



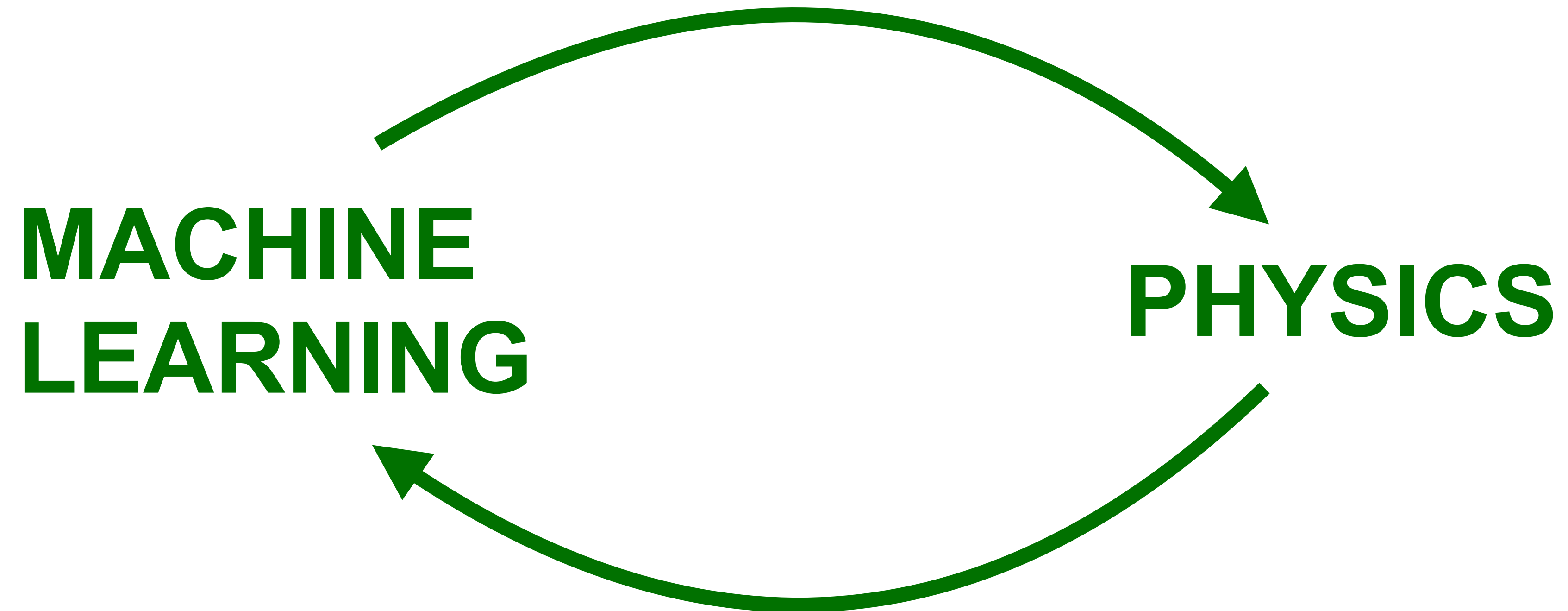
FELLINI MEETING 30/05/2022 FERRARA

STATISTICAL PHYSICS OF MACHINE LEARNING WITH GEOMETRICALLY STRUCTURED DATA

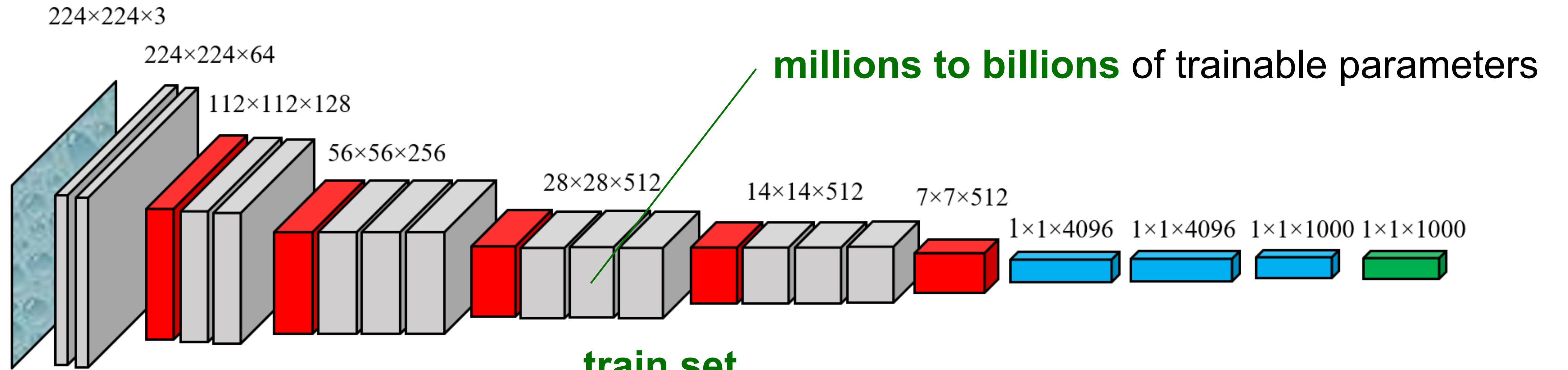
Pietro Rotondo — INFN, University of Milan



