

## SEMINARIO INFN

Giovedì 20 Marzo 2025  
Aula Rutherford – ore 15  
Plesso di Fisica, Parma

### **Bilinear Sequence Regression: A Model for Learning from Long Sequences of High-dimensional Tokens**

**Vittorio Erba** (EPFL, Losanna)

Current progress in artificial intelligence is centered around so-called large language models that consist of neural networks processing long sequences of high-dimensional vectors called tokens. Statistical physics provides powerful tools to study the functioning of learning with neural networks and has played a recognized role in the development of modern machine learning. The statistical physics approach relies on simplified and analytically tractable models of data. However, simple tractable models for long sequences of high-dimensional tokens are largely underexplored. Inspired by the crucial role models such as the single-layer teacher-student perceptron (aka generalized linear regression) played in the theory of fully connected neural networks, in this paper, we introduce and study the bilinear sequence regression (BSR) as one of the most basic models for sequences of tokens. We note that modern architectures naturally subsume the BSR model due to the skip connections. Building on recent methodological progress, we compute the Bayes-optimal generalization error for the model in the limit of long sequences of high-dimensional tokens, and provide a message-passing algorithm that matches this performance. We quantify the improvement that optimal learning brings with respect to vectorizing the sequence of tokens and learning via simple linear regression. We also unveil surprising properties of the gradient descent algorithms in the BSR model.