Spring Workshop on Physics of Data

Report of Contributions

Contribution ID: **1**   Type: **not specified**

Registration

Wednesday 6 April 2022 14:00 (30 minutes)

Contribution ID: **2**   Type: **not specified**

Welcome and Introduction to the workshop

Wednesday 6 April 2022 14:30 (15 minutes)

Presenters: RINALDO, Andrea; SUWEISS, Samir (University of Padova); SENO, flavio (università di Padova)

Contribution ID: **3** Type: **not specified**

test

Contribution ID: **4**   Type: **not specified**

Timescales of neural activity, their inference and relevance.

Wednesday 6 April 2022 14:45 (45 minutes)

Timescales characterize how fast the observables change in time. In neuro science, they can be estimated from the measured activity and can be used, for example, as a signature of the memory trace in the activations. Inferring the
timescales seems to be an easy task; however, I will show you how the timescales are subject to a statistical bias that is impossible to remove by a simple mathematical transformation. Instead, I will advertise using a Bayesian method that infers the timescales by matching the statistics of the data. I will use the set of generating models with known timescales and search for the parameters that
give me the sample autocorrelation closest to the one from the data.
As a next step, I will use the method on the data recorded from a local population of the cortical neurons from the visual area V4. I will demonstrate that the ongoing spiking activity unfolds across at least two distinct timescales - fast and slow - and the slow timescale increases when monkeys attend to the
location of the receptive field. Finally, I will discuss this change's relevance for behavior and cortical computations.

Presenter: LEVINA, Anna (University of Tuebingen)

Session Classification: Machine Learning, Data Science & Interdisciplinary applications

Contribution ID: **5**   Type: **not specified**

Physics of data for social impact

Wednesday 6 April 2022 15:30 (45 minutes)

In a rapidly changing world, facing an increasing number of socioeconomic, health and environmental crises, physics of data and complex systems can help us to assess and quantify vulnerabilities, and to monitor and achieve the UN Sustainable Development Goals. In this talk, I will provide an overview of the main areas of applications where physics of data and complex systems has shown its potential for social impact. I will then review the challenges and limitations related to data, methods, capacity building, and, as a result, research operationalization, and conclude with some suggestions for future directions.

Presenter: OMODEI, Elisa (Central European University)

Session Classification: Machine Learning, Data Science & Interdisciplinary applications

Contribution ID: **6**   Type: **not specified**

Failure and success of the spectral bias prediction for Kernel Ridge Regression: the case of low-dimensional data

Wednesday 6 April 2022 16:45 (15 minutes)

Presenter: TOMASINI, Umberto (EPFL)

Session Classification: Machine Learning, Data Science & Interdisciplinary applications

Contribution ID: **7**   Type: **not specified**

Language-dependent model-based deep reinforcement learning

Wednesday 6 April 2022 17:00 (15 minutes)

Presenter: DAINESE, Nicola (Aalto University)

Session Classification: Machine Learning, Data Science & Interdisciplinary applications

Contribution ID: **8**   Type: **not specified**

Rethinking the Geometry of Machine Learning: "Physics of Data" Five Years Later

Wednesday 6 April 2022 17:15 (45 minutes)

In physics the data that is acquired in experiments are highly-controlled and often taken with specific goals in mind. However, the notion of a "Physics of Data" is about using mathematical tools developed in physics to understand data acquired in more open-ended and uncontrolled environments. As the data acquisition process becomes more opaque and distant from any particular purpose, we have to be

more careful about our assumptions. We are no longer protected by the remarkable intuition of experimental physicists to find the right thing to measure. Instead, the data is often whatever can be measured about some process in the world and our task is to sift through the resulting mess. The key concept we need is how data and relevance to some specific task interact to create a predictive model.

Note, the data may or may not support the chosen task and different tasks lead to different models even with the same data. This talk will outline how one of the crucial concepts in theoretical physics, fiber bundles or gauge theories, provides a framework for understanding how data, relevance and

models are related. For the particular example of computer vision, principal fiber bundles can play a key

role since the fibers are Lie groups much like in particle physics. These connections demonstrate the importance of understanding the geometric structure of the data using symmetry transformations well-known to fundamental physics. This may offer a more sensible way forward than the endless tinkering with neural network models in Deep Learning that more often stumble into success and failure without

much insight into how the data is actually organized in high-dimensional spaces. Building new tools

adapted from physics for this pursuit is an opportunity to fundamentally change how data is analyzed

and well-suited to those in the "Physics of Data" Masters program.

