Advanced Computing in INFN CSN5

Alessandro Lonardo

Istituto Nazionale di Fisica Nucleare

CSN5 Coordinator for Roma - La Sapienza

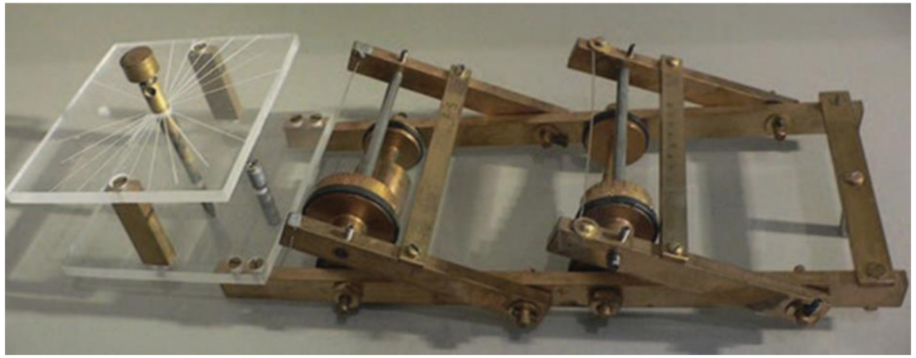
Human Technopole (Milano) - February 20th, 2024

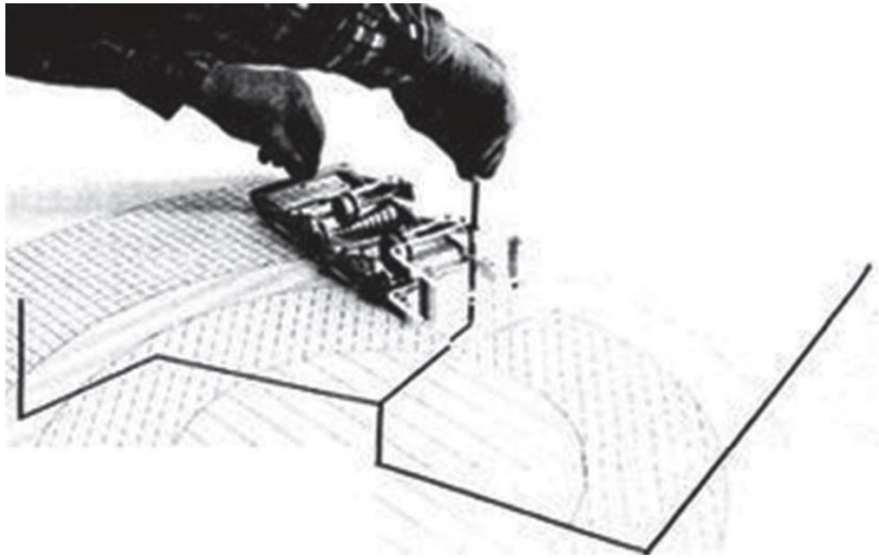
Outline

- A brief history of computing at INFN
- INFN computing infrastructure
- Technological research in computing in CSN5
- (incomplete list of) advanced computing applications in CSN5

Fermiac: Physicists' Attitude Toward Computing

- Conceived by Enrico Fermi in 1947 while the ENIAC was unavailable for a long shutdown due to maintenance and memory upgrade.
- A 30 cm long hand-operated analog computer to study the evolution in time of the neutron popolation in a nuclear device via the Monte Carlo method.
- Operated on a scale drawing of the nucleare device under study.
- Follow the history of each neutron, once fixed the initial conditions.
- Used until 1949.



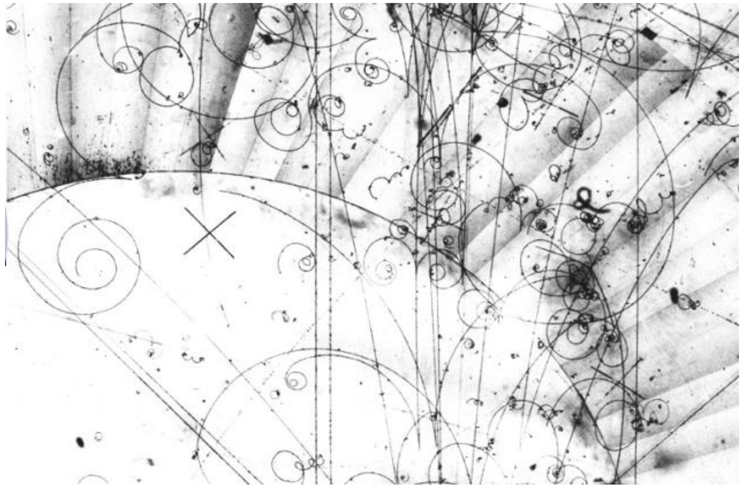


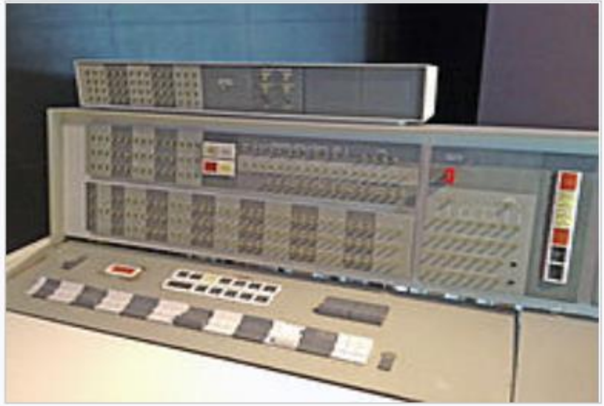
F. Coccetti, The Fermiac or Fermi's Trolley, 2016
DOI: 10.1393/ncc/i2016-16296-7

A brief history of computing @ INFN

The 60s

CNAF ("*Centro Nazionale Analisi Fotogrammi*") in Bologna was founded in 1962, dedicated to what was at the time the most technologically challenging analysis method: bubble chambers images.
This needs computers!





IBM 7094 operator's console showing additional index register displays in a distinctive extra box on top. Note "Multiple Tag Mode" light in the top center.

IBM System/360 Model 44

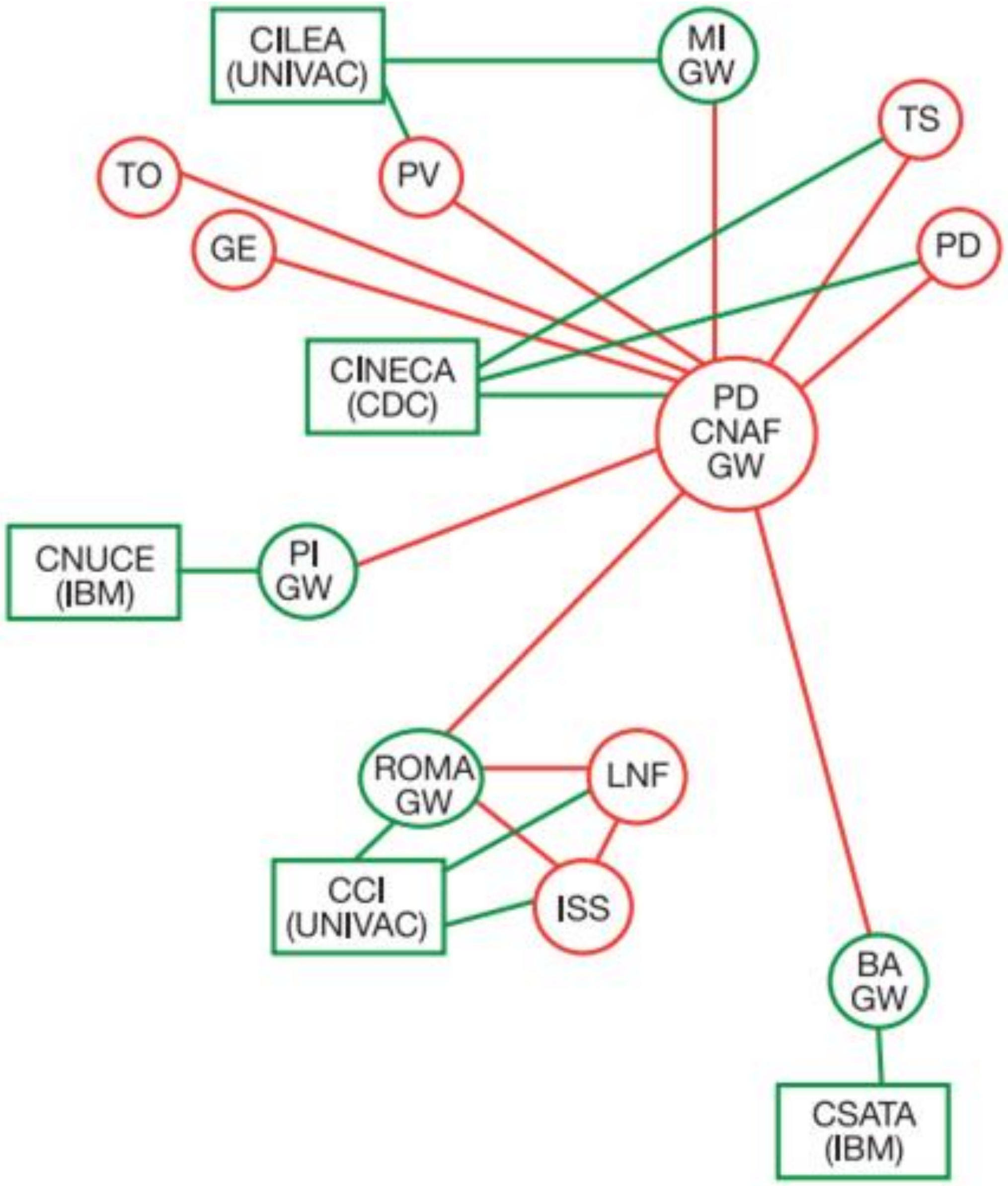
IBM

System/360 Model 44 front panel

Manufacturer	International Business Machines Corporation (IBM)
Product family	System/360
Release date	August 16, 1965
Discontinued	September 23, 1973
Memory	32–256 KB Core

A brief history of computing @ INFN

The 60s → The 70s-80s



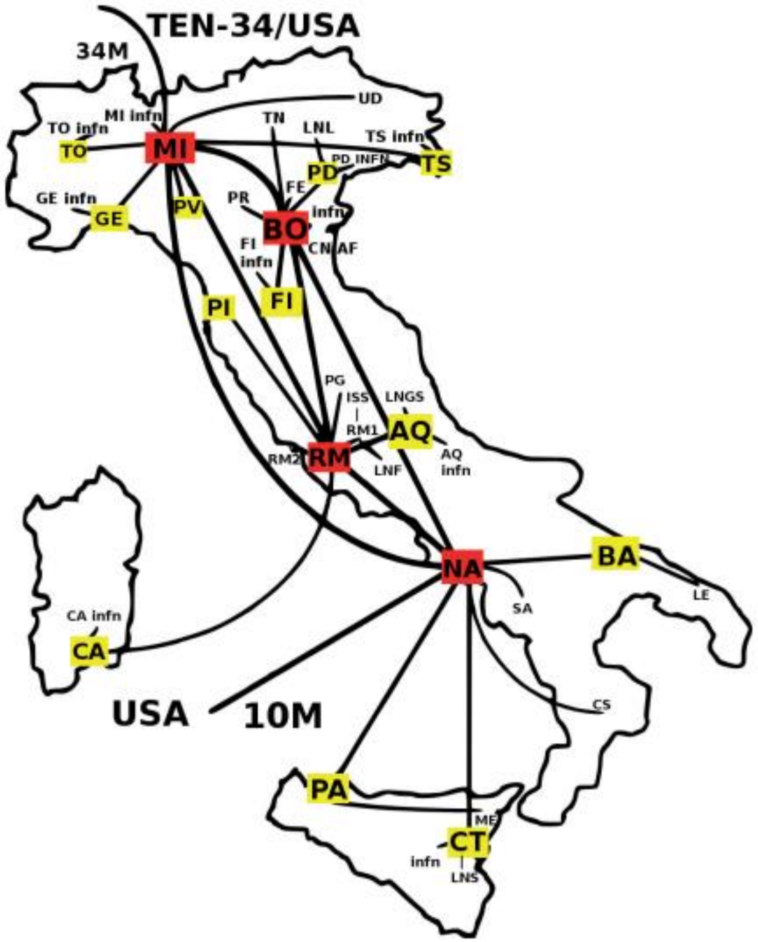
Computers became more and more popular among physicists, and due to the distributed nature of INFN, they were sitting in different structures, mostly handled independently.
From the need to allow an intercommunication, INFNet project was started using dial-up connections. CNAF, with its technology-related mission, became the central node of the effort.
In early 80s, a connection was built to CERN (via CERNet) for direct access and later to FNAL.
At the end of the 80s, INFNet was topping 64 kbit/s

A brief history of computing @ INFN

The 60s → The 70s-80s → 1988-1995

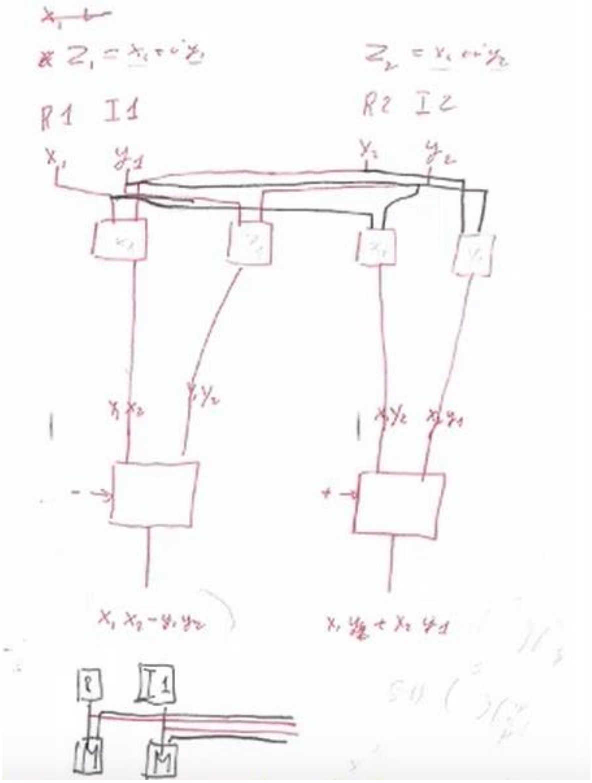
Remote access to computers was quickly becoming a need in other scientific domains; GARR was pioneering 2 Mbps connections by 1988, starting with a CERN-CNAF and later a connection to CINECA, Rome and Milan. This is the infrastructure which handled LEP, TeVatron, SLC computing.

That has with time become the backbone of the research networking in Italy, which reached 34 Mbps by 1995. Still today, it is handled by GARR.
By that time, we were in the planning for the “LHC” era and it was clear how the Computing would have been a major effort for HEP and for INFN. CNAF was again having a central role for INFN Computing.

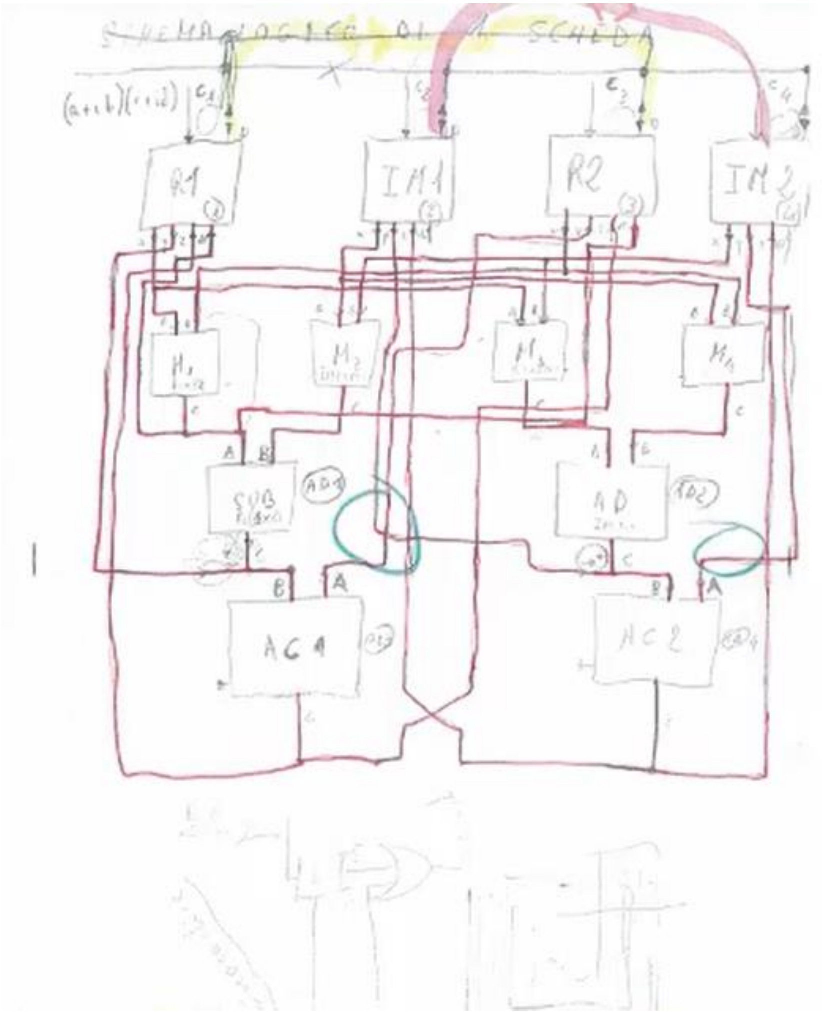


APE (1988-2004): INFN Supercomputers for LCQD

Design of the Floating-Point Unit
(G. Parisi)



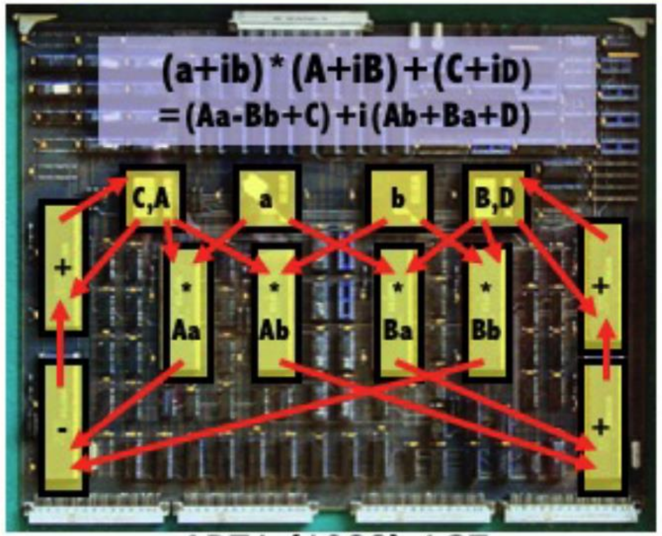
Technical Elaboration of the FPU
(G. Salina)



Left to Right: V. Marinari, P.S. Paolucci,
G. Salina, N. Cabibbo

4 Generations of APE: 1 GFlops to 10 TFlops

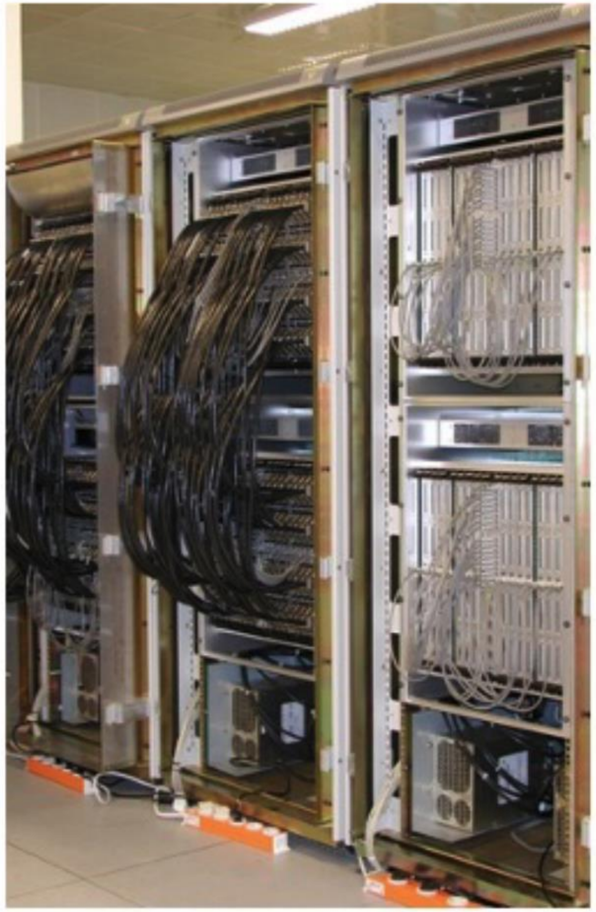
https://doi.org/10.1016/0010-4655(87)90172-X



APE1 (1988) 1GF

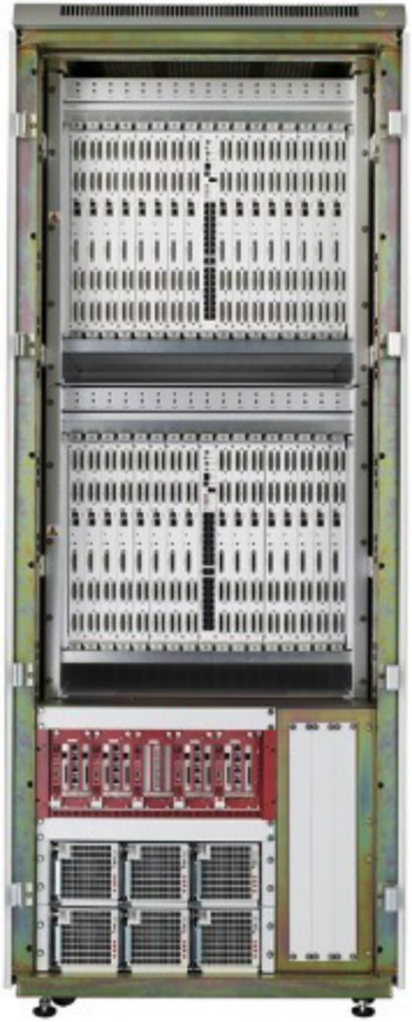
APE100 (1992) 25GF, SP, REAL

https://doi.org/10.1063/1.39557



APEmille (1999) 128GF, SP, Complex

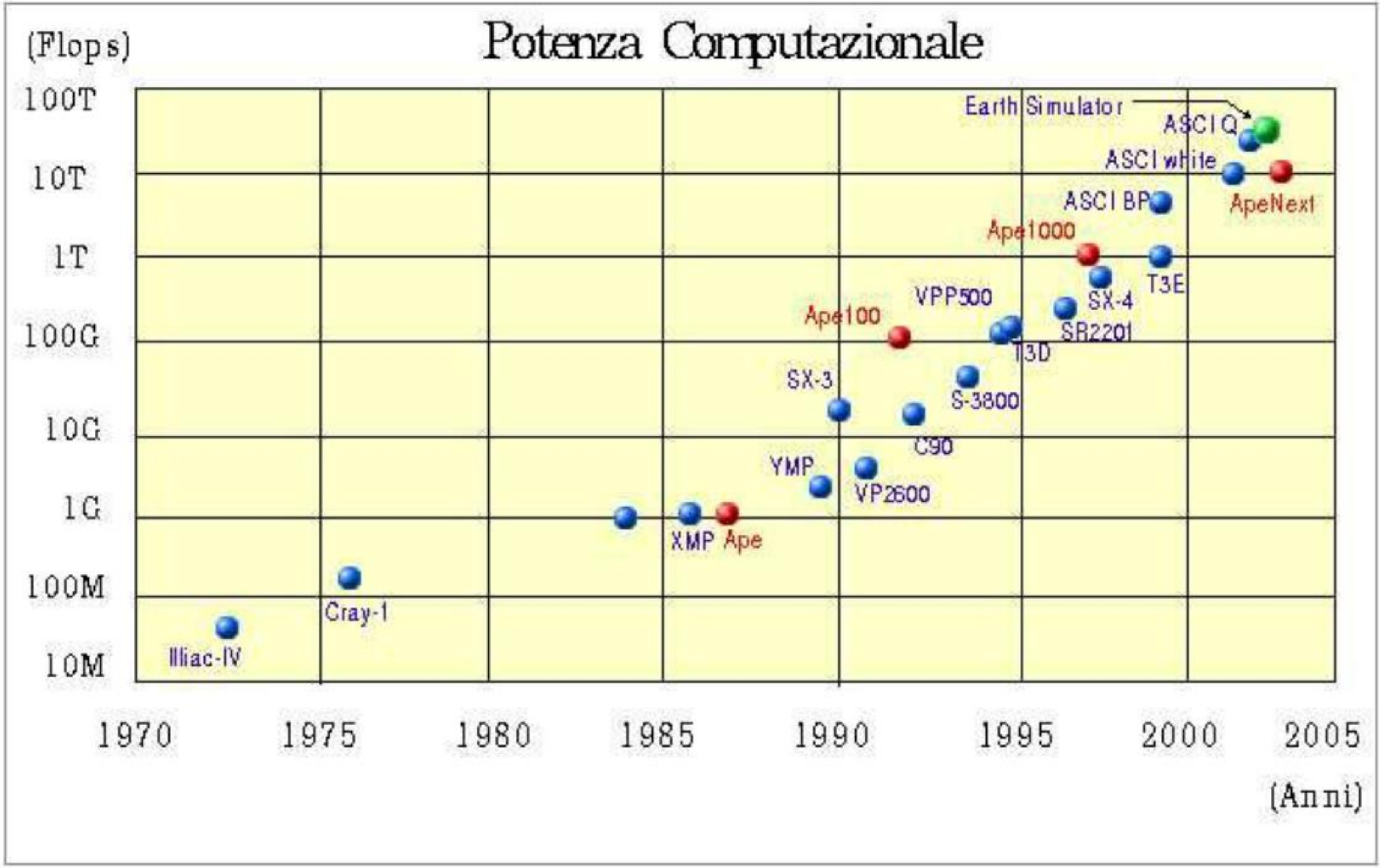
https://doi.org/10.1016/S0920-5632(97)00485-4



apeNEXT (2004) 800GF, DP, Complex

https://doi.org/10.1016/S0920-5632(01)01656-5

4 Generations of APE (1988-2004)



A brief history of computing @ INFN

The 60s → The 70s-80s → 1988-1995 → 2000-2005

Ten centers were selected to host WLCG Computing:

- 1 Tier-1 at CNAF (red in the picture)
- 9 Tier-2s at LNL, LNF, Turin, Milan, Pisa, Rome, Naples, Bari, Catania (yellow in the picture)
- Then came the GRID, the Cloud, …

- They are all still operational, even if their size has increased O(1000x) since then and their interconnectivity (thanks to GARR-X) reaches multiples of 100 Gpbs


Milano
Rome
Catania
Bari
ReCaS
CNAF
Napoli
T2 Catania
Potenziamento
IBISCO
In allestimento
LNF
Turin
Pisa
2022
LNL

A History of Collaborations …

The development of Scientific Computing in INFN was driven by the needs of its theoretical/experimental communities.

