimental validation and architectural design, it bridges the gap between quantum communication theory and the engineering constraints of real telecommunication systems. The results identify practical deployment strategies for the next generation of highly-secure PONs, capable of offering both quantum-safe data protection and infrastructure monitoring in future access networks.

ANALYSIS AND DEVELOPMENT OF AUGMENTED REALITY AND COMPUTER VISION ALGORITHMS FOR AUTOMOTIVE APPLICATIONS

Luca Gambarotto - Supervisor: Matteo Corno
Tutor: Fabio Dercole

Augmented Reality (AR) is increasingly regarded as a valuable technology for industrial and automotive applications, since it enables digital information to be superimposed onto the real world in a direct, intuitive, and context-aware way. In recent years, AR has shown strong potential in several domains, including human-machine interaction, advanced driver support, training, and smart farming. Nevertheless, its adoption in moving vehicles is still limited by several technical challenges. In these contexts, the environment is highly dynamic, the reference frame is non-inertial, and the relative motion between the user, the vehicle, and the external world must be estimated accurately in real time. Even small errors in localization or delays in visualization may severely reduce the quality, coherence, and usefulness of the augmented experience.
This thesis addresses these issues by developing methods and algorithms for reliable AR in automotive environments, with particular focus on Head-Mounted Displays (HMDs).
The research is built around two main objectives. The first is the estimation of the six-degree-of-freedom pose of an HMD inside a moving vehicle. This problem is particularly challenging because the motion measured by the sensors mounted on the headset is affected not only by the movement of the user's head, but also by the motion of the vehicle itself. To tackle this issue, the thesis proposes a visual-inertial fusion architecture designed to estimate the motion of the user relative to the cockpit while compensating for vehicle dynamics. The framework combines inertial sensing and vision-based observations to improve robustness and accuracy in real operating conditions. Within this architecture, two original computer-vision algorithms are introduced, exploiting both monocular and stereo cameras integrated in modern HMDs. These methods are designed to achieve robust head localization onboard a vehicle and to overcome some of the limitations of standard off-the-shelf tracking solutions, which are typically conceived for static or quasi-static environments rather than for highly dynamic automotive scenarios.
A second key aspect of the work is the management of hardware and computational latency, which represents one of the major obstacles to an effective AR experience in moving platforms. In fact, even when the pose estimation is sufficiently accurate, the delay between sensing, computation, rendering, and visualization can generate a mismatch between virtual and real elements. For this reason, the thesis also investigates prediction and compensation strategies aimed at reducing the perceptual impact of latency. This aspect is essential to guarantee stable and convincing visualization, especially in demanding applications such as motorsport or fast off-road operations, where rapid changes in vehicle attitude and user motion can quickly degrade AR alignment.
The second major objective concerns vehicle localization for geo-referenced AR visualization. In order to place virtual content consistently in the surrounding world, it is necessary to estimate the vehicle position and orientation with sufficient accuracy and continuity. To achieve this goal, the thesis develops two dedicated localization pipelines tailored to different application scenarios. For on-track applications, a method is proposed that combines GPS-RTK information and attitude estimation to enable accurate mapping and positioning along a racetrack. For off-road scenarios, where GPS may be unavailable, degraded, or unreliable, stereo-vision-based approaches are developed for vehicle localization in agricultural environments such as open fields and vineyards. These solutions are designed to extend the use of AR beyond conventional road environments and to support real operating conditions in which infrastructure is limited and environmental variability is high. To evaluate the quality of localization also when an accurate ground truth is not available, the thesis introduces a reprojection-based validation approach, allowing the consistency of AR alignment to be assessed directly from the application perspective.
The proposed methods are validated through heterogeneous case studies that demonstrate both the versatility of the developed algorithms and the breadth of potential AR applications in automotive contexts. Among these, the thesis presents a driver-coaching tool for racetrack experiences, where AR can support the driver by visualizing relevant information directly in the field of view. A second application is related to smart farming, where AR technologies can enhance the efficiency of in-field operations through improved interaction and localization capabilities.
Finally, the work also explores an automatic pipeline for synthetic dataset generation and labeling for neural-network training, exploiting AR techniques not only as a visualization tool, but also as an enabling technology for machine-learning workflows.
These applications show that the methodologies developed in the thesis are not limited to a single niche use case, but can support multiple domains requiring precise registration between digital content and the real world
All the solutions developed in this thesis were tested in collaboration with HMDrive s.r.l., and most of the proposed algorithms were integrated into the company's software stack. This demonstrates not only the technical feasibility of the proposed approaches, but also their practical relevance and industrial impact. Overall, the thesis contributes to bridging the gap between AR research and real-world automotive deployment, showing that dedicated perception, fusion, and localization algorithms can make immersive and reliable AR feasible even in complex dynamic environments..

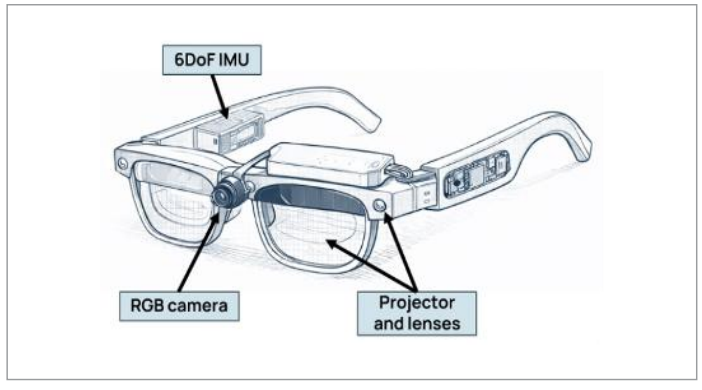

Fig. 1 - Head-Mounted Display Prototype

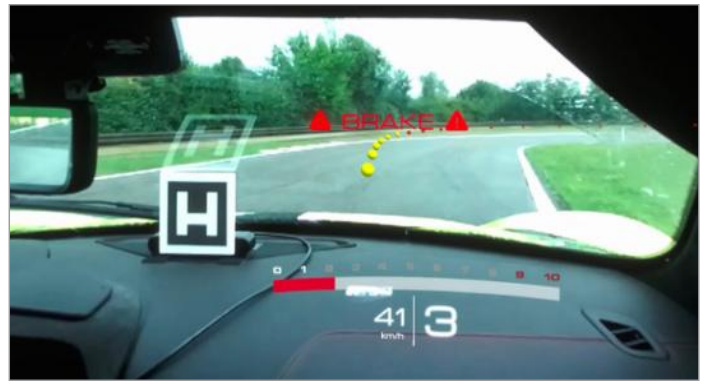

Fig. 2 - Driver Coaching application demo

DESIGN OF ENERGY-EFFICIENT NEURAL NETWORK ACCELERATORS WITH EMBED- DED PHASE CHANGE MEMORY DEVICES

Artëm Glukhov - Supervisor: Daniele Ielmini

AI and Neural networks are experiencing unrestricted growth in both public adoption and size. While we begin to glimpse the great capability, potential, and impact these algorithms may have on the social fabric of our times, a closer look shows that there is no free lunch in life, and the forever-growing neural network models hide in plain sight an unprecedented energy demand.

This novel, still evolving, and fastly growing computing application happens in conjunction with a constantly slowing evolution rate of traditional computing architectures. Among the many reasons behind this deceleration, the most evident are the (long) end of Moore's law, the rapidly increasing cost behind further technology node scaling, and an intrinsic difficulty in optimizing the traditional von Neumann architecture to the newly emerged data-intensive applications.

In this framework, the new paradigm of in-memory computing (IMC) finds its place as a promising candidate to complement the traditional architectures, accelerating those computations that are so expensive on typical machines yet so requested. IMC is made possible thanks to the flourishing research on emerging nonvolatile memory devices that can be organized in crosspoint arrays, memory matrices with high integration density, low-power, and highly parallel structures that enable trivial analog computation of matrix-vector multiplication, the main building blocks of the whole neural network framework.

As with almost everything in electronics, as we move further from an idea and get closer to a practical implementation, an infinite series of undesired and parasitic effects start polluting the outcome, complicating the task.

In-memory computing is not exempt from such problems, and indeed, between the nonidealities of the memory devices, the circuital limitations, and the algorithmic compatibility issues, the task of developing a standard-setting and fully-featured IMC accelerator is still far from completion, despite the high amount of interest revolving around it.

In this scenario, this doctoral thesis ultimately focuses on the design and development of a fully-custom PCM-based ASIC for IMC acceleration, but before that, it retraces all the discussions and analysis starting from PCM devices and their programming algorithms, a differential readout scheme proposal that eventually has seen the light as another testchip, and the description of an analog-to-digital readout chain that was at the core of the last ASIC.

The research work tackled various challenges of the in-memory computing paradigm, exploring and addressing them from different standpoints, pivoting and exploiting the intrinsic multidisciplinarity of the field.

READ-OUT INTEGRATED CIRCUITS AND SYSTEMS FOR HIGH TIME RESOLUTION NEAR INFRARED SPAD ARRAYS

Kevin Andres Herrera Pesantes - Supervisor: Alberto Tosi

Tutor: Giorgio Ferrari

Single-photon detectors have become indispensable instruments across a wide spectrum of scientific and industrial applications. In materials science and biology, they enable the measurement of ultrafast fluorescence decays with high sensitivity; in range finding systems, they allow precise time-of-flight measurements of laser echoes; and in quantum optics, they form the technological backbone of emerging photonic technologies.

Modern single-photon detection technologies are primarily dominated by Single-Photon Avalanche Diodes (SPADs) and Superconducting Nanowire Single-Photon Detectors (SNSPDs). While SNSPDs can deliver outstanding performance in terms of photon detection efficiency, timing resolution, and noise, they depend on superconducting materials and therefore require cryogenic cooling systems, which are bulky and costly, restricting their use to laboratory environments. SPADs, in contrast, stand out for their compactness, robustness, and compatibility with standard microelectronic fabrication. These features make them practical for integration in scalable systems.

In the visible spectral range and up to approximately 1 µm wavelength, silicon SPADs can be fabricated in standard CMOS processes, allowing monolithic integration with dedicated electronics into large-scale arrays. Beyond this limit, at longer wavelengths, alternative materials such as III–V compounds or germanium can be employed. By hybridizing these SPADs onto dedicated Read-Out Integrated Circuits (ROICs), compact and high-performance detectors can be realized. The extension of SPAD technologies into the infrared region is driven by several advantages, operation at longer wavelengths benefits from reduced scattering and absorption in various media, increased transmission efficiency in both fiber and free-space channels, and improved eye-safety conditions, which permit the use of higher optical power levels. Furthermore, the infrared domain offers reduced background noise under ambient illumination and opens new opportunities for advanced imaging and sensing applications. These combined factors make the NIR and SWIR spectral ranges strategically important for the development of next-generation single-photon detection systems.