Presenter: BYERS, Jeff (Naval Research Laboratory)

Session Classification: Machine Learning, Data Science & Interdisciplinary applications

Contribution ID: **9** Type: **not specified**

TBD

Thursday 7 April 2022 15:30 (45 minutes)

Presenter: CALDARELLI, Guido (University of Venice)

Session Classification: Physics of Matter / Quantum

Contribution ID: **10**

Type: **not specified**

What is typical in microbial communities?

Thursday 7 April 2022 09:45 (45 minutes)

Microbial communities are high-dimensional systems, with many co-evolving heterogeneous species and many environmental factors, that vary across time and space. Due to the sequencing advancements of the last 30 years, ecology, a traditionally data-poor discipline, has transitioned to become a data-rich one. In the first part of the talk, I will discuss some open questions that drive modern quantitative microbial ecology research. In the second part of the talk, I will present our recent research results regarding the statistical properties of the dynamics of empirical microbial communities.

Presenter: JACOPO, Grill (ICTP)

Session Classification: Complex/Biological Systems

Contribution ID: **11**   Type: **not specified**

Knot classification in polymers through deep learning

Thursday 7 April 2022 11:00 (15 minutes)

One of the fundamental open problems in knot theory is their classification, which aims to discriminate whether two given closed curves are topologically equivalent or not. The
problem might be tackled with knot invariants, such as the Alexander polynomial, quantities that are the same for equivalent knots. Nevertheless, algorithms implementing
knot recognition through invariants might take extremely large time or even fail.
In this work, we study the problem of knot classification in polymers by using deep learning. In particular, we resorted to long-short term memory (LSTM), a recurrent neural
network architecture usually used to process time-series data.
We simulated polymers, including different chain lengths and knots types. After the simulation, we computed a set of generalized dihedrals along the polymer chains and we
used them to train the LSTM.
Our preliminary results are encouraging and seem to lead to a flexible and quick method for detecting knots in polymers.

Presenter: BRAGHETTO, Anna (University of Padova)

Session Classification: Complex/Biological Systems

Contribution ID: **12**   Type: **not specified**

Computational modeling of information encoding in the primary visual cortex

Thursday 7 April 2022 11:15 (15 minutes)

The visual cortex is the sensory area of the brain responsible for the information processing that underlies our visual perceptions. The first part of the cortex that receives input from visual stimuli is called primary visual cortex, or V1. It is the most studied area of the visual cortex, and probably the most studied sensory area of the brain in general.
Since the Nobel prize works of Hubel and Wiesel, understanding V1 neurons in terms of edge detectors, many experimental findings has been collected over the years, shedding light on many structural and functional aspects of V1 and of its neurons.

Our understanding of V1 information processing is, however, currently limited in two ways:
1) the experimental findings differs in nature and give us a "fragmented" picture of how v1 works.
2) until recently, most modeling approaches aimed at explaining only a small number of phenomena discovered through experiments, most of the times based on the presentation of simple synthetic stimuli, highlighting specific aspects of information encoding, but not engaging neural responses in ecological settings and therefore unable to capture the richness of v1 information encoding.

In this talk, I'll present the two computational approaches pursued my research group contributing to a better understanding of v1 neurons information encoding:
1) the development of deep learning models to predict neural responses and of tools to inspect them
2) the developments of large scale spiking neural networks simulations strongly constrained by biology that aim at achieving a cohesive understanding of v1 information encoding through an integrative approach of experimental findings

Presenter: BARONI, Luca (Charles University)

Session Classification: Complex/Biological Systems

Contribution ID: **13**   Type: **not specified**

How contact pattern influence epidemic outbreaks

Thursday 7 April 2022 11:30 (45 minutes)

I will present a data-driven approach to identify from physical proximity data features of human contact patterns that determine crucial properties of epidemic outbreaks. From the physical proximity data, we construct for each individual
a point-process-like representation of their contacts, from which we estimate the distribution of potential secondary infections for different disease models.
The resulting distributions drastically differ from randmized surrogate data. Building branching processes from this empirical data, we demonstrate how the clustering of contacts decreases the robustness of disease outbreaks and how the
cyclostationarity of contacts modulates the pace of epidemic spread.

