deeppp

DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

Machine Learning to learn Physics?

Trials and questions

Marco Cristoforetti
27-9-2021

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Me & Lattice QCD

deeppp

My last contribution to the lattice QCD community.

Some time ago …

DSIP: Data Science for Industry & Physics

deeppp

Machine Learning

Learn from data

Predictive models

Predict on novel data

Industry

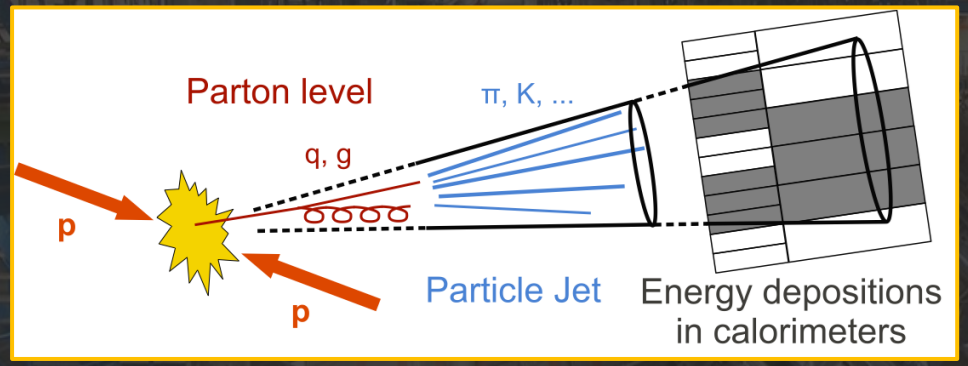
Agriculture

Physics

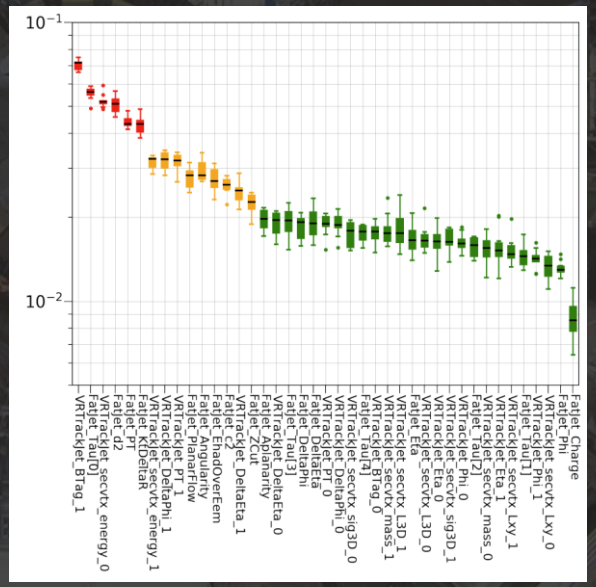
ATLAS experiment

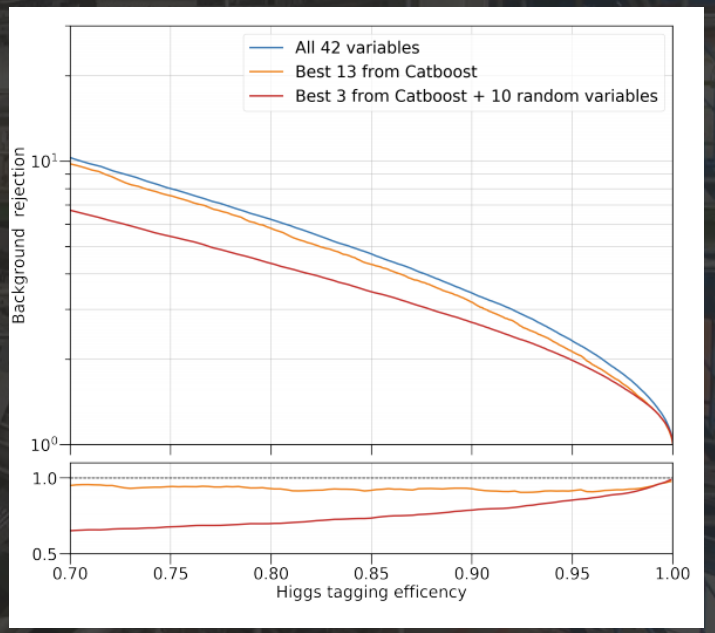
Flavor tagging

The **separation of *b*-quark initiated jets** from those coming from **lighter quark flavors** is a **fundamental tool** for the ATLAS physics program.



Automated feature selection A. di Luca





CSES - Limadou

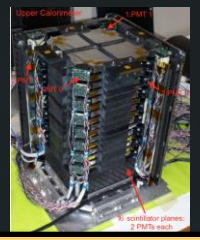
DeepPP
Deep Particle physics

CSES - CHINA SEISMO ELECTROMAGNETIC SATELLITE

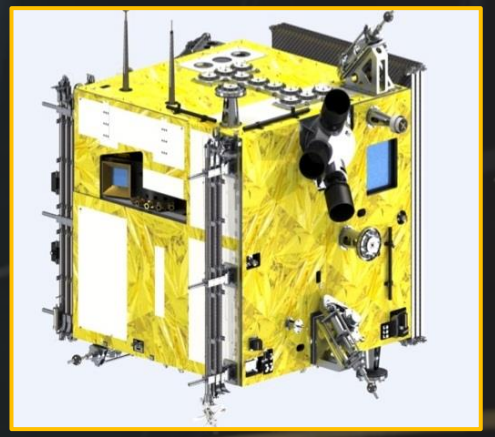
Italian Collaboration - LIMADOU PROJECT

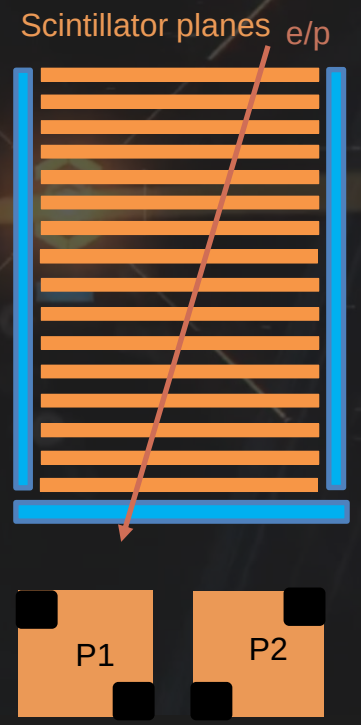
Particle identification and characterization