Still, being at the forefront of computing in research seeded many projects which have a larger scope.

DataGRID
EGEE Enabling Grids for E-sciencE
LCG
EGI-Engage
EGI-InSPIRE

"preparing the GRID"

EOSC-hub
EOSC Future
interTwin
Skills4EOSC
INDIGO - DataCloud
eXtreme DataCloud
DEEP Hybrid DataCloud
EOSCpilot The European Open Science Cloud for Research Pilot Project
EOSC-Pillar
DICE
HELIX NEBULA THE SCIENCE CLOUD
EGI-ACE

"preparing the Cloud"

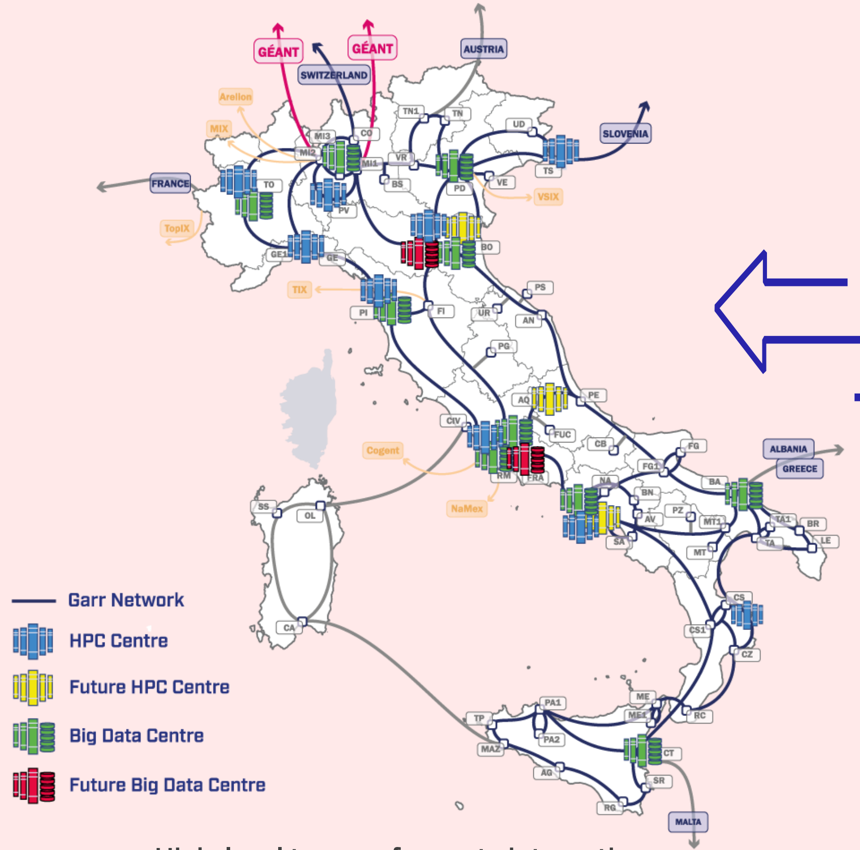
EuroHPC Joint Undertaking
APE
ExaNeSt
textarossa
EUROEXA
RED SEA

"developing HPC Tech"

INFN in the Italian Research Center on HPC, Big Data and Quantum Computing (ICSC)

- Maintenance and upgrade of the Italian HPC and Big Data infrastructure, as well as on the
- Development of advanced methods and numerical applications and software tools to integrate computing, simulation, collection, and analysis of data of interest for research, manufacturing, and society.
- Cloud and distributed approaches.
- 25 Universities,
  12 Research Institutes,
  14 Private companies.

0 SUPERCOMPUTING CLOUD INFRASTRUCTURE



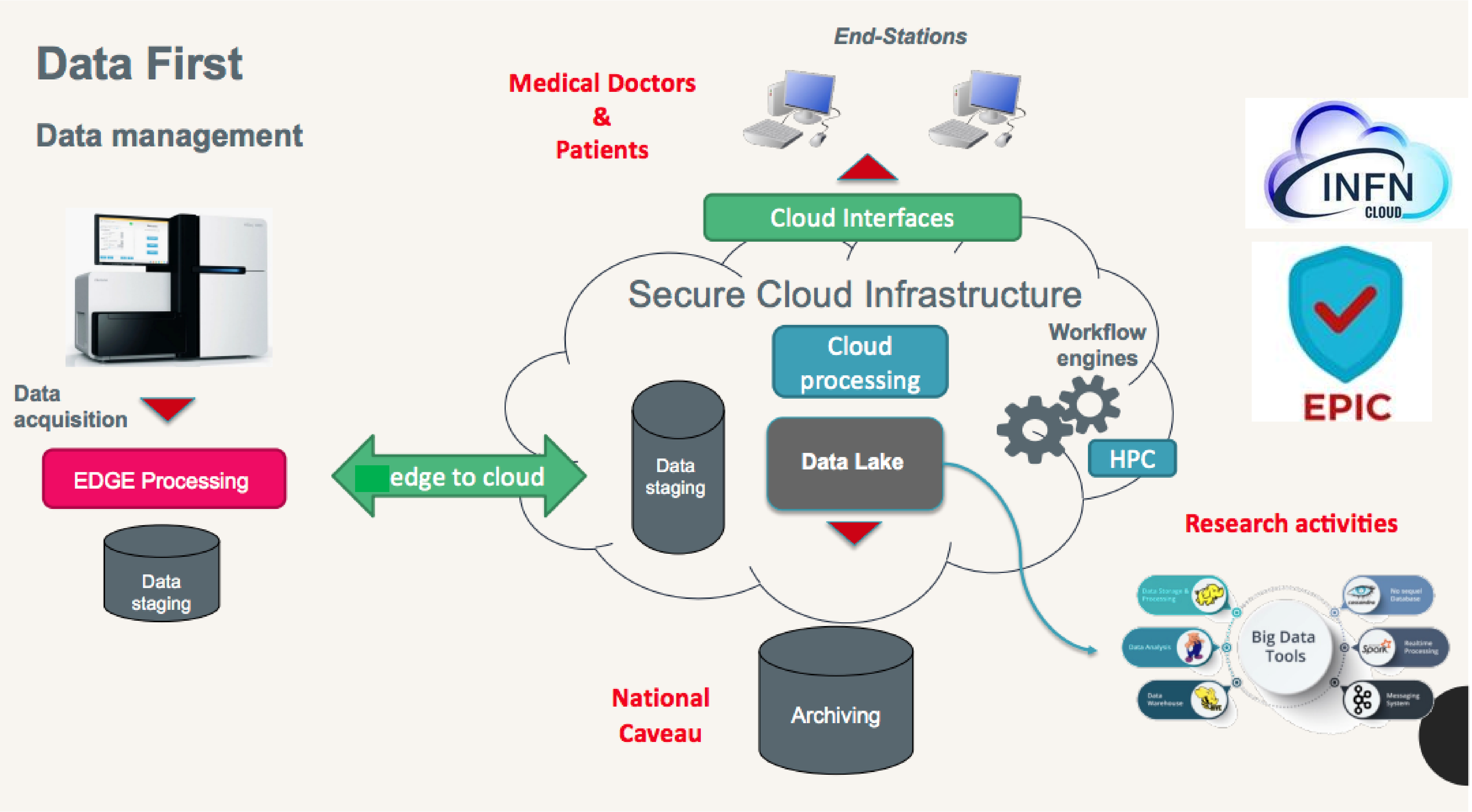
High-level teams of experts integrating the Spokes working groups (mixed cross-sectional teams)

ISTRUZIONE E FORMAZIONE, IMPRENDITORIALITÀ, TRASFERIMENTO DI CONOSCENZE, POLICY, OUTREACH

- **1** FUTURE HPC & BIG DATA
- **2** FUNDAMENTAL RESEARCH & SPACE ECONOMY
- **3** ASTROPHYSICS & COSMOS OBSERVATIONS
- **4** EARTH & CLIMATE
- **5** ENVIRONMENT & NATURAL DISASTERS
- **6** MULTISCALE MODELING & ENGINEERING APPLICATIONS
- **7** MATERIALS & MOLECULAR SCIENCES
- **8** IN-SILICO MEDICINE & OMICS DATA
- **9** DIGITAL SOCIETY & SMART CITIES
- **10** QUANTUM COMPUTING

ICSC Spoke 8 – In-silico Medicine & Omics Data

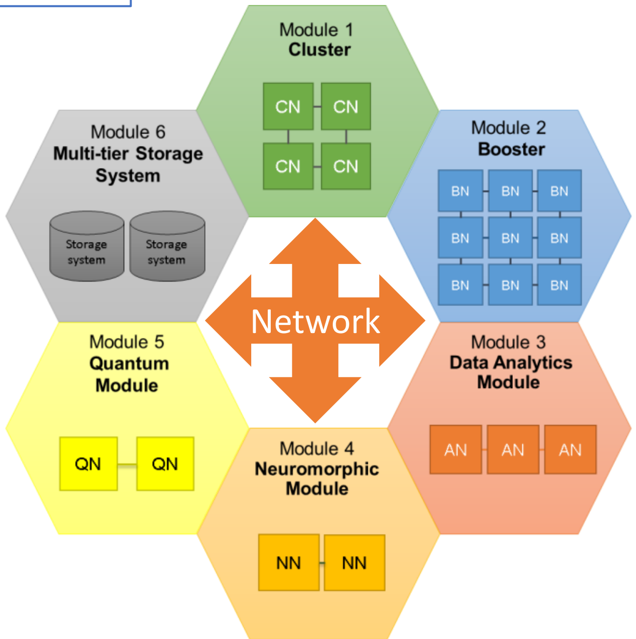
EPIC Cloud (Enhanced PrIvacy and Compliance Cloud) is the cloud service developed and managed by CNAF to fulfill the requirements of projects and experiments dealing with clinical, biomedical and genomic data.

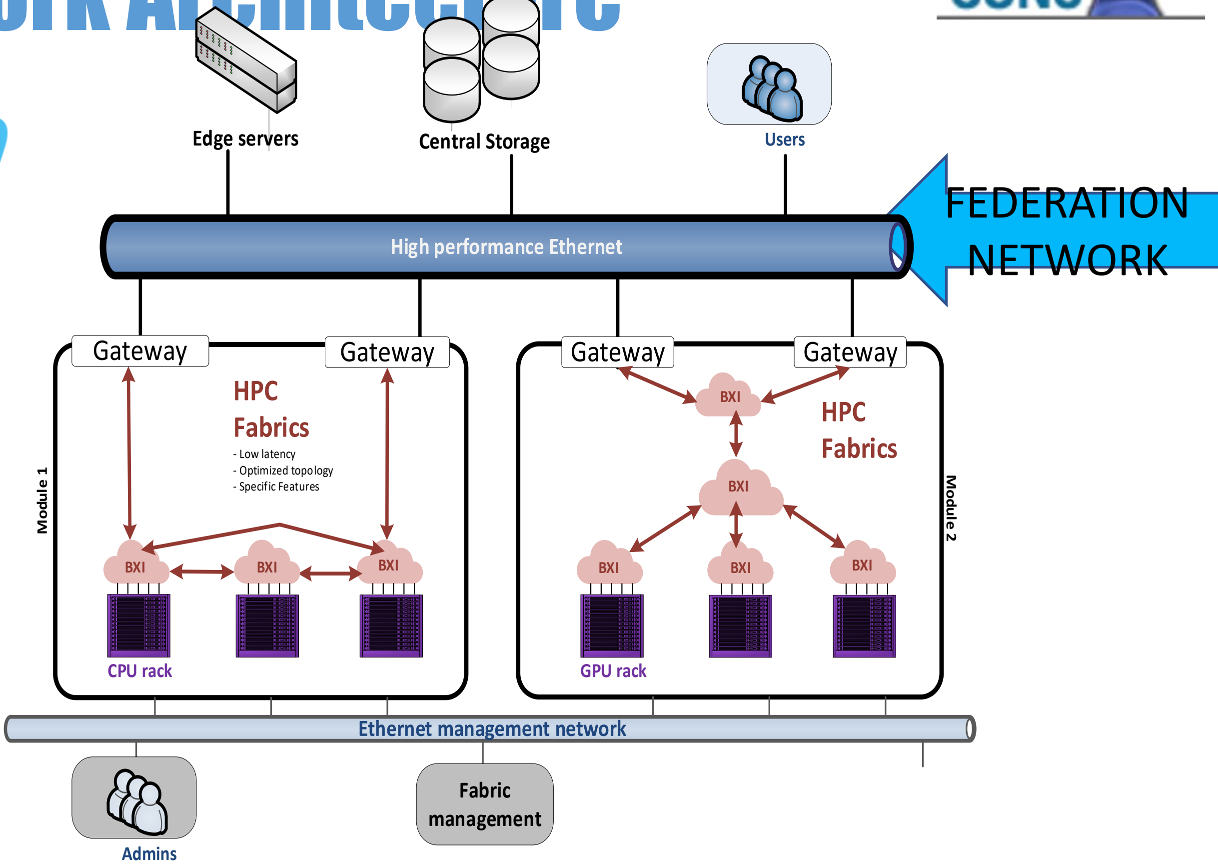


RED-SEA: MSA Network Architecture

EuroHPC Joint Undertaking

RED-SEA





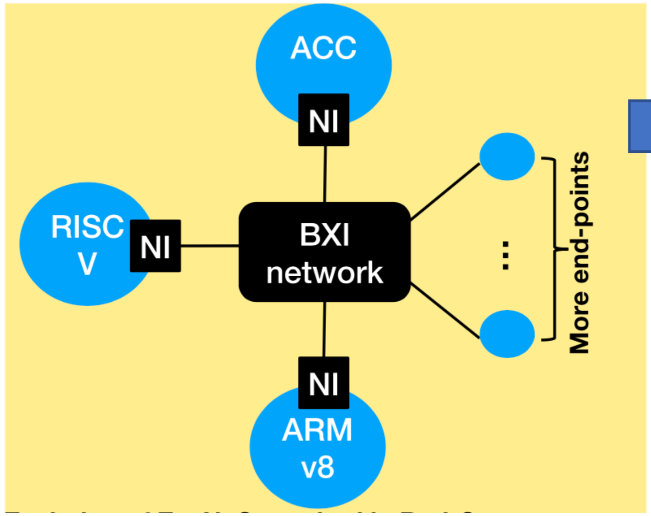
- HPC (High Performance Computing) ; HPDA (High-Performance Data Analytics); AI (Artificial Intelligence)
- Supercomputer: aggregation of resources that are organized to facilitate the mapping of applicative workflows
- HPC is part of the continuum of computing

- High performance Ethernet as federation network featuring state-of-the-art low latency RDMA communication semantics;
- BXI as the HPC fabric consisting of two discrete components, a BXI NIC plus a BXI switch, and the BXI fabric manager.

INFN
Istituto Nazionale di Fisica Nucleare

High Performance Computing
BXI ECOSYSTEM: INFN APEnetX

CSN5



END-POINT: INFN APEnetX

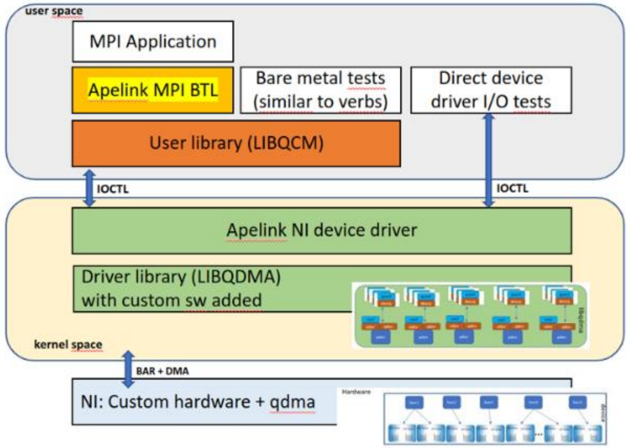
- to tightly **integrate** the network interfaces (NIs) **to RISC-V and ARMv8 cores and to FPGA-based accelerators and GPUs**
  - To prepare a number of EPI-related IPs
  - To create a highly heterogeneous programmable platform connected with state-of-the-art interconnect technologies.

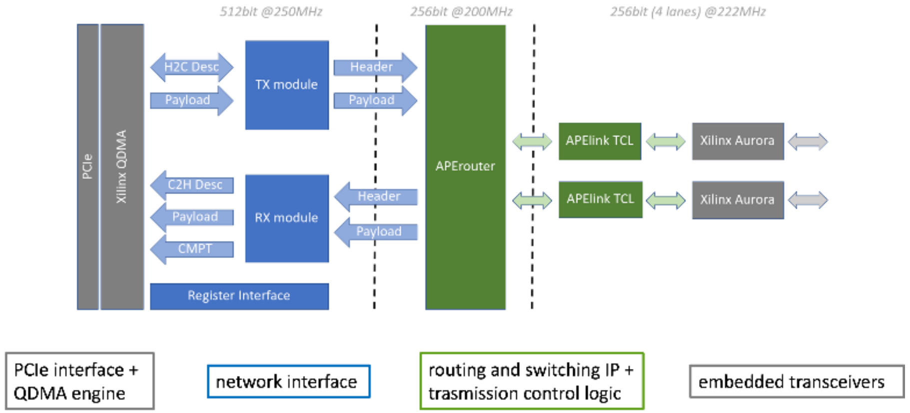
- Network Interface Card (APEnetX)
  - PCIe gen4 (GPU+CPU) + BXI link (Xilinx Alveo FPGA)
- Co-Design through applications (NEST)
- Developing network IPs to optimize spiking neural network communication

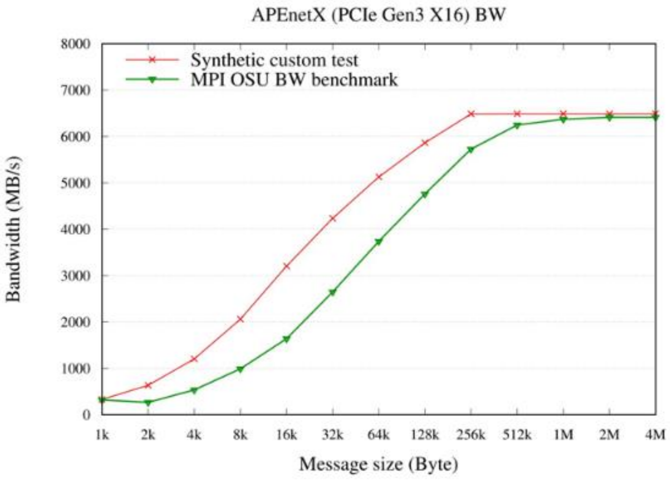
High Performance Computing

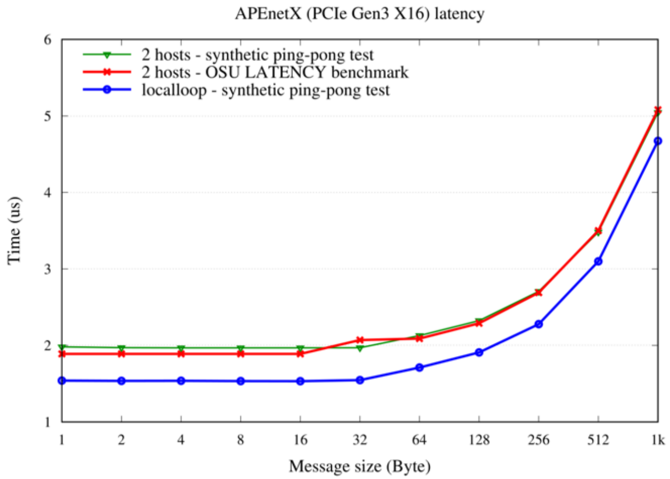
APEnetX Overview

- Xilinx Alveo Board DMA engine
- Matching requirement for the communication generated by NEST
- Providing proprietary software driver and low-level communication library
- NVIDIA GPUDirect RDMA
- Custom OpenMPI BTL
- Bandwidth per channel 57.6 Gbps
- Latency 1.9us
- Validated through HPC-benchmark
- Large-scale simulation environment (NEST traces)
- Interoperability with the BXI interconnect
- Proprietary priority management mechanism to improve QoS of the data transmission system









textarossa

The TEXTAROSSA project aims, among other objectives, to reduce both energy consumption and execution time of an HPC application also through the seamless integration of FPGA accelerators

High-Level Synthesis (HLS) has been considered as a mature enough and promising way to pursue this goal.

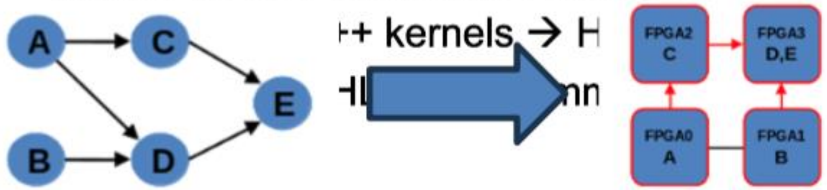
Using multi-FPGA accelerators to implement complex algorithms, not fitting within a single FPGA, is a way to further expand the possibility to exploit FPGA capabilities and to broaden the class of addressable algorithms.

How? We developed a HW/SW framework (APEIRON) extending the HLS workflow to multi-FPGA systems.

