Major scientific achievements

All references clickable, coloured [refs] are from my group

I started my scientific career in the early 2000 as a **theoretical physicist**, with an interest in condensed matter theory, statistical physics and in particular disordered amorphous systems such as magnets & glasses, a field where I obtained a number of remarkable results. Starting from my postdoc, however, my interests have gradually shifted toward **computer science & applied mathematics** and my activity has been increasing moving toward this direction. The first stop on this journey has been the theoretical studies of random instances of **NP-complete constraint satisfaction problems.** In [KMR07] I described a series of phase transitions that crucially influence the difficulty of **Boolean satisfiability** and **graph colouring**. This paper, cited more than 400 times, has set up a whole new direction of research and a considerable part of this program has been proven rigorously by leading mathematicians & computer scientists (and has been the subject of several ERC grants). The latest of those, by myself and collaborators, was published in *Advances in Mathematics* [CPZ18]. The field that I have contributed creating in [KMR07] was the subject of a semester in the Simons Institute of computing in Berkeley in 2016, where I was invited. After this visit, I concentrated on **statistical inference** & **learning problems,** a direction that is leading to **OperaGOST**.

There have been an increasing convergence of interest and methods between theoretical physics and probability theory, machine learning, and optimisation, and I have been at the forefront of this transformation. I am considered as one of the leader of this emerging new field as attested -for instance- by the many visiting scholar positions (Los Alamos, Tokyo, Berkeley, Duke, Santa Barbara), frequent invitations to conferences and departments in physics, mathematics, and computer-science. **Since September 2020 I hold a full professor position *both* in theoretical physics & Applied Mathematics departments in EPFL.** In particular, I edited a book [KRZ13], wrote a review [ZK16] on the subject, and organised a number of schools & international events (*see cv*). Over the last 10 years only, I have written more than 100 articles, spanning all subfield of computer science & engineering in high-impact selective conferences such as NeurIPS, COLT, ICML, (learning) STOC (Computer science), ISIT (information theory), and in journal as diverse as Annals of statistics, Advances in mathematics, IEEE transition on information theory, or PNAS. Nevertheless, a common thread, the use of mathematical statistical physics and probabilistic methods, links these different activities more than the titles would let appear at first sight. On these topics, I am also particularly well known for my contributions to the following problems:

—- **Community detection**. I studied the stochastic block model, a fundamental model in statistics, machine learning, and network science where it serves as a useful benchmark for the task of recovering community structure in graph data. My work [DKM11], cited more than 600 times, predicted strict limits in the possibility to learn the community form graphs, and has been the source of an intense collaboration between mathematicians ever since, literally created a sub-field of statistical inference. Subsequently, I showed in another high-impact paper (more than 450 citations) how to reach these limits efficiently in many cases with spectral algorithms [KMM13,SKZ14].

—- Most signals can be reconstructed from far fewer measurements than was generally considered necessary, saving time, cost and accuracy, and this **Compressed Sensing** approach has revolutionised signal processing. Financed by ERC (*SPARCS 307087, PE7*) on this subject, I developed physics-inspired algorithms and new theoretical approaches that solved long standing open problems. In particular, [KMS12] (more than 250 citations) showed how to further improve compressed sensing procedures and has been the source of subsequent breakthroughs. I was blessed to see David Donoho, the inventor of compressed sensing, mention my contribution to the field during his Gauss prize lecture at the International Congress of Mathematics in 2018.

—- A key contribution I made to **Quantum Computting** has been the development of the quantum cavity method, a powerful mathematical tool which I then use to demonstrated the failure of the *adiabatic quantum algorithm* that was proposed by a group of researchers at MIT (see grant list) for the most difficult optimisation problems. These results put a strong limitation on the performance of future quantum computers. Among many publications, I wrote a monograph on these topics in Physics Reports [BFK12].

I will now summarise some of the **major scientific achievements in the last five years:**

—- **Error Correcting Codes/informationt theory**. Using the both insights from compressed sensing and theoretical physics, designed [BK17] a new communication error correction scheme (sparse spatially-coupled superposition codes with message passing) that efficiently reached the optimal Shannon limit and that has been shown to be universal with respect to the coding channel. It is a candidate for 6G communication, and has been the subject of a tutorial at ISIT **[link].**

—- I discover new ways of performing **Optical Computations:** Through a collaboration with the Kastler Brossel laboratory in Ecole Normale, we used an interdisciplinary approach between signal processing, algorithms & physics to study optical wave propagation in complex environments In particular, we have shown how to use a scattering medium to make compressed imaging under the Shannon boundary and to facilitate the measurement of the transmission matrix [DK15], but also to optically implement Machine Learning algorithms [SC+16,OWD19,LPBK20], which we have patented [GDC16]. This has led, in particular, to the creation of the startup LightOn and to the attribution of the Joseph Fourier Price.