F. Follega



INPUT32 PMT signal





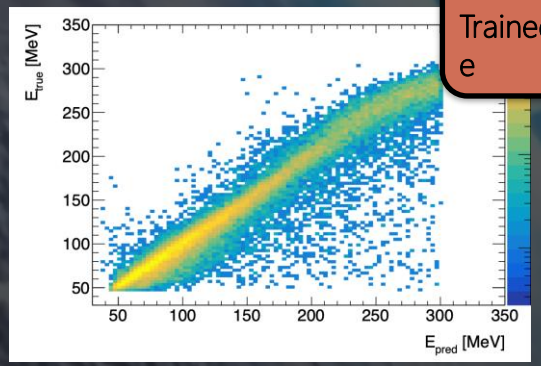
FCNN$_{KIN}$ Trained on e

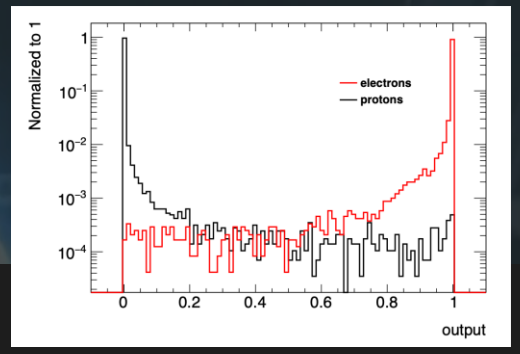
Electron

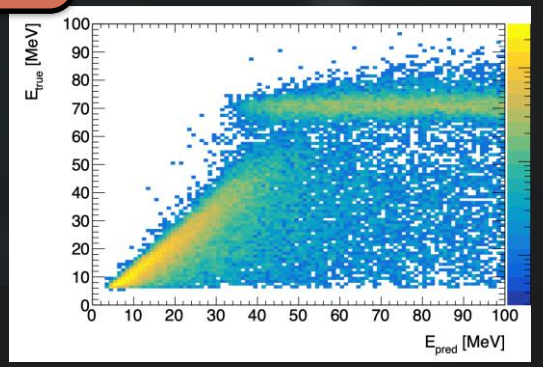
FCNN$_{PID}$

Proton

FCNN$_{KIN}$ Trained on p







DSIP
Data Science
for Industry and Physics

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

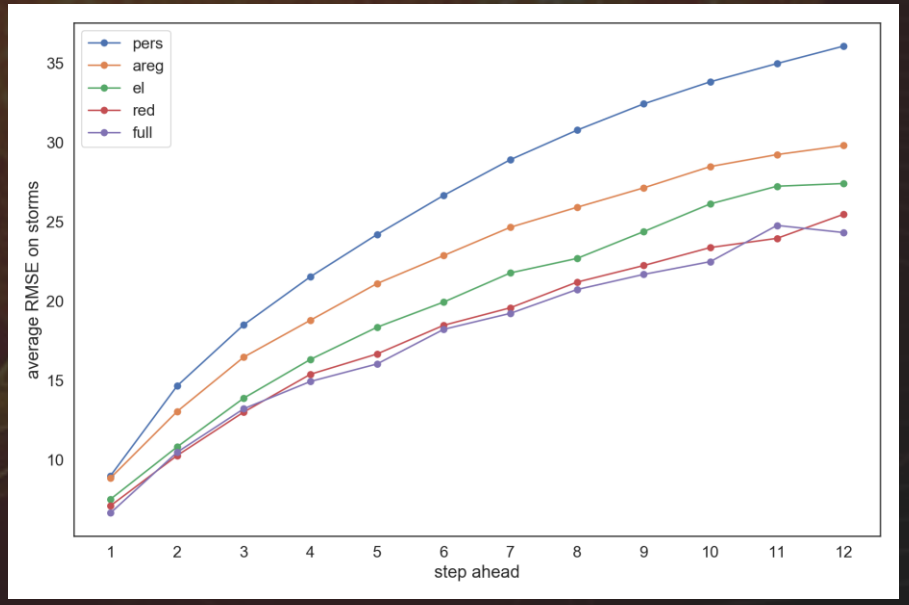
Dst – Geomagnetic Equatorial index

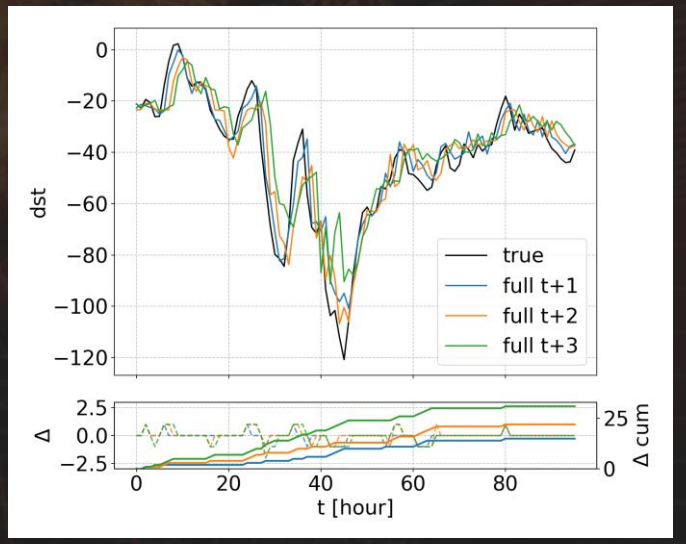
DeepPP
Deep Particle physics

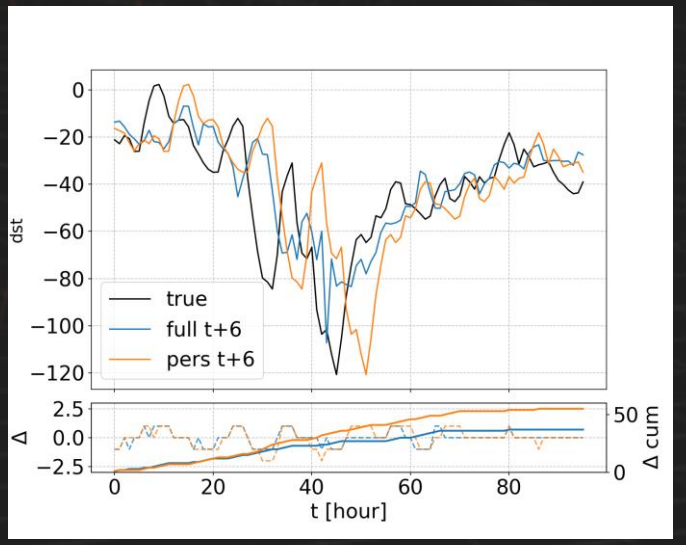
Forecast of Geomagnetic storm

1989 Geomagnetic storm caused a **nine-hour outage** of Hydro-Québec's electricity transmission system

ML/DL for predicting a few hours in advance the next storm







DSIP
Data Science
for Industry and Physics

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Proton CT

DeepPP
Deep Particle physics

arXiv:2010:00427

Proton path reconstruction for pCT using Neural Networks

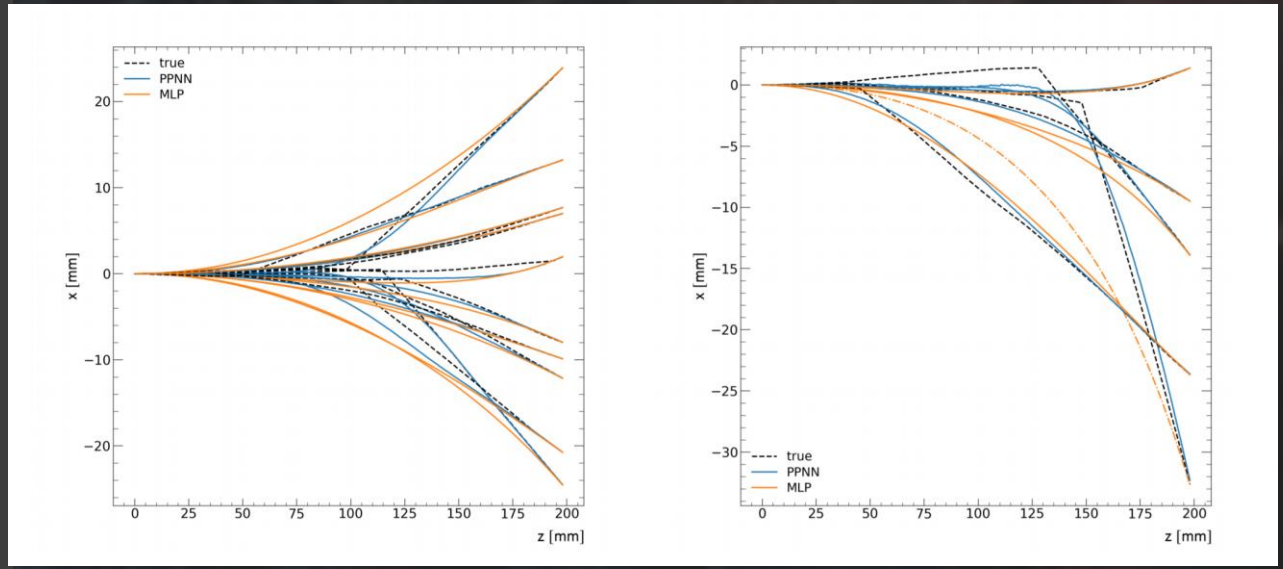
T. Ackernley[1,2], G. Casse[1,2] and M. Cristoforetti[2]

[1]Oliver Lodge, Department of Physics, University of Liverpool, Oxford Street, L69 7ZE Liverpool, United Kingdom

[2]Fondazione Bruno Kessler (FBK), Via Sommarive, 18, Povo, 38123, Trento, Italy

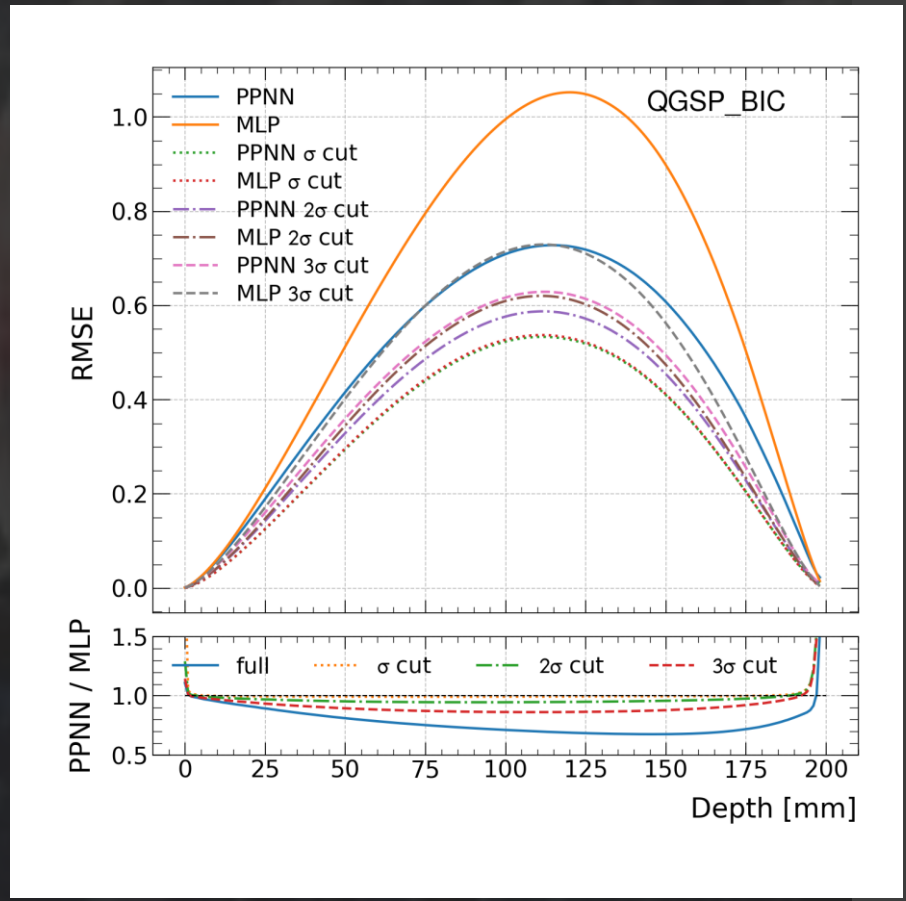
E-mail: mcristofo@fbk.eu

September 2020



NN outperforms standard reconstruction method

T. Ackernley



DSIP
Data Science
for Industry and Physics

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Theoretical Physics

deeppp

$$Z = \int \mathcal{D}A \, \det(\not{\partial} + m) \, e^{-S[A]}$$

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}A \, \mathcal{O} \, \det(\not{\partial} + m) \, e^{-S[A]}$$

DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Machine Learning

deeppp

What is machine learning (formally)?

$\mathcal{X}$ Domain set. Set of objects
One point is a vector of features

$\mathcal{Y}$ label set. Is the target.
Can be a discrete set or
a continuous variable

$f : \mathcal{X} \to \mathcal{Y}$

Correct function

Mapping x to y *

UNKNOWN !!

$h : \mathcal{X} \to \mathcal{Y}$

Predictor

Output of the

LEARNING process

* The existence of this function is a strong assumption

DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Machine Learning

deeppp

What is machine learning (formally)?

The learner want to reduce the probability that given a x randomly picked $h(x) \neq f(x)$

Minimize the error function $\mathcal{L}(h)$

This is what guides the learning

DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Machine Learning

deeppp

Minimize
the error function $\mathcal{L}(h)$

The learner want to reduce the probability that given a x randomly picked $h(x) \neq f(x)$

Need $\mathcal{D}$ probability distribution to extract samples from X than compute $f(x)$

$\mathcal{L}_{\mathcal{D}f}(h)$

$\mathcal{D}$ is unknown
f is unknown

DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Machine Learning

deeppp

Minimize
the error function $\mathcal{L}_{\mathcal{D},f}(h)$

$\mathcal{D}$ is unknown
f is unknown

We have **ONLY** the training samples

$S = ((x_1, y_1), \dots (x_m, y_m)) \subset \mathcal{X} \times \mathcal{Y}$

Training data:
observations as can be from an experiment

EMPIRICAL RISK
(minimization)

$$L_S(h) = \frac{|\{i \in [m] : h(x_i) \neq y_i\}|}{m}$$

DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Machine Learning

deeppp

Problem recap.:

We want $h(x)$, the predictor, to be equivalent to $f(x)$ on the whole $\mathcal{X}$ domain

BUT

We have access only to a finite number of samples $S = ((x_1, y_1), \dots (x_n, y_n))$

Without knowing how the $(x_1, \dots, x_n)$ have been extracted ($\mathcal{D}$ unknown)

1. How do we choose $h(x)$?
2. After the predictor $h(x)$ can reproduce $f(x)$ on the samples available, how do we know that it will work on new samples?

DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Machine Learning

deeppp

1. How do we choose $h(x)$?

   Many different algorithms available to approximate the target function

   Algorithms work fixing parameters to reproduce $f(x)$: we have $h(\vec{w}, x)$

   We need a **learner**: a way to fix the $\vec{w}$ exploring the space of all the possible $h(\vec{w}, x)$

2. After the predictor $h(x)$ can reproduce $f(x)$ on the samples available, how do we know that will work on new samples?

Overfitting, bias, REPRODUCIBILITY

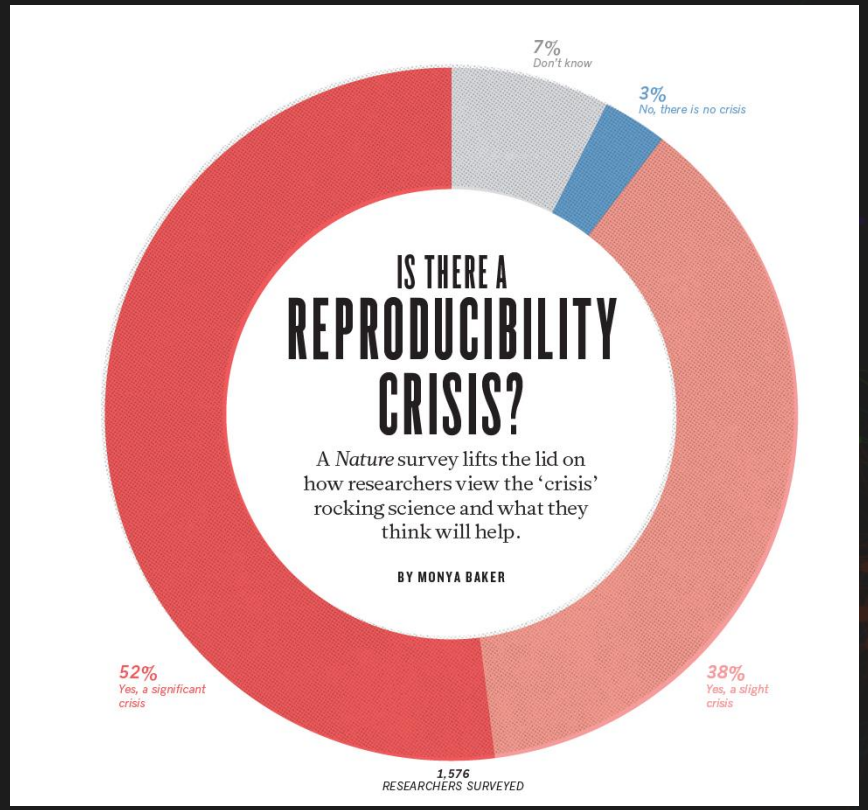
DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

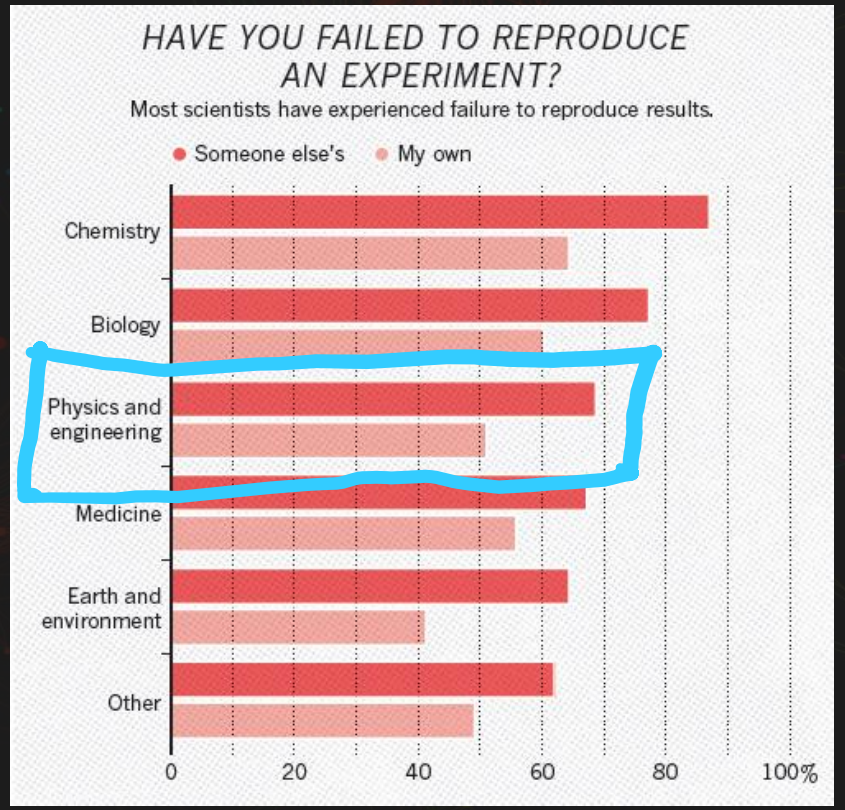
FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

(short) digression on reproducibility

deeppp

NATURE | VOL 533 | 26 MAY 2016





DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Machine Learning

deeppp

1. How do we choose $h(x)$?

   Many different algorithms available to approximate the target function

   Algorithms work fixing parameters to reproduce $f(x)$: we have $h(\vec{w}, x)$

   We need a **learner**: a way to fix the $\vec{w}$ exploring the space of all the possible $h(\vec{w}, x)$

2. After the predictor $h(x)$ can reproduce $f(x)$ on the samples available, how do we know that will work on new samples?

Overfitting, bias, REPRODUCIBILITY

DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Physics

deeppp

$$Z = \int \mathcal{D}A \, \det(\not{D} + m) \, e^{-S[A]}$$

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}A \, \mathcal{O} \, \det(\not{D} + m) \, e^{-S[A]}$$

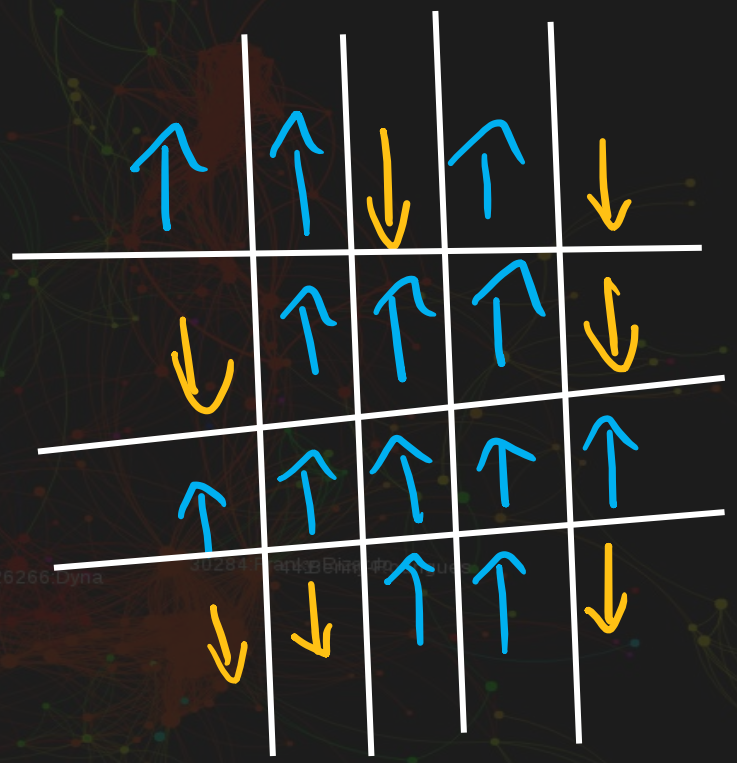
DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Ising

deeppp

$$H = \frac{J}{k_B T} \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$



Phase transition at $T_c \simeq 2.269$

Magnetization is the order parameter $M = \langle \sigma \rangle$

Susceptibility $\chi = \frac{N}{T}\left(\langle m^2 \rangle - \langle M \rangle^2\right)$

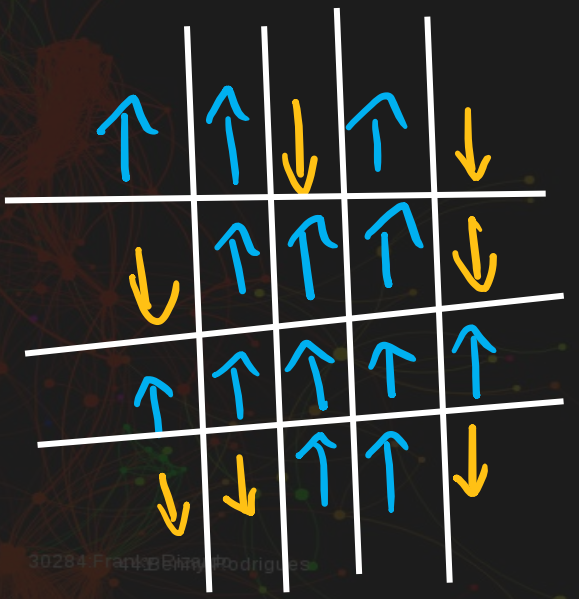
DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Ising

deeppp

$$H = \frac{J}{k_B T} \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$



Susceptibility $\chi = \frac{N}{T}\left(\langle m^2 \rangle - \langle M \rangle^2\right)$

Finite size system phase transition characterized by scaling law

$\chi \sim L^{\gamma/\nu}$

Critical exponents

Universality class

Spin-spin correlation function

$G(r) \sim r^{-(d-2+\eta)}$

DSIP DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER FUTURE BUILT ON KNOWLEDGE

Ising - MCRG

deeppp

PHYSICAL REVIEW LETTERS

VOLUME 42 — 2 APRIL 1979 — NUMBER 14

Monte Carlo Renormalization Group

Robert H. Swendsen
Brookhaven National Laboratory, Upton, New York 11973
(Received 15 December 1978)

A simplified method of applying a renormalization-group analysis to Monte Carlo simulations of general systems is presented and illustrated with applications to the Ising model and the three-state Potts model.