High Performance Computing/High Performance Data Analytics
APEIRON

Goal: to offer hardware and software for development and execution of real-time dataflow applications on a system composed by directly interconnected FPGAs by **extending the HLS workflow to MULTI-FPGA systems**

- To map the dataflow graph of the application on the distributed FPGA system offering runtime support
  for its execution
- Allows users with little experience in hw design tool,
  to develop their applications on such system:
- **Kahn Process Networks** Paradigm

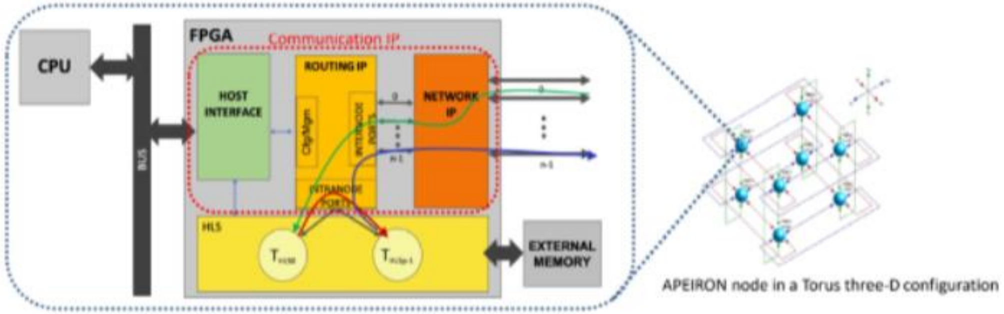


-send(msg, size, dest_node, task_id, ch_id)
-receive(ch_id)

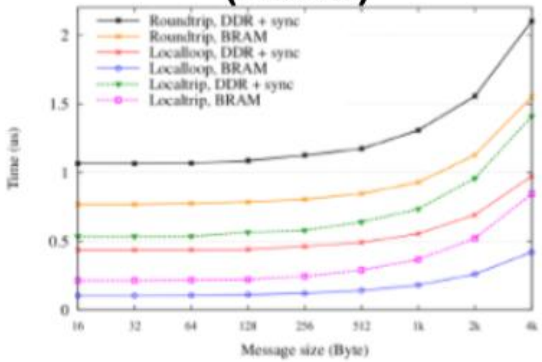
Where :
dest_node are the n-Dim coordinates of the destination node (FPGA) in a n-Dim torus network.
task_id is the local-to-node receiving task (kernel) identifier (0-3).
ch_id is the local-to-task receiving fifo (channel) identifier (0-127).

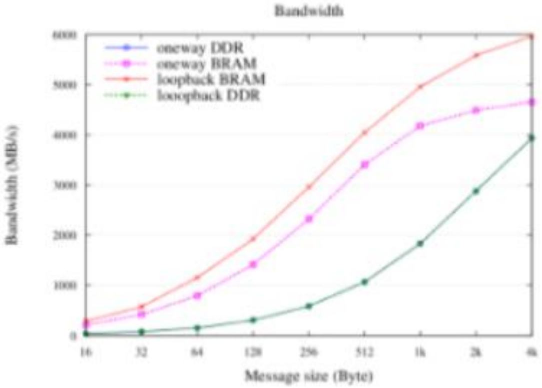
APEIRON is based on the Xilinx Vitis HLS framework and on the INFN Communication IP

- **Direct network of processing tasks (intra/inter communication)**
- **Customized and application dependent I/O: APElink 20/40 Gbps, UDP/IP 10/25 Gbps**



- **Intranode latency: 553ns (DDR); 213ns (BRAM)**
- **Internode latency: 1065ns (DDR); 768ns (BRAM)**

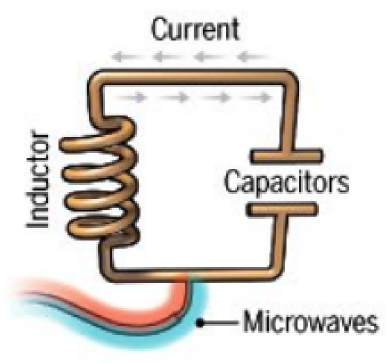




Quantum Computing

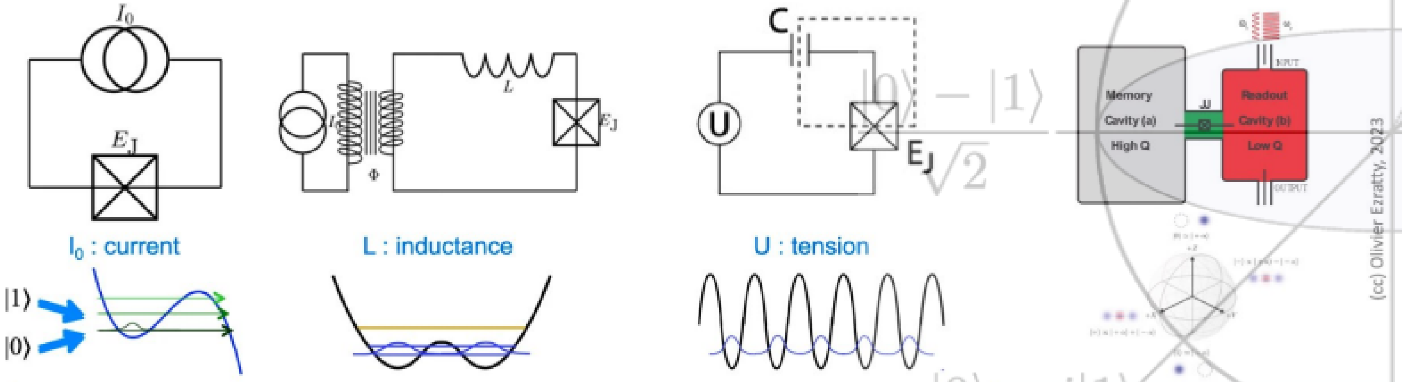
- not yet a standard way to implement qubits, unlike for classical bits encoded in transistors
  - physically, qubits can be any two-level systems: the spin of an electron, the polarization of a proton, …
  - current leading technology in the quantum computing commercial space: superconducting qubits

Superconducting loops



A resistance-free current oscillates back and forth around a circuit loop. An injected microwave signal excites the current into super-position states.

Longevity (seconds)	0.00005
Logic success rate	99.4%
Number entangled	9



	phase qubit	flux qubit	charge qubit - transmon	cat-qubits
	Josephson junctions handle the qubit degree of liberty			Josephson junctions prepare, couple and correct the cat-qubits
|0⟩ and |1⟩ qubits	two energy levels in a potential well	two superconducting current directions	two levels of charge of Cooper pairs	pairs of entangled microwave photons in a cavity
quantum gates	micro-waves	magnetic field	micro-waves	micro-waves
qubits readout	resonator and micro-waves	magnetometer (SQUID)	resonator and micro-waves	resonator and micro-waves
commercial vendors	abandonned	D-Wave The Quantum Computing Company, rigetti, Qilimanjaro Quantum Tech, Alibaba, bleximo, Atlantic Quantum	IBM, Google, IQM, OQC	Alice & Bob, qci, amazon

⊕ **Pros** — fast operation (ns), built on existing semiconductor industry

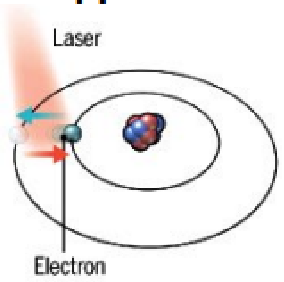
⊖ **Cons** — quickly experience decoherence so requires error correction, have to be kept cold (T<100 mK), only 2D topology (swaps)

O.Ezratty, Eur. Phys. J. A 59, 94 (2023)

Quantum Computing

- multiple other technologies used to implement current quantum processing units

Trapped ions or neutral atoms arrays

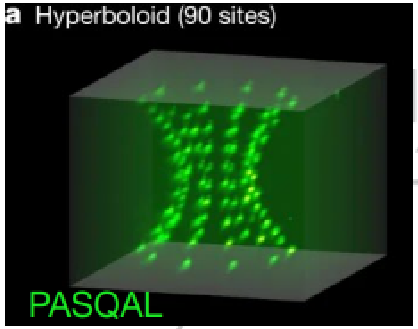


use the energy levels of electrons in neutral atoms or ions as qubits. In their natural state, these electrons occupy the lowest possible energy levels. Using lasers, we can “excite” them to a higher energy level. We can assign the qubit values based on their energy status

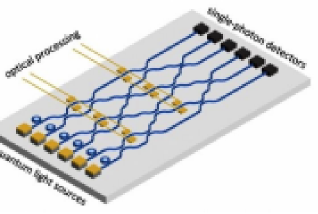
Company support
IonQ, PASQUAL, AQT, Atom Computing, …

⊕ **Pros** very stable, longer decoherence time, high gate fidelity, 2D and 3D, many qbits

⊖ **Cons** slow operations, hard to program, many and sophisticated laser technology needed



Linear / non-linear optical QC

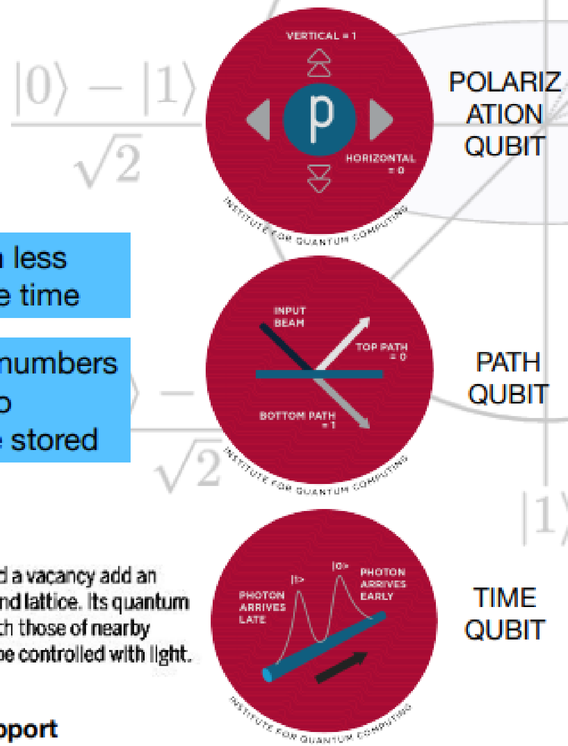


use particles of light to carry and process information. Qubits realised by processing states of different modes of light through both linear (mirrors, beam splitters, phase splitters, …) and nonlinear element (quantum microprocessor based on laser photonics at room temperature)

Company support
Xanadu, PsiQuantum, …

⊕ **Pros** can operate at room temperature, photons much less sensitive to the environment, longer decoherence time

⊖ **Cons** emerging technology, difficult to construct large numbers of gates and connect them in a reliable fashion to perform complex calculation, photons cannot be stored

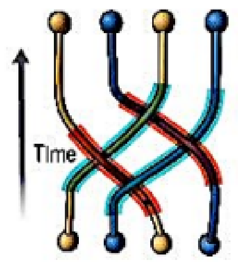


Silicon quantum dots

These “artificial atoms” are made by adding an electron to a small piece of pure silicon. Microwaves control the electron’s quantum state.

Company support
Intel, SQC, HRL, …

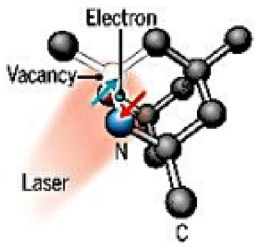
Topological qubits



Quasiparticles can be seen in the behavior of electrons channeled through semi-conductor structures. Their braided paths can encode quantum information.

Company support
Microsoft

Diamond vacancies



A nitrogen atom and a vacancy add an electron to a diamond lattice. Its quantum spin state, along with those of nearby carbon nuclei, can be controlled with light.

Company support
Quantum Diamond Technologies

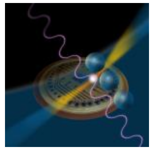
images/text adapted from: C. Bickle/G. Popkin

INFN
Istituto Nazionale di Fisica Nucleare

Quantum Computing

Superconduting Qubits at INFN

DEMETRA 2018-2020 CSNV
Radioactivity effects on SC qubits



SIMP 2019-2021 CSNV
Josephson junction and nano TES for quantum sensing

SUPERGALAX

SUPERGALAX 2020-2024 H2020 FET
Array of superconducting qubits for quantum sensing

DART WARS 2020-2024 Call CSNV
Traveling Wave Parametric Amplifiers for quantum sensing and computing

Qub-IT

QubIT 2021-2024 CSNV
Superconducting qubits and JPA amplifiers for quantum sensing and computing

SQMS
SUPERCONDUCTING QUANTUM MATERIALS & SYSTEMS CENTER

SQMS 2021-2024 DOE
Quantum Computing and Sensing

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

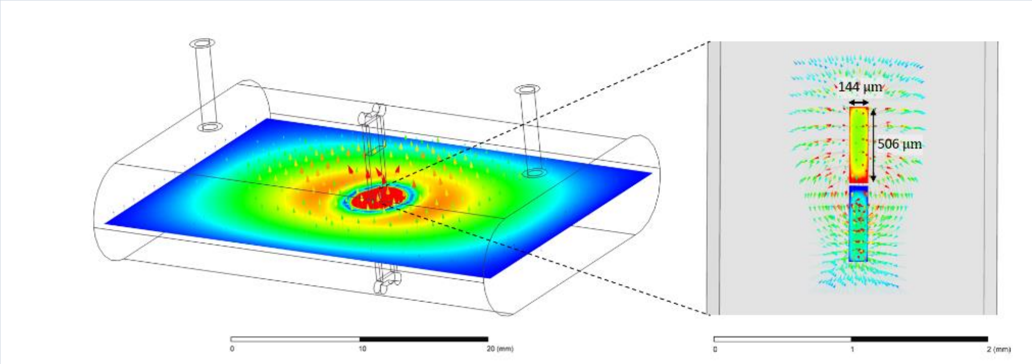
NQSTI
National Quantum Science and Technology Institute

ICSC and NQSTI 2023-2025 PNRR
Quantum Computing and Sensing

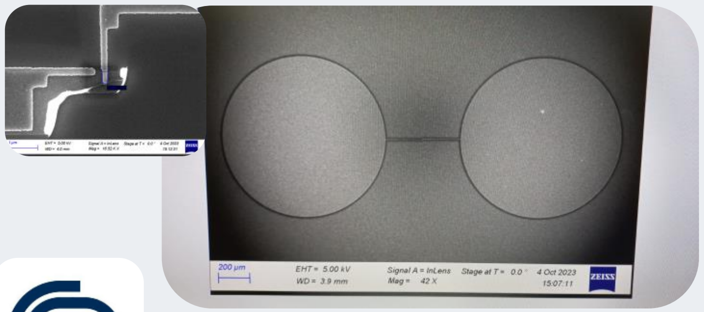
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

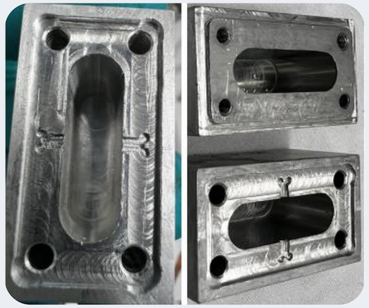
Istituto Nazionale di Fisica Nucleare

Superconducting Qubits in 3D Cavity

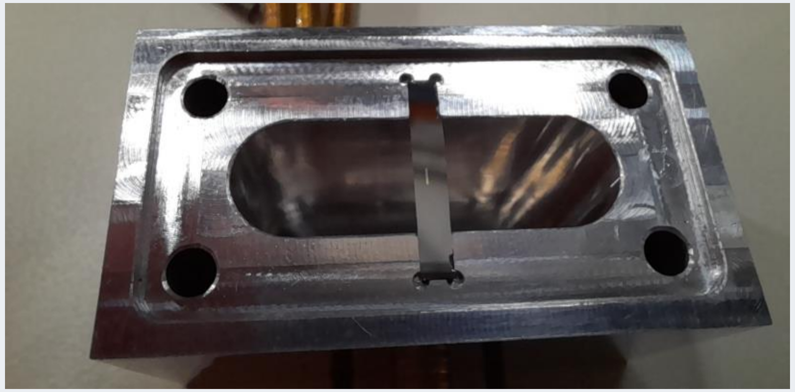


Design and simulation of qubit in 3D cavity

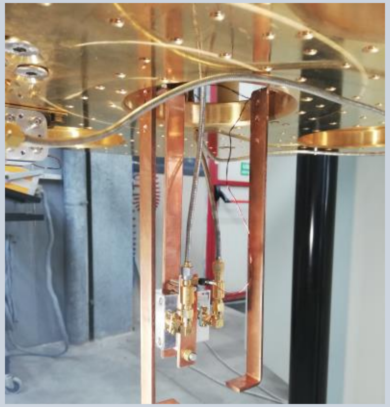




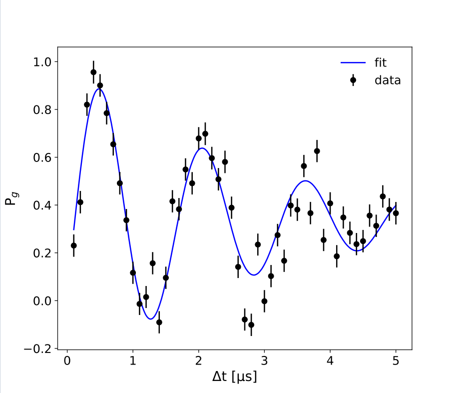
Fabrication within the Collaboration



Qubit in 3D cavity from external collaborations



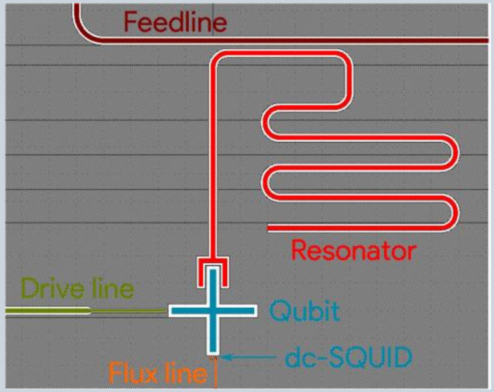
Qubit characterization

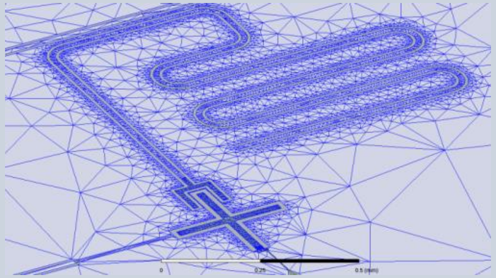


Appl. Sci. 2024, 14, 1478.

Superconducting Qubits on Planar Chip

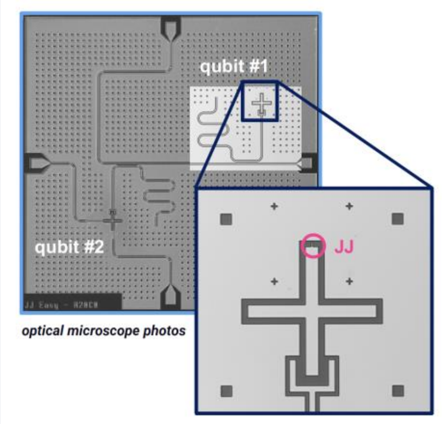
IEEE Transactions on Applied Superconductivity (Volume: 34, Issue: 3, May 2024)



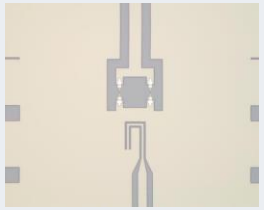


Design and simulation of qubits on planar chip

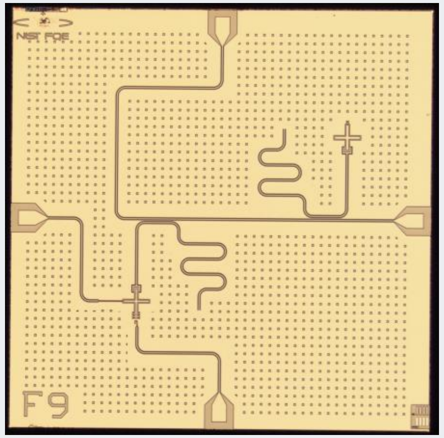
FONDAZIONE BRUNO KESSLER



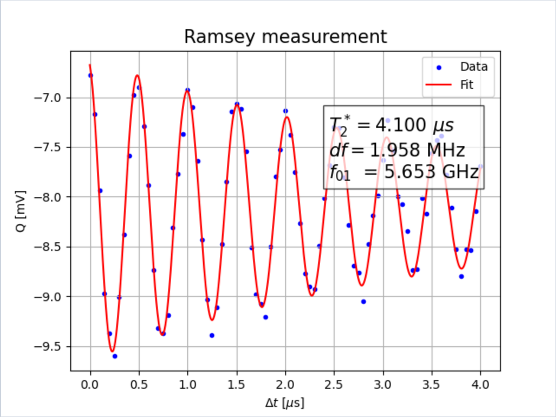
Fabrication within the Collaboration



NIST



Qubit fabrication from external collaborations

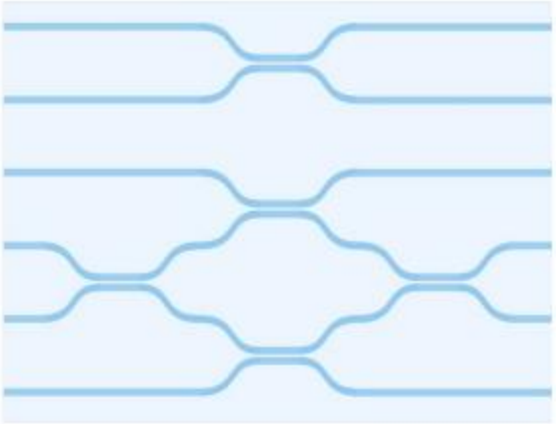


Qubit characterization

QUANTEP: Optical QC (2021-2024)

INFN Sections and Laboratories involved: **LNL**, **MI**, **PG** (Camerino), **PI**, **PV** (Modena e Reggio Emilia), **RM2**, **SA**, **TO**

- Interest and support from: LNGS (LUNA-MV), LABEC (DEFEL), NEST, TYNDALL, Institut Ruđer Bošković (RBI), Micro Photon Devices (MPD), University of Leipzig, Chalmers University of Technology, Physikalisch-Technische Bundesanstalt (PTB).
- 15-17 FTE/year, ~ 800 kEuro budget
- Creation of a common **Silicon Photonics** platform for development and characterization of
  - quantum computing circuits;
  - single photon sources;
  - single photon detectors;
  - polarization control circuits.