A further distinguishing feature of SPADs is their ability to operate in fast-gated mode, whereby they can be rapidly switched from OFF to ON within a few hundred picoseconds. This capability allows precise temporal selection of photons, effective noise suppression, and dynamic range extension, as well as enabling high-resolution time-of-flight measurements. This thesis builds upon these strengths to explore the design of advanced Read-Out Integrated Circuits and systems for high time-resolution near-infrared SPAD arrays, paving the way for future developments in next-generation photonics, imaging, and communication technologies. The work follows two complementary directions: i) the realization of single-pixel and array cameras based on InGaAs/InP SPADs, with the objective of extending Non-Line-of-Sight (NLOS) imaging into the Short-Wave Infrared Range (SWIR); ii) the development of high-throughput single-photon detectors for quantum communication applications, in particular Quantum Key Distribution (QKD). This involves the design of an 8×8 silicon SPAD array operating as a single high-rate detector with an aggregate count rate of 2 Gcps and operating at 800 nm wavelength, as well as the design of a compact single-pixel ROIC capable of operating InGaAs/InP SPADs at gigahertz gating frequencies for QKD at 1550 nm wavelength in both fiber-based and free-space scenarios.

SELF-ADAPTIVE INTEGRATED PHOTONIC TRANSCEIVER FOR FREE SPACE OPTICS COMMUNICATIONS

Andrés Martínez - Supervisor: Francesco Morichetti

Free-space optical (FSO) communication is a promising technology for high-capacity and low-latency wireless links, enabling fiber-like data rates without the need for physical fibers. Its potential spans a wide range of applications, including terrestrial backhaul, data center interconnects, satellite communications, and next-generation 5G/6G networks. However, the practical deployment of FSO systems remains constrained by atmospheric channel impairments. Among these, attenuation due to fog, haze, rain, and aerosols limits link availability, while pointing errors caused by mechanical vibrations and platform instability introduce additional power fluctuations. Under clear-sky conditions, the dominant limitation arises from atmospheric turbulence, which induces random refractive-index variations along the propagation path (see Fig. 1). These fluctuations lead to amplitude scintillation, phase distortions, beam wander, and modal coupling, ultimately degrading signal-to-noise ratio and increasing outage probability. This thesis investigates photonic integrated circuit (PIC)–based solutions to mitigate turbulence-induced impairments in FSO systems, with a particular focus on adaptive, scalable, and hardware-efficient receiver architectures. The central approach is based on multi-aperture optical reception combined with coherent optical processing. The incoming distorted wavefront is spatially sampled using an optical antenna array, and the collected optical fields are recombined through a programmable mesh of Mach–Zehnder interferometers (see Fig. 2). By performing adaptive coherent combining directly in the optical domain, the receiver compensates for turbulence-induced phase distortions and suppresses scintillation without relying on high-speed digital signal processing.
The proposed architecture is analyzed through theoretical modeling and validated experimentally under controlled turbulence conditions. The results demonstrate a significant reduction of power fluctuations and stable operation at symbol rates up to 25~Gbaud, even in regimes characterized by strong turbulence. Design guidelines are established by linking the receiver aperture geometry to the atmospheric coherence width, providing a practical framework for scaling the architecture to different propagation conditions.
Beyond conventional intensity-modulated signals, the thesis demonstrates the signal-agnostic nature of the proposed receiver. Adaptive coherent combining is shown to preserve the dynamical properties of chaos-based optical communication, maintaining synchronization under turbulent conditions. In addition, the architecture enables robust radio-over-free-space-optical (RoFSO) transmission, supporting high-order modulation formats up to 1024-QAM at millimeter-wave frequencies, thereby bridging optical and wireless communication domains.
Compatibility with wavelength-division multiplexing is also investigated, showing negligible wavelength-dependent penalties across the C-band.
The thesis further explores bidirectional PIC-to-PIC FSO links, demonstrating that channel reciprocity enables simultaneous optimization of transmission and reception without requiring a dedicated feedback channel.
The impact of back reflections on control stability is experimentally characterized, and mitigation strategies based on pilot-tone labeling and wavelength multiplexing are proposed and validated, significantly extending the operational power margin.
Finally, the role of in-line optical power monitoring is addressed as a key enabler of adaptive control. A compact photo-thermal plasmonic sensor is demonstrated for on-chip power monitoring, enabling robust feedback and stabilization in complex photonic circuits.
Overall, this work establishes adaptive photonic integrated receivers as a viable and scalable solution for turbulence-resilient FSO communication systems. The proposed approach supports high data rates, advanced modulation formats, and multi-channel operation, while maintaining low latency and reduced system complexity compared to purely digital compensation techniques.

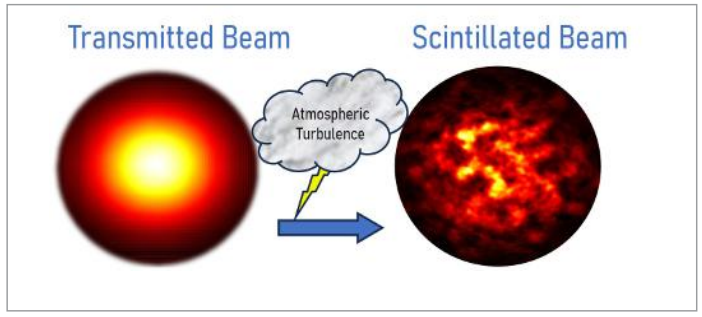

Fig. 1 - A light beam that travels through the atmosphere gets distorted due to turbulence. If not compensated, most of the information transmitted in the beam will be lost

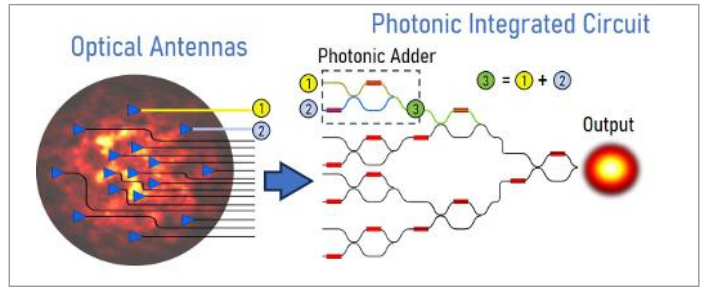

Fig. 2 - An integrated FSO photonic receiver recovers the transmitted beam in 2 steps. Step 1: Optical antennas sample the scintillated beam. Step 2: A Photonic Integrated Circuit sums all the samples using many Photonic Adders. The output signal contains all the data transmitted with less than 0,0001% of errors.

DESIGN AND DEVELOPMENT OF ADVANCED AIRCRAFT GROUND HANDLING CONTROL SYSTEMS FOR SAFE HUMAN-IN-THE-LOOP CONTROL

José Joaquín Mendoza Lopetegui - Supervisor: Mara Tanelli

Co-Supervisors: Matteo Corno, Sergio Matteo Savaresi

Ground maneuvering is a crucial part of an aircraft's operational envelope, encompassing operations such as taxiing, take-off, and rollout after landing. During these operations, the pilot must maintain full control despite disturbances such as wind, uneven pavement grip, and the multiple actuators used for maneuvering. The high-speed rollout phase after touchdown is particularly critical, as aircraft may experience stability issues in both the longitudinal and lateral directions that can lead to catastrophic runway excursions. Developing methods to assist pilots in effectively managing these phases is therefore essential.

Ground maneuvering involves not only aircraft dynamics but also interaction with the pilot. Under certain conditions, pilot control actions can couple with the aircraft's response, giving rise to pilot-induced oscillations (PIOs), defined as sustained or uncontrollable oscillations resulting from the pilot's attempts to control the aircraft. This phenomenon is highly dangerous and, in extreme cases, may cause total loss of control. Although considerable research has focused on mitigating PIOs in flight, studies addressing them during ground operations remain scarce.

Another topic receiving increasing attention is the development of autonomous aircraft capabilities on the ground, allowing pilots to focus on higher-level tasks. While autonomous flight has been investigated for nearly 70 years, highly autonomous ground operations are a more recent development. Aircraft may spend 10% - 30% of mission time in taxiing conditions, presenting significant opportunities for reducing pilot workload.

This thesis addresses all the aforementioned challenges by developing a coherent theory of aircraft ground-handling dynamics, encompassing modeling, simulation, analysis, planning, perception, and control. The work focuses on fixed-wing aircraft with tricycle landing gear, the most common configuration for manned aircraft. The problem is studied during solid ground-contact regimes, including rollout after touchdown, taxiing, and acceleration before take-off, including rejected take-offs.

The first contribution addresses modeling and simulation of aircraft ground-handling dynamics, which is a complex multidisciplinary problem involving mechanics, aerodynamics, contact dynamics, tribology, hydraulics, wear dynamics, ergonomics, and human-machine interaction. An object-oriented modular simulation framework is developed, and a multibody model is constructed to serve as a realistic prototyping environment for the control systems proposed later in the thesis. The modeling techniques employed are general, but a specific application case of a reference aircraft platform is illustrated, employing data made available from an industrial partner. The simulator is validated against experimental data from this platform, demonstrating strong capability in reproducing aircraft ground-handling dynamics. To illustrate the model's versatility, a study is conducted to examine brake-pad wear dynamics across their operational life, producing the first control-oriented model capturing these effects.

The thesis then addresses the design and implementation of effective, safe, and industrially applicable anti-skid systems capable of ensuring strong longitudinal braking performance while improving directional stability. The anti-skid design process is studied from low-level pressure control to experimental flight testing. Model-based gain-scheduling controllers are developed to prevent wheel locking and maximize braking performance, with stability proven in a Linear Parameter-Varying framework. The system is verified in simulation, validated using an industrial-grade dynamometer, and evaluated in flight tests with professional test pilots on an advanced supersonic aircraft. This represents the first complete treatment of the problem in the literature. In addition, Artificial Intelligence-based health monitoring and prognostics modules are developed to augment the anti-skid system and enable active performance improvement by identifying key braking parameters online, which is the first development of its kind.

The thesis also investigates active coordination of anti-skid control laws between the right and left sides of the aircraft to enhance directional stability. No such laws have been previously presented, since previous work has largely addressed anti-skid design solely in the longitudinal direction with limited attention to directional dynamics. A novel coupling algorithm is proposed, and its well-posedness is established using contraction theory. The control law is experimentally verified in a dedicated flight-test campaign with positive pilot evaluations, demonstrating that longitudinal braking performance can be effectively traded for improved directional stability.