Two years ago, Ma[1] suggested combining Monte Carlo (MC) simulations of statistical-mechanical models[2] with a renormalization-group (RG) analysis of the critical properties.[3] The particular method which he suggested was based on a direct simulation of the fixed-point Hamiltonian, from which he calculated matrix elements for the linearized RG equations. The eigenvalues of these matrices then gave estimates of the critical exponents. Ma applied his method to the two-dimensional Ising model with encouraging results.

Ma's method has some drawbacks which have prevented its general application to problems of interest. The main difficulty was that the *direct* simulation of the fixed-point Hamiltonian involved a severe truncation in the number of coupling constants, while leaving a large parameter space to be scanned for the fixed point.

In this Letter, I would like to present a somewhat different approach that eliminates these difficulties: (1) It is only necessary to simulate the original Hamiltonian, *not* the fixed point or any renormalized Hamiltonian; (2) the truncation is small (and systematically improvable) if the range of interactions for the fixed-point Hamiltonian is small with respect to the lattice size; and (3) the parameter space to be scanned is only that of the original Hamiltonian. The result can be viewed as an extension of the standard methods of analyzing MC simulations that extracts useful information from correlation functions that are normally discarded. It can equally well be viewed as a systematically improvable real-space renormalization-group approximation that includes the effects of many interactions in the renormalized Hamiltonians.

Consider a lattice model in d dimensions with N^d sites. A "spin" σ_i (discrete or continuous) is associated with each site and the Hamiltonian has the general form

$$\mathcal{H} = \sum_\alpha K_\alpha S_\alpha , \tag{1}$$

where each S_α is some combinations of the σ's that is translationally invariant subject to periodic boundary conditions. For example, $\mathcal{H}$ could be a two-dimensional Ising model, $\sigma_i = \pm 1$, and

$$S_1 = \sum_{\langle ij \rangle} \sigma_i \sigma_j ,$$

where the sum extends over all nearest-neighbor pairs. The RG transformation will, of course, generate effective interactions between more-distant neighbors as well as many-spin couplings.

Once a particular renormalization transformation, $H^{(n+1)} = R_b H^{(n)}$ (with scale factor b), is chosen, the asymptotic critical properties will be determined in the usual way by the eigenvalues of the linearized RG transformation matrix,

$T_{\alpha\beta}^*$, in the vicinity of the fixed-point Hamiltonian, H^*. Specifically,[3]

$$K_\alpha^{(n+1)} - K_\alpha^* = \sum_\beta T_{\alpha\beta}^* (K_\beta^{(n)} - K_\beta^*) , \tag{2}$$

where

$$T_{\alpha\beta}^* = \left[\partial K_\alpha^{(n+1)} / \partial K_\beta^{(n)} \right]_{H^*} \tag{3}$$

and the eigenvalue equation is

$$\sum_\alpha \varphi_\alpha T_{\alpha\beta}^* = \lambda \varphi_\beta . \tag{4}$$

The critical exponents are obtained from the eigenvalues in the usual way [$\nu = \ln b / \ln \lambda_1^e$, $\eta = d + 2 - 2 \ln \lambda_1^o / \ln b$, etc., where $\lambda_1^{e(o)}$ is the largest even (odd) eigenvalue].[3]

Note that, for these equations to be useful, the derivatives $\partial K_\alpha^{(n+1)} / \partial K_\beta^{(n)}$ must change slowly near the fixed point. To evaluate $T_{\alpha\beta}^*$, it is therefore only necessary to calculate the derivatives somewhere in the "linear region," where they are essentially constant. Furthermore, only the eigenvalues have physical significance. The location of the fixed point (and even the space of Hamiltonians) depends on the choice of renormalization transformation and need not be calculated.

A sequence of approximations for $T_{\alpha\beta}^*$ can be obtained if the *exact* RG transformation had been performed on H (a difficult calculation involving an extremely large number of coupling parameters) and the renormalized Hamiltonian had been simulated (also a difficult calculation). The RG transformation can, of course, be applied to the block spins repeatedly with the limitation that the last transformation leaves a lattice that is large with respect to the range of the corresponding renormalized Hamiltonian.

The chain rule

$$\frac{\partial \langle S_\gamma^{(n)} \rangle}{\partial K_\beta^{(n-1)}} = \sum_\alpha \frac{\partial K_\alpha^{(n)}}{\partial K_\beta^{(n-1)}} \frac{\partial \langle S_\gamma^{(n)} \rangle}{\partial K_\alpha^{(n)}} , \tag{5}$$

together with the identities

$$\frac{\partial \langle S_\gamma^{(n)} \rangle}{\partial K_\beta^{(n-1)}} = \langle S_\gamma^{(n)} S_\beta^{(n-1)} \rangle - \langle S_\gamma^{(n)} \rangle \langle S_\beta^{(n-1)} \rangle \tag{6}$$

and

$$\frac{\partial \langle S_\gamma^{(n)} \rangle}{\partial K_\alpha^{(n)}} = \langle S_\gamma^{(n)} S_\alpha^{(n)} \rangle - \langle S_\gamma^{(n)} \rangle \langle S_\alpha^{(n)} \rangle ,$$

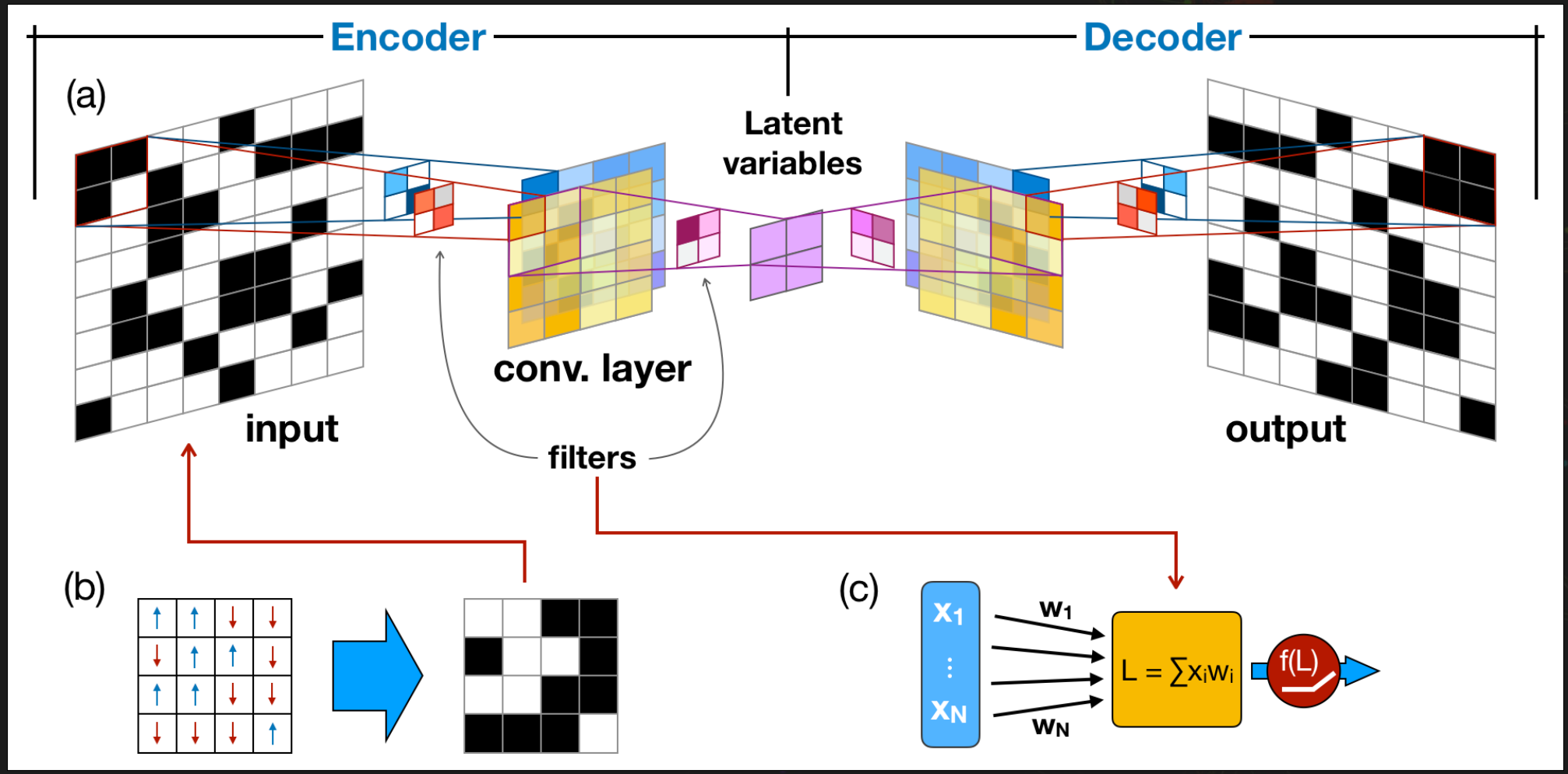
provides a direct way to calculate a sequence of approximations to $T_{\alpha\beta}^*$ as the renormalization transformation is iterated towards the fixed point.