Just an example: the CNOT quantum gate

1 qubit: $\alpha_0|0\rangle + \alpha_1|1\rangle, \ |\alpha_0|^2 + |\alpha_1|^2 = 1$

Some 1 qubit elementary gates

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad R_\phi = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\phi} \end{pmatrix} \quad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

Pauli-X (NOT) gate — Pauli-Z gate — Phase shift gate — Hadamard gate

2 qubits: $a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle \quad |a|^2 + |b|^2 + |c|^2 + |d|^2 = 1$

The prototype (universal) 2 qubits gate is the Controlled NOT (CNOT) gate

control bit

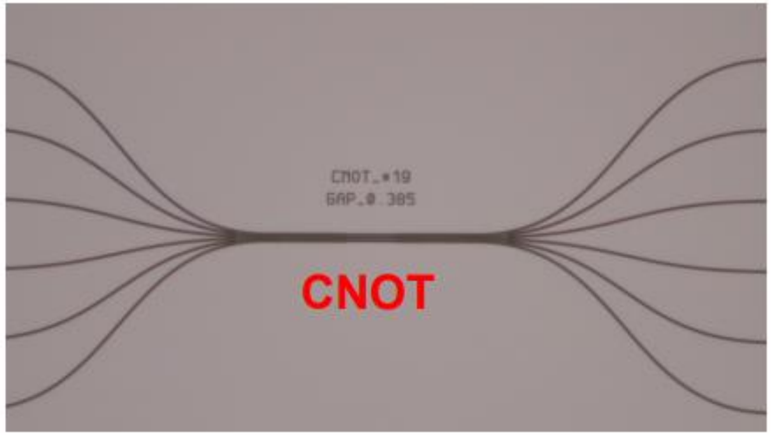
$$\text{CNOT} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

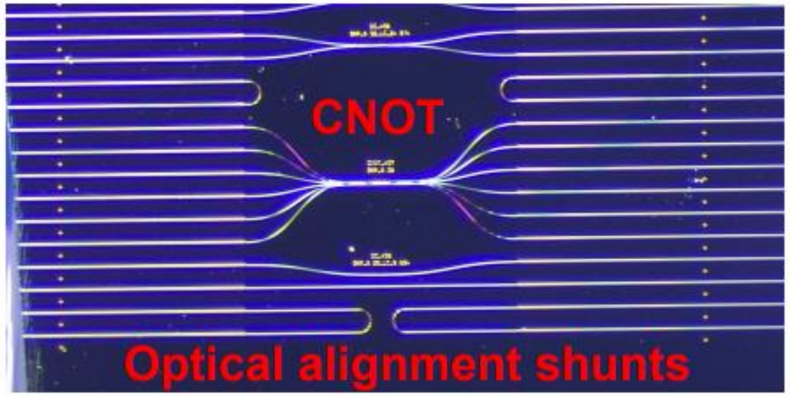
target bit

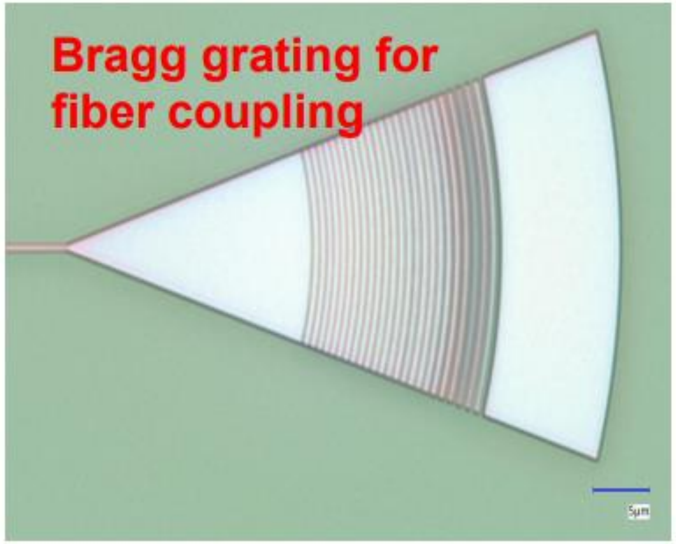
- the control bit is left unchanged
- the output target bit is the XOR of the input control and target bits
- but of course it does much more: it works on the wave function

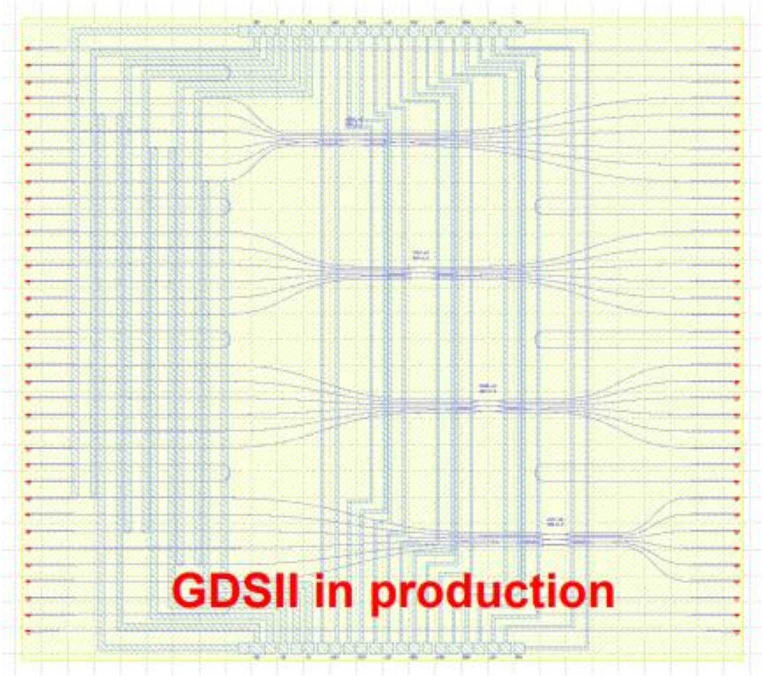
$$a|00\rangle + b|01\rangle + c|10\rangle + d|11\rangle \longrightarrow a|00\rangle + b|01\rangle + c|11\rangle + d|10\rangle$$

QUANTEP: Optical QC (2021-2024)



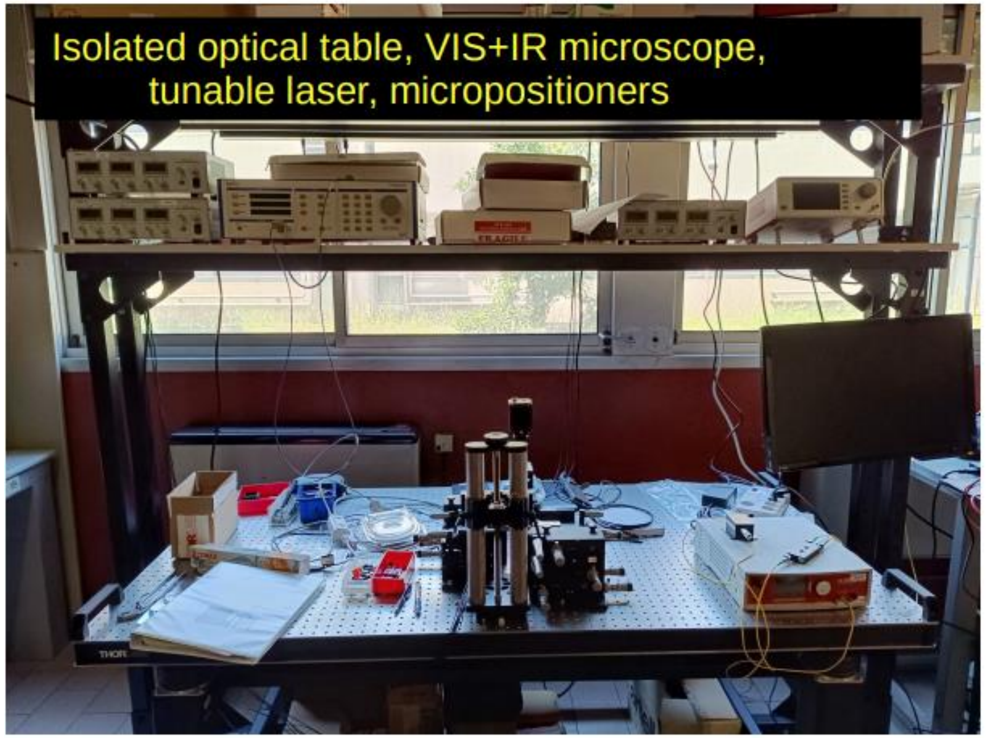


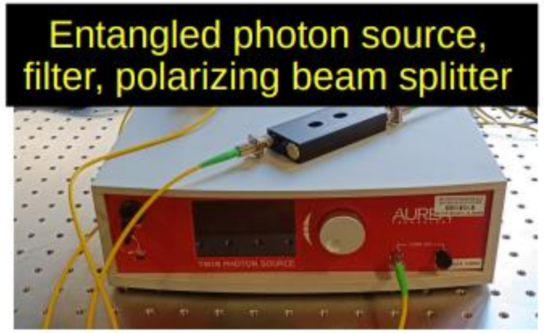


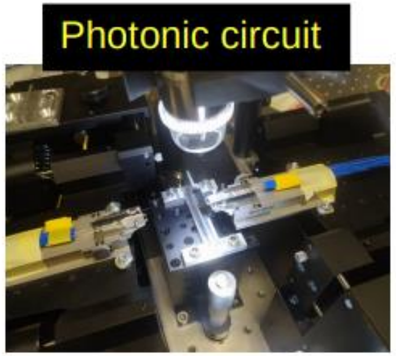


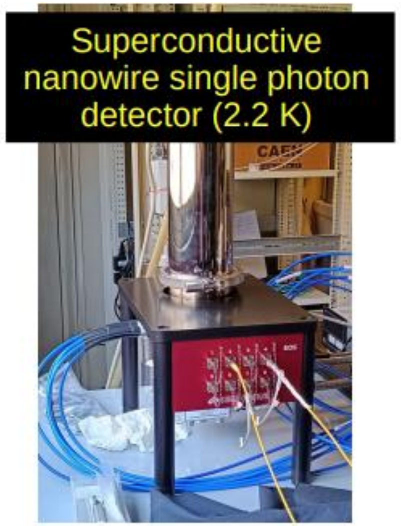
Quantum Computing

QUANTEP: Optical QC (2021-2024)









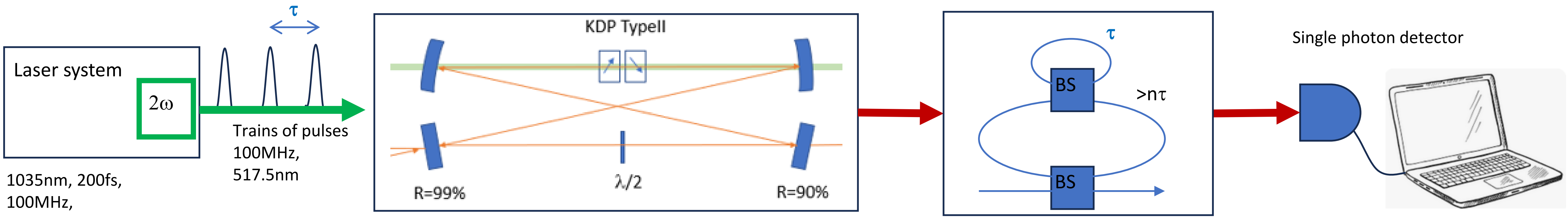
T4QC

Technologies for Quantum Computing

Simone Cialdi

T4QC: Optical quantum computer based on Gaussian Boson Sampling implemented with high-frequency loop technology.

The experimental setup consists of 4 parts: pump laser system, optical system for squeezed state generation, loop-based computing unit, and detection.



Single-mode temporal squeezed state generator

Loop-based computing unit

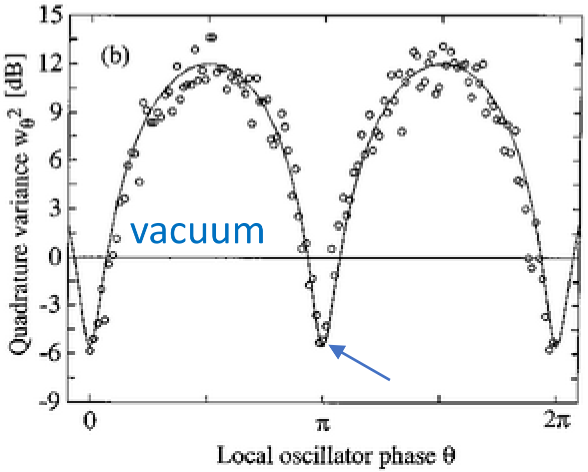
The Gaussian Boson Sampling (GBS) involves sending squeezed states into a network of beam splitters (BS) and measuring the photon distribution at the output. The beam splitter network 'entangles' the qubits (squeezed pulses), making the problem hard (ideally, the computational complexity grows exponentially with the number of qubits).

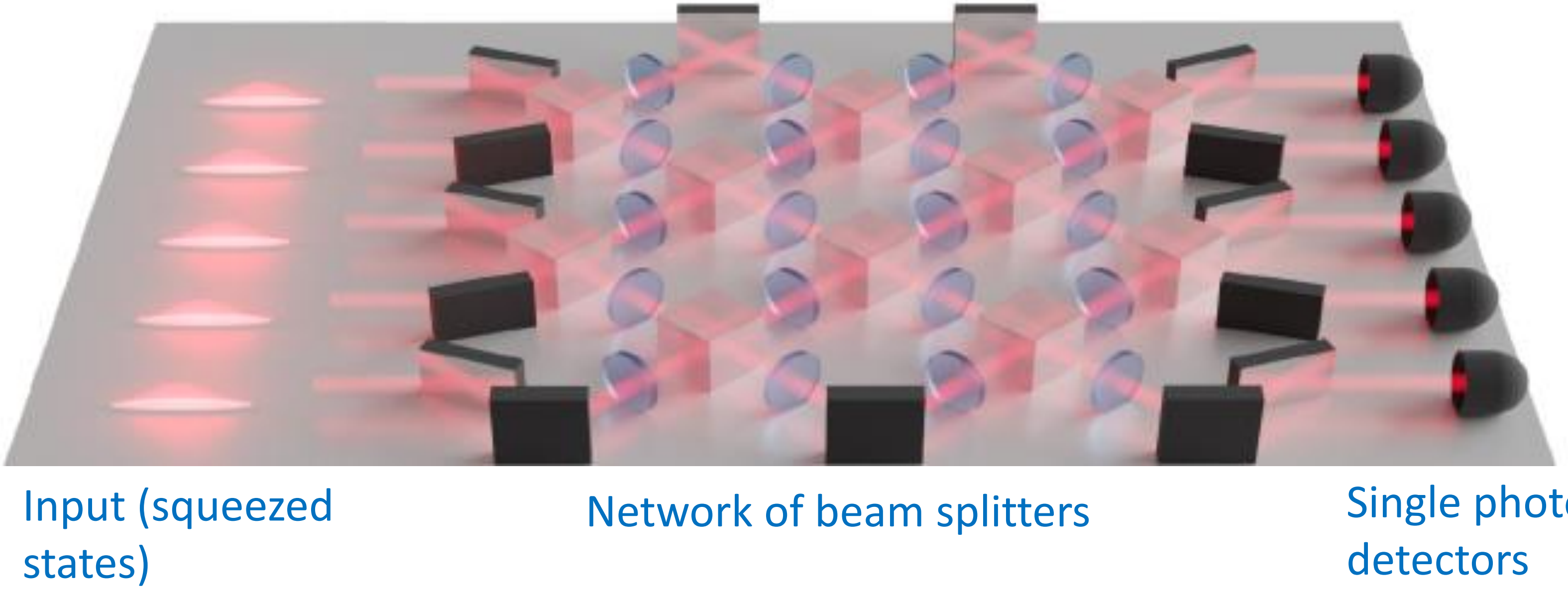
With the loop architecture, it's possible to increase the number of BS in the network simply by increasing the number of input pulses without changing the system's structure. Furthermore, in the loop structure, there's only one BS, so all the BSs in the network are identical.

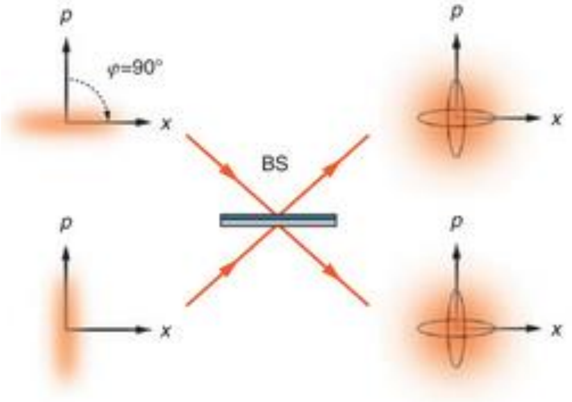
Gaussian Boson Sampling (GBS)

The Gaussian Boson Sampling (GBS) involves sending squeezed states into a network of beam splitters (BS) and measuring the photon distribution at the output. The beam splitter network 'entangles' the qubits (squeezed pulses), making the problem hard (ideally, the computational complexity grows exponentially with the number of qubits).

The squeezed state is a specific quantum state in which the fluctuation of the electric field is lower than that of the vacuum for a particular phase.







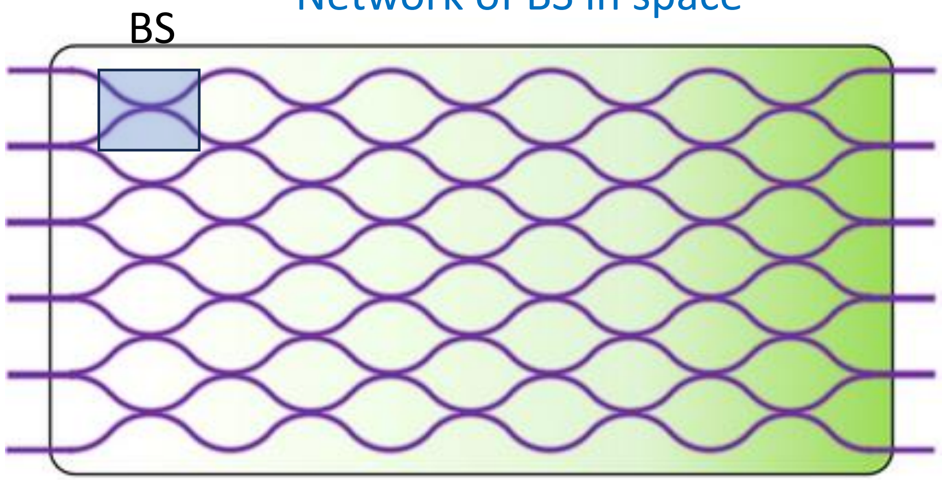
Specific problems can be **mapped** onto the network of **programmable BS** (programmable in the sense that reflectivity and phase of one of the outputs can be arbitrarily set). For instance, studies on the distribution of rovibrational levels of molecules and graph problems have been implemented. In principle, a programmable network of BS and PNR detectors allows for the realization of a universal computer.

Note: while the squeezed state is generated on-demand (I generate it for each pump laser pulse), 'single-photon' states do not always contain only 1 photon. The best sources (see Quandela) have a brightness of 60%. **Therefore, using squeezed states avoids error correction procedures on the source**.

GBS implemented with loop architecture

With the loop architecture, it's possible to increase the number of BS in the network simply by increasing the number of input pulses without changing the system's structure. Furthermore, in the loop structure, there's only one BS, so all the BSs in the network are identical.

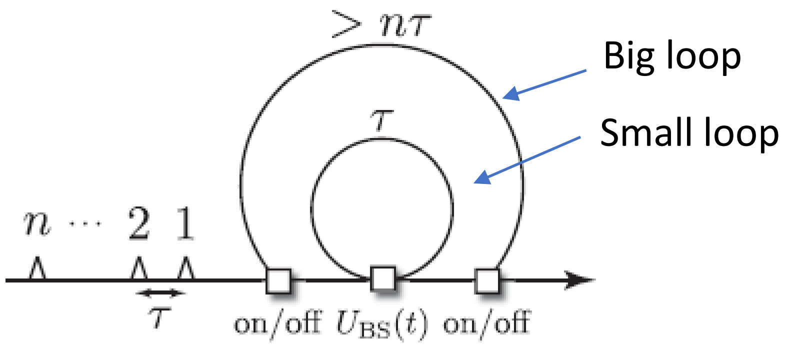
Network of BS in space



Equivalent to

Loop architecture

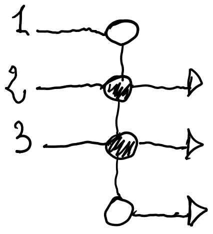
PRL 113, 120501 (2014)



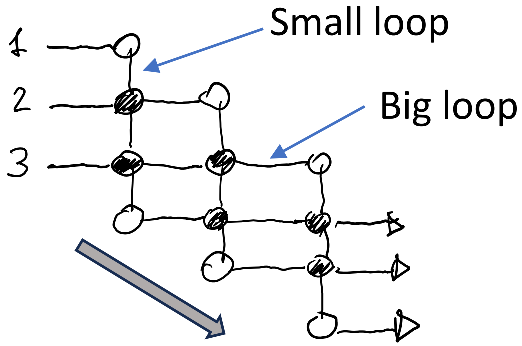
Example of a network with 3 inputs and a depth of 3 steps (3-1 Big loops): the input switch selects 3 pulses, since the time of the Small loop is exactly equal to the time between two pulses, the pulses enter the BS simultaneously. After a number of Small loops equal to the number of selected pulses, the pulses are sent to the Big loop, which returns them to the entrance of the Small loop.

○ = R 100%

● = R programmable



Then I repeat the Big loop n-1 times



AI_INFN: AI Technologies for INFN Research

INFN produces (acquires or simulate) digital data for most of its activities.

Artificial Intelligence provides new techniques to process, interpret and visualize digital data.

Work-packages:

- **Infrastructure**
  provisioning of shared resources
- **Stewardship**
  supporting the adoption of AI with training
- **Scientific use-cases**
  harmonize access to shared resources
- **Hardware acceleration**
  study hardware solution beyond GPUs,
  e.g. FPGAs and Quantum Processors

AI_INFN
Artificial Intelligence technologies for INFN research

Cloud-native Infrastructure

Strong sinergy with INFN Cloud

INFN CLOUD

Cloud-native solutions

jupyterhub

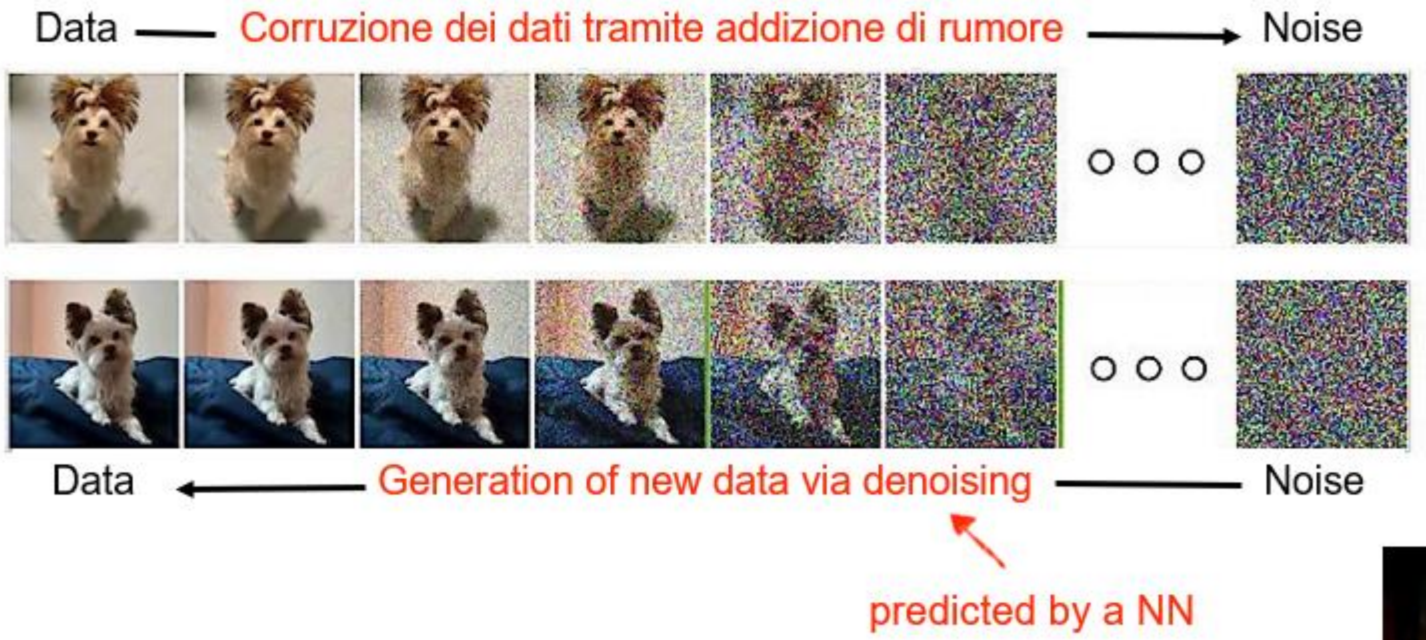
INDIGO - DataCloud

Soon adopted by the Italian Center for Super Computing

ICSC
Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

Quantum Machine Learning: Diffusion Models

- Generative AI models, inspired by non-equilibrium thermodynamics, that use artificial neural networks to gradually add and then remove noise from data, with the goal of generating or reconstructing high-quality data samples