The work further addresses the design of stability augmentation systems to mitigate PIOs during ground handling. The effects of the rate-limiting element in the Steering Control Unit are analyzed using a describing-function approach, revealing its potential to generate PIOs in the yaw degree of freedom. This is the first time such an explicit connection has been made. A stability augmentation method based on phase anticipation is proposed to mitigate this issue, providing significant phase recovery and reducing PIO likelihood. Additionally, the coupling between longitudinal and lateral dynamics during differential braking usage is characterized. Conventional pedal braking laws are shown to introduce strong couplings and discontinuous sensitivity variations that increase PIO susceptibility. To mitigate this effect, a decoupling logic is proposed that shapes the aircraft's longitudinal and lateral responses to pilot braking commands. Simulation results demonstrate its effectiveness in reducing PIO likelihood.

The thesis also develops an active control-assistance system to help pilots recover from directional instabilities during high-speed rollout after touchdown. The system uses all available moment-generating actuators, is robust to actuator failures, and tolerates adverse environmental conditions. No existing assistance system in the literature simultaneously satisfies these conditions, as most operate within restricted regimes or actuator sets. The system processes pilot commands and introduces corrective actions when incipient instabilities are detected. Particular attention is devoted to the control allocation problem, ensuring optimal distribution of commands to achieve the desired yaw-moment generation. Pilot-in-the-loop experiments using the simulation environment demonstrate improved maneuverability during challenging scenarios involving subsystem failures and adverse conditions.

Finally, the thesis develops a highly autonomous system aimed at almost complete pilot workload reduction during ground handling. The goal is to ensure safety across all ground phases by enabling autonomous taxi-in/taxi-out, autobraking, and both high-speed and low-speed autonomous rollout. The approach addresses the associated control, planning, and perception challenges. The control layer consists of high-performance longitudinal and lateral error controllers enabling accurate runway centerline capture and tracking during both rollout and taxiing. The autonomous system also includes a self-localization module capable of operating with or without GNSS (Global Navigation Satellite System) to support robust autonomous operations with minimal human supervision. The GNSS-denied scenario is examined, and a visual-inertial solution is proposed for operation within a reduced envelope. The developments are validated in simulation through autonomous high-speed rollout and low-speed taxiing maneuvers.

LEARNING-BASED SOLUTIONS FOR ACTIVE MONITORING AND ANOMALY DETECTION FOR ADVANCED INDUSTRIAL APPLICATIONS

Sabrina Milani - Supervisor: Mara Tanelli

Tutor: Simone Garatti

Modern engineering systems are increasingly characterized by high complexity, strong variability in operating conditions, and the availability of large amounts of high-frequency sensor data. In this context, advanced monitoring and diagnostic tools are becoming essential to guarantee safety, reliability, efficiency, and adaptability. Traditional physics-based approaches remain fundamental because they offer interpretability and robustness, but their effectiveness may decrease when systems operate under non-nominal or highly variable conditions. On the other hand, data-driven techniques based on Machine Learning can capture complex patterns directly from data and often achieve superior predictive performance, although they may suffer from limited interpretability and reduced physical consistency.
This thesis is motivated by the need to bridge these two perspectives and to develop learning-based solutions that are accurate, robust, and informed by domain knowledge.
The work addresses this challenge through two main industrial application domains. The first concerns wind turbine monitoring, with a particular focus on the detection, classification, and severity assessment of blade pitch misalignment anomalies. These faults are especially relevant because they can reduce power production, increase structural loads, and negatively affect the long-term health of the turbine. To tackle this problem, the thesis develops learning-based frameworks that exploit mechanical loads and acceleration signals to identify anomalous behavior and quantify its severity. The proposed approaches are designed not only to detect the presence of a fault, but also to distinguish among different fault configurations and, in some cases, to support localization of the misaligned blades. This enables a more informative diagnostic process and demonstrates how sensor-based Machine Learning can contribute to active monitoring strategies in renewable energy systems.
The second application domain is automotive engineering, where the thesis investigates the early classification of road surfaces, grip conditions, and tyre types using vehicle dynamic signals, with particular emphasis on braking scenarios. In such contexts, the ability to recognize the operating condition of the vehicle in real time can significantly improve control, safety, and decision-making. Two complementary perspectives are explored: model-based methods, which leverage the physical interpretation of the vehicle dynamics, and learning-based approaches, which exploit data patterns to improve classification capabilities under realistic conditions. The work shows how these methods can be applied to identify road condition and tyre characteristics from measurable signals, and extends the analysis to more challenging situations such as split-surface conditions and reduced signal availability. In this way, the thesis contributes to the development of intelligent monitoring and classification tools for advanced automotive applications.
Building on the results obtained in these two domains, the core methodological contribution of the thesis is the development of a domain-informed **Mixture of Experts (MoE)** framework. Rather than relying on a single monolithic model, the proposed approach combines multiple specialized experts within a unified decision-making architecture. Each expert is designed to operate in specific local conditions or to solve a particular subtask, such as classification or regression, while the overall framework coordinates their contributions in a coherent and interpretable way. This structure makes it possible to simultaneously classify operating conditions and reconstruct unmeasured variables, while preserving robustness and improving explainability with respect to purely black-box models. The MoE framework therefore represents a general methodological contribution that goes beyond the individual case studies and offers a flexible strategy for monitoring and diagnosis in structurally different engineering domains.
The proposed methodologies are validated on both simulated and real datasets, demonstrating their effectiveness across heterogeneous scenarios. In the wind energy domain, the results show accurate and informative detection of pitch misalignment anomalies, including severity assessment and fault characterization. In the automotive domain, the developed models achieve reliable classification of road surfaces and tyre types, supporting early recognition of operating conditions relevant for vehicle dynamics and safety. Taken together, these results confirm the potential of hybrid, domain-informed learning architectures to improve monitoring, anomaly detection, and predictive decision-making in modern engineering systems. More broadly, the thesis shows that combining physical insight with Machine Learning is a promising direction for the design of practical, scalable, and interpretable solutions for advanced industrial applications.

Fig. 1 - Main industrial application domains

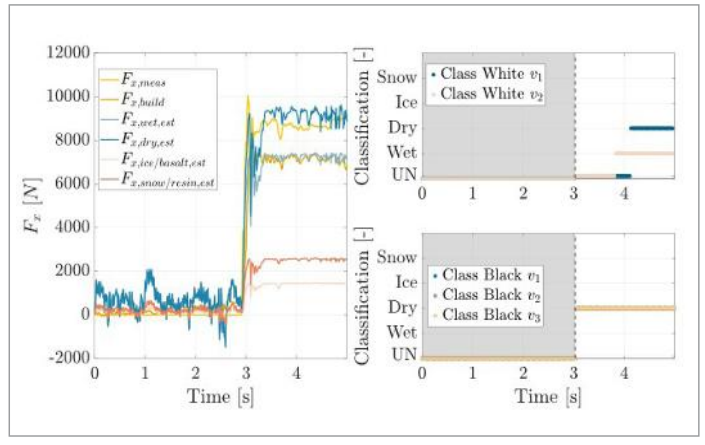

Fig. 2 - Dry Asphalt braking interval, comparison between white and black box models output

ARTIFICIAL INTELLIGENCE-BASED CYBERATTACK MITIGATION TECHNIQUES

Tommaso Paladini - Supervisor: Michele Carminati

Tutor: Nicola Gatti

Cyberattacks are growing in sophistication and impact, compelling organizations to continuously evolve their defensive strategies. As the actions of threat actors often exhibit recurring patterns, defenders have turned to Artificial Intelligence (AI)-based methodologies to automatically analyze data collected from targeted software systems, both to detect anomalous behavior in real time and to learn from previously recorded intrusions to inform future strategic decisions. This thesis explores the challenges of two popular paradigms in AI-driven cyber defense: real-time detection systems, focusing on their application in the banking fraud domain, and proactive approaches based on Cyber Threat Intelligence (CTI). Accordingly, this thesis is structured in two parts.

In part one (Fig. 1), we investigate the robustness of Machine Learning-based banking fraud detection systems against attackers attempting to subvert the intended behavior of the learning algorithm to evade detection. First, we show how attackers may manipulate banking transactions to evade or poison fraud detectors and propose a corresponding mitigation tailored to realistic financial fraud scenarios. We evaluate its effectiveness under different levels of attacker knowledge of the target system. Our results show that the economic damage of the attack can be extensively reduced with a small increase of false positives. Then, we introduce an Online Learning-based detection framework leveraging the Multiplicative Weights Update algorithm to dynamically adapt to adversarial behavior. We demonstrate its capacity to minimize economic loss and maintain robustness against evolving fraudulent strategies. We conduct our experimental evaluation on real-world banking data provided by an Italian financial institution.

In part two (Fig. 1), we examine the challenges of CTI-based proactive defense strategies, focusing on the quality of unstructured intelligence sources and the applications of Natural Language Processing (NLP) required to extract actionable insights from them. First, we present a large-scale longitudinal analysis correlating discussions in underground forums (over 88 million posts) with CTI reports spanning two decades. Our findings reveal that hacker forums have historically contributed to malware later used in real-world attacks, and that systematic CTI monitoring has only caught up within the past decade. Finally, we provide a comprehensive systematization and empirical comparison of NLP-based Tactics, Techniques, and Procedures (TTP) extraction methods, uncovering limitations in current research and frameworks, and highlighting the need for future work on ontology refinement and large-scale dataset development.

Despite their constant evolution, both paradigms should be employed in a complementary manner to cover their respective limitations. Future real-time detection systems could progress toward the identification of adversaries and automatic deployment of optimal defense strategies, possibly informed by CTI analysis. To advance and increase cooperation between the two paradigms, CTI research should not only cover the data sources that anticipated cyberattacks, but also refine existing frameworks, which still exhibit ambiguities that hinder automatic interpretation of intelligence.