Presenter: ZIERENBERG, Johannes (Max Planck Institute for Dynamics and Self-Organization)

Session Classification: Complex/Biological Systems

Contribution ID: **14**   Type: **not specified**

D-Wave as a generator of structural models for prototypical problems in materials science

Thursday 7 April 2022 14:45 (45 minutes)

The promise of quantum computing is to provide new methods to unveil the physics of molecules and materials that has been inaccessible to the conventional numerical modeling. Over the past few years, quantum annealers have grown in complexity to the point that the computation of molecular energies has
become a feasible application. Whilst typical approaches use quantum annealers to extract the ground state solution of an optimization problem, we suggest a new application as generator of structural models for disordered materials, where disorder appears from the competition between the different degrees of freedom. Starting from the representation of the crystal in terms of network, we map the relevant interactions
into Ising Hamiltonians and encode the disordered phases in the excited states spectrum of the target Hamiltonian. In our approach the quantum annealer is used to explore the energy surfaces and to identify stable and metastable phases of prototypical disordered materials.

Presenter: SILOI, Ilaria (University of Padua)

Session Classification: Physics of Matter / Quantum

Contribution ID: **15**   Type: **not specified**

Unraveling the mechanism of ice nucleation through rare events sampling and free energy calculations

Thursday 7 April 2022 09:00 (45 minutes)

Many interesting physical processes occur on timescales that are very long compared to the shortest significant timescale involved. For example, timescales for folding the smallest of proteins are in the range of microseconds to milliseconds, while small-amplitude motions of amino acid side chains occur within 1 fs.

This large difference of timescales can present serious computational challenges: 1 second of computational time (within one or two orders of magnitudes) would be needed to advance the simulation of 1 fs of physical time. Therefore, the investigation of processes like chemical reactions, diffusion in solids, protein folding, and nucleation processes require the use of rare events sampling methods.

Here I will demonstrate the suitability of rare events techniques to investigate the nucleation mechanism of ice. The freezing of water affects the processes that determine Earth's climate, therefore, accurate weather and climate forecasts hinge on good predictions of ice nucleation rates. Such rate predictions
are based on extrapolations using classical nucleation theory (CNT). CNT assumes that the nucleation mechanism is one-step, that the reaction coordinate is the size of the critical nucleus and that the thermodynamic properties of the
crystallite at the top of the barrier are the same as for the bulk. Transition path sampling and free energy calculations will be used to test these assumptions and to unravel the mechanistic pathways leading to ice nucleation in the atmosphere.

Presenter: LUPI, Laura (Roma Tre University)

Session Classification: Complex/Biological Systems

Contribution ID: **16**   Type: **not specified**

Rethinking classical optimization in the age of quantum computers

Thursday 7 April 2022 16:45 (15 minutes)

The development of quantum computers is one of the most intriguing and motivating challenges of the current century. Thanks to their inborn quantum nature, these machines are expected to offer an unprecedented computational advantage over classical machines in solving highly complex computational problems that span from the simulation of quantum systems to quantum chemistry and material science.

In addition to all this, the future availability of these powerful machines is opening new scenarios in the areas of global optimization and machine learning. Standard computers have been used successfully on a large variety of problems of this kind over the past decades. However, increasing interest has been spreading throughout the scientific community to understand whether using quantum resources may provide a computational advantage over classical ones in this context, too. Since large-scale quantum computers are still under development and their availability is still limited,

different quantum-inspired methodologies able to mimic some aspects concerning the functioning of quantum computers have been developed in the past years to bridge the gap toward the development of algorithms that can exploit this type of resources.

During this brief talk, I will show how to adapt a general optimization problem to quantum hardware by presenting some recent applications of two of the most popular quantum-inspired approaches available, namely Tensor Network Methods and the D-Wave annealers.