DSIP DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER FUTURE BUILT ON KNOWLEDGE

Ising - autoencoder

deeppp

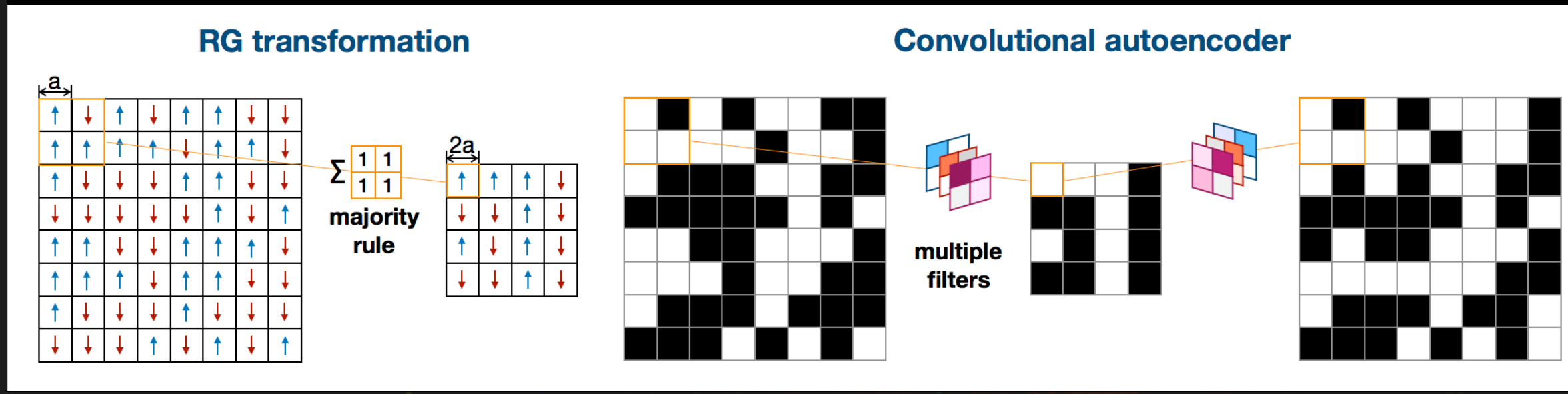


DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Ising: RG-autoencoder

deeppp



Sampling confs using MC for large lattice size

$$H = \frac{J}{k_B T} \sum_{\langle ij \rangle} \sigma_i \sigma_j$$

Use autoencoder as RG trafo to compute the exponents
Inspired by MCRG method [Swendsen 1979; Ron, Brandt, Swendsen 2017]

$$G(r) \sim r^{-(d-2+\eta)} \qquad \chi \sim L^{\gamma/\nu}$$

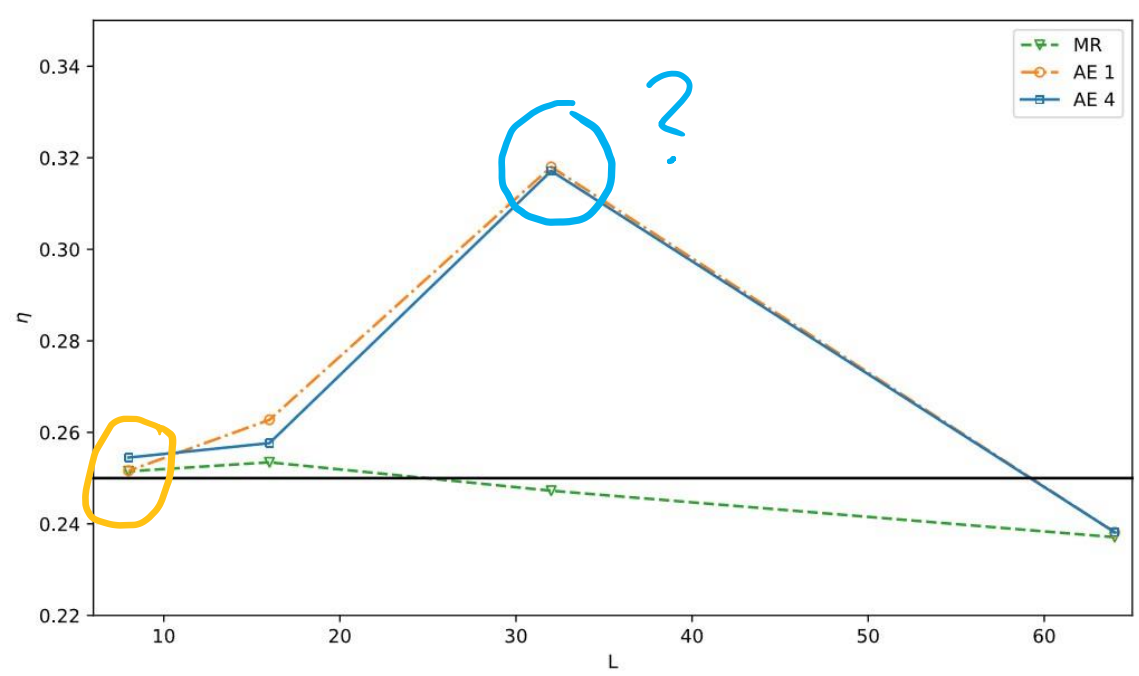
DSIP DATA SCIENCE FOR INDUSTRY & PHYSICS

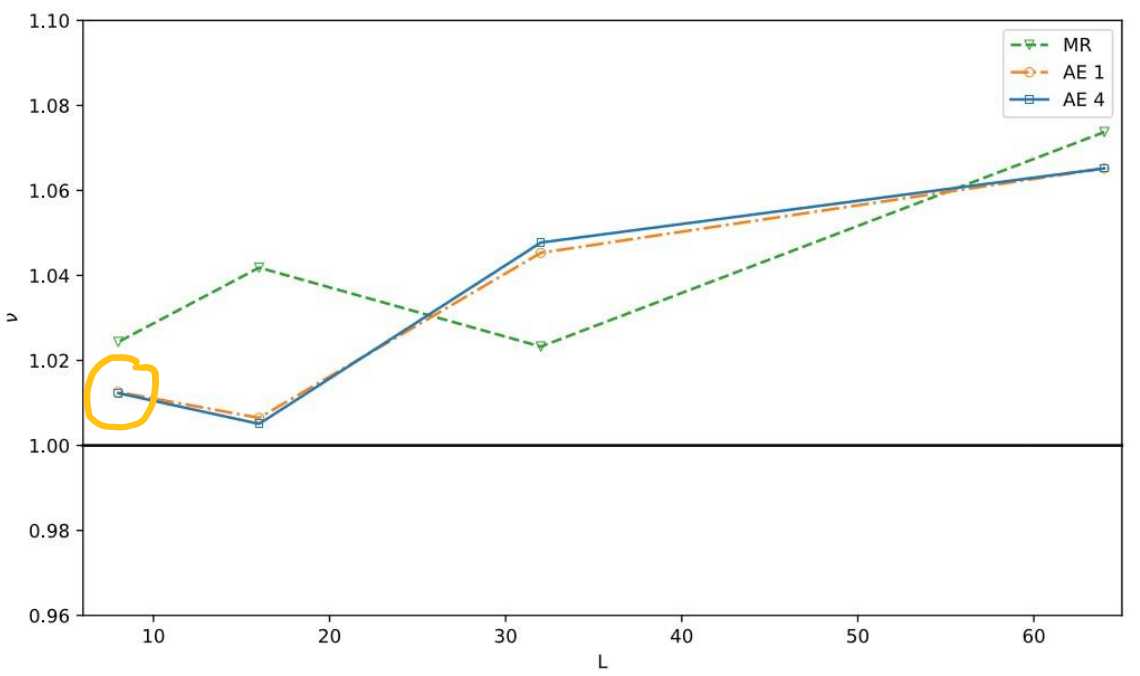
FONDAZIONE BRUNO KESSLER FUTURE BUILT ON KNOWLEDGE