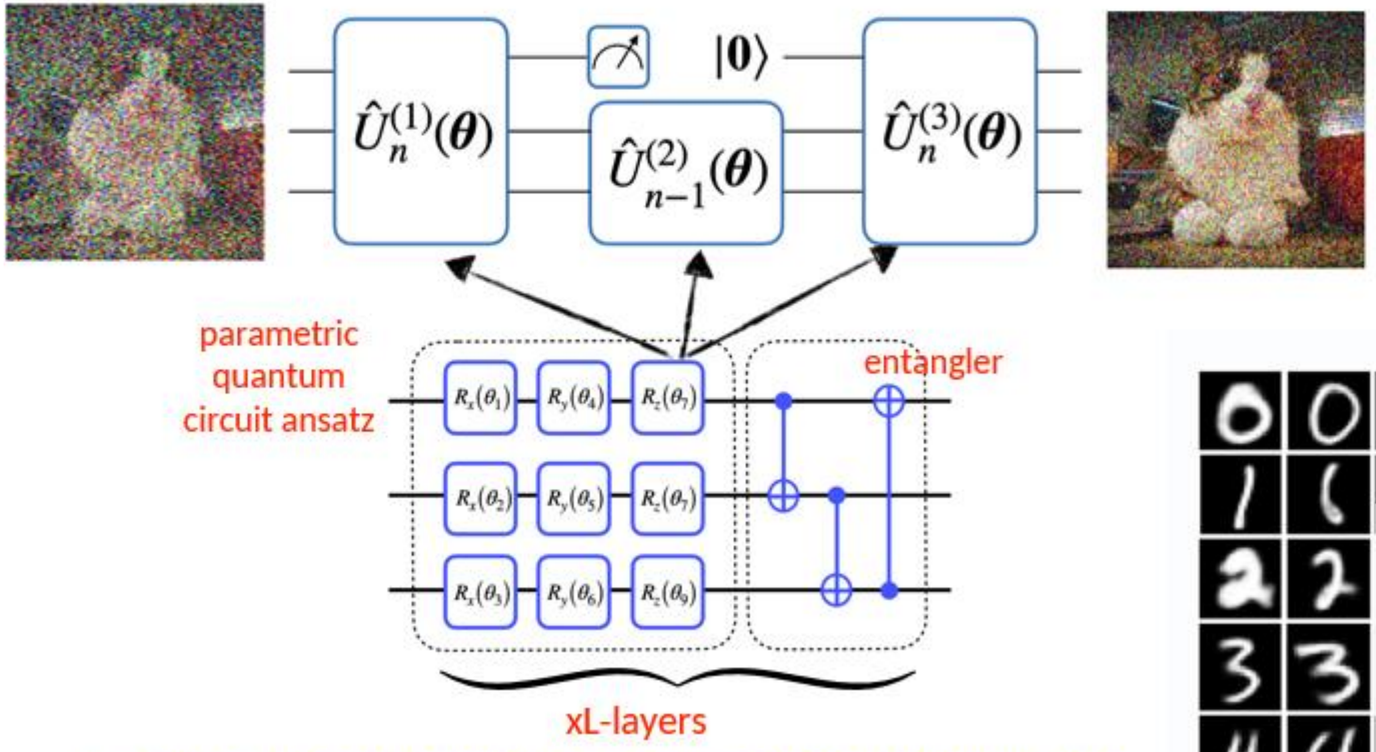


- today used in several tasks: image denoising, inpainting, super-resolution, image generation (ex. text-conditional image generators like DALL-E, Stable Diffusion, …)

QML: Quantum Diffusion Models

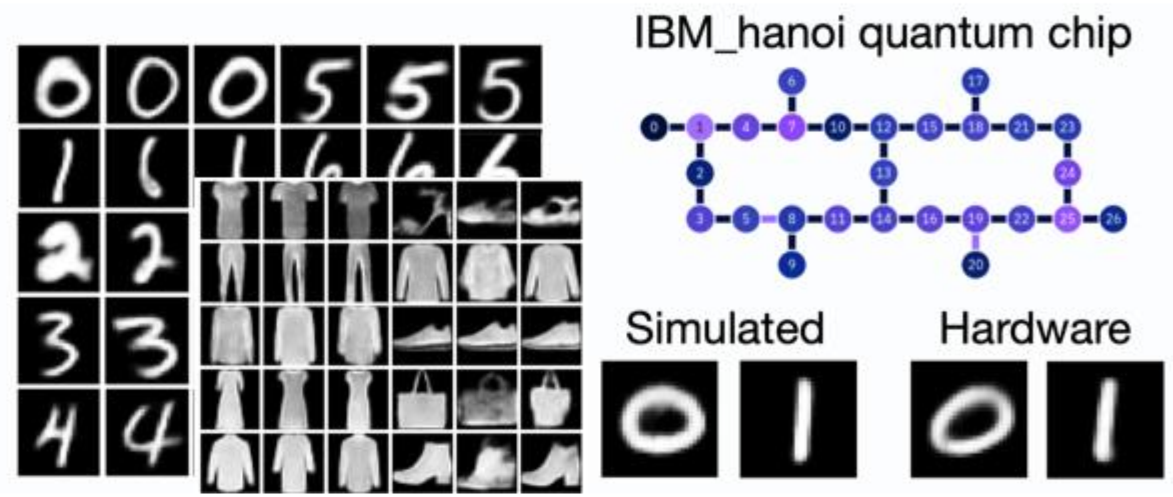
- leverage the ability of variational quantum circuits to efficiently represent the solution space of the problem and to identify complex correlations in the data to implement a quantum denoiser
- can be used in a full quantum or in a hybrid mode, where the quantum circuit is trained in the latent space of a classical Auto-Encoder
- conditioning achieved by adding ancillary qubits to encode labels

Quantum Denoiser



A. Cacioppo, L.Colantonio, S.Bordoni, S.Giagu, arXiv:2311.15444 [quant-ph]

implemented and tested on real quantum hw

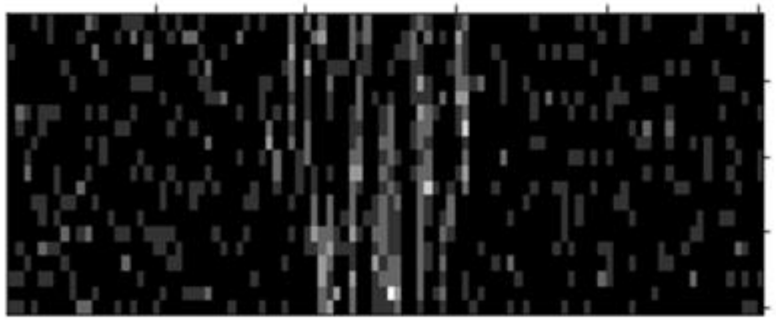


QML: Quantum Anomaly Detection

- Design and train a Quantum-AutoEncoders able to identify highly displaced decays using the ATLAS muon spectrometer information

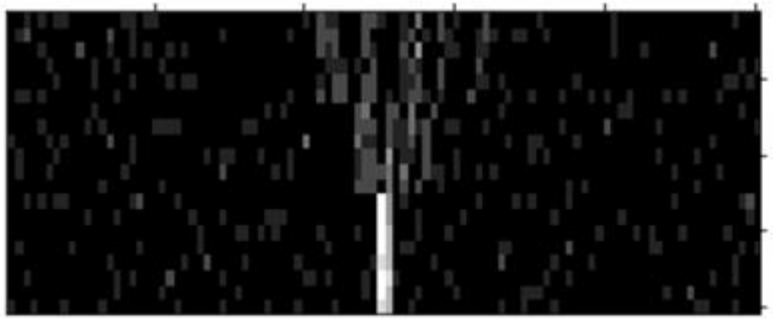
NORMAL event

"image" representation of a prompt decay in multi-muons

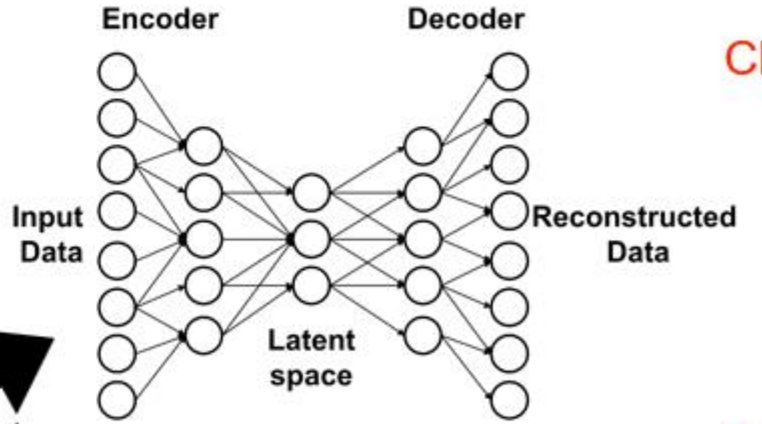


ANOMALOUS event

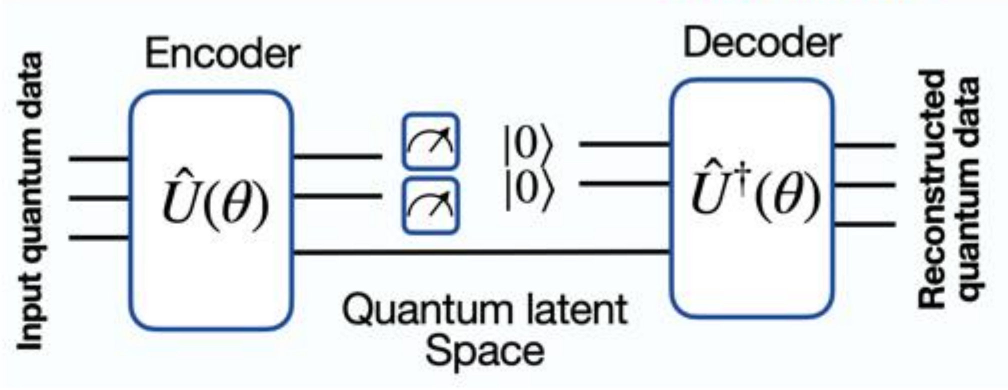
"image" representation of a highly displaced decay in multi-muons



Classical AE



Quantum-AE

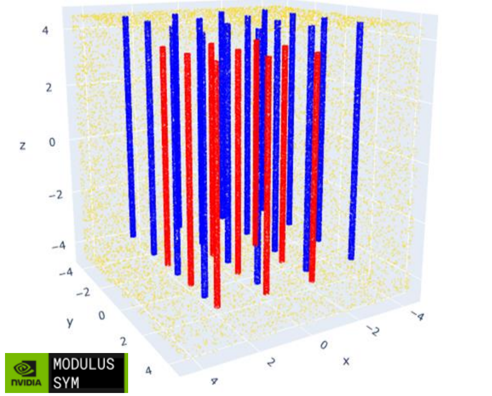


S.Bordoni et al, Particles 2023, 1, 1–15

Examples of scientific use-cases

Detector numerical modelling

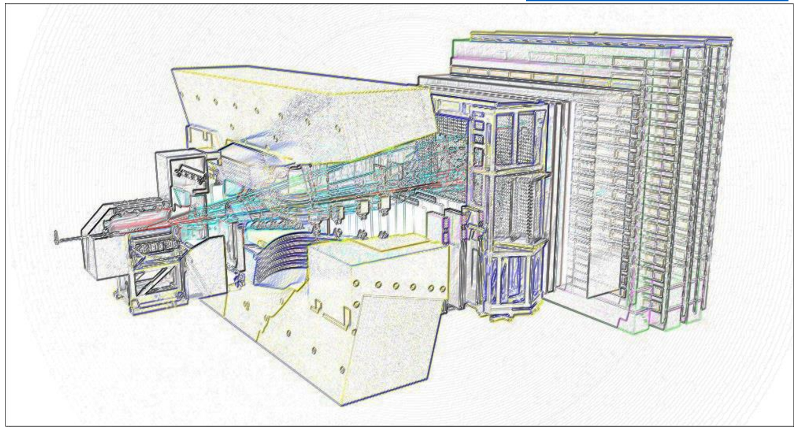
10.1016/j.nima.2022.167230



Physics Informed Neural Networks; Neural Operators...

Simulation of High-Energy Physics experiments

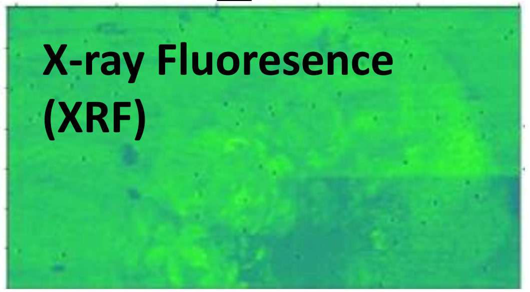
arXiv:2309.13213



Generative models, domain adaptation, Normalizing Flows...

Digital Cultural Heritage

ICIAP (2022) 685

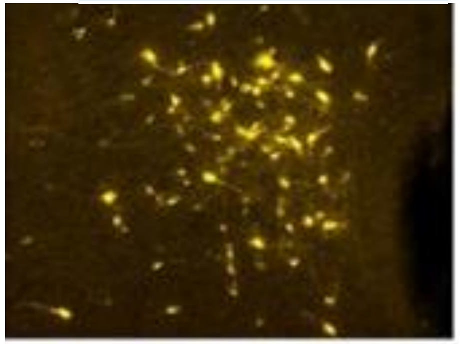


Artificial Intelligence
Microscopy

Machine learning and GPUs are used to automate processing of microscope images for both detector studies and life sciences.

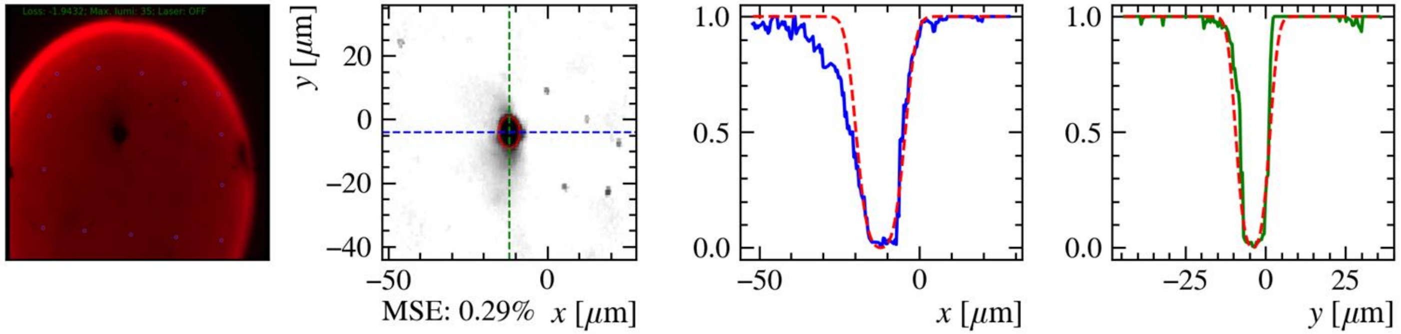
Automation in image processing is an ingredient for automated and reproducible measurements and tests.

10.1038/s41598-021-01929-5



Fluorescent Microscope Image Processing used to count the number of cells.

Convolutional Neural Networks

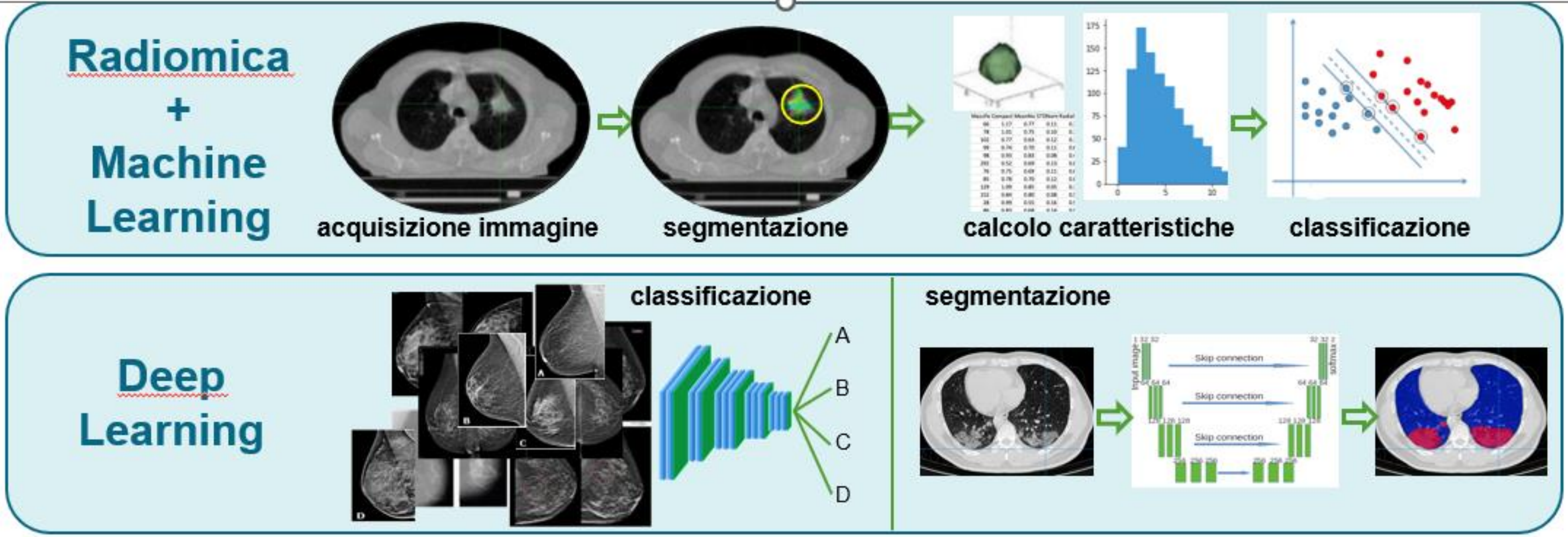


Differential geometry techniques used to correct for optical aberrations and measure the spot size and profiles.

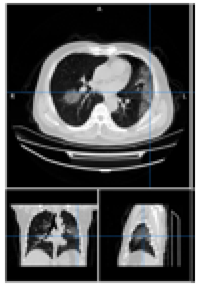
Artificial Intelligence
next_AIM: Medicine

Artificial Intelligence in Medicine: main analysis approaches

In **medical image analysis** a large variety of approaches based on AI can be developed, according to different goals, e.g. image segmentation, image classification, building predictive models based on images and additional patient information.

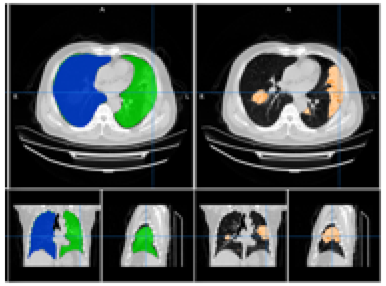


LungQuant: open-access SW tool for COVID-19 lesion detection and structured reporting [https://www.openaccessrepository.it/record/76937]



LungQuant

[Lizzi F et al Quantification of pulmonary involvement in COVID-19 pneumonia by means of a cascade of two U-nets: training and assessment on multiple datasets using different annotation criteria. IJCARS 2022;17:229–37. doi.org/10.1007/s11548-021-02501-2.

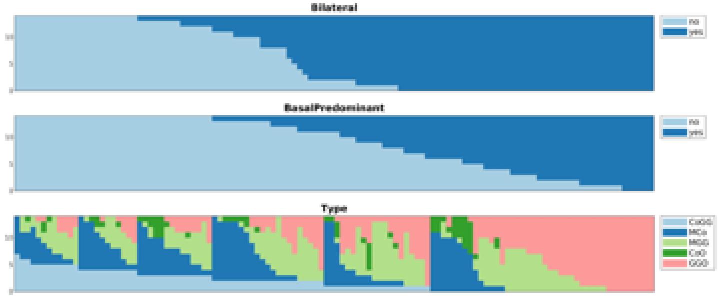


SW output:
- segmented masks
- qualitative parameters to describe the lesions

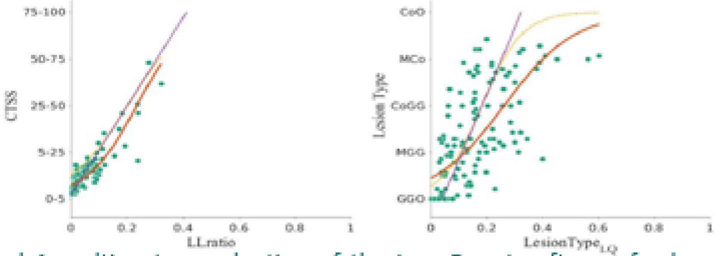
ID	LESION_TYPE_INDEX	BILATERAL_INDEX	BASAL_INDEX
A-0037	0,137	0,447	37
A-0311	0,198	0,041	61
A-0291_0	0,224	0,193	31
A-0327	0,292	0,351	60
	$V_{Consolidation} / V_{Lesion}$	0: unilateral 1: bilateral	0: basal 100: apical

The validation of the LungQuant software output against the qualitative assessment of 14 radiologists from 5 University Hospitals (Pisa, Pavia, Firenze, Palermo, Milano) has shown:

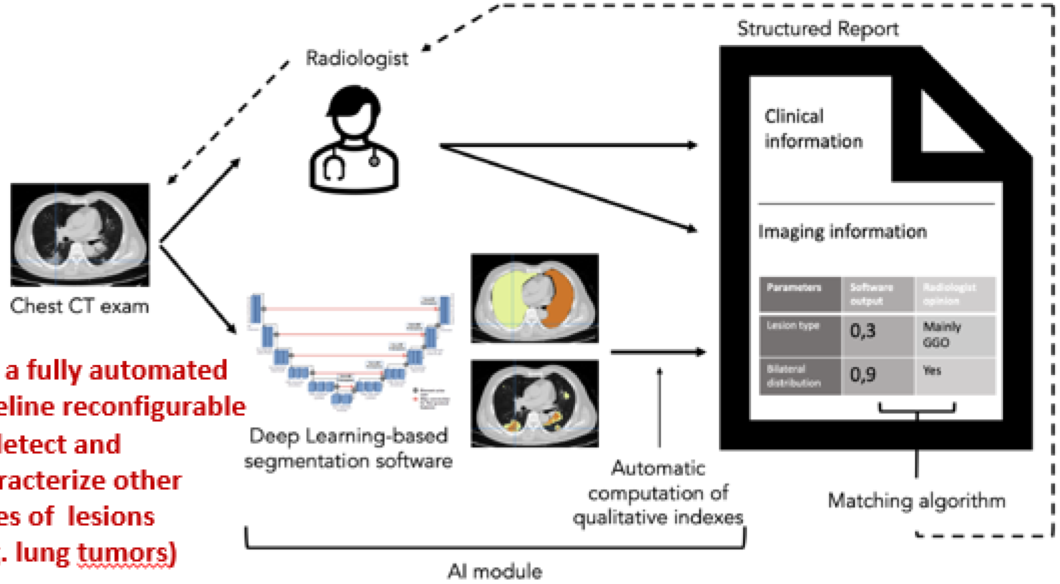
- a poor agreement among the opinions of radiologists



- a good correlation between average radiologists' opinions and the equivalent software output metrics



[Chincarini A, Scapicchio C et al A multicenter evaluation of the LungQuant software for lung parenchyma characterization in COVID-19 pneumonia, European Radiology Experimental, https://doi.org/10.1186/s41747-023-00334-z]



It is a fully automated pipeline reconfigurable to detect and characterize other types of lesions (e.g. lung tumors)

[Scapicchio C, et al. Integration of a Deep Learning-Based Module for the Quantification of Imaging Features into the Filling-in Process of the Radiological Structured Report. Int. Jt. Conf. Biomed. Eng. Syst. Technol., SCITEPRESS 2023, p. 663–70. https://doi.org/10.5220/0011921900003414.]