MACHINE LEARNING & PHYSICS: DISCLAIMER



OVERPARAMETRISATION IN DEEP NETS: A BLESS FOR PRACTITIONERS, A CURSE FOR THEORISTS



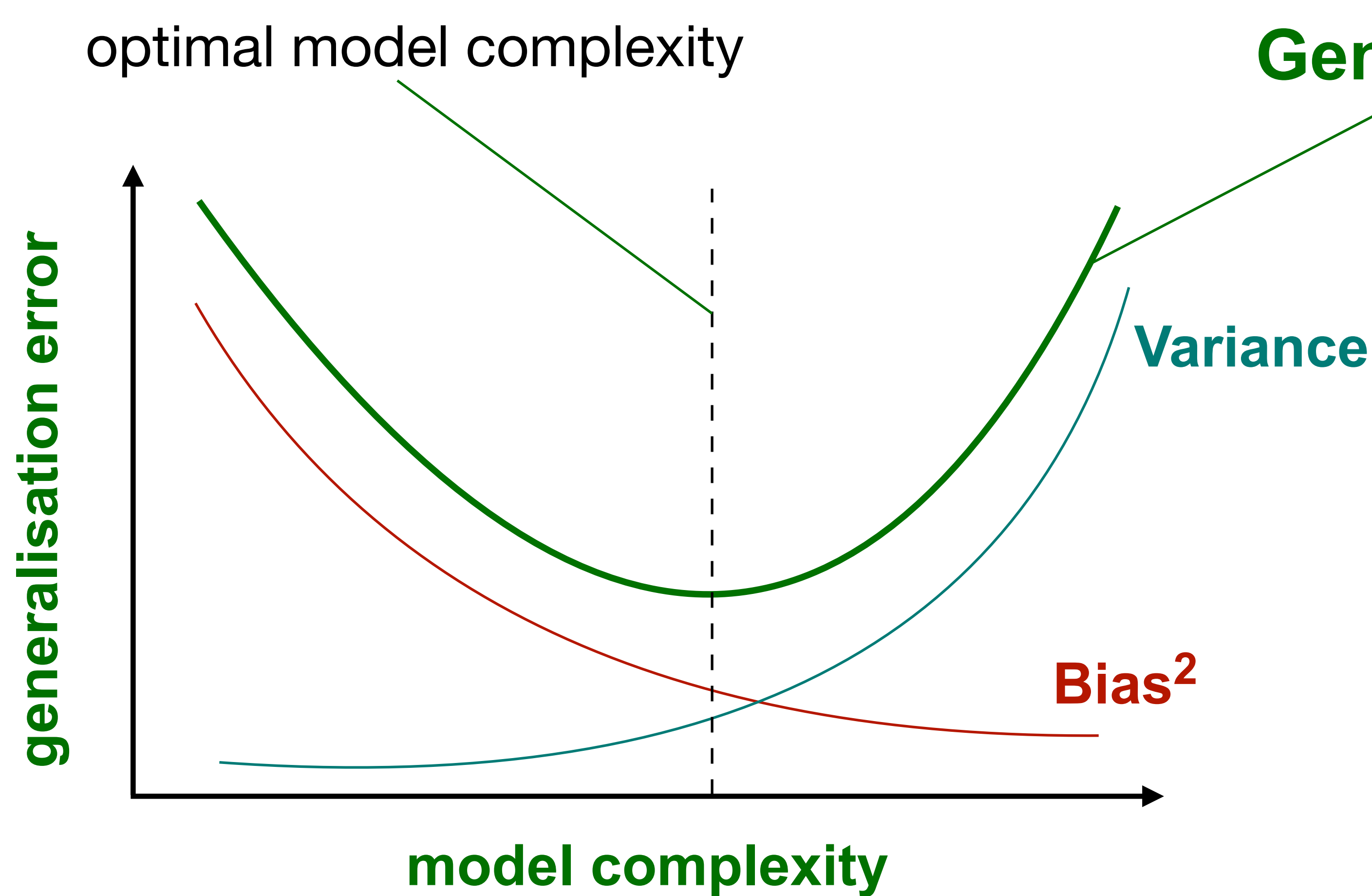
overparametrised regime

number of trainable parameters \gg size of the training set

0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9	9	9	9	9	9

generalisation

BASIC INTUITION AND THEORY SUGGEST THAT DEEP LEARNING SHOULD NOT WORK



Vapnik-Chervonenkis upper bound

$$\Delta \epsilon \lesssim \sqrt{\frac{d_{VC}}{P}}$$

gen error - train error

the VC dimension of a DNN is (very) roughly proportional to the number of trainable parameters