Ising: RG-autoencoder

deeppp

Truth: $\eta = 0.25$ $\nu = 1$



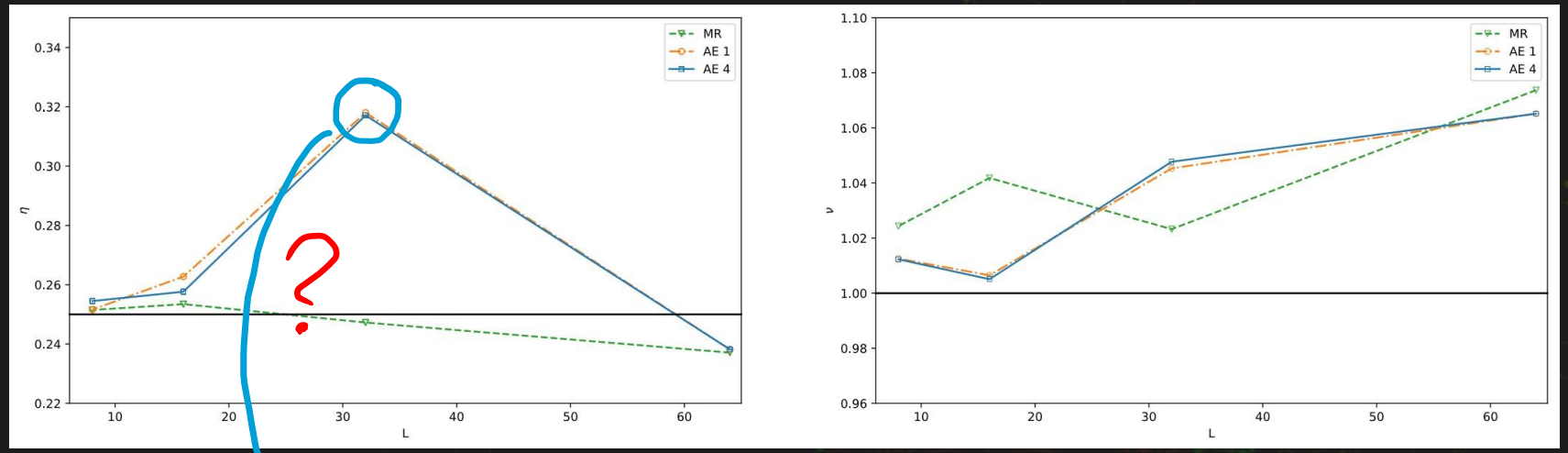


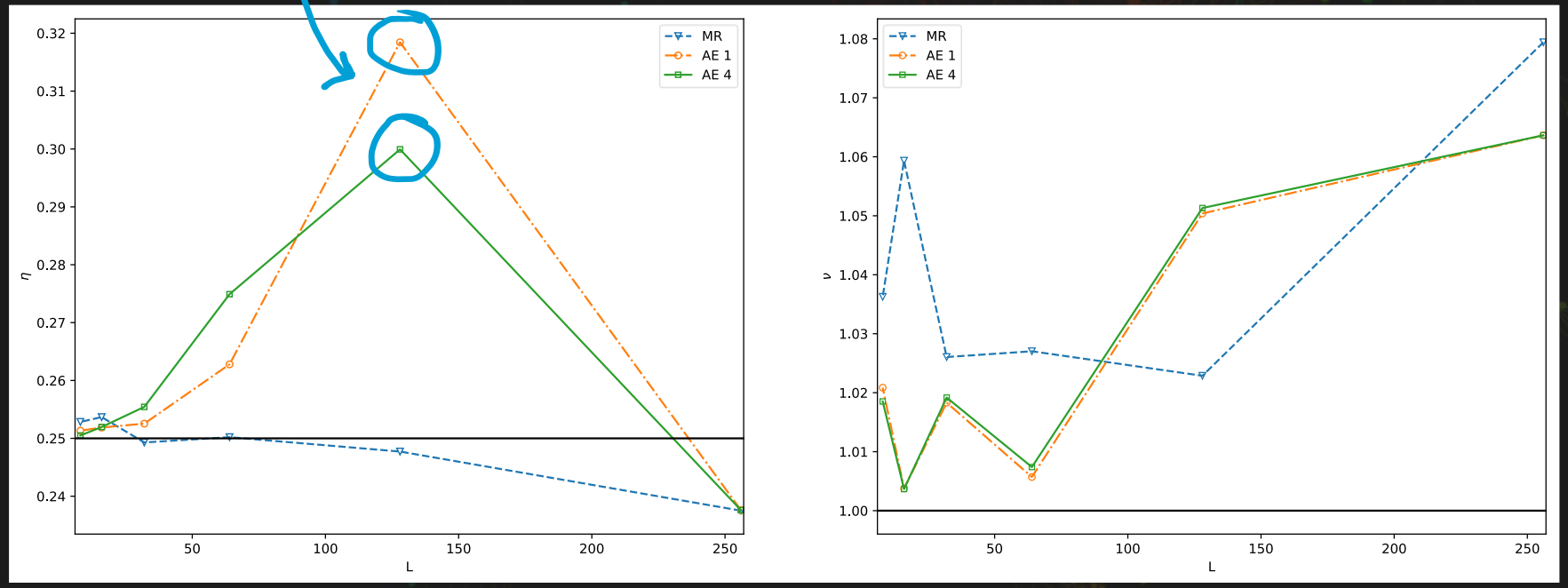
Ising: RG-autoencoder

deeppp

$L = 128$

$L = 512$





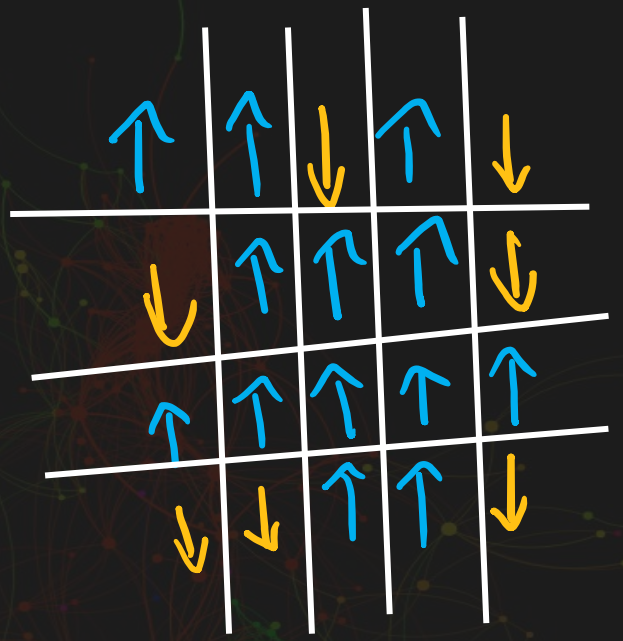
DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Ising: phase transition (with AE)

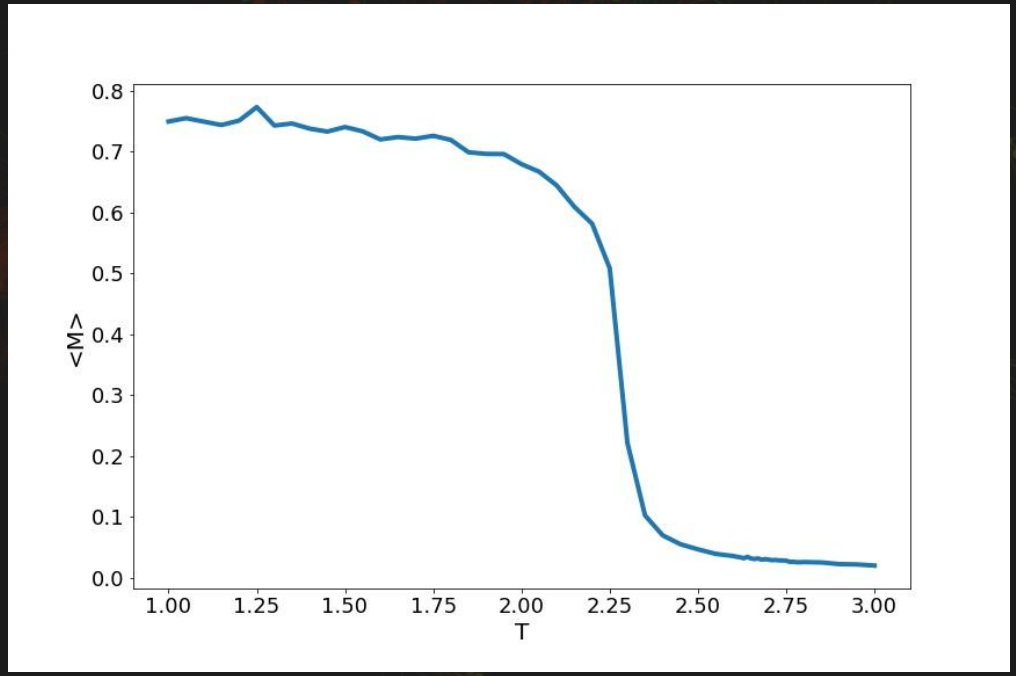
deeppp

$$H = \frac{J}{k_B T} \sum_{\langle i,j \rangle} \sigma_i \sigma_j$$



Phase transition at $T_c \simeq 2.269$

Magnetization is the order parameter $M = \langle \sigma \rangle$



DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

Ising: phase transition (with AE)

deeppp

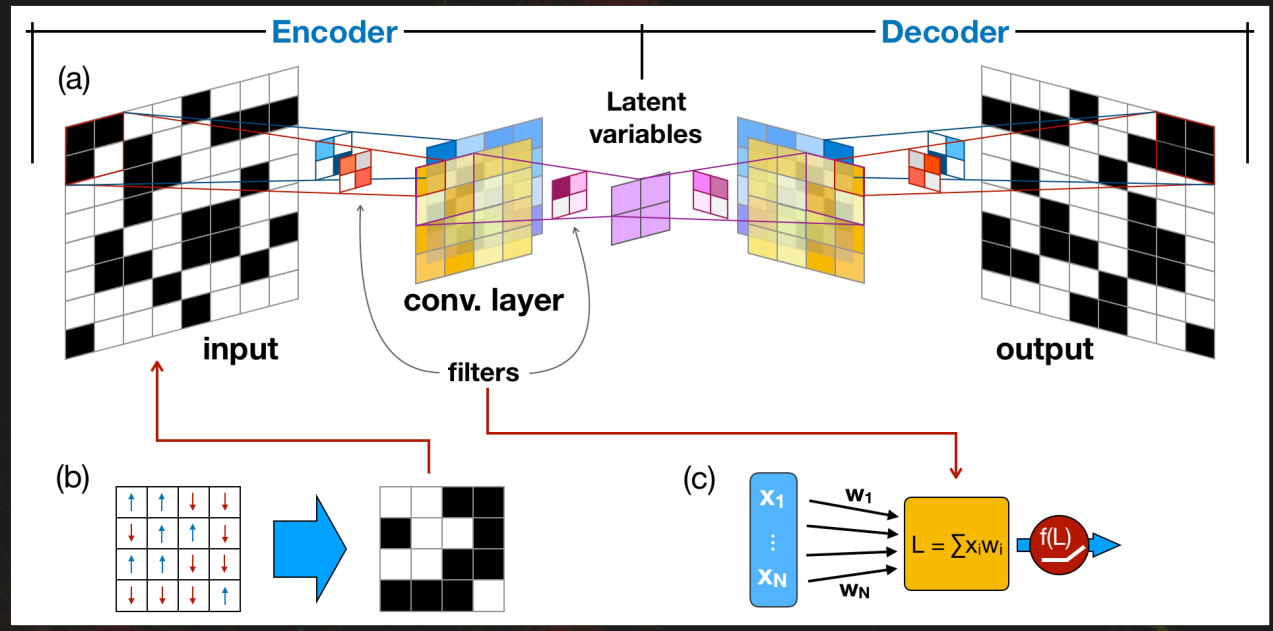
Learn phase transitions with Autoencoders

Convolutional layers 2x2 → 1 latent variable

$L = 128$

Network trained with confs at $T = \infty$ random

MCMC to generate confs at T in the range $T = [1.0, 3.0]$



Look at the latent representation z of the confs

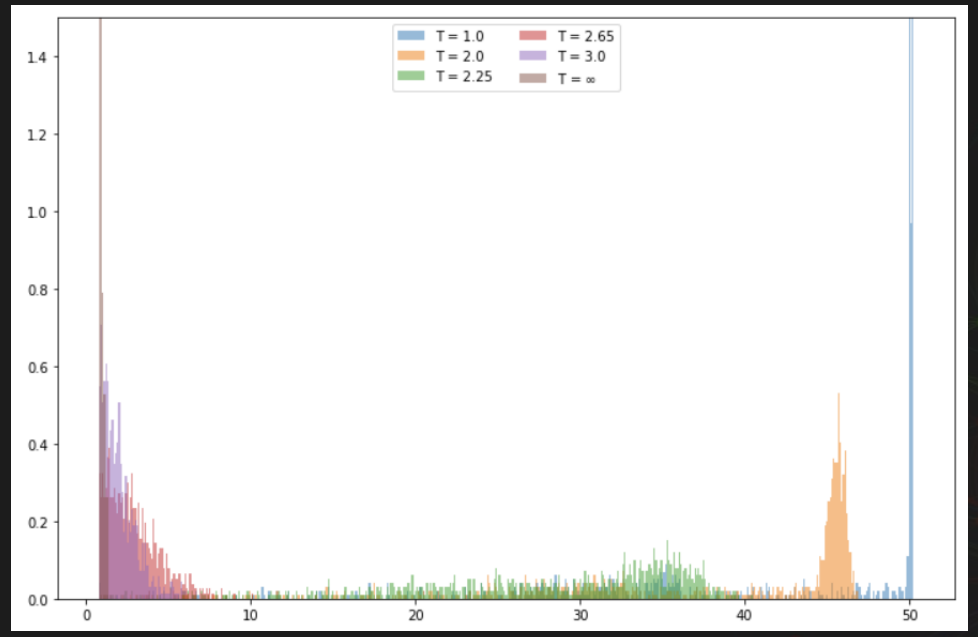
DSIP DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER FUTURE BUILT ON KNOWLEDGE