Presenter: CAVINATO, Samuele (IOV)

Session Classification: Physics of Matter / Quantum

Contribution ID: **17**   Type: **not specified**

TBD

Thursday 7 April 2022 17:00 (15 minutes)

Presenter: SCHIMMENTI, Vincenzo Maria (Univ. Paris-Saclay)

Session Classification: Physics of Matter / Quantum

Contribution ID: **18**   Type: **not specified**

Question answering in the medical field

Thursday 7 April 2022 17:15 (15 minutes)

Natural language processing (NLP) is the ability of a computer to understand human languages. In both the academic and the industrial world, NLP has been widely used for different purposes such as Sentiment Analysis, Semantic
Text Similarity (STS), Text Translation, and Question Answering (QA), to cite a few.
With the advent of the Transformer model architectures like BERT, the performances of these tasks had huge improvements, by continuously reaching better accuracy levels. Despite these models requiring huge training
resources, their ability to learn the human language makes them astate-of-the-art solution for a plethora of NLP tasks.
A QA task is the ability of a deep learning model to extract an answer from a text (e.g.: a medical history of a patient), given an input question (e.g: is a patient suffering from this symptom?).
QA in clinical and medical notes of patients has gained a lot of attention in recent years. Extracting important features for the diagnoses, or helping the doctors to figure out symptoms and correlations between them, is in fact becoming essential to properly take care of a patient and to help doctors in
their diagnoses.
In this talk, we will have a brief introduction to how Transformers models can be used to extract answers to medical questions from clinical patient notes.
Moreover, we will analyze an example of how the Semantic Text Similarity can be used to search for different patients that have rare symptoms, and that can be related to the same diagnoses.

Presenter: CAMPESE, Stefano (stefano.campese.90@gmail.com)

Session Classification: Physics of Matter / Quantum

Contribution ID: **19**   Type: **not specified**

Predicting solar activiy

Friday 8 April 2022 09:30 (45 minutes)

Presenter: ALBERT, Carlo (Eawag Institute)

Session Classification: Astrophysics and Cosmology

Contribution ID: **20**   Type: **not specified**

Data Challenges in "Gravitational-Wave Paleontology"

Friday 8 April 2022 10:15 (45 minutes)

The rapidly increasing population of detected gravitational wave sources carries valuable information about the properties of black holes and neutron stars, such as their rates,
masses and spins, that we aim to use to probe their progenitors and answer two of the big open questions in Astronomy today: "How do these sources form?"and "What can we learn from their gravitational waves about the birth, lives and explosive deaths of stars?"New gravitational-wave observing runs and next generation detectors will rapidly provide data with ever increasing precision and volume. However, on the theory side we are limited in answering these questions due to "the great gravitational-wave formation channel challenge": uncertainties within the modeling of the formation channels leading to gravitational-wave sources are so large, that disentangling formation channels, and learning about their progenitors is computationally expensive and seems completely out of reach for the gravitational-wave field in the coming decades. In this talk we will interactively discuss these challenges, and how to overcome them using modern-day machine learning techniques.

Presenter: BROEKGAARDEN, Floor

Session Classification: Astrophysics and Cosmology

Contribution ID: **21**   Type: **not specified**

Pulsars spin-up: a semi-analytical approach

Friday 8 April 2022 12:15 (15 minutes)

Pulsars are powerful probes of our universe: thanks to their extraordinary long-term rotational stability and their fast rotation, they allow extremely precise timing measurements. However, the physics behind their spins and magnetic fields evolution is still poorly understood. A particular interest resides in the process of spin-up: neutron stars in binary systems
can be spun-up by accreting matter from the companion. The correct modelling of these processes is of fundamental importance in order to accurately reproduce the observed population of pulsars. I will talk about how we implemented the evolution of spins and magnetic fields of neutron stars in SEVN, a C++ based binary population synthesis code, focusing on the main results and discussing the main criticalities of this approach.

Presenter: SGALLETTA, Cecilia (SISSA)

Session Classification: Astrophysics and Cosmology

Contribution ID: **22**   Type: **not specified**

Gravitational Waves and Machine Learning

Friday 8 April 2022 12:30 (15 minutes)

The LIGO-Virgo collaboration has detected dozens of gravitational wave signals so far, and will do so at an increasing rate in the following years with detector
upgrades. These signals are extremely faint and arrive to us buried in noise; measuring and analyzing them is a hard computational challenge.
I will discuss how machine learning can help in this task, mostly focusing on the theoretical/modelling side: a neural network can learn a computationally expensive function and reproduce its results "by memory", speeding up Bayesian
inference.