size of the training set

$\sim 10^6 - 10^9$ $\sim 10^4 - 10^6$

**Let's employ Statistical Physics
to investigate Deep Learning!**



Ludwig Boltzmann



Elisabeth Gardner



Haim Sompolinsky



Giorgio Parisi

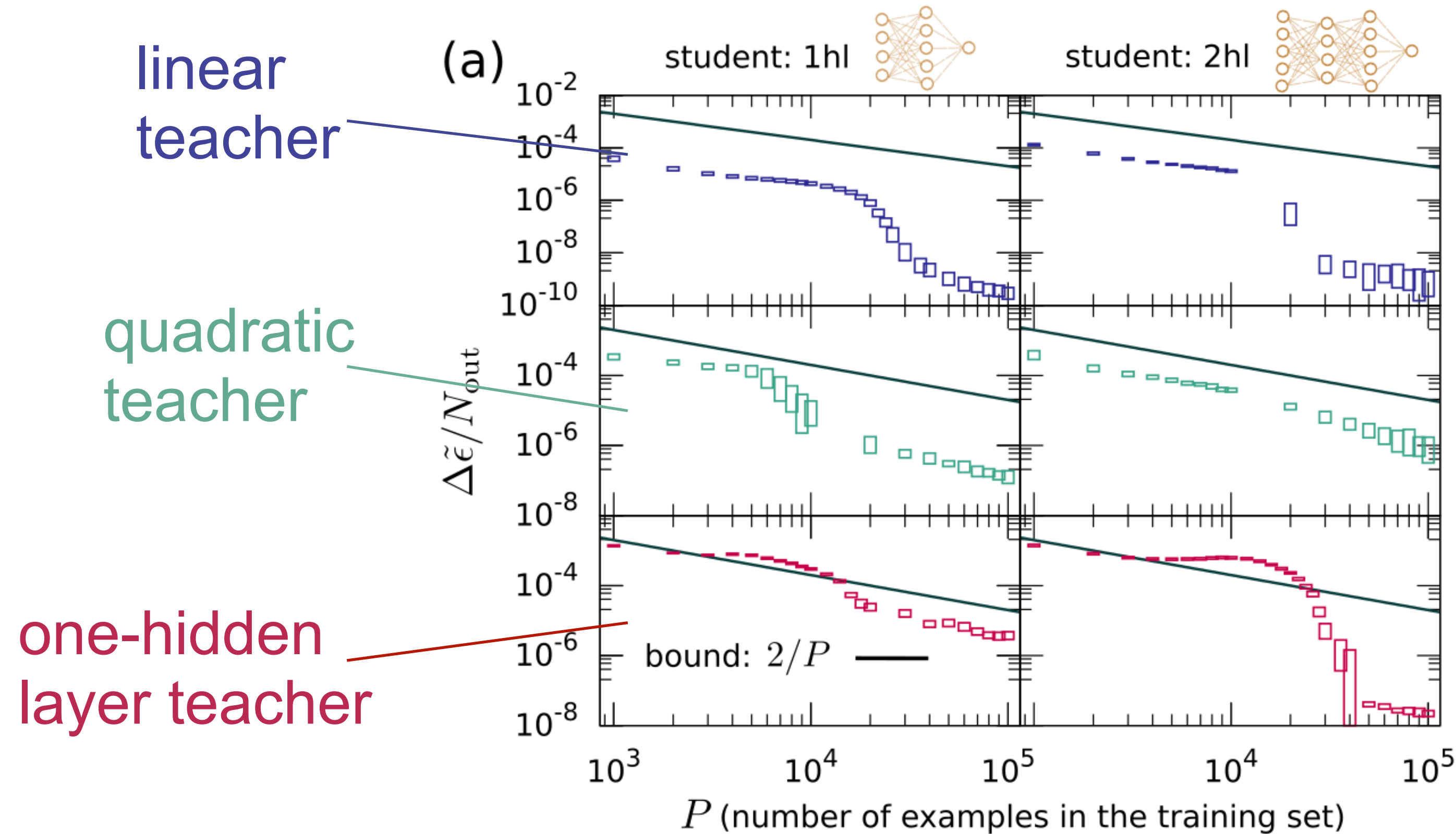
(VERY) RECENT RESULTS: GENERALISATION IN DEEP NETWORKS IS (MAINLY) CONTROLLED BY THE SIZE OF THE LAST LAYER

$$\Delta\epsilon \leq \frac{2N_{\text{out}}}{P}$$

size of the last layer

It is a **universal result**, i.e. it does not depend on the details of the architecture and on the learning task

It is derived by combining **Statistical Physics of disordered systems** with **Statistical Learning Theory**



THANKS!



Sebastiano Ariosto



Rosalba Pacelli



Francesco Ginelli



Marco Gherardi



Mauro Pastore



Vittorio Erba