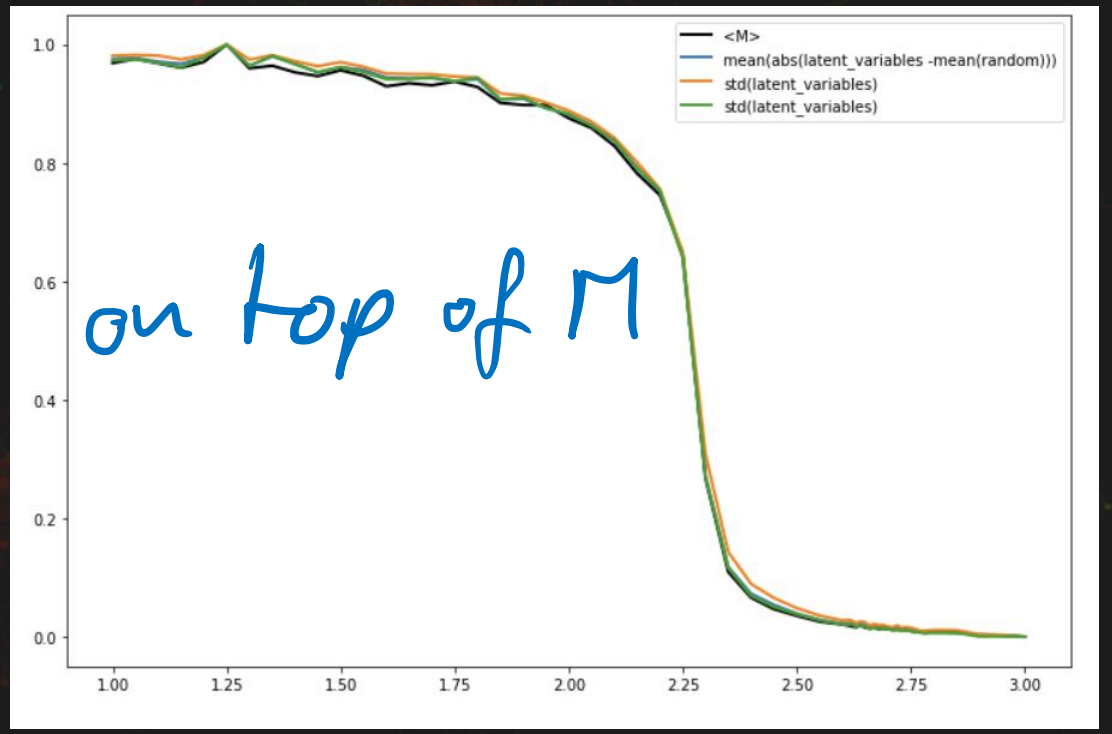
Ising: phase transition (with AE)

deeppp

The distribution of the latent variable clearly change with T



Define "expectation values" from this latent representation



What is the network learning?

DSIP DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER FUTURE BUILT ON KNOWLEDGE

Ising: phase transition (with AE)

deeppp

What is the network learning?

Generate configuration with fixed ratios of up and down spins and look at the value of the latent variable.

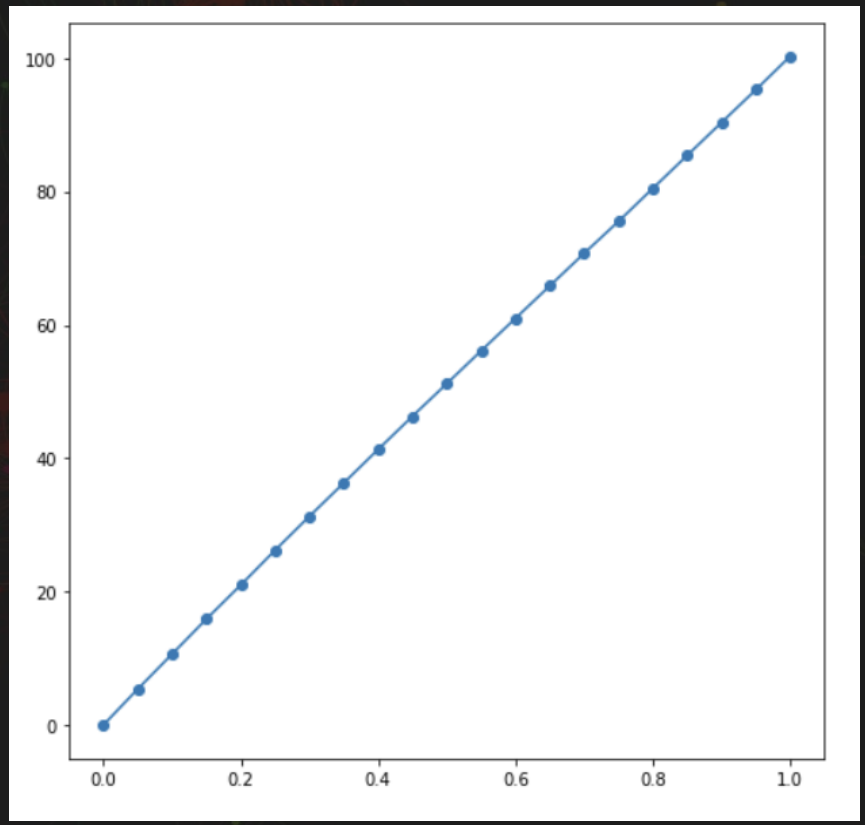
Confs used to train the network are of **0** and **1** not **-1, 1**

fixed ration means abundance of ones with respect to zeros.

The network is learning $\sum_i \sigma_i$ for $\sigma_i \in [0, 1]$

$\hookrightarrow \equiv M = \langle \sigma \rangle$ for $\sigma_i \in [-1, 1]$

Fraction of positive spins vs latent value is perfectly linear.



DSIP DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER FUTURE BUILT ON KNOWLEDGE

XY: phase transition (with AE)

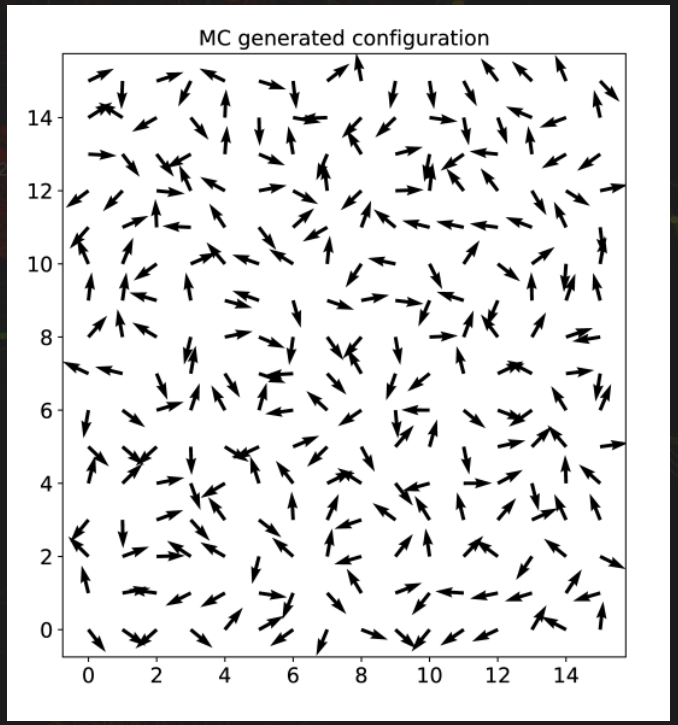
deeppp

$$H_{XY} = -J \sum_{\langle i,j \rangle} \vec{\sigma}_i \cdot \vec{\sigma}_j = -J \sum_{\langle i,j \rangle} \cos(\theta_i - \theta_j)$$

$\theta \in [0, 2\pi)$

2D XY has Kosterlitz-Thouless transition
Topological charges (vortices)

$$\Pi = \frac{1}{V} \sum \vec{\sigma}$$



DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

XY: phase transition (with AE)

deeppp

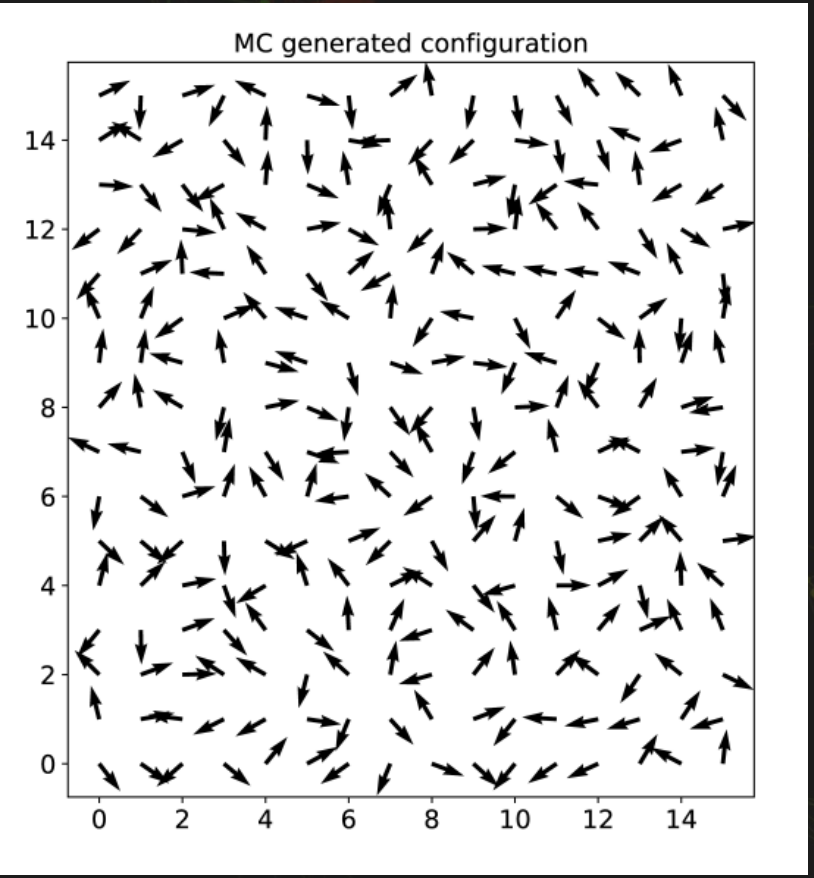
Learn phase transitions with Autoencoders