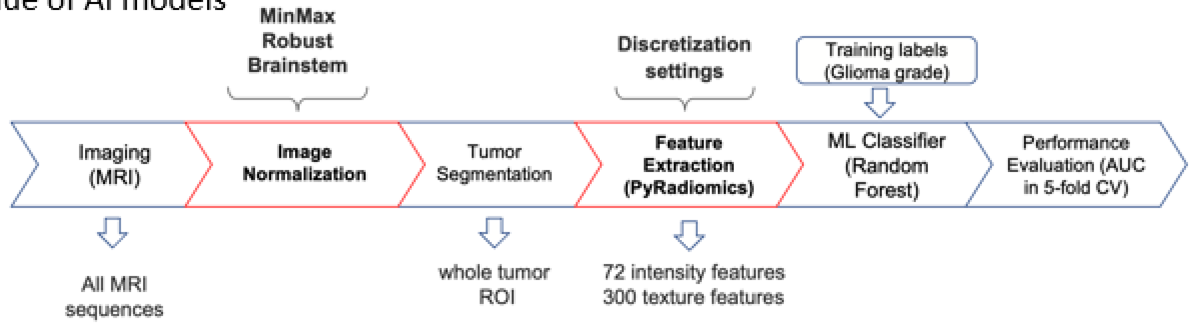
Predictive model to discriminate low-grade vs. high-grade gliomas

Evaluation of the robustness of radiomic features in multiparametric MRI and its impact on predictive value of AI models

Multiparametric MRI scans (T1, T1-Gd, T2, FLAIR) of:

- **61 patients with** Low-Grade Gliomas (**LGG**)
- **97 patients with** High-Grade Gliomas (**HGG**)

The analysis pipeline



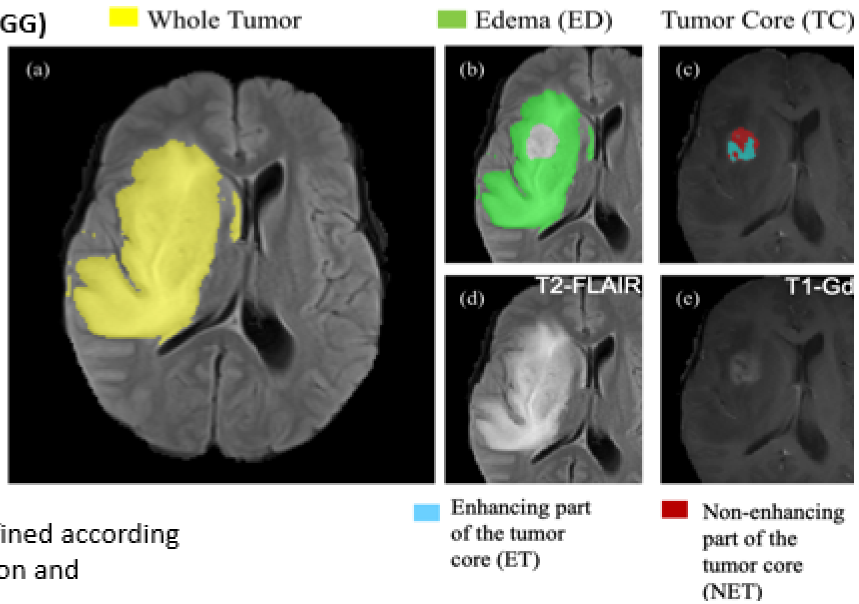
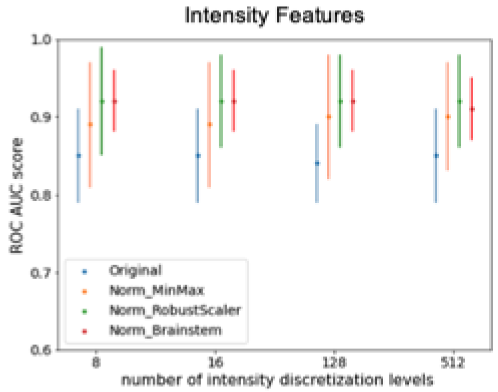
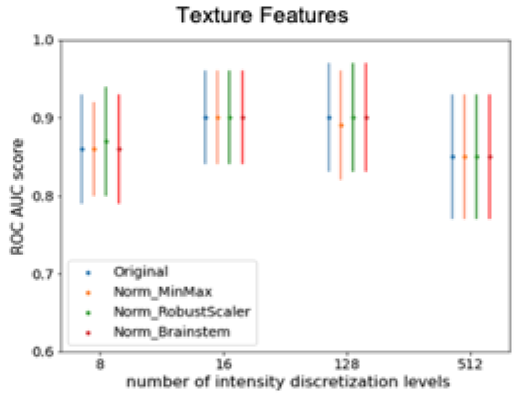


Image normalization and intensity discretization have an impact on the performance of ML classifiers based on radiomic features.





Random forest (RF) classification

- target: LGG vs HGG discrimination
- features: **MRI-reliable features** defined according to the most appropriate normalization and discretization settings.

Conclusions

- The **complementary information of multi-parametric MRI** has to be taken into account
- The **image preprocessing** step is relevant for radiomic and ML analysis

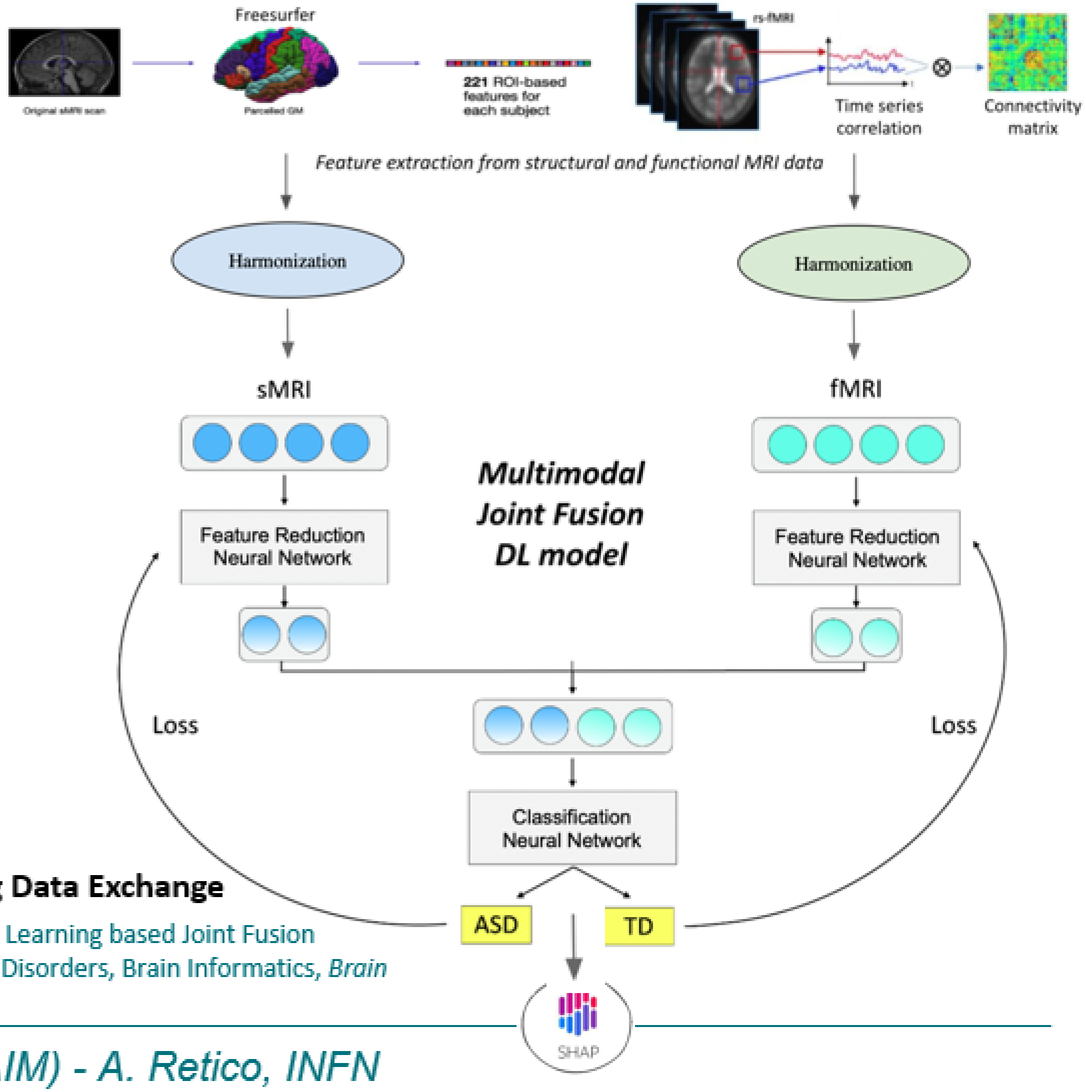
Modality	Raw feature Set (372 Features complessive per le 4 modalità)	MRI-reliable feature Set (372 Features) [Norm_Brainstem] (bin counts = 128)
T1	0.73 ± 0.05	0.69 ± 0.04
T1-Gd	0.89 ± 0.05	0.93 ± 0.05
T2	0.76 ± 0.08	0.75 ± 0.06
T2 FLAIR	0.76 ± 0.08	0.76 ± 0.06
All sequences	0.88 ± 0.08	0.93 ± 0.05

Ubaldi L, Saponaro S, Giuliano A, Talamonti C, Retico A. Deriving quantitative information from multiparametric MRI via Radiomics: Evaluation of the robustness and predictive value of radiomic features in the discrimination of low-grade versus high-grade gliomas with machine learning. Phys Medica 2023:107:102538. https://doi.org/10.1016/j.ejmp.2023.102538

Joint fusion approach to exploit both structural and functional data

Brain imaging features

- **sMRI –** The Freesurfer *recon-all* pipeline has been implemented to extract 221 structural features for each subject
- **rs-fMRI –** The CPAC processing pipeline for fMRI data has been implemented:
  - The Harvard-Oxford atlas has been used, thus generating 103 temporal series for each subject
  - The functional connectivity matrix has been computed for each subject implementing the Pearson correlation, thus obtaining 5253 functional features for each subject



Joint fusion approach:

- The Feature Reduction and the Feature Classification Neural Networks are trained using a single cost function, thus the most meaningful features for the classification are extracted
- The model was trained with 150 epochs within a 10-fold cross validation scheme

Explainability framework:

- SHpley Additive exPlanations (**SHAP**)

ABIDE
Autism Brain Imaging Data Exchange

Saponaro S, Lizzi F, Serra G, Mainas F, Oliva P, Giuliano A, Calderoni S, Retico A. Deep Learning based Joint Fusion approach to exploit anatomical and functional brain information in Autism Spectrum Disorders, Brain Informatics, *Brain Informatics 2023*. https://doi.org/10.1186/s40708-023-00217-4.

Computational Neuroscience @ INFN

- **Simulating the activity of significant portions of the human brain**
  - Scales ranging from the microscopic level of individual neurons/synapses to the macroscopic level of measurements with tools such as fMRI (functional magnetic resonance imaging) and EEG (electroencephalogram)
- **The INFN has played and continues to play an important role in this development**
  - skills of physicists in modeling, calculation and electronics
- **Development of technologies for the analysis & simulation of biological neural networks**
- **Study of the link between synaptic mechanisms and high-level cognitive processes**
  - Short-term synaptic plasticity & working memory
  - Sleep-Awakeness interplay in learning
  - Spike Timing Dependent Plasticity (STDP), structural synaptic plasticity & learning
- **INFN participated to the Human Brain Project and to its follow-up project, e-Brains**
  - was the leader of a sub-project of the Human Brain Project, **WaveSCALES**
  - carried out numerous computational projects in the HPC infrastructure of the Human Brain Project, e.g. Computational Neuroscience Collaborative Brain Wave Analysis Pipeline (**Cobrawap**).
- **Development of the spiking neuronal network simulator NEST GPU (NEural Simulation Tool GPU)**
  - NEST is one of the two most used simulators to simulate the activity of biological neurons and neuron networks, and is considered one of the pillars of the Human Brain Project and the e-Brains project that followed it
  - NEST GPU is developed in collaboration with the researchers of INM-6, Jülich Research Center. Germany

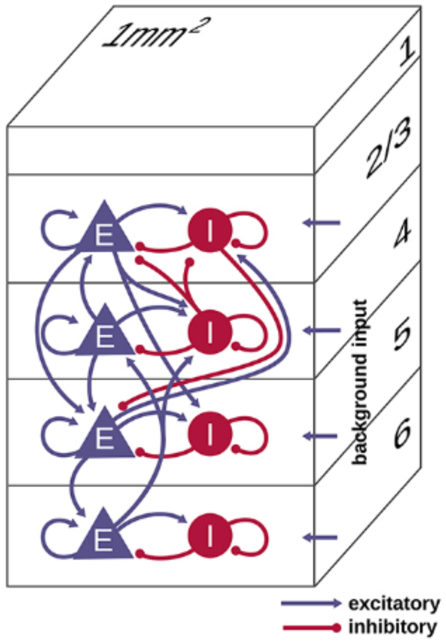
Large scale SNN simulations (1 of 2)

We implemented spiking models that can be simulated on single GPUs or on MPI-GPU systems, and compared the results of the simulations against the implementations of the respective models in the NEST simulator. In particular we worked on

- Cortical microcircuit model
  (single GPU, ~80k neurons, ~3x10^8 synapses)

  Golosio et al., ***Front. Comput. Neurosci.***, 15:627620, 2021

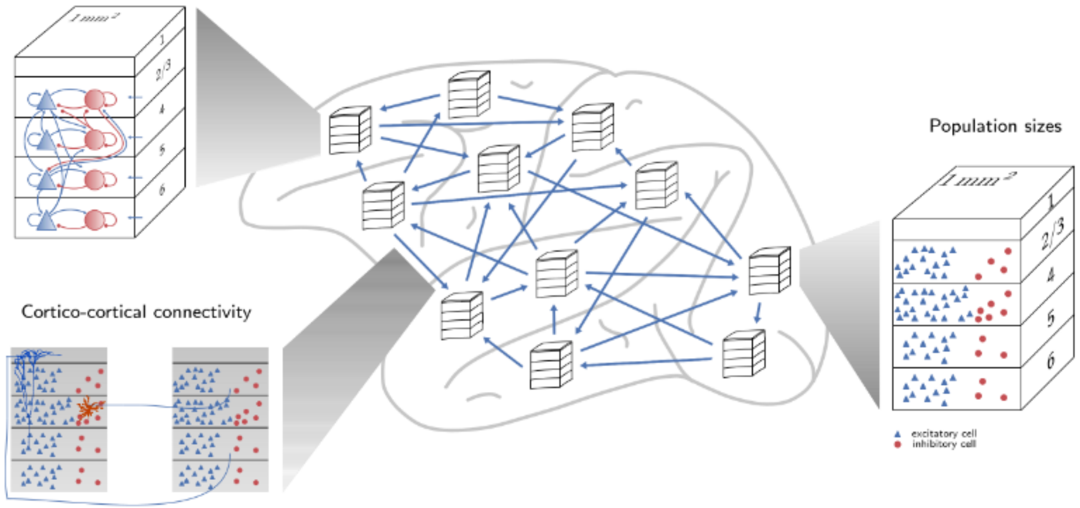
  Golosio et al., ***Appl. Sci.***, *13*, 9598, 2023



Potjans and Diesmann, ***Cerebral Cortex***, 24(3), 2014

- Multi-area model of the macaque cortex
- (multi-GPU, ~4x10^6 neurons,
  ~24x10^9 synapses)

  Tiddia et al., ***Front. Neuroinform.***, 16:883333, 2022

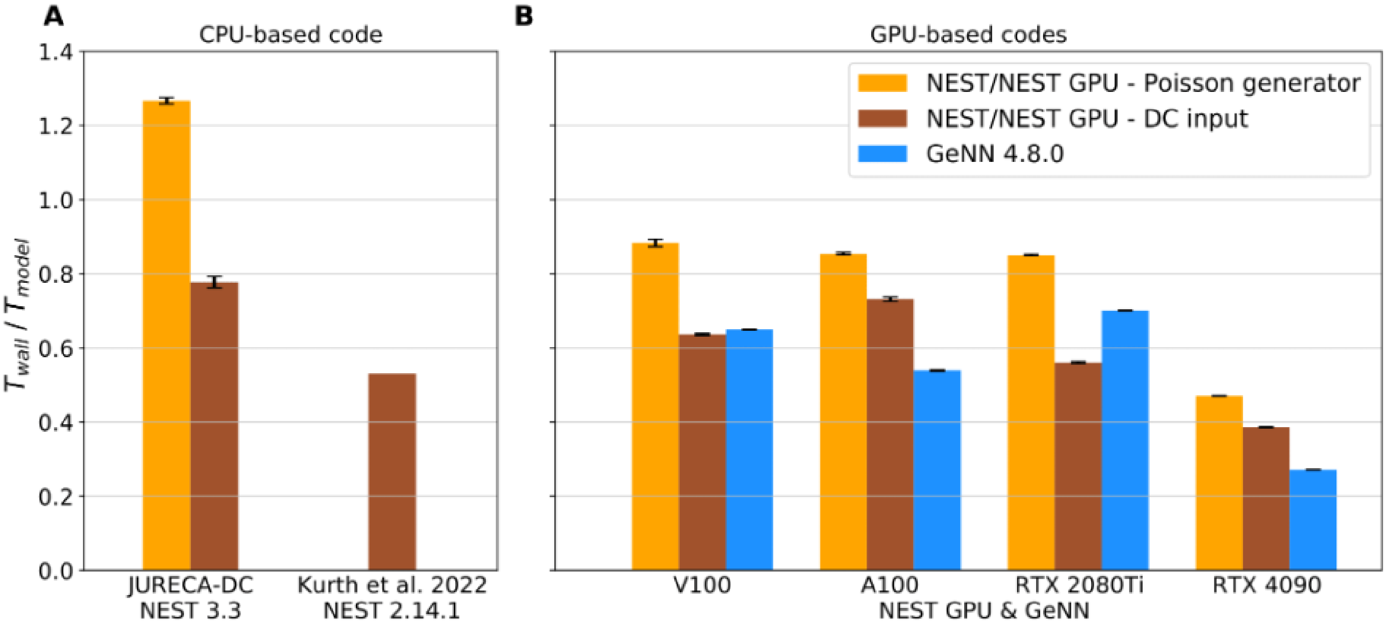


Schmidt et al., ***Brain Struct. Funct.***, 2018

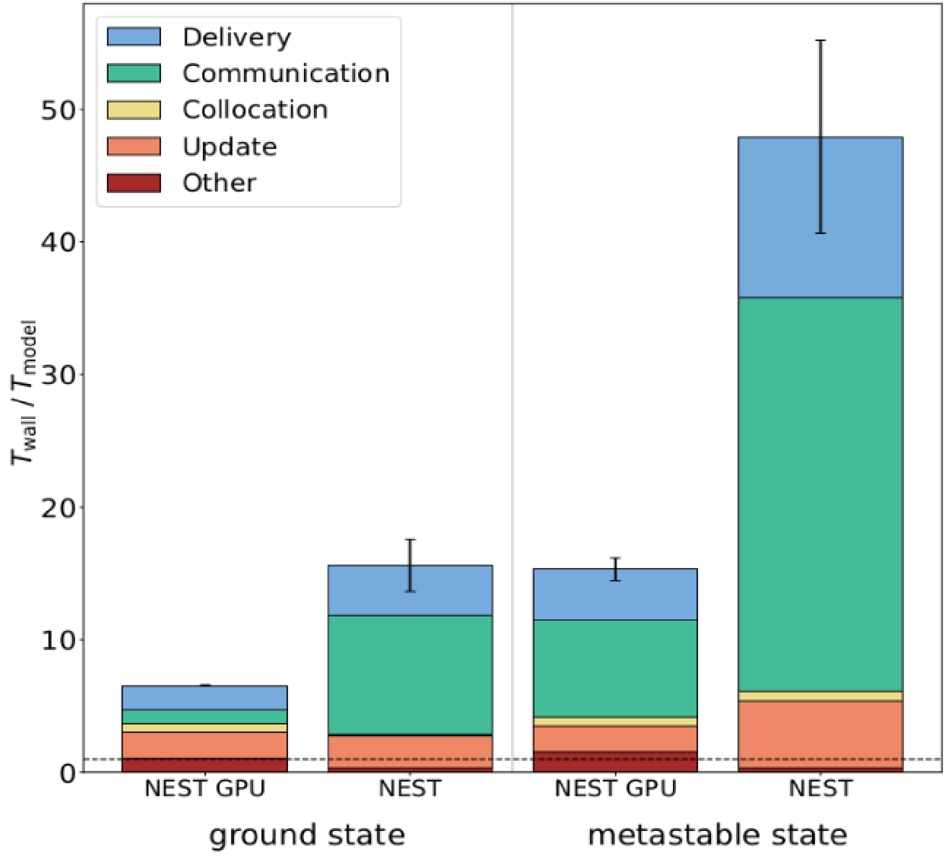
Large scale SNN simulations (2 of 2)

Simulator performance on the described network models (using NVIDIA GPUs, both consumer and data center)

Cortical microcircuit



Multi-area model (cluster MPI-GPU w/NVIDIA V100)



Short-term synaptic plasticity & Working Memory

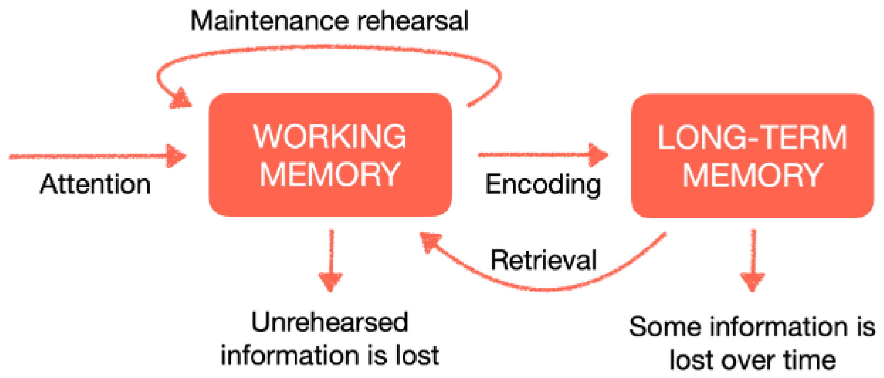
Working Memory (WM) is the cognitive mechanism responsible for temporarily maintaining and processing information in short-term memory and controlling the flow of information between this and long-term memory.

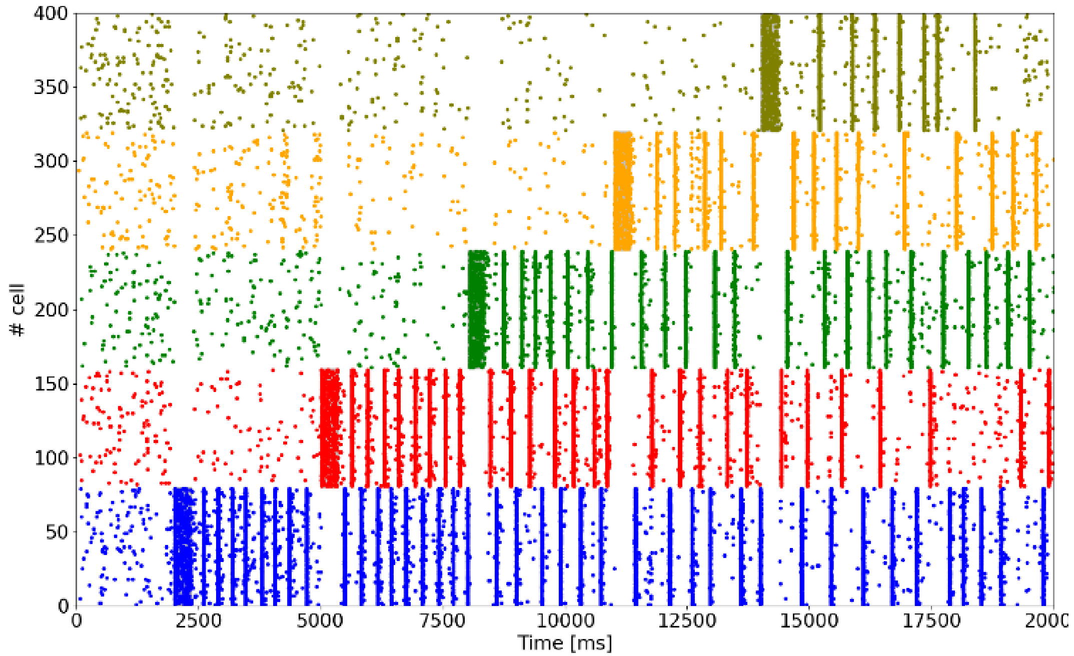
Spiking model of Working Memory entirely maintained by a short-term plasticity mechanism initially proposed in

Mongillo et al., ***Science***, 319, 2008

The model can store various items in memory thanks to the presence of facilitated synapses, present to a large extent in the prefrontal cortex.