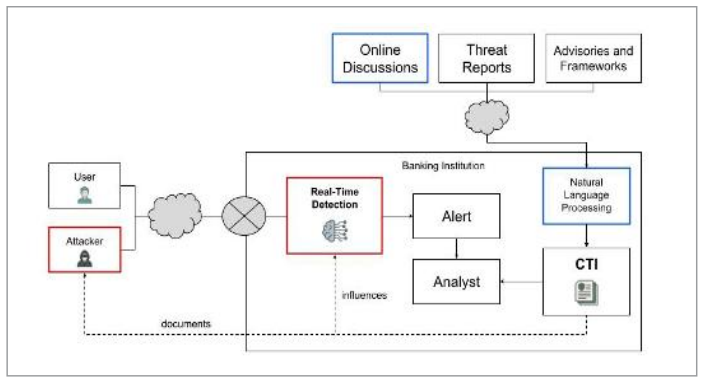

Fig. 1 - Simplified overview of Real-Time Detection approaches (in red), and Proactive defenses based on Cyber Threat Intelligence (CTI) (in blue), and the research questions explored by this work

SYNTHETIC MODELS FOR SIDE CHANNEL ATTACK ANALYSIS AND COUNTERMEASURE DEVELOPMENT

Isabella Piacentini - Supervisor: Alessandro Barenghi

Side-Channel Attacks (SCAs) represent a persistent threat to the security of cryptographic devices, exploiting physical leakages, such as power consumption, to retrieve secret information. These attacks fall into two categories: non-profiled attacks, which rely on synthetic models of device behavior, and profiled attacks, which build highly accurate, data-driven models from a target device instance.

This thesis presents a comprehensive methodology to address both threats. It contributes a novel horizontal attack for security assessment and a unified defensive countermeasure, Computational Interleaving (CI), implemented across both hardware and software domains.

The research introduces a powerful single-trace horizontal attack methodology designed to evaluate the vulnerability of iterative cryptographic primitives. Unlike traditional power analysis, which often requires large datasets or specific implementation knowledge, this technique operates autonomously and requires no prior knowledge of the internal structure of the algorithm. It identifies statistical similarities between intermediate values processed piecewise across successive iterations. The effectiveness of this approach was validated on a constant-time square-and-multiply-always RSA implementation, successfully recovering the entire secret exponent from a single power trace in a scenario where traditional attacks typically fail. Beyond its utility as an attack, this methodology serves as a systematic blind testing framework. It can allow designers to map leakage patterns back to specific source code segments, enabling targeted rather than global hardening, which optimizes the trade-off between security and performance.

The core defensive contribution is Computational Interleaving, a countermeasure providing unified protection against both non-profiled threats, such as Correlation Power Analysis, and profiled threats, such as Template Attacks. CI utilizes temporal computational redundancy by interleaving real cryptographic operations with fake computations using device-dependent keys. This process injects algorithmic noise that obscures side-channel leakage patterns and systematically renders profiled models non-portable across different device instances.

The hardware realization of CI, validated on FPGA platforms with AES as a case study, demonstrated superior security by achieving an improvement of over 333x in the Measurements-to-Disclose metric compared to unprotected implementations. Furthermore, it reduced template attack accuracy to approximately 50.7%, effectively rendering information-theoretically optimal attacks equivalent to random guessing. From an implementation perspective, CI proved highly efficient, incurring only a 66% overhead in logic utilization. This is significantly lower than the 219% to 274% overheads typical of conventional masking schemes. Notably, CI requires 38x to 104x less randomness throughput than masking, effectively removing a common performance bottleneck in resource-constrained systems. Recognizing that practical impact depends on deployment accessibility, this thesis presents a compiler-based methodology to fully automate the application of Computational Interleaving to software implementations. This approach eliminates the need for manual code modification or specialized cryptographic expertise, allowing for the integration of security countermeasures into the software development workflow for symmetric ciphers.

Experimental validation on an ARM Cortex-M4 target confirmed that this automated approach effectively neutralizes both Correlation Power Analysis and Template Attacks. The CI-protected implementation required at least 5,000x more measurements to disclose secret information compared to unprotected baselines. While the software implementation showed a 4.6x computational overhead, this remains favorable compared to state-of-the-art masking schemes, which often incur higher penalties while offering lower protection.

In conclusion, this thesis presents approaches both to designing and evaluating secure cryptographic implementations. The attack methodology enables identification of vulnerable code regions, while the Computational Interleaving countermeasure, validated in both hardware and software realizations, provides unified protection against profiled and unprofiled attacks with practical resource requirements. Furthermore, the compiler-based automation eliminates manual hardening effort, making the countermeasure accessible for widespread deployment.

CONVERSATIONAL USER INTERFACES FOR AN INCLUSIVE WEB: FROM GUIDELINES TO DESIGN METHODS AND DEVELOPMENT TOOLS

Ludovica Piro - Supervisor: Maristella Matera

This research investigates how conversational interaction should be designed to enhance web accessibility for blind and visually impaired (BVI) users. Currently, despite accessibility standards the web remains often not accessible for BVI users. Furthermore especially in complex tasks such as form filling, even when technical accessibility is met, the literature documents multiple usability challenges faced by BVI users on the web. These challenges include making sense of what is available on the page, understanding what is possible to do, or re-finding information. Conversational AI, could offer an alternative paradigm to access the web. However, conversational web browsing remains an understudied task in the field of conversational interaction and current guidelines have yet to be tested on it.

This work contributes to the advancement of the state of the art in conversational interaction by providing conversational interaction designers and developers with a new methodological approach to design conversationally accessible websites. As an initial step towards a design methodology in this area, this thesis introduces a design framework and a set of guidelines for inclusive conversational interaction with web content, grounded in established accessibility principles, informed by the perspectives of blind and visually impaired (BVI) users, and tested with practitioners (i.e., designers and developers) to assess their usability in a practical setting. Practitioners can use the methodological tools presented in this work to understand how to structure conversational interaction and translate it into web-design choices. Since the guidelines are grounded in accessibility principles, this methodology poses itself as an accessibility-first methodology to design conversational web interactions.

To achieve this, first, existing web accessibility and conversational interaction guidelines, were mapped and analyzed to identify gaps. Then, the selected guidelines were refined through workshops involving designers and developers. Building on this foundation, a prototype LLM-driven conversational agent, ConWeb, was developed and evaluated with BVI participants, to elicit new requirements for navigation, data entry, and input verification. Lastly, to bridge the gap between guidelines and practice, the research further explores how AI coding assistants can be augmented with conversational accessibility guidelines to support developers, deriving initial requirements for tools that facilitate the creation of conversationally accessible websites.

The main contributions of this work were combined in a structured methodology, following the software lifecycle, presented in Figure 1. In this framework, guidelines hold a central role, ensuring awareness of user needs and alignment between phases. Further, they not only offer high-level design principles, but also provide concrete considerations that can be used to inform ideation, guide development, and support systematic testing of conversational interactions.

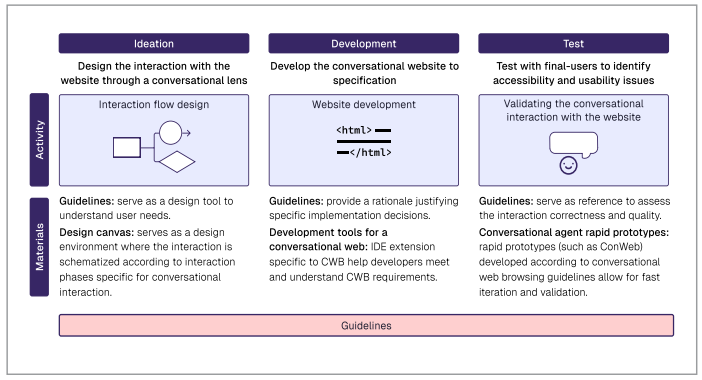

Fig. 1 - Diagram synthesizing the framework structure. It is structured in three phases: ideation, development, and testing. Guidelines serve as a cohesive element that guides practitioners' work in all the phases of CWB design

ELECTROCHEMICAL RANDOM ACCESS MEMORY (ECRAM) FOR ANALOG IN-MEMORY COMPUTING

Matteo Porzani - Supervisor: Daniele Ielmini

The rapid growth of artificial intelligence and the increasing complexity of deep neural network (DNN) architectures are pushing the limits of conventional computing hardware. Modern AI training workloads involve the continuous exchange of enormous amounts of data between processor and memory, resulting in a severe performance bottleneck known as the "memory wall." In-memory computing (IMC) has emerged as a promising paradigm to overcome this limitation by performing computation directly where data is stored, exploiting analog resistive memory devices organized in crosspoint arrays to accelerate key operations such as matrix-vector multiplication with massive parallelism and low power consumption.
Among emerging resistive memory devices investigated for IMC, the Electrochemical Random Access Memory (ECRAM) is receiving increasing attention for its unique combination of favorable properties. Unlike two-terminal resistive memories such as RRAM, the ECRAM is a three-terminal device with a transistor-like structure: a semiconducting channel, contacted laterally by source and drain electrodes, is covered by a gate stack composed of a solid-state electrolyte layer and an ion reservoir layer, topped by a metal gate electrode.
Applying a gate voltage drives ion migration through the electrolyte, modulating the channel conductance electrochemically. The decoupling between the read and write paths provided by the three-terminal structure results in highly linear and symmetric conductance updates, a high number of addressable analog states, and reduced stochastic variability — all properties critical for accurate analog IMC. This thesis focuses on ECRAM devices based on tungsten oxide (WO_3) as the switching channel, a transition metal oxide offering full CMOS back-end-of-line (BEOL) compatibility, enabling monolithic integration with standard logic circuits.
The fabrication of ECRAM devices and crosspoint arrays was entirely carried out at Polifab, the cleanroom of Politecnico di Milano, employing photolithography, reactive plasma sputtering, e-beam evaporation, and reactive ion etching. An 8-mask process flow was developed for the realization of WO_3-based ECRAM crosspoint arrays on SiO_2/Si substrates. A systematic material study of reactively sputtered WO_3 thin films established the role of oxygen partial pressure and annealing temperature in controlling the film conductance via oxygen vacancy concentration. The ECRAM device stack comprises Pt source/drain electrodes contacting the WO_3 channel laterally, a gate oxide bilayer of electrolyte (Ta_2O_5 or Al_2O_3) and WO_3 reservoir deposited on top, and a Ti/Pt gate electrode, all interconnected in crosspoint arrays via three metal layers separated by SiO_2 spacers (Fig. 1).
Electrical characterization of the fabricated devices revealed robust and reproducible analog switching behavior. DC transfer measurements of ID and IG vs VGS showed progressive hysteretic conductance modulation with increasing sweep amplitude. Pulsed programming experiments demonstrated consistent incremental conductance steps ΔG per pulse, controlled by gate voltage amplitude and pulse width (Fig. 2). Retention measurements showed multi-state stability over timescales exceeding 10^3 seconds, tunable through a gate hold voltage. Two-terminal programming was also investigated, identifying a compact analytical model to describe conductance modulation in the small-signal regime.
A physics-based model of the ECRAM switching mechanism was derived from first principles, describing the nonlinear migration of oxygen vacancies through the electrolyte via an ion-hopping drift-diffusion transport equation. The model was solved both analytically and numerically, yielding a closed-form expression for ΔG that correctly reproduces the nonlinear sinh-like dependence on gate voltage amplitude across pulse durations from 5 to 100 ms, in excellent agreement with experimental data (Fig. 3). A compact analytical formula for ΔG was also derived, providing a computationally efficient physics-based description for IMC system-level simulations.
Finally, ECRAM devices were employed in array-level demonstrations of in-memory computing. Closed-loop programming of crosspoint arrays was demonstrated, followed by experimental realizations of matrix-vector multiplication and its transpose, including rotation matrix operations. Scalar and outer products were demonstrated on 2×2 arrays and formally derived from the physical model. System-level DNN training simulations incorporating realistic ECRAM device models confirmed the potential of the technology for next-generation AI hardware acceleration.