Convolutional layers 2x2 → 1 latent variable

$L = 128$

Network trained with confs at $T = \infty$ random

MCMC to generate confs on a range of T that include the transition



Look at the latent representation z of the confs

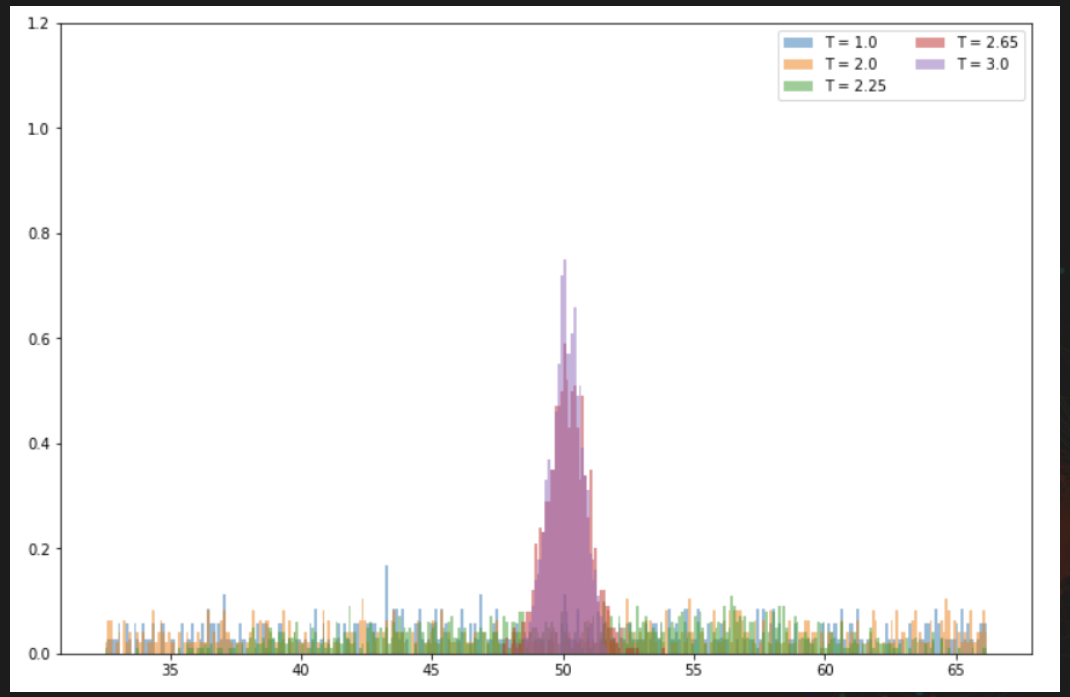
DSIP DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER FUTURE BUILT ON KNOWLEDGE

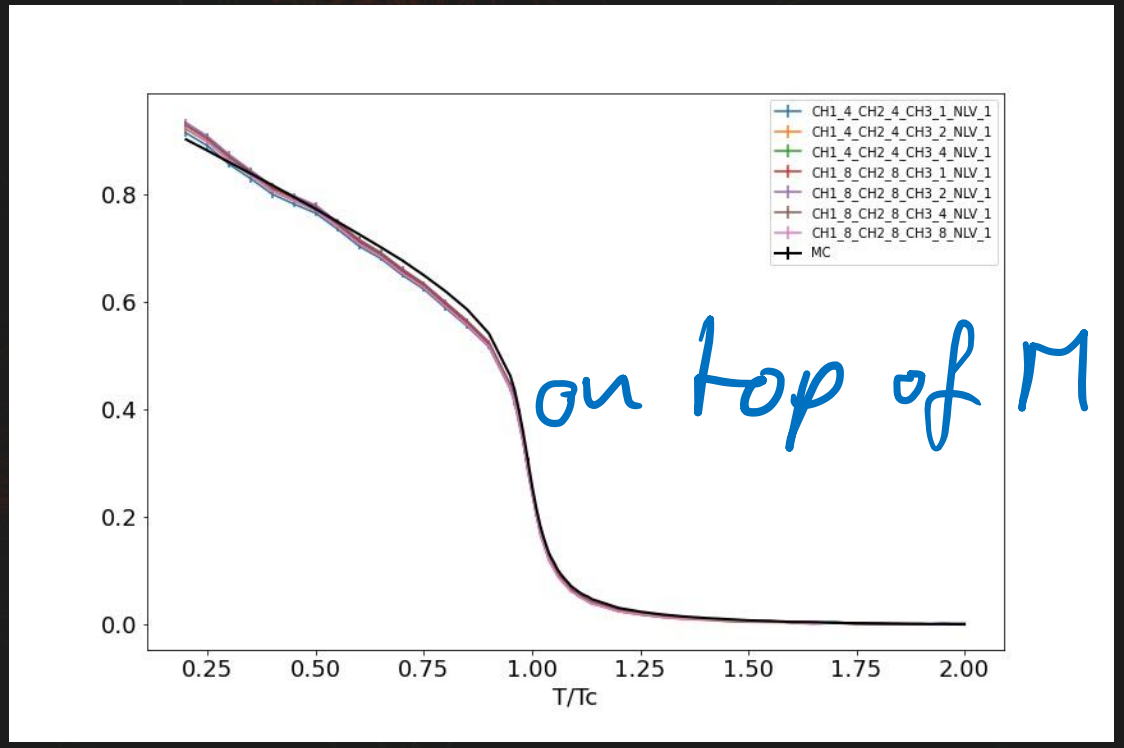
XY: phase transition (with AE)

deeppp

Even in this case the distribution of the latent variable clearly change with T



Use the standard deviation of z as order parameter



What is the network learning? NO IDEA

DSIP DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER FUTURE BUILT ON KNOWLEDGE

Learn Physics - provocation

deeppp

Many ways of using ML in physics

Can we use ML to develop a NEW way of making predictions on physical phenomena WITHOUT using all the "classical" language?

$$\vec{v} \quad \vec{a} \quad \mathcal{L} \quad H \quad \phi$$

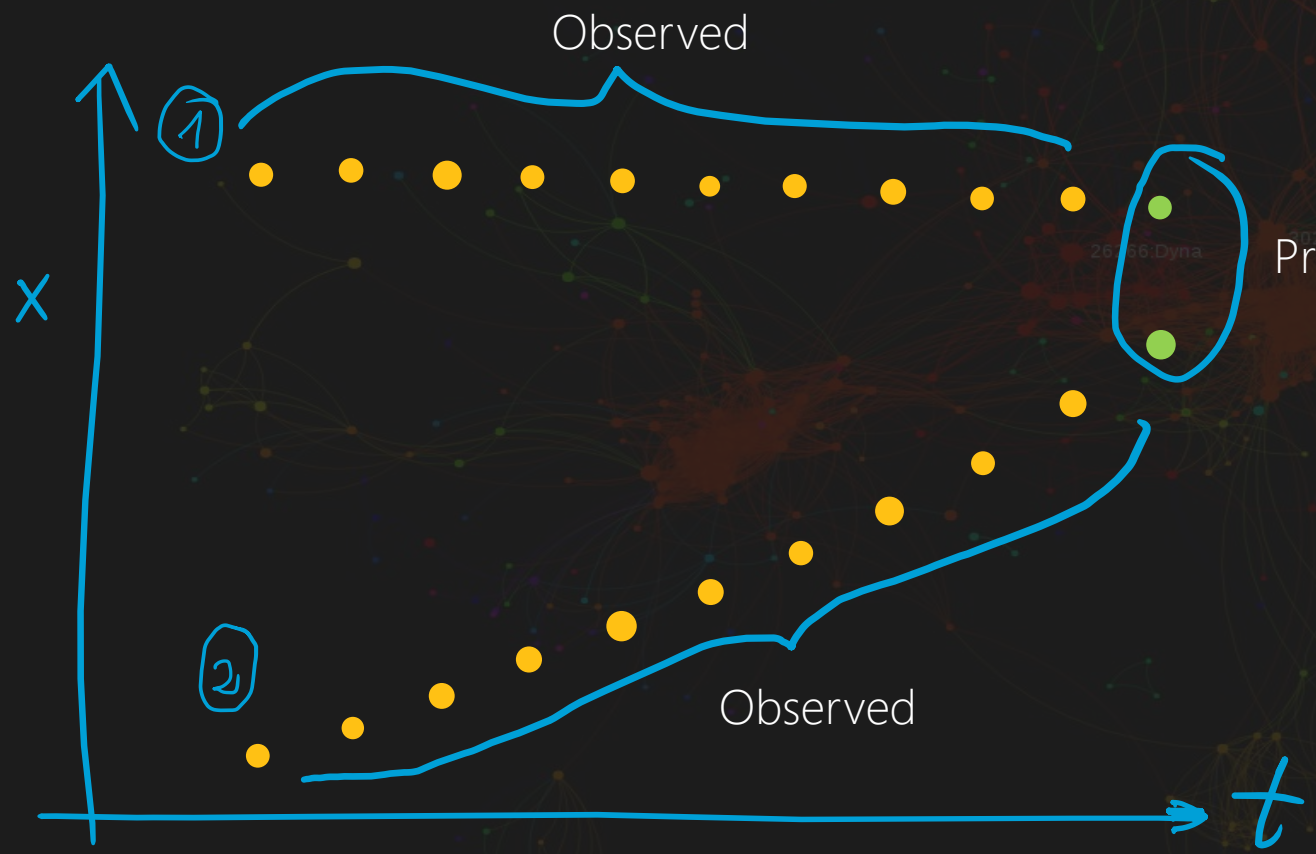
Does this correspond to an analogous comprehension of the physical world?

Learn Physics - provocation

deeppp

$$F = -G\frac{m_1 m_2}{r^2}$$

Is the unique/best way to predict the trajectory of two "massive" (?) bodies?



DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE

deeppp

Thank you

DSIP
DATA SCIENCE FOR INDUSTRY & PHYSICS

FONDAZIONE BRUNO KESSLER
FUTURE BUILT ON KNOWLEDGE