I am currently deeply interested by the **intersection between statistics, probability, machine learning and computation**, in particular in understanding the fundamental limits of extracting information from noisy data, and the algorithmic feasibility considerations surrounding this task. A large part of my activity over the last few years has focused over such problems in fields such as signal processing and statistical learning, where I have been notably productive (see full cv), publishing during the last five years in all major conferences of computer science (*STOC*), information theory (I*SIT, ITW*), Statistical Learning (*COLT*), signal processing (*ICASSP*) & machine learning (NeurIPS, ICML). In the last 6 months alone, I have published 6 papers at the highly selective NeurIPS [MLKZ,MKUZ,AKLZ,MBC+,DORK,LPBK] and 3 at ICML [GLK+,MKLZ,ARBK], thus ranking among the most productive scholar in machine learning. Given the shear amount of work, I will only highlight some of the most striking results:

—- **Mathematically rigorous approach to statistical physics predictions:** An important aspect of OperaGOST (WP3 in particular) concerns this part. Given a large part of my scientific reasonning have been exploiting powerful yet heuristic methods from statistical physics, an important aspect of my research has been to put a firm rigorous basis over these methods to turn them into proper mathematic tools. I have obtained trailblazer results in this direction that allows to close some conjecture open for decades in information theory [BMDK17], Optimisation [CPZ18], machine learning [BKM18,AKG20] and neural networks [AMB+18]. Starting from my seminal work [KXZ16], have been at the forefront of a set of work establishing a series of equivalence between formal statistical physics results and rigorous mathematical predictions. I do believe that, for a successful transfer of technology between physics and applied mathematics, rigorous results are a fundamental and much needed asset.

— Deeply related to the subject of the *OperaGost* proposal as well are the **Limits of statistical learning in high-dimension:** I have obtained many fundamental results on these fundamental limits in a range of problems such as statistical estimation in generalised linear models [BKM19], planted colouring and constraint satisfaction problems [CPKZ17] and neural networks [AMB+18] compressed sensing and phase retrieval [MSML20]. I certainly consider this one of the most exciting direction in my research. Such high-dimensional problems were the subject of a full semester in the Simons Institute of computing in Berkeley in 2020 (where I have been invited and gave a set of lecture on the subject).

—- **Matrix and Tensor factorisation and random matrices:** In data science and machine learning, large, high dimensional datasets collected across multiple modalities can be organised as a matrix or a higher order tensor. Low-rank matrix and tensor decomposition thus arises as a powerful and widely used tool to discover low dimensional structures underlying the data. Using deep connection with information theory and random matrices, I have written a series of works given a systematic and powerful way to derive optimal algorithms, & statistical theory results, for the matrix [BDM+16,LKZ17] and tensor [LML+17].

—- **Phase retrieval,** the recovery of a complex-valued object from amplitude measurement only, is certainly one of the most important applied problem in signal processing and experimental physics. I have been mostly concentrating on the problem of phase retrieval from random measurement, as appeared in compressed phase retrieval as well as in optical imaging with complex media. I was able to derived fundamental theoretical limits to the problem [ALB+20,MLK+20], as well as to propose new powerful algorithmic powerful methods [DK15,RT+16,RGK17,RGK17b,DKG20].

—- **Generative models are the new sparsity:** Using a low-dimensional parametrisation of signals is a generic and powerful way to enhance performance in signal processing and statistical inference. For a while, the most widely used type of dimensionality reduction was sparsity. A new, powerful approach — that I have been among the early proponents [TDK16,TMC16] —- is to use generative modelling of signal distributions via neural networks such as GANs or Auto-encoders, that allows to drastically reduce the number of measurement needed. While extremely powerful, this methods lacked theoretical guarantees. I have been able to produced pioneering theoretical works showing the amazing fundamental performance of using neural networks as priors for statistical reconstruction [ALMKZ19,ALBKZ20].

—- **Neural Networks:** One of the main question I have been trying to answer over the last years is to understand the "Unreasonable effectiveness of learning neural networks": why do they work so well? I have been at the forefront of the statistical physics approach to these questions, deriving key results on the structure of two layer neural nets [AMB+18], their dynamics of learning [GAS19], and the question of over-parametrization [ARBK20]. Recently, I have been trying to explore the loss (or energy) landscape of these non convex problems [SKU19,SBC19], and the deep connection I established between energy landscape in physics, random matrices and the theory of glasses are a core part of the OperaGOST project [MBC20+,SBC20].