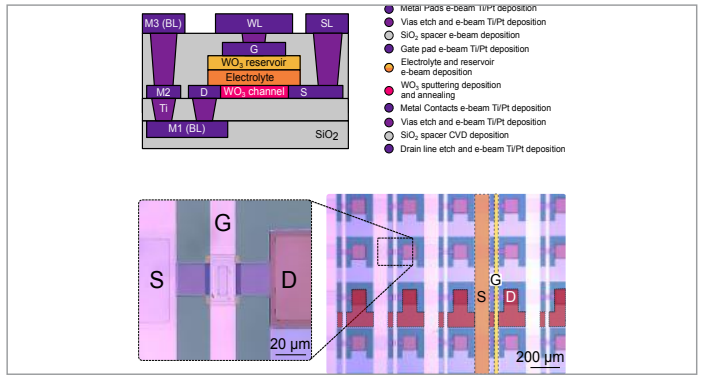

Fig. 1 - Cross-sectional schematic of the fabricated WO_3-based ECRAM cell in crosspoint array configuration, showing the full 8-step BEOL-compatible process flow. The device stack comprises a WO_3 semiconducting channel contacted laterally by Pt source (S) and drain (D) electrodes, with an electrolyte/WO_3 reservoir oxide bilayer and a Ti/Pt gate (G) on top, interconnected by three metal layers (M1–M3) separated by SiO_2 spacers. Optical microscopy images show a close-up of a single fabricated cell (left) and an overview of a crosspoint array (right).

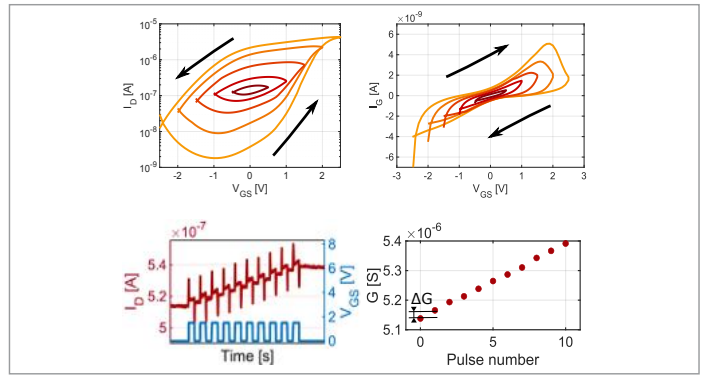

Fig. 2 - Electrical characterization of ECRAM devices. (top left) DC transfer hysteresis of drain current ID vs gate voltage VGS for increasing sweep amplitudes, showing progressive conductance modulation. (top right) Corresponding gate current IG vs VGS, capturing the ionic current during programming. (bottom left) Time-domain response of ID to a sequence of gate voltage pulses. (bottom right) Extracted conductance G as a function of pulse number, showing the incremental ΔG per pulse step

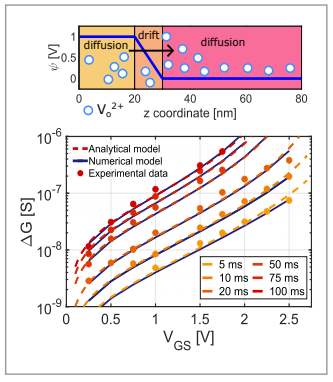

Fig. 3 - Physics-based model of ECRAM switching. (top) Schematic of the electrostatic potential ψ across the device layers, illustrating drift-dominated oxygen vacancy (Vo^{2+}) migration through the electrolyte and diffusion-dominated dynamics in the reservoir and channel. (bottom) Conductance update ΔG as a function of gate voltage amplitude VGS for pulse durations from 5 to 100 ms: experimental data (dots) are accurately reproduced by both the analytical model (dashed lines) and the numerical model (solid lines), capturing the nonlinear sinh-like dependence on voltage.

CONVERSATIONAL AI FOR INCLUSIVE WEB BROWSING: THE CONWEB APPROACH

Emanuele Pucci - Supervisor: Maristella Matera

The evolution of the Web from static pages to dynamic, conversational interfaces is reshaping how users interact with digital content. Despite decades of progress in web standards, the digital world remains largely inaccessible to the more than one billion people worldwide living with some form of disability. For Blind and Visually Impaired (BVI) users in particular, conventional web browsing poses persistent and significant challenges. Web pages are structured around visual hierarchies (menus, buttons, banners) that are inherently difficult to navigate with traditional assistive technologies such as screen readers. These tools, though valuable, tend to provide a linear and cognitively demanding experience, often failing to offer users a meaningful sense of orientation or control.
This thesis addresses these challenges by introducing the Conversational Web (ConWeb) framework, a novel approach that reimagines web accessibility by enabling users to navigate and interact with websites through natural language conversation. Rather than retrofitting accessibility onto existing systems, ConWeb integrates inclusivity into the very foundation of web interaction, turning the browsing experience into an ongoing, voice-driven dialogue between the user and digital content.
The research followed a human-centered design (HCD) methodology throughout. The process began with preliminary interviews aimed at understanding the unique challenges BVI users face when navigating the Web. These were followed by questionnaires and iterative co-design sessions that brought together BVI individuals, accessibility experts, and developers. Through this participatory approach, the research identified key pain points and design dimensions, ultimately leading to the definition of a comprehensive library of conversational patterns for voice-based web interaction.
Central to the ConWeb framework is the Conversation-oriented Navigation Tree (CNT), a hierarchical data structure that organizes web content into logical nodes navigable through simple voice commands. This abstraction provides BVI users with a clear mental map of a website's content, greatly reducing the disorientation that often accompanies screen reader use. Building on this foundation, ConWeb defines four categories of conversational patterns: orientation patterns (which help users understand their current context), navigation patterns (which guide users through the browsing journey), content reading patterns (which deliver information in a digestible format), and conversation control patterns (which offer scaffolding and repair mechanisms). Together, these patterns constitute a reusable design vocabulary for developers creating accessible conversational interfaces.
The technical architecture of ConWeb follows a client-server model. The ConWeb Server manages Natural Language Understanding (NLU), session state, and response generation, while the ConWeb Client handles voice input and output on the user's device. The system incorporates a suite of specialized conversational bots, including a Navigation Bot, a Scaffolding Intents Bot, a Link Bot, and a Content Reading Bot, each responsible for a different dimension of the web interaction experience. Over time, the NLU pipeline was extended to leverage pre-trained Large Language Models (LLMs), significantly improving intent recognition, entity extraction, content summarization, and direct question-answering capabilities.
This integration enabled ConWeb to support more natural, multi-turn dialogues and to adapt its responses to individual user needs.
User studies conducted with BVI participants demonstrated ConWeb's effectiveness in real-world browsing scenarios. Participants reported reduced cognitive load, improved orientation, greater ease in accessing information, and a more satisfying overall web experience compared to conventional screen readers. These findings validate the participatory design decisions made throughout the project and confirm that conversational AI holds transformative potential for creating more inclusive digital environments.
A dedicated chapter of the thesis evaluates the quality of AI-generated text summaries in Italian, a language underrepresented in accessibility research. The study compared extractive and abstractive summarization models using both automatic metrics (ROUGE, BLEU, BertScore) and human crowdsourced evaluation. Results revealed that traditional automatic metrics, which rely on lexical overlap, systematically underestimate the quality of generative summaries, which prioritize semantic coherence and readability. Human evaluators consistently rated GPT-generated summaries higher in fluency and informativeness. These findings underscore the need for richer, semantics-aware evaluation methodologies, and highlight important accessibility implications: while AI summarization can significantly reduce cognitive load for BVI users, it must be deployed with care to avoid oversimplification and loss of nuance.
The thesis also extends ConWeb to address the needs of users with dysarthria, a motor speech disorder that severely limits the effectiveness of standard voice-based interfaces. Through a new round of co-design sessions conducted directly with dysarthric participants, the research identified a set of multimodal conversational patterns that combine limited-vocabulary voice commands with visual augmentation and graphical navigation cues. The resulting system, built on a dysarthric-speech-trained ASR model, demonstrated that purely voice-based or purely visual approaches are insufficient: effective accessibility requires hybrid interaction modes tailored to individual impairment levels. This extension broadens ConWeb's scope and contributes to the emerging field of inclusive conversational AI.
Looking beyond web browsing, the thesis explores two further frontiers. The first concerns the integration of ConWeb's conversational patterns into eXtended Reality (XR) environments. As virtual and augmented reality technologies become increasingly embedded in everyday life, the need for accessible immersive experiences grows urgent. ConWeb offers a promising foundation for voice-driven navigation of 3D content spaces, where traditional web design paradigms no longer apply. The second frontier concerns LLM-powered chat interfaces. As these systems become more widespread, they introduce new accessibility barriers for BVI users, particularly around crafting effective prompts and verifying AI-generated responses. The thesis demonstrates how ConWeb's pattern library can be adapted to this new context, paving the way for more inclusive human-AI interaction.
In conclusion, this thesis makes several interconnected contributions: the definition and validation of the ConWeb paradigm through participatory design; the implementation of a fully functional conversational web architecture; the empirical evaluation of AI summarization for Italian-language accessibility; the design and testing of multimodal patterns for dysarthric users; and early explorations into XR and LLM accessibility. Taken together, these contributions advance the state of the art in web accessibility and conversational AI, and lay the groundwork for a more equitable digital future, one in which technology is designed not merely to accommodate diversity, but to be genuinely empowering for all users, regardless of ability.