Tiddia et al., **Front. Integr. Neurosci.**, 16:972055, 2022

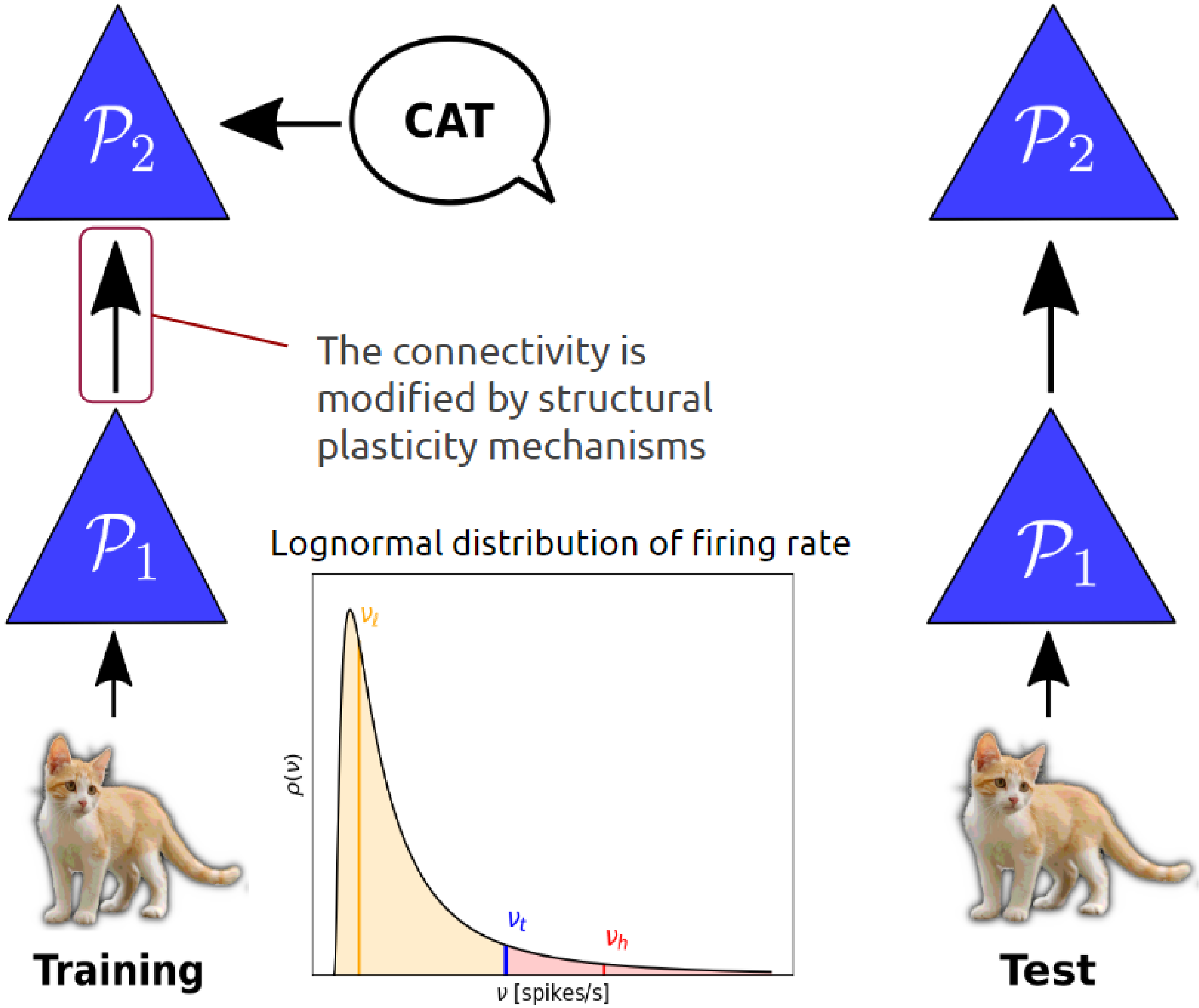




Spiking neural networks (SNN)

Model of learning mediated by structural plasticity

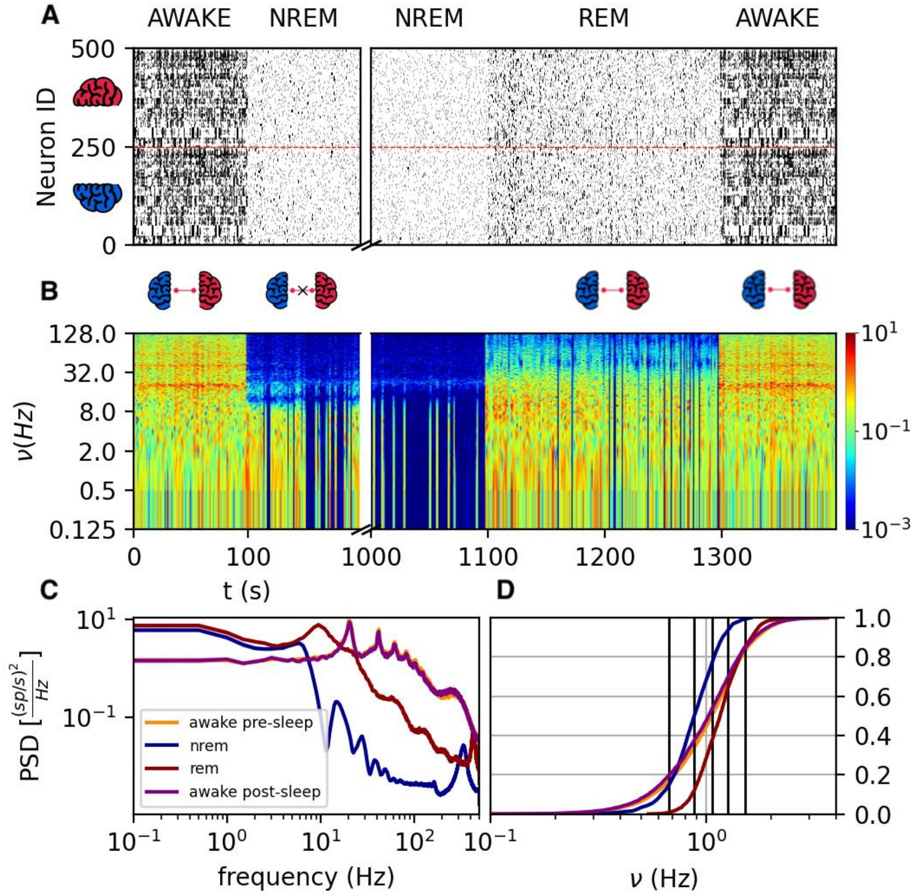
- ❖ During learning, structural plasticity modifies the connectivity $P_1 \rightarrow P_2$
- ❖ In the test phase we evaluate the signal in input to each neuron of P_2
- ❖ Theoretical model capable of predicting the value of the input signal to selective or non-selective neurons with respect to a given input pattern



Thaco: multi-areal plastic cognitive model for thalamo-cortical spiking network simulations

- ❑ Capable of incremental learning
- ❑ Able to enter different brain states (wakefulness, REM dreaming and NREM deep-sleep)

- Showing the beneficial cognitive and energetic effects of the interplay among sleep and memories, learned by combining contextual and perceptual information
- Combining prior knowledge with novel evidence using brain-state specific apical-amplification, apical-drive and apical isolation mechanisms.
- Reducing energy consumption and time to response using spiking mechanisms
- Spiking Plastic Models & exploration of Hardware IPs on FPGA and neuromorphic



Golosio et al (2021)Thalamo-cortical spiking model of incremental learning combining perception, context and NREM-sleep ***PLoS Computational Biology***

Capone et al. (2019) Sleep-like slow oscillations improve visual classification through synaptic homeostasis and memory association in a thalamo-cortical model ***Scientific Reports***

Toward Thaco next generation, including multi-compartment neuron models

- Implementation in NEST of multi-compartment customizable neuron models supporting apical-amplification, apical-isolation, apical-drive dynamics

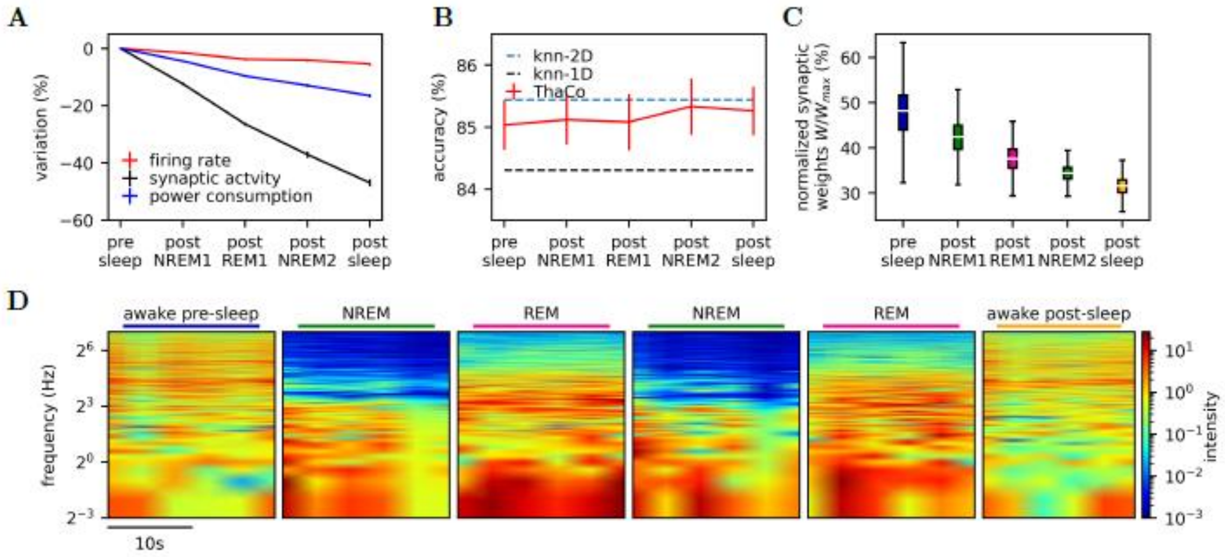
- Adoption of the L2L (Learning 2 Learn) framework to search best fitting multi-compartment neurons using evolutionary algorithms applied to single neuron tasks
  - ~ 500K core-h on HPC systems

- Insertion of the multi-compartment neuron in Thaco

- Optimization of Thaco parameters using the L2L approach applied to the whole network model in different brain states
  - ~ 1500K core-h on HPC systems

Beneficial effects on:

- incremental learning of large training sets
- implementation of learning and sleep cycles that are expected to efficiently reorganize the synaptic representation



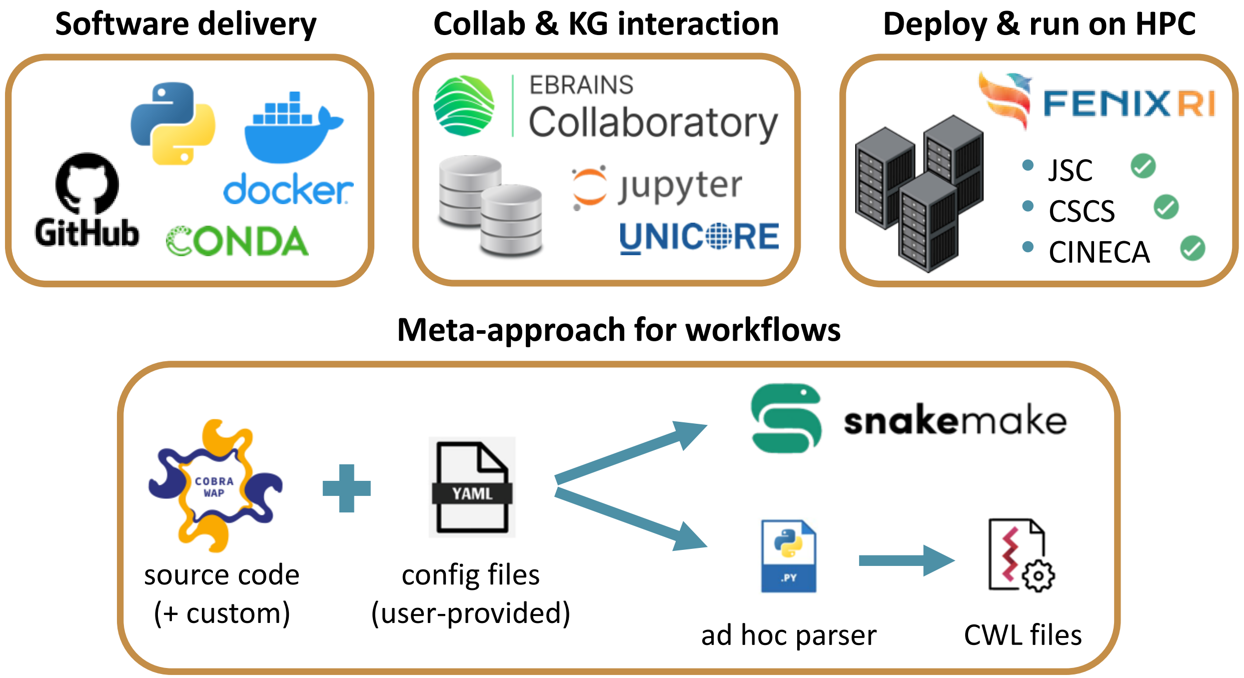
Under development, in strong cooperation with L2L and NEST teams

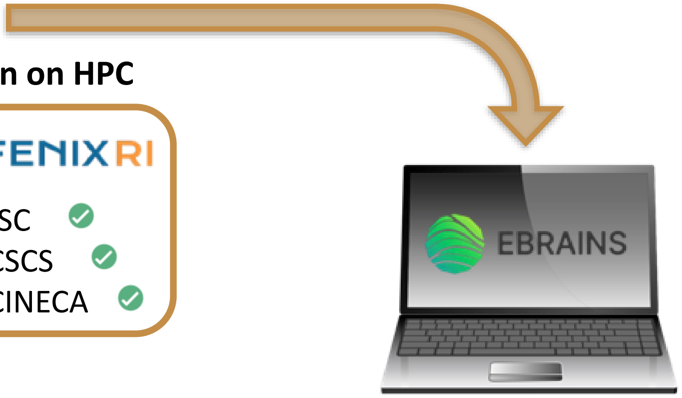
nest-simulator.org

Collaborative Brain Wave Analysis Pipeline (Cobrawap)

Developed in HBP/EBRAINS; Gutzen, De Bonis et al (2023): https://doi.org/10.1016/j.crmeth.2023.100681

Intercepting the demand for resource scalability & usability





Cobrawap as a service

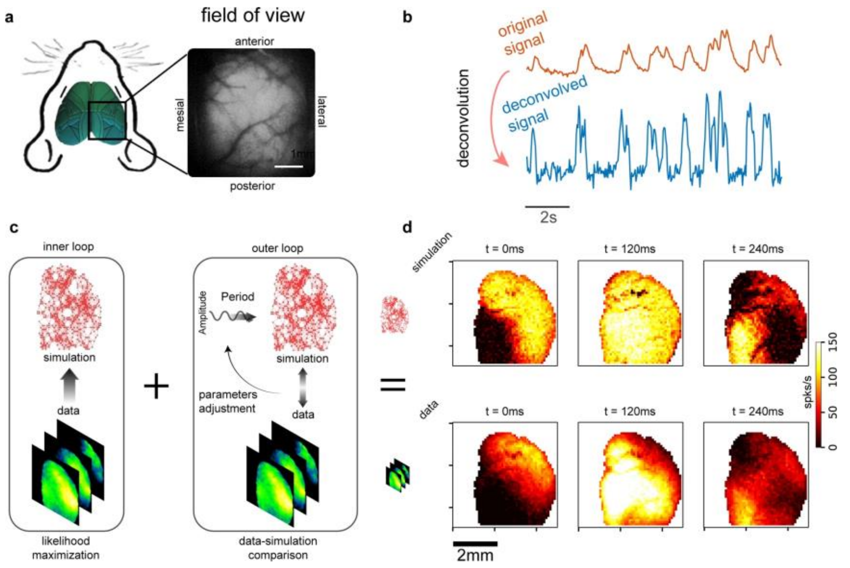
Target tasks

- Model calibration & validation
- Large-scale data analysis
- Metrics for clinical applications
- Buildout of methods & algorithms

Currently funded within EBRAINS-Italy (PNRR)
Maintained in GitHub with FZ Jülich:
NeuralEnsemble/cobrawap
Main improvements: optimization, novel features, parallel computing and acceleration

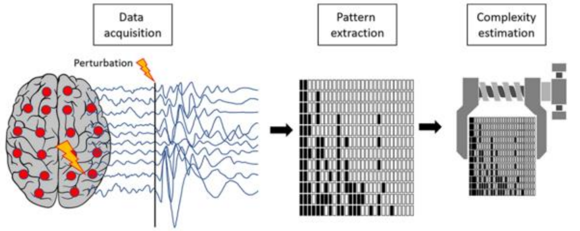
Computational Neuroscience
Collaborative Brain Wave Analysis Pipeline (Cobrawap)

From 2024 RESEARCH in the BRAINSTAIN CSN5 PROJECT

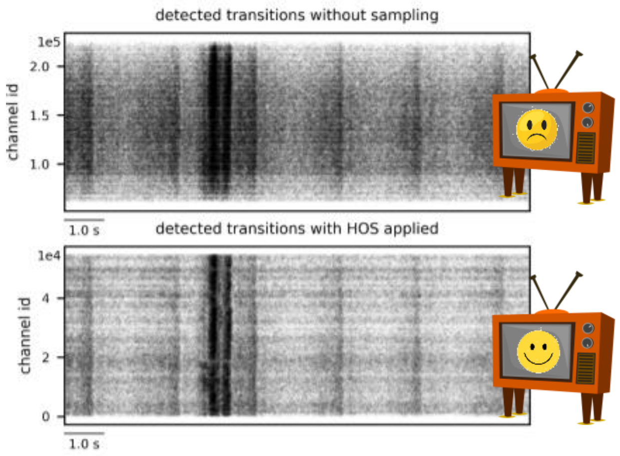


Initially developed on mice data from LENS, IDIBAPS
(more invasive techniques, simpler analysis)
doi: [10.1038/s42003-023-04580-0](https://doi.org/10.1038/s42003-023-04580-0)

Now moving to human data (simulations & EEG)
- TVB simulator; collaboration with UniMi
doi: [10.5281/zenodo.10361054](https://doi.org/10.5281/zenodo.10361054)

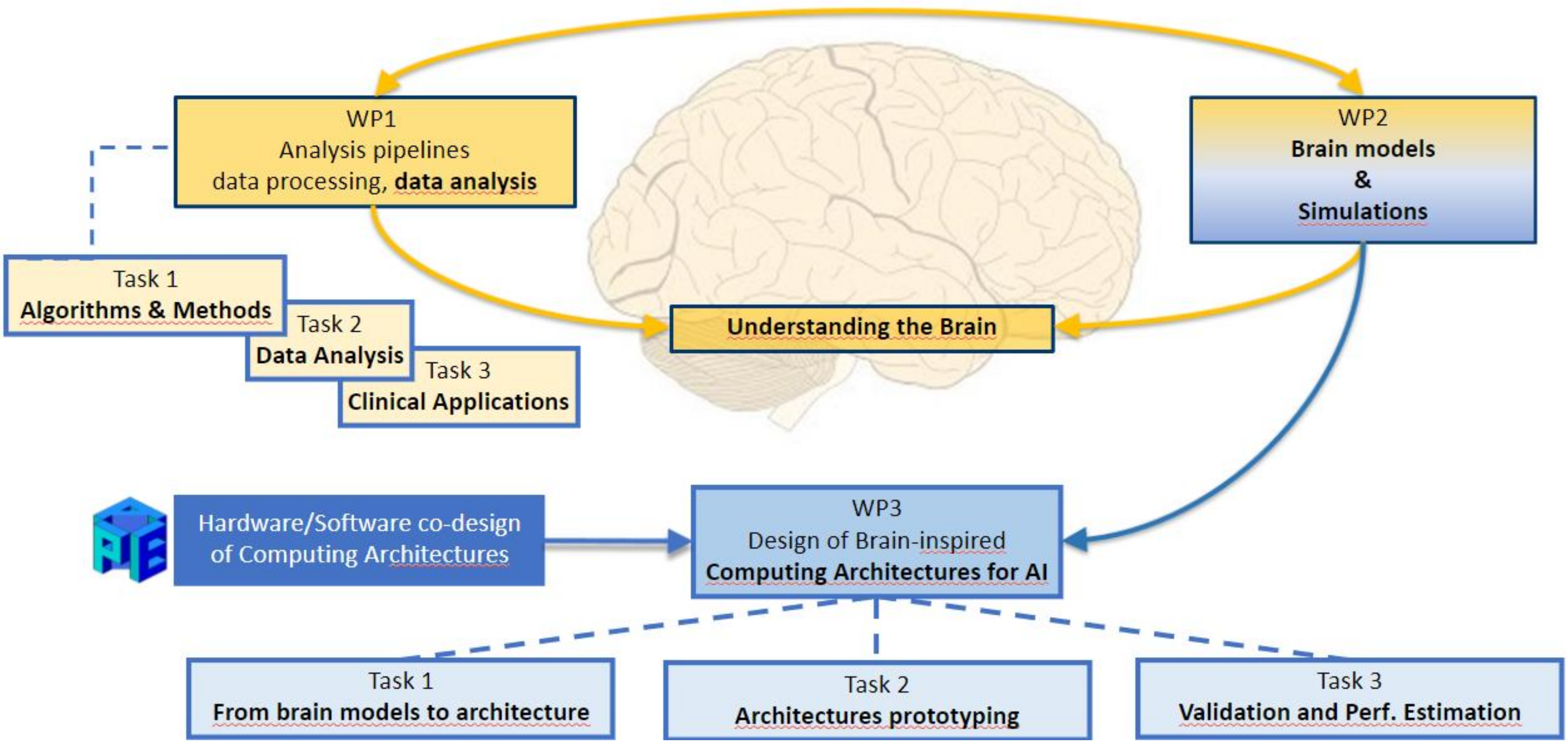


THEVIRTUALBRAIN.



High-res imaging data require smart approaches for optimal processing
Hierarchical Optimal Sampling (HOS):
- Heterogeneous downsampling, improved signal-to-noise ratio
- Smaller data size, faster processing

Computational Neuroscience & Neuromorphic Computing
BRAINSTAIN: BRAIN Studies and Technologies for Artificial Intelligence and Neuroscience



Simulation Toolkit

Geant4 (GEometry ANd Traking)

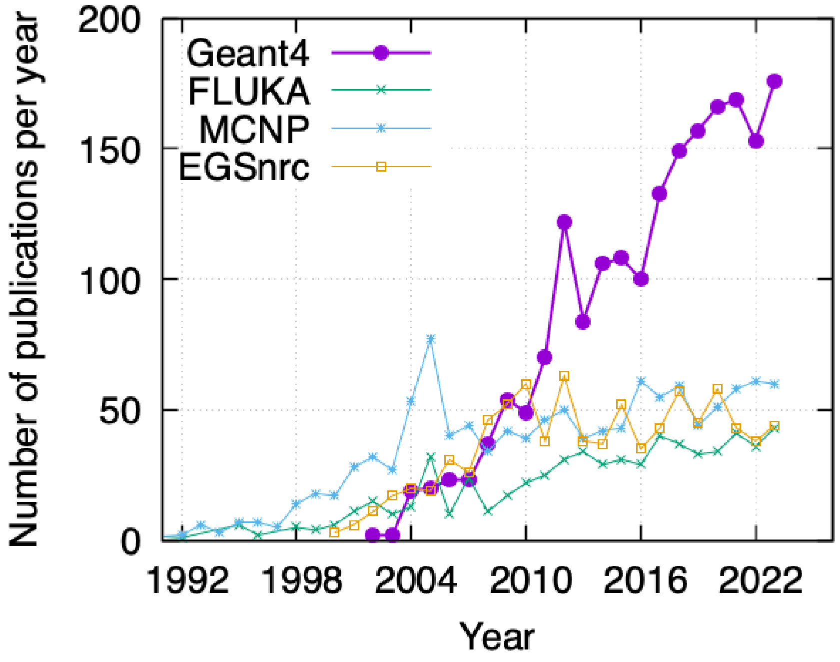
- MC Simulation Toolkit
- Developed by an International Collaboration
  - Established in 1998
  - INFN contribution from the beginning (to the kernel, the development of EM physics, the advanced examples …)
  - Approximately 100 members, from Europe, US and Japan
- Open source
  - Written in C++ language
  - Takes advantage from the Object Oriented software technology
- The most used MC tool for research in medical applications
- http://geant4.org

GEANT4
A SIMULATION TOOLKIT

[Geant4, a simulation toolkit Nucl. Inst. and Methods Phys. Res. A, 506 250-303

Geant4 developments and applications Transaction on Nuclear Science 53, 270-278]

Publications related to the most used MC tools on PubMed



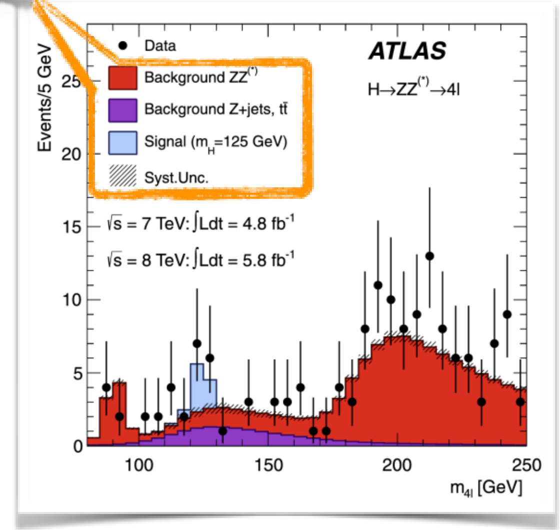
Geant4 app s

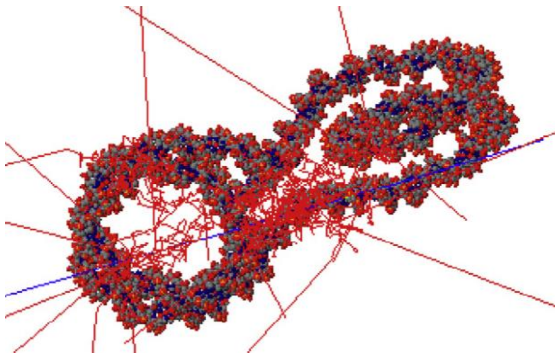
Geant4 output!