Presenter: TISSINO, Jacopo (GSSI)

Session Classification: Astrophysics and Cosmology

Contribution ID: **23** · Type: **not specified**

A new tool on the workbench: studying GW progenitors with SEVN

Friday 8 April 2022 11:30 (45 minutes)

In 2015, the LIGO/VIRGO interferometers detected the first gravitational wave (GW) signal coming from the merger of two black holes. Since then, about 90 merging binary compact
objects (BCOs), namely binary neutron stars and black holes, have been detected through GW signals. This wealth of new data provides us with crucial insight on the populations
of BCOs. For this reason, numerical tools to simulate the evolution of stars and binary processes leading to the formation of BCOs are needed.
In this talk, I present SEVN (Stellar Evolution N-body), a state-of-the-art population synthesis code we are developing in our group. The stellar evolution is implemented interpolating evolutionary tracks on the fly, while binary processes are simulated with analytic and semi-analytic prescriptions.
I will highlight what are the novelties and the key differences of SEVN with respect to other population synthesis code, especially regarding the computation of stellar evolution
and the prediction/correction adaptive time step schema. Examples of scientific exploitation of SEVN in the investigation of BCOs/GWs progenitors will be shown.
Finally, I will describe how the synergy between our group and Physics of Data students is helping in exploiting Machine Learning algorithms to assist SEVN and boost its capabilities.

Presenter: IORIO, Giuliano (University of Padua)

Session Classification: Astrophysics and Cosmology

Contribution ID: **24**   Type: **not specified**

Artificial Intelligence and Quantum Computing for High Energy Physics: examples from CERN Openlab

Friday 8 April 2022 15:20 (45 minutes)

Presenter: VALLECORSA, Sofia (CERN)

Session Classification: Fundamental Physics

Contribution ID: **25**   Type: **not specified**

Data science for fundamental physics at the Large Hadron Collider

Friday 8 April 2022 14:35 (45 minutes)

The Large Hadron Collider (LHC) at CERN is one of our most powerful tools to probe the fundamental particles of nature and their interactions. By colliding protons at extremely high energies (13 TeV centre of mass), the LHC can probe conditions of the early universe just after the Big Bang. Particle detectors, such as the Compact Muon Solenoid (CMS) experiment are designed to reconstruct the proton collisions from the complex system of particles that are produced in such collisions and recreate the fundamental interaction that
occurred. Detectors like CMS are capable of recording huge quantities of data to do this and as experimental particle physicists, our job is to analyse these data to determine whether or not some new particle (like the Higgs boson) or process can be seen in the data. Producing and collecting data at the LHC is an expensive task, so making the most out of the data we have is vital in our field. In this talk, I will discuss the way we analyse data from experiments like CMS and how analysis of data led to the discovery of the Higgs boson. I will discuss techniques from the fields of machine learning and data science that have been used to analyse our data and new methods being proposed to analyse data from CMS and potentially discover new physics in future data taking runs of the LHC.

Presenter: WARDLE, Nicholas (Imperial College of London)

Session Classification: Fundamental Physics

Contribution ID: **26**   Type: **not specified**

Model-independent search for New Physics at the LHC

Friday 8 April 2022 16:20 (20 minutes)

Experimental observations and convincing conceptual arguments indicate that the present understanding of fundamental physics is not complete, motivating the search for physics beyond the Standard Model at collider experiments. The most common searching strategy is to test the data for the presence of one candidate new theory at a time and therefore optimise the data analysis to be sensitive to the specific features predicted by that theory. This model-dependent approach is in general insensitive to sources of discrepancy that differ from those considered. There is therefore a strong effort in developing analysis strategies that are instead agnostic about the nature of potential new physics and thus complementary to the former ones. Signal-model-independent analysis aim at detecting any departures from a given reference hypothesis, like the Standard Model. In practice, this is a challenge given the complexity of the experimental data in modern experiments and the fact that the new physics is expected to be "small"and/or located in a region of the input features which is already populated by standard events. Recently, there has been a strong push towards developing solutions based on machine learning for (partial or full) model-independent searches in high energy physics. In this talk I am going to review some of the newest machine-learning- based techniques pushing the frontiers of model independent searches at collider experiments.

Presenter: GROSSO, Gaia (University of Padua, CERN)

Session Classification: Fundamental Physics

Contribution ID: **27**   Type: **not specified**

Conclusions

Friday 8 April 2022 16:40 (10 minutes)

Presenter: ZANETTI, Marco (DFA, Università di Padova)

Session Classification: Fundamental Physics