REAL-TIME AND ROBUST VIRTUAL SENSING FOR LASER MANUFACTURING VIA LOW-LATENCY DEEP LEARNING AND CROSS-MACHINE ADAPTATION

Alessandro Restifo - Supervisor: Mara Tanelli

Context and objective

Advancements in high-power laser cutting have positioned this technology as an indispensable component within the contemporary manufacturing landscape. Despite the widespread adoption of this technology, maximizing output quality and cutting efficiency remains challenging due to process variability and constraints arising from static parameters. This doctoral research addresses these limitations by developing an end-to-end virtual quality sensing pipeline for industrial laser cutting based on coaxial process-emission imaging, spanning the full chain from synchronized multimodal data acquisition to real-time, machine-deployable inference. Our virtual sensors continuously monitor the melt pool and kerf region through a coaxial high-speed camera to estimate quality indicators, including degraded process states such as plasma formation and loss of cut, thereby enabling diagnostic monitoring and prescriptive actions to optimize both cut quality and factory productivity in real time.

Infrastructure and dataset

The initial stage of the research project involved developing a unified hardware and software platform, including lens auto-focus and auto-exposure routines and a real-time processing architecture suitable for online operation. This foundational work enabled the generation of one of the largest multi-machine laser cutting datasets in the literature to date, with over 500 GB of synchronized data collected across multiple industrial machines, materials, thicknesses, and process conditions. In parallel, dedicated preprocessing and annotation procedures were developed to obtain quantitative labels, with particular emphasis on the main industrially relevant defects affecting cutting quality and plant productivity. This infrastructure and dataset provided the necessary basis for the subsequent development of real-time virtual sensors under realistic industrial constraints.

Causal spatiotemporal neural architecture for low-latency state classification

Building upon this foundation, the research first improved and optimized traditional machine learning baseline pipelines to enable sub-millisecond inference latency and establish a performance benchmark. The work then introduced *CUT-Net*, a novel hardware-aware neural architecture designed to solve the real-time performance gap of neural approaches in the literature. Thanks to its innovative architecture and training process, CUT-Net achieves 153 µs inference latency on a single CPU core with a three-state laser cutting classification accuracy of 94.11%. It also achieves up to 450× higher pixel throughput and up to 590× lower memory usage compared to prior high-accuracy approaches, while introducing for the first time an innovative module for temporal integration and stability improvement, resulting in a 67% reduction in prediction jitter.

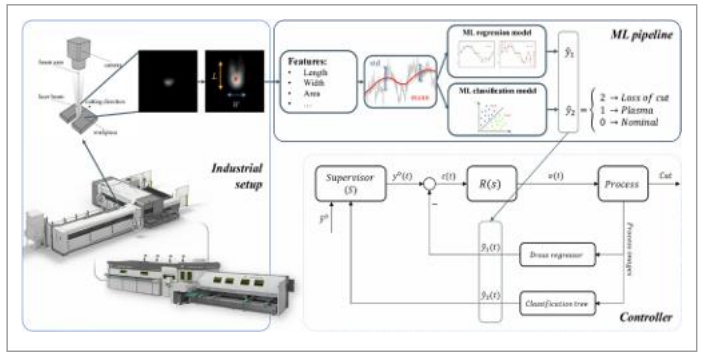

Fig. 1 - Overview of camera-based virtual sensing for industrial laser cutting. Coaxial process-emission images are processed into feature representations and mapped by data-driven models to quality estimates

Zero-shot generalization and weakly-supervised adaptation for cross-machine monitoring under domain shift

While CUT-Net solves the real-time performance challenge, a critical industrial problem remains the domain shift across different machines and setups. To address this challenge, the thesis introduces *GEN-Cut*, a completely novel neural framework for cross-machine generalization. GEN-Cut builds upon the CUT-Net backbone and augments it with architectural and training modifications specifically designed for robustness under strong exposure and scale variations. In particular, the framework introduces a *Scale-Invariant Log Normalizer (SILN)*, together with neural network optimizations and weights merging, to enable stable zero-shot transfer across machines. As a result, GEN-Cut attains 85.5% zero-shot accuracy on unseen cutting machine data, a substantial improvement over previous baselines, and provides a highly effective framework for efficient data usage and weakly supervised adaptation. Using only two unlabeled calibration cuts gathered in situ, the system increases its operational accuracy to 92.7%, enabling monitoring models that can be transferred and recalibrated with minimal target-domain data.

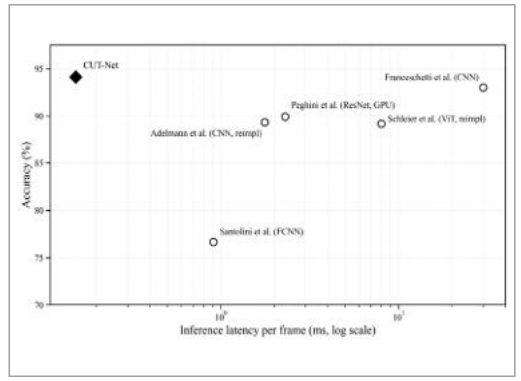

Fig. 2 - Speed vs accuracy comparison highlighting the hardware-aware design of CUT-Net and its suitability for CPU-based real-time deployment

Impact

In summary, this research provides the state-of-the-art frameworks and technologies needed for building scalable, adaptive, and extremely efficient virtual sensing solutions for laser cutting. The work addresses the lack of robustness of traditional machine learning approaches based on handcrafted feature extraction, the trade-off between accuracy and latency in current deep learning models and, most importantly, the industrial challenge of cross-machine generalization and domain shift. The findings provide the foundational infrastructure for the upcoming generation of autonomous and self-optimizing laser manufacturing systems, able to adapt to different working conditions and materials, by targeting optimal cutting quality and maximal industrial throughput in real time.

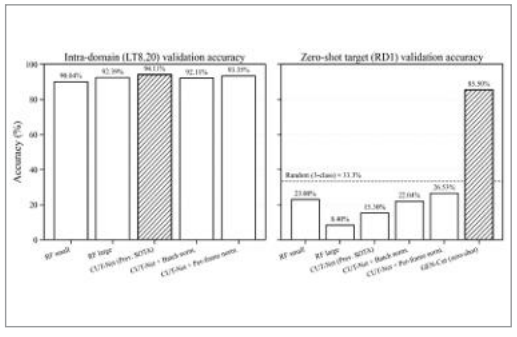

Fig. 3 - Intra-domain accuracy on the source machine versus zero-shot accuracy on the target machine. Machine-specific baselines retain high performance on the source domain but collapse under domain shift, while GEN-Cut maintains strong zero-shot generalization

AN ADVANCED FRAMEWORK FOR FAULT RESILIENCY IN HPC APPLICATIONS FOCUSING ON NOVEL MPI FEATURES AND ENERGY IMPLICATIONS

Roberto Rocco – Supervisor: Gianluca Palermo

Over the years, the HPC field changed its shape, moving from supercomputers to clusters of machines. These changes were necessary to overcome the limitations that arose during the years, such as the end of Moore's law and Dennard scaling. Those empirical laws guided the growth of computational capabilities, but physical constraints eventually made it impossible to sustain the same trend. Despite these challenges, we reached and overcame the ExaFLOPS boundary with the advent of the Frontier cluster, a remarkable achievement that was unthinkable only a few years ago. While computational efficiency has significantly improved and remains a key focus of modern HPC systems, further progress is required to move beyond Exascale. It is necessary to address all sources of inefficiency: some degrade performance (e.g., load imbalance or improper memory allocation), while others waste resources, increasing the energy-to-result. Fault presence is a paramount example of the latter.

The presence of faults in HPC execution has been an overlooked and usually underrated issue. The main reasoning behind these considerations is that HPC executions involve reliable and tested components running in a relatively isolated environment in the best conditions. This statement is accurate, and it has always been valid. However, alongside the growth in performance, we see an increase in the number of components part of an HPC cluster. All these additional resources imply additional connections, thus additional components that may feature a fault. So, while all being reliable, the probability that any of these components feature a fault becomes non-negligible, going from days to theoretically minutes for the most potent HPC clusters. The problem also comes from the usage of the Message Passing Interface (MPI). MPI has been the de facto standard for inter-process communication since its first definition in 1994, and it evolved by introducing functionalities like dynamic executions, I/O operations and even more. Despite all these efforts, MPI cannot handle faults, and they often lead to undefined behaviour. While this was not a significant limitation 30 years ago, it is starting to constrain MPI usability at scale now.

The most adopted approach to deal with faults in MPI applications is leveraging Checkpoint and Restart (C/R), thus saving the execution status periodically so that we can continue from it upon failure. While C/R reduces the amount of work lost due to faults (and, consequently, resource and energy waste), it comes with some significant drawbacks: first, saving the state of an application introduces overhead and additional I/O load, even in the absence of faults. Moreover, with the increase in the execution size, the size of the state grows as well, introducing scalability concerns. Nonetheless, C/R adapts to all applications, and the execution reverts to a consistent state, completely masking the effect of faults on the result.

Alongside C/R, the literature features many other techniques to deal with fault presence in MPI executions. These techniques include message logging, process migration, graceful degradation, process replication and algorithm-based fault tolerance (ABFT). These techniques are not as universal as C/R but may better fit the characteristics of the applications, so it is essential to consider them. However, the area of the literature receiving the most attention is one of the tools to continue MPI execution even in the presence of faults: in this category, we find the User Level Fault Mitigation (ULFM) extension and the Reinit proposal. These tools are compatible with all the other fault management techniques and are mandatory to ensure that the execution continues without needing to restart, thus avoiding initialisation costs and, potentially, runtime scheduler queuing times.

Since ULFM and Reinit are compatible with the other fault management mechanisms, many efforts combined them to produce all-in-one frameworks. Most of these frameworks employ C/R strategies, while our previous work called Legio adopts graceful degradation, thus achieving faster recovery at the cost of impacting result accuracy. This tradeoff is not always viable but perfect for embarrassingly parallel applications and yields better results in terms of overhead and, therefore, energy consumption when the impact on accuracy is moderate.

Despite these efforts, several key challenges remain. Existing works mainly evaluate fault management strategies in terms of temporal overhead, which, while useful as a proxy, does not provide a direct quantification of the energy cost associated with faults and recovery mechanisms. This limits the ability to assess their real impact on resource waste. Moreover, there is a lack of general methodologies to introduce fault resilience (i.e., the ability to continue execution despite faults) into arbitrary MPI applications without requiring extensive redesign. Another limitation lies in the limited support for recently introduced MPI features, whose behaviour in the presence of faults is still largely unexplored. Finally, the literature provides little insight into the validity and usability of approximate results produced during fault-affected executions, despite their potential to avoid costly recomputations and reduce overall energy consumption.