- Almost all particle and nuclear physics experiments have a Monte Carlo simulation developed with Geant4
- But it's also used for
  - medical applications
  - Radiobiology
  - Radio-protection
  - Shielding
  - Single event upset and radiation damages to electronics
  - Simulations for nuclear spallation sources

Invariant mass for the Higgs boson discovery in the decays golden channel (4 leptons) by ATLAS

Image from Physics Letters B 716 (2012) 1–29



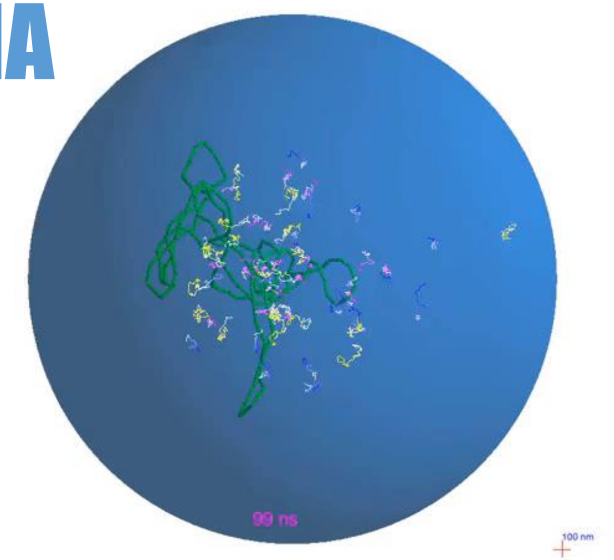


atomistic view of a dinucleosome irradiated by a single 100 keV proton

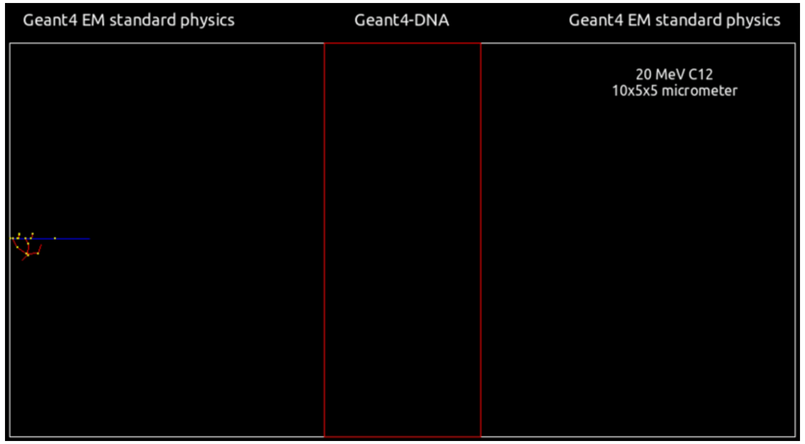
Image from M. A. Bernal et al Physica Medica, vol. 31, no. 8, pp. 861–874, Dec. 2015.

Geant4-DNA

- A Geant4 extension to perform radiation-matter simulations in the scale 1µm-10nm (cell to DNA)
- All elementary interactions are simulated on an event-by-event basis (no average approach)
- Description of target molecular properties
- Allows physico-chemical simulations
- http://geant4-dna.org



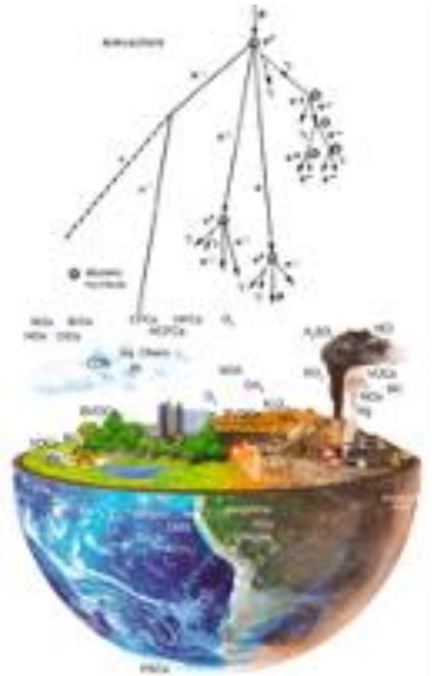
Irradiation of a pBR322 plasmid, including radiolysis
movie courtesy of V. Stepan (NPI-ASCR/LP2iB-CENBG/CNRS/IN2P3/ESA)



20 MeV C12 on water
movie from Y. Perrot XI International Geant4 school

INFN contribution to Geant4 development

- **Coherent interactions in crystals** 10.1088/1742-6596/898/4/042041
  - Photons: coherent scattering and reflection/refraction
  - Charged: channeling, volume reflection and coherent bremsstrahlung
- **Electromagnetic models to simulate cosmic rays ionisation of the atmosphere** 10.1016/j.ejmp.2023.102661
  - Using Geant4-DNA approach
  - e- on O2 and N2 (ionisation, scattering, and excitation)
- **Nuclear reaction models** 10.1016/j.ejmp.2019.10.026
  - Interface of models developed by INFN theoreticians to simulate nuclear reaction below 100 MeV/u
  - Testing Deep Learning to emulate the most cpu intense part of these models
- **Extended/advanced examples**
  - Internal dosimetry, compact crystal calorimeter, hybrid positron source, crystal deflector, medical linac

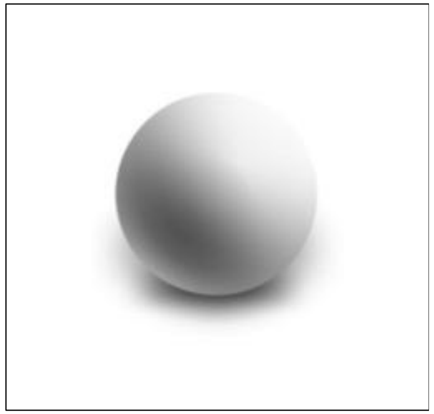


Computational studies for Particle Beam Radiation Biophysical Modeling

Why we need models in radiation biology?

- To make predictions on different radiation effects on cells/tissue
- To implement in Treatment Planning
- To understand and explain phenomena on physics bases (*computational microscopy*)

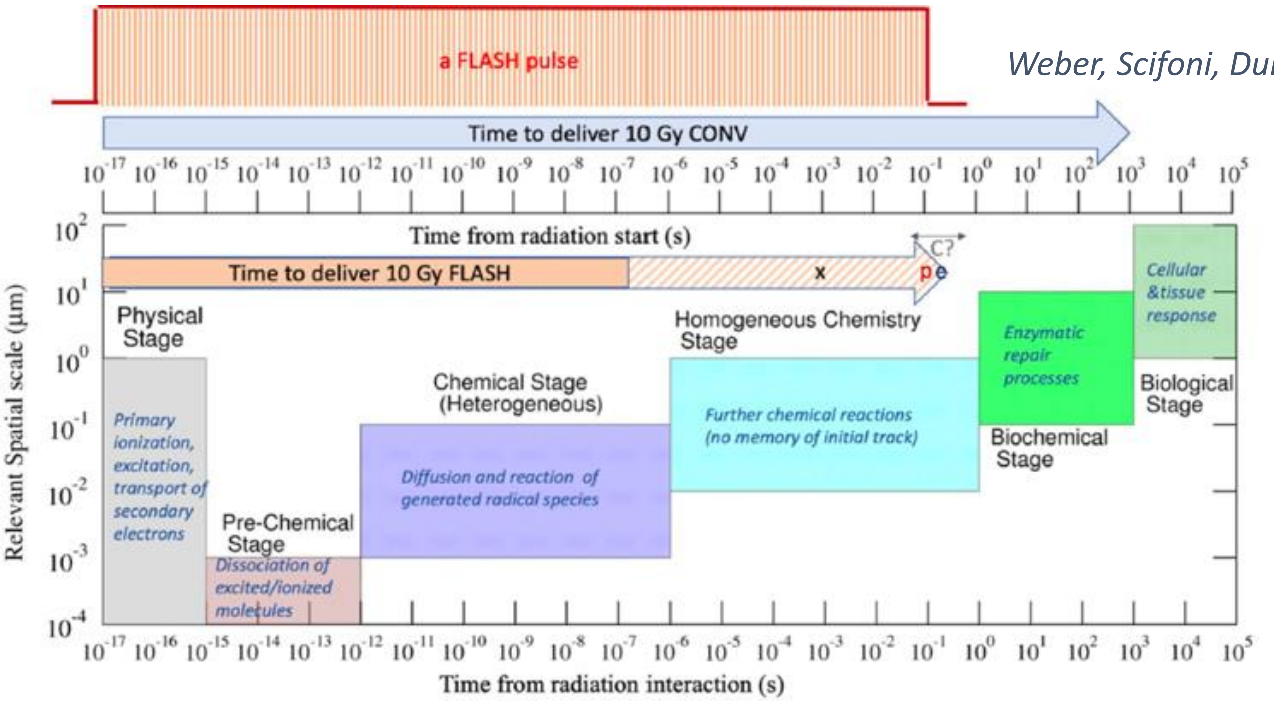
"This is not a cow"
--- René Magritte



"This is a cow"
--- Anonymous physicist

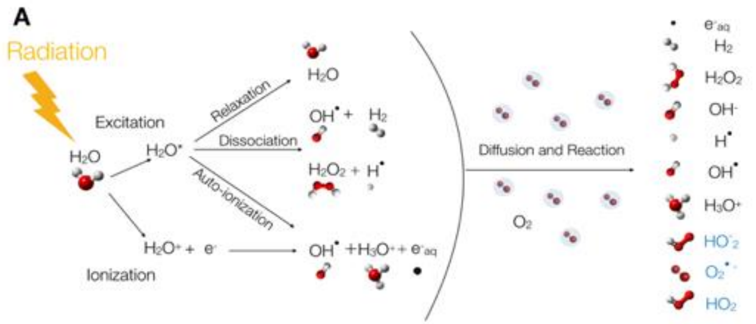
Courtesy from A.Attili

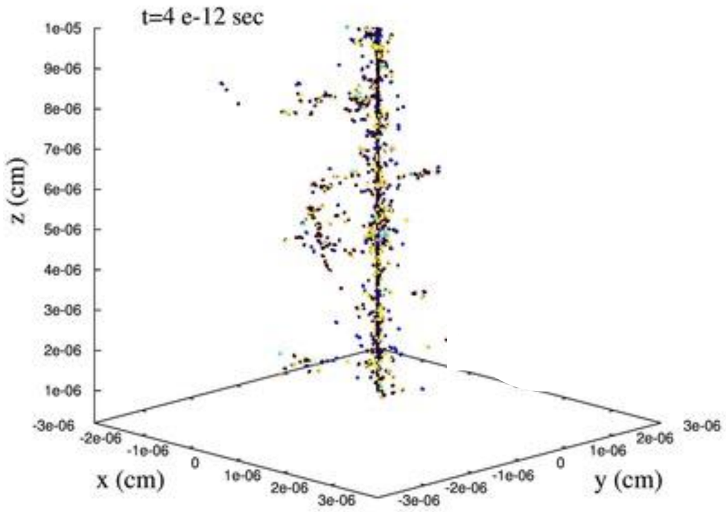
Multiscale modelling of the Ultrahigh dose rate radiation response



Weber, Scifoni, Durante 2021

Monte Carlo chemical Track Structure based

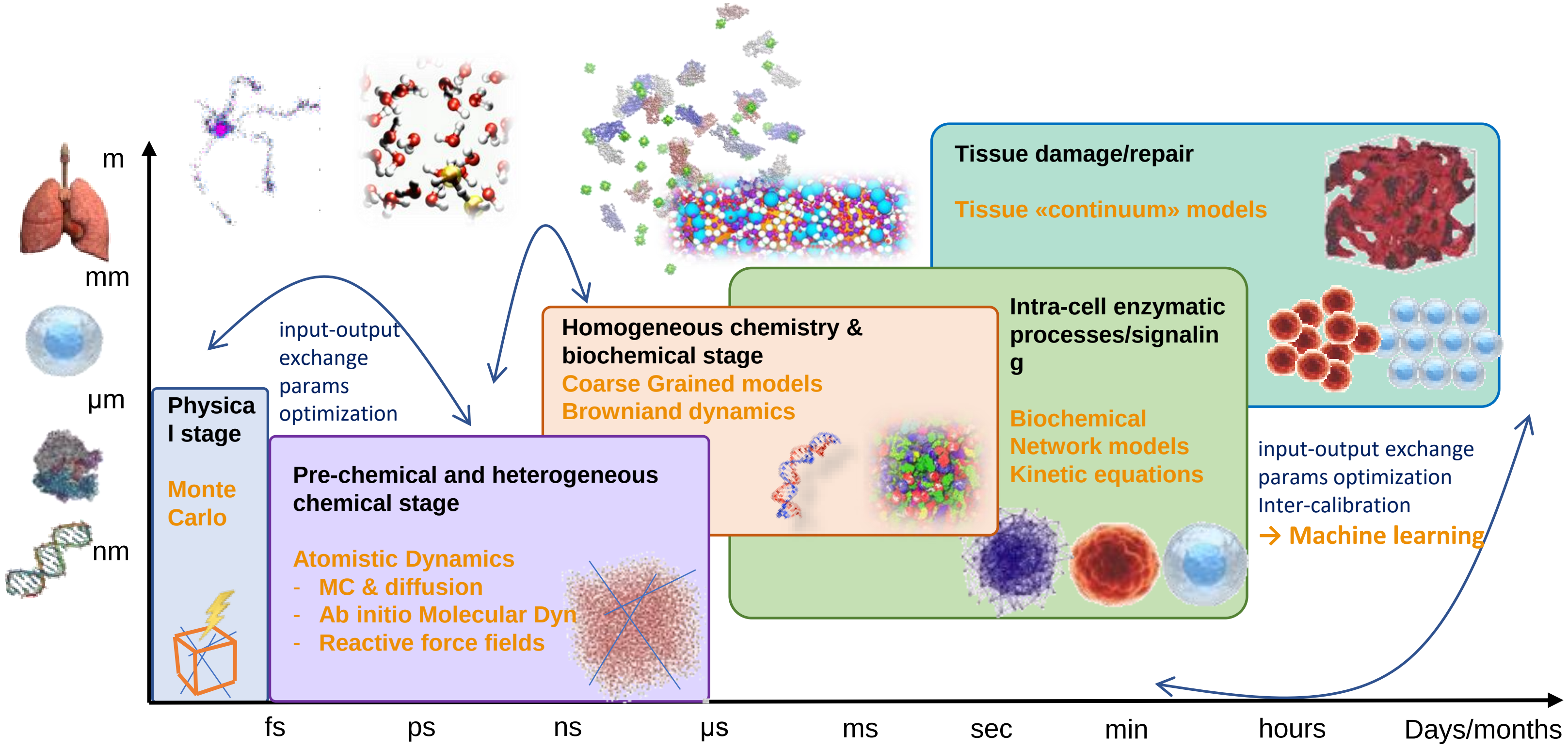


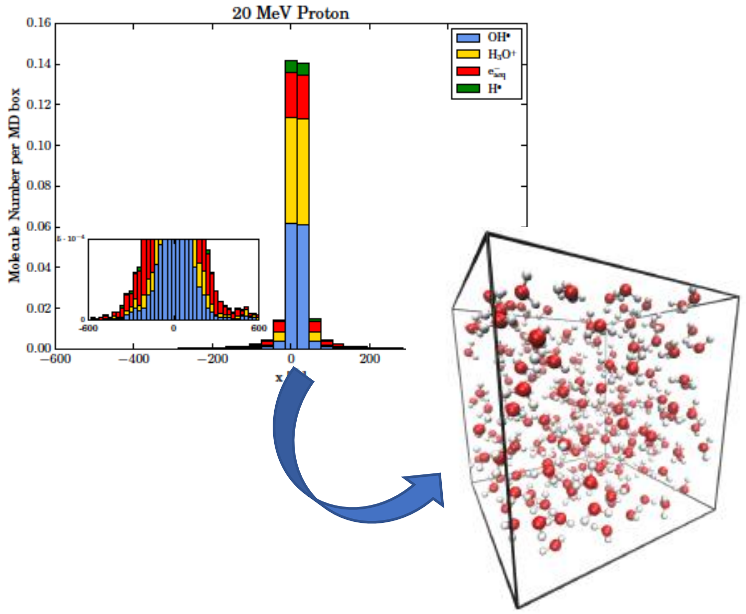


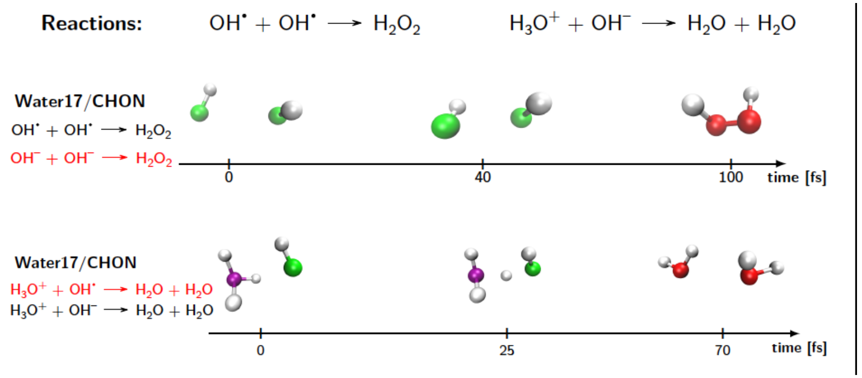
Understanding of the Biological response observed, e.g. in **FLASH radiotherapy** requires deep analysis of the full spatiotemporal cascade of events following the primary radiation events

Boscolo Scifoni, Kramer Drurante Fuss 2022

WIP: Joining Molecular Dynamics to MC Track structure







Need of reactive force fields (REAXFF, CHARMM)

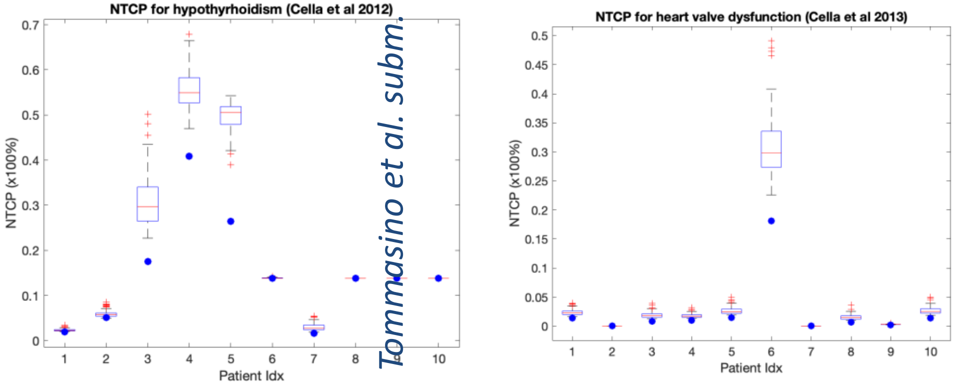
L. Castelli, V. Tozzini & E. Scifoni in prep.

Treatment planning and modeling studies: proton radiation toxicity

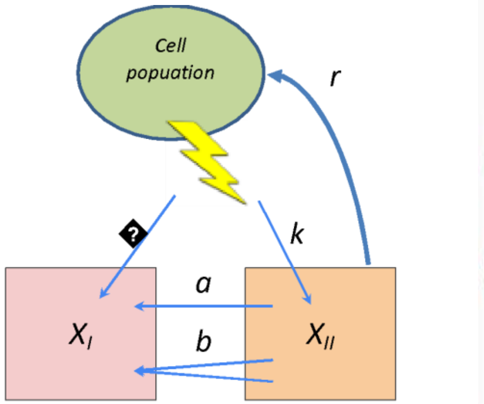
Strong Collaboration with Clinical staff @TPTC

- NTCP protons vs photons
  - impact of variable RBE on NTCP (fig)
  - NTCP optimization
  - NTCP for retreatments p+X

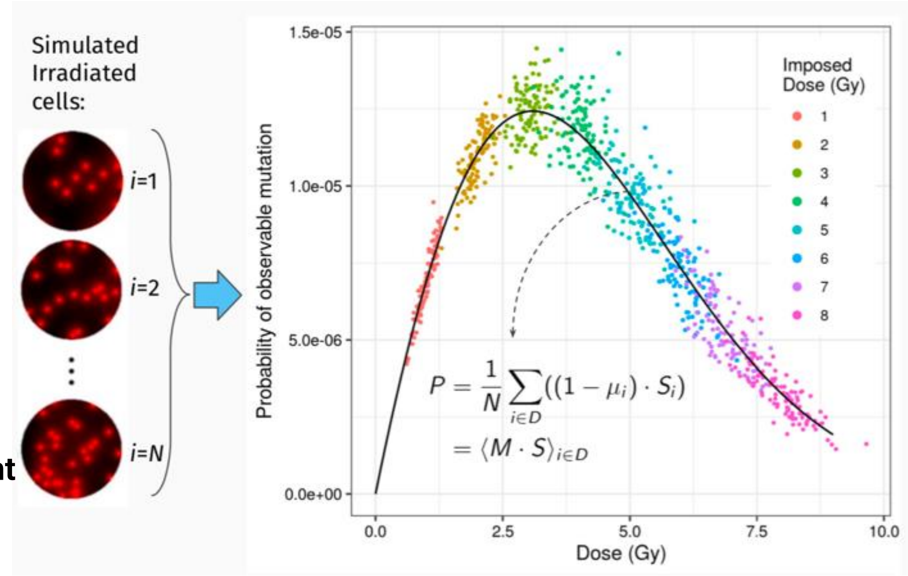
Tommasino et al. *PMB 2023*

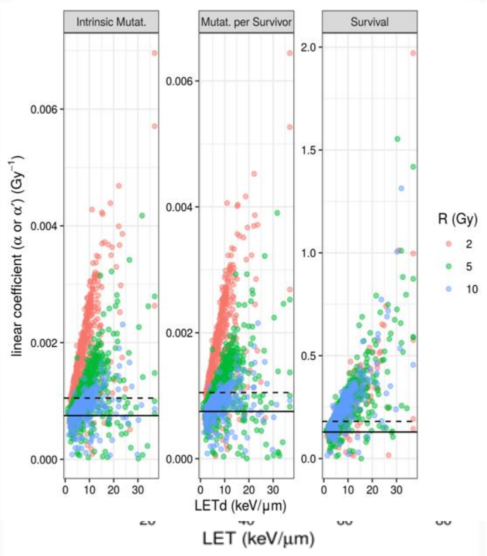


- Secondary Cancer Induction: protons vs photons



directly mutant **potentially mutant**





Attil et al. *PMB 2022*