This thesis addresses these challenges by extending the Legio fault resilience framework beyond its original scope. First, we introduce support for critical processes, whose correct completion is necessary to obtain meaningful results. This extension enables the application of graceful degradation techniques to a broader class of MPI applications, including those that are not embarrassingly parallel. Second, we extend the framework to support both widely used and recently introduced MPI functionalities, including Cartesian topologies, group-based communicator creation, and the Session initialization model, providing one of the first analyses of their behaviour under fault conditions. Third, we evaluate the integration of the framework in real-world applications with different communication patterns, assessing the impact of faults on result accuracy and analysing the usability of the outputs obtained in faulty executions. Finally, we model the energy impact of faults and fault management strategies, comparing their effects on both execution time and energy consumption, and providing a systematic evaluation of their benefits in realistic scenarios.

Overall, this work aims to improve the resilience and energy efficiency of HPC systems by providing practical solutions and insights for fault management in MPI applications, contributing to the advancement of next-generation large-scale computing systems.

A REINFORCEMENT LEARNING FRAMEWORK FOR MULTI-OBJECTIVE FORECAST-INFORMED RESERVOIR OPERATION

Davide Spinelli - Supervisor: Andrea Castelletti

Multi-Objective Forecast-Informed Reservoir Operation. Managing multi-purpose water reservoirs requires balancing competing social, economic, and environmental goals. Climate-driven shifts in hydrological patterns increasingly render traditional operating rules ineffective, necessitating a shift toward more adaptive, data-driven paradigms. In the Adda River basin, Lake Como's management is defined by the conflict between local and downstream water requirements and the market-driven behavior of the upstream hydropower network. This dissertation develops a Reinforcement Learning framework for Forecast-Informed Reservoir Operation (FIRO) to enhance operational flexibility and resilience. By formalizing management as a multi-objective control problem, the research provides a methodology to natively process ensemble forecast uncertainty, automate the selection and interpretation of predictors, and design coordinated governance mechanisms based on diverse hydro-meteorological information.

Reservoir Control using Ensemble Forecast Uncertainty and Cross-Validation. This research addresses the dual challenge of utilizing probabilistic forecast information while ensuring policy robustness against overfitting. To address the common loss of information when ensemble forecasts are reduced to a single deterministic mean, the Parallel Ensemble foreCAst coNtrol (PECAN) algorithm is developed to process full probabilistic distributions natively. PECAN simulates system responses for each ensemble member in parallel via virtual reservoir copies, applying nonlinear constraints to each scenario before aggregating the virtual responses into a final control decision. To ensure out-of-sample validity, this methodology integrates a Blocked K-Fold Cross-Validation framework and a dynamic early stopping criterion, preserving temporal integrity by using contiguous sequential blocks to prevent overfitting.
Results across diverse climatic regimes, ranging from multi-year droughts to heavy flooding, demonstrate that this combined approach outperforms traditional ensemble-average methods. By remaining sensitive to outlier members and extreme events, the methodology facilitates a proactive hedging stance that is typically suppressed in deterministic strategies. Furthermore, PECAN achieves performance levels competitive with specialized, high-accuracy local forecasts despite using more readily available global products. This confirms that the integrated design and validation process extracts significant operational value from standard forecast products while ensuring that the learned control logic generalizes effectively to unseen hydrological conditions, providing a reliable basis for operational deployment.

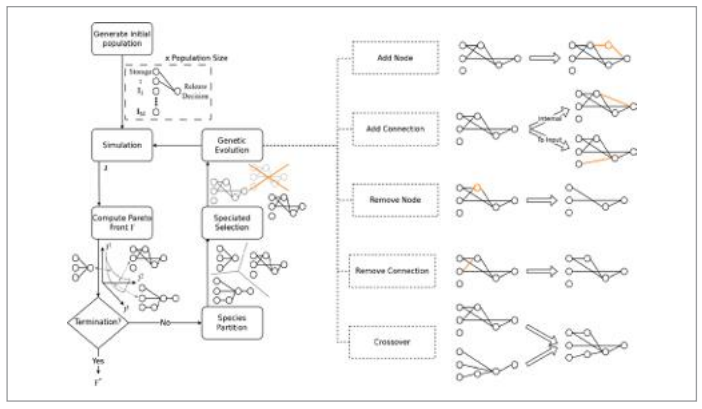

Fig. 1 - Methodological workflow of the extended NEMODPS framework

Automated Input Selection and Policy Interpretation. The optimal set of inputs for a control policy is highly basin-specific and varies according to the desired trade-off between conflicting objectives. Traditional design methods often require predefined input sets, creating a dependency on expert knowledge that may be unavailable or limited in many real-world applications. This research addresses the dual challenges of automated feature selection and model interpretability by extending the Neuro-Evolutionary Multi-Objective Direct Policy Search (NEMODPS) algorithm. As illustrated in Fig. 1, the algorithm concurrently optimizes a policy's parameters, neural architecture, and input set from a broad pool of candidate variables. By treating input connections as evolvable components of the network topology, the framework automatically discovers the most valuable information for each trade-off, achieving performance comparable to state-of-the-art methods without requiring prior expert knowledge.
To render these complex policies transparent, an adapted Time-Varying Sensitivity Analysis (TVSA) employs an absolute, perturbation-driven metric to identify which specific forecasts are selected and visualize how and when they influence release decisions throughout the annual cycle. This diagnostic tool reveals sophisticated, objective-specific strategies: as shown in Fig. 2, policies prioritizing flood control rely on short-range forecasts to manage immediate risks, while leveraging medium-range information during the spring filling season to anticipate and mitigate longer-term deficits. Conversely, agricultural-focused policies adopt reactive strategies, using reservoir state and seasonal timing as primary drivers while utilizing short-term forecasts for corrective adjustments. When using real-world forecasts, the framework demonstrates a rational adaptation to forecast uncertainty, learning to prioritize reliable local information over less certain global ensemble products. By "opening the black box," this methodology provides clear visualizations of how, when, and for which objectives different information sources influence decisions, bridging the gap between advanced artificial intelligence and the practical requirements of water management.

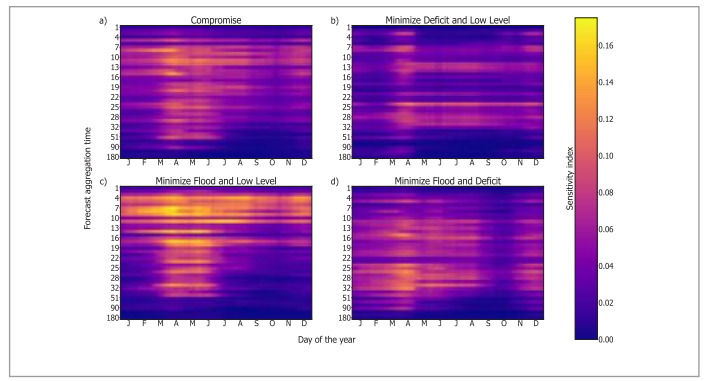

Fig. 2 - Time-Varying Sensitivity Analysis of forecast input importance for four representative policy trade-offs

Managing Multi-Authority Conflicts via Parametrized Restitution Rules. Geographical and institutional fragmentation in the Adda River basin causes operational conflicts between upstream hydropower and Lake Como, externalizing hydrological risks to the agricultural sector. Using an extended basin-wide model, this research quantifies that full system integration could reduce the agricultural deficit by half at a minimal cost to hydropower revenue. To facilitate practical coordination, a novel parametrized restitution rule is developed as a semi-cooperative negotiation instrument based on a proportional storage balance principle. Extensive simulations demonstrate that this rule generates win-win solutions for downstream water security while identifying historical storage exemptions as inefficient legacies that degrade system performance.

AI-DRIVEN ROBOTICS FOR MANUFACTURING AND WAREHOUSE ITEM HANDLING

Alessandra Tafuro - Supervisor: Paolo Rocco

This thesis explores the development of AI-powered robotic systems that transcend the traditional role of rigid, pre-programmed machines. It addresses the growing demand for intelligent and adaptive automation in modern industrial environments, which are increasingly defined by high variability and product customization. As manufacturing and logistics shift toward high-mix, low-volume production, conventional automation, reliant on precise models and fixed routines, has become progressively insu!cient. In response, this research advances the field of intelligent robotics by integrating artificial intelligence, multi-modal sensing, and human-inspired cognitive capabilities, enabling robots to perceive, reason, and learn within complex and dynamic settings. The pursuit of transforming robots into more autonomous agents is examined across three application domains: robotic surface polishing, deburring of deformable shoe soles, and object grasping in warehouse environments. Each scenario demands a high degree of perception, dexterity, and adaptability, qualities typically associated with human skill. In manufacturing, a polishing path-planning algorithm is developed for free-form polysurfaces, allowing robots to perform finishing operations even when 3D models are unavailable. In the footwear industry, where rubber soles vary in shape, color, and material, two deburring setups are investigated: one with a fixed deburring tool and moving workpiece, and another one with a fixed workpiece and moving tool. For the first configuration, a hybrid learning-from-demonstration framework is proposed, enabling robots to imitate expert actions through visual segmentation and motion modeling. This is further extended with a self-supervised vision-motion learning method that leverages reinforcement learning to associate visual inputs with tool trajectories, eliminating the need for labeled data or CAD models. For the second setup, a complete processing pipeline is developed, encompassing defect detection and execution, and incorporating online vision-based path correction to compensate for workpiece deformation. In the warehouse domain, a grasp learning framework is proposed that enables robots to acquire grasping strategies through experience and adapt them to novel yet similar objects. Overall, the proposed methodologies integrate AI, vision, and control to foster flexible and robust robotic behavior, reducing dependence on expert programming while enhancing robotic autonomy.

ASSESSING THE EFFECTIVENESS OF ANSATZ SELECTION METHODS IN VARIATIONAL QUANTUM ALGORITHMS

Gloria Turati - Supervisor: Paolo Cremonesi
Co-Supervisor: Maurizio Ferrari Dacrema

Quantum computing represents a revolutionary computational paradigm that exploits the properties of quantum mechanics to tackle complex problems that are intractable with classical computers.
Despite its promise, quantum computing remains in its early stages and faces several challenges. Significant obstacles include environmental noise and qubit decoherence, which hinder the ability to retrieve the correct solution of a computation. Additional hurdles include limited qubit counts, efficient problem encoding, hardware optimization, and qubit connectivity constraints. Hardware characterized by these limitations is collectively referred to as noisy intermediate-scale quantum (NISQ) devices. Consequently, quantum algorithms designed for NISQ devices must account for a limited number of qubits and utilize shallow circuits to mitigate error propagation.

In this context, Variational Quantum Algorithms (VQAs) have recently emerged as a promising approach to solving complex problems using quantum computing in the NISQ era. VQAs are hybrid quantum-classical algorithms that employ a parameterized quantum circuit, known as ansatz, in combination with a classical optimizer. The objective of these algorithms is to identify the optimal circuit parameters that minimize a cost function representing the problem being addressed. These optimal parameters enable the quantum circuit to effectively approximate the desired solution. A key advantage of VQAs is their reliance on low-depth quantum circuits, which improves robustness to noise and accelerates computation, making them well-suited for current NISQ devices. However, a critical challenge in VQAs is designing an effective ansatz that leverages the specific properties of the problem being addressed while remaining tailored to the adopted hardware. This balance is crucial, as the ansatz structure significantly impacts both computational efficiency and solution quality. To address this challenge, the thesis investigates the ansatz search problem and presents methodologies for designing and selecting effective ansatzes.

After introducing the foundational principles of quantum computing and variational quantum algorithms, the thesis presents an empirical study of the Variational Quantum Linear Solver (VQLS), a variational algorithm for solving systems of linear equations. The study examines whether VQLS can effectively solve linear systems, including large, real-world systems related to fluid dynamics, which have significant applications in industry.
The results reveal that while VQLS performs well on small problem instances or those with a particular problem structure, it struggles in more general scenarios. The primary cause of these low-quality results is likely the insufficient ansatz exploration capability. This observation motivates the search for effective ansatz design strategies throughout the thesis.

The first class of methodologies explored consists of adaptive variational quantum algorithms. These methods iteratively build the quantum circuit during execution, adapting its structure to the problem and hardware requirements. However, a review of the literature revealed that adaptive algorithms have typically been tested on arbitrarily selected problem instances using varying experimental protocols. To address this discrepancy, the thesis undertakes a comprehensive benchmarking study to evaluate the effectiveness of adaptive algorithms under consistent conditions, comparing them to the fixed-structure Quantum Approximate Optimization Algorithm (QAOA).
The results show that adaptive algorithms exhibit similar performance to QAOA in terms of solution quality on the problems considered, but discover circuits with fewer gates, making them more resilient to noise and qubit decoherence. However, their high computational cost often renders them impractical for larger or more complex problem instances.

To explore alternative methodologies and overcome these computational hurdles, the thesis also investigates reinforcement learning (RL) for ansatz design. RL provides a general framework for learning circuit structures without requiring extensive domain expertise or the existence of a large dataset of circuits for building a supervised method, an impractical requirement given the large space of potential circuits.
Specifically, a novel RL algorithm designed for ansatz construction is introduced. This algorithm was able to successfully identify circuits capable of producing high-quality solutions and, notably, discovered a family of circuits particularly well-suited for solving the Maximum Cut problem. While these results highlight the algorithm's potential in this domain, the RL approach's scalability and computational cost remain significantly challenging, indicating the necessity for further improvement.

Given the limitations of adaptive algorithms and reinforcement learning approaches, the thesis also investigates a specific circuit property known as expressibility to better understand its applicability in guiding circuit selection and design. Expressibility measures a circuit's capability to explore the solution space or the energy landscape associated with potential solutions.
Specifically, state expressibility quantifies how uniformly a circuit explores quantum states, while Hamiltonian expressibility evaluates how closely the exploration of the energies corresponding to potential solutions approximates a uniform distribution.
Unlike state expressibility, Hamiltonian expressibility is problem-dependent and has received limited attention in the literature. To address this gap, the thesis employs a Monte Carlo approach to estimate Hamiltonian expressibility across various circuits and problem instances. The study examines its properties and explores its correlation with the quality of solutions obtained with the Variational Quantum Eigensolver (VQE) algorithm. Ultimately, the analysis reveals that Hamiltonian expressibility is a valuable metric for guiding circuit selection, particularly for problems where the Hamiltonian is non-diagonal or, more generally, where the problem solution is in a superposition of states.

After exploring these methods, the thesis compares the ansatz design methodologies (i.e., adaptive algorithms and reinforcement learning) by evaluating their effectiveness on the same set of problems, including specific QUBO and VQLS instances. Alongside this, an analysis of Hamiltonian expressibility on notable ansatzes for these same problem instances is presented.
Among the results, we identify EVQE as the most effective algorithm for the analyzed QUBO problems, while RL shows the highest potential for finding suitable ansatzes for VQLS. Regarding the Hamiltonian expressibility analysis, we find a low-to-moderate correlation with solution quality when evaluated based on Pearson, Spearman, and Kendall Tau coefficients. However, we also observe a moderate mutual information score. This suggests the possibility of a more complex, non-linear relationship between Hamiltonian expressibility and solution quality, which requires further investigation.
This thesis concludes by identifying possible future research directions, among which the enhancement of the RL method for circuit construction is the most promising. The scalability limitations of the RL algorithm could be overcome by shifting from monolithic circuit generation to modular synthesis. In this approach, the RL agent is trained to discover parameterized circuit blocks on small-scale systems. These modular sub-structures can then be used to construct effective ansatzes for larger, computationally demanding problems that share similar properties. Ultimately, this scalable approach would significantly improve the practical applicability of VQAs on currently available quantum devices.

COBOTS UNDERSTANDING SKILLS PROGRAMMED BY DEMONSTRATION

Isacco Zappa - Supervisor: Andrea Maria Zanchettin

Collaborative robots have become a key enabling technology for flexible automation within factories, as they combine precision and repeatability with the ability to operate in environments that remain accessible to human workers. Compared with traditional industrial robots, cobots are designed to support closer human-robot collaboration and are typically equipped with more intuitive programming interfaces, including proprietary high-level languages, graphical block-based environments, and hand-guiding modalities. These characteristics have made collaborative robotics increasingly attractive not only for large manufacturers but also for small and medium-sized enterprises, where ease of deployment, adaptability, and reduced investment barriers are particularly important.

Despite these advances, significant obstacles still limit the effective use of cobots in industrial settings. Many enterprises continue to face difficulties whenever robot tasks must be adapted to product changes, process updates, or evolving production requirements. Although current programming interfaces simplify initial deployment, they often do not eliminate the need for robotics expertise during reconfiguration. As a result, companies frequently remain dependent on external system integrators, which increases cost, prolongs downtime, and discourages broader adoption. This gap between the promised accessibility of collaborative robotics and its practical use highlights the need for programming methodologies that allow non-expert operators to transfer their task knowledge to the robot directly, without requiring extensive coding skills or specialized technical support.

Among the approaches proposed to simplify robot programming, skill-based programming and Programming by Demonstration are particularly relevant. Skill-based programming structures robot behavior around modular and reusable actions rather than low-level motion commands, thereby expressing tasks in a form that is closer to the way human operators reason about them.

A skill is typically defined as an object-centered unit of behavior equipped with preconditions and postconditions, making it reusable across different settings. This paradigm has shown considerable value in industrial applications. However, it still presents important limitations, since the definition of new skills generally requires robotics expertise and the manual specification of skill sequences does not readily accommodate dynamic environments, product variability, or changing goals.

Programming by Demonstration addresses this limitation by enabling robots to learn tasks directly from examples provided by a user rather than from explicit programming. Demonstrations may be delivered through teleoperation, passive observation, virtual reality, or kinesthetic teaching. Among these modalities, kinesthetic demonstration is especially suitable for collaborative robots because it allows the user to guide the robot physically through the desired motion, thus reducing the distance between domain expertise and robot programming.

Nevertheless, existing Programming by Demonstration approaches often require multiple demonstrations to achieve generalization, while industrial implementations frequently remain limited to the recording and replay of waypoints or trajectories. This creates a major constraint in production environments where even small variations in object placement, workspace configuration, or task objectives would require repeated teaching.

For industrial deployment, an effective demonstration-based framework must therefore infer both the meaning of a skill and its execution policy from minimal user input, ideally from a single demonstration.

A fundamental step toward this objective consists in introducing semantic representations that connect low-level robot actions with high-level task understanding. In this perspective, semantics provides a structured description of objects, relations, and actions, allowing the robot to reason about the state of the environment and the consequences of its behavior. Dynamic semantic networks offer a particularly suitable representation, since they model scene entities as nodes and their relationships as labeled edges whose evolution over time captures the transformations induced by an action. By comparing the semantic state of the scene before and after a demonstration, it becomes possible to identify which relations have been activated, deactivated, or preserved, and therefore to infer the semantics of a demonstrated skill in terms of preconditions and effects. This type of representation is especially valuable in industrial contexts because it preserves interpretability and explainability, allowing robot reasoning to remain understandable and predictable to the user. Once formalized in a symbolic planning language, these learned skills can be combined to generate executable action sequences that achieve new goals under different initial conditions.

A further challenge lies in the autonomous parameterization of skill execution. For many manipulation tasks, transferability can be achieved by recording waypoints during demonstration and expressing them relative to meaningful reference entities in the scene rather than to a fixed world frame. This strategy supports the reuse of skills when object positions vary. However, waypoint-based representations are insufficient for continuous tool skills such as welding, polishing, or gluing, where the detailed geometry and continuity of the trajectory are essential to the quality of the result. In these cases, the robot must not only understand the high-level objective of the skill but also extract the relevant low-level parameters from the demonstration in a form that remains executable and generalizable. Existing data-driven solutions often require multiple demonstrations, large datasets, or extensive optimization, which limits their suitability for industrial contexts that demand reliability, rapid deployment, and minimal interruption of production.

Within this framework, three main directions emerge in the doctoral thesis. The first is the extension of semantic Programming by Demonstration to complex industrial tasks requiring richer symbolic descriptions than those employed in simplified laboratory scenarios. The second is the broader exploitation of learned skill descriptors, not only for symbolic planning but also for the segmentation and classification of continuous demonstrations, enabling the robot to reconstruct the sequence of skills shown by the user. The third is the autonomous parameterization of continuous tool skills from a single demonstration, particularly for applications such as robotic welding. These directions converge toward a common objective: enabling robots to acquire from intuitive demonstrations both the logical structure and the execution details of a task, while maintaining transparency, robustness, and practical usability for non-expert operators.

The resulting framework supports a no-coding approach to collaborative robot programming in which non-expert users can teach reusable skills, communicate task goals, and rely on the robot to reason about action sequencing and execution under changing conditions. Symbolic descriptions of relative poses extend the expressive power of semantic reasoning for complex manipulation tasks; multi-agent extensions allow coordinated skill learning and planning in scenarios involving more than one robot; semantic segmentation and classification make it possible to interpret continuous demonstrations and detect unknown actions; and trajectory parameterization methods enable the transfer of continuous motions into robot code. Taken together, these elements contribute to a more flexible, interpretable, and deployable form of industrial automation, in which the knowledge of the operator becomes the central resource for programming collaborative robotic systems.