Machine Learning based Unfolding to reduce noise on lattice QCD observables

Simran Singh

(University of Bielefeld)

work done with Christian Schmidt

41st Lattice Conference @ University of Liverpool

July , 2024

Outline

I. Proposed solution : Machine Learning based unfolding

II. The problem : Noisy trace estimation with few sources

III. Distribution of noise vectors and their correlations

IV. The method applied to mock matrices data

V. Using lattice data

VI. Summary

Machine Learning based Unfolding I

- Unfolding : Inversion problem aimed at getting *genuine physics* data from *measured smeared* data

$$X_{observed} = P \circledast Y_{true}$$

1. Identify P^{-1} for given (x_i, y_i)
2. Apply P^{-1} X_i to get Y_i

- Used extensively in particle physics experiments to de-correlate *detector defects* from measurements

- Both ML and non-ML (IBU, TUnfold, SVD …) algorithms exist for data unfolding

- Main advantage of ML based unfolding : *no binning of data required*!

check out for a recent overview


SciPost Physics Submission

The Landscape of Unfolding with Machine Learning

Nathan Huetsch[1], Javier Mariño Villadamigo[1], Alexander Shmakov[2], Sascha Diefenbacher[3], Vinicius Mikuni[3], Theo Heimel[1], Michael Fenton[2], Kevin Greif[2], Benjamin Nachman[3,4], Daniel Whiteson[2], Anja Butter[1,5], and Tilman Plehn[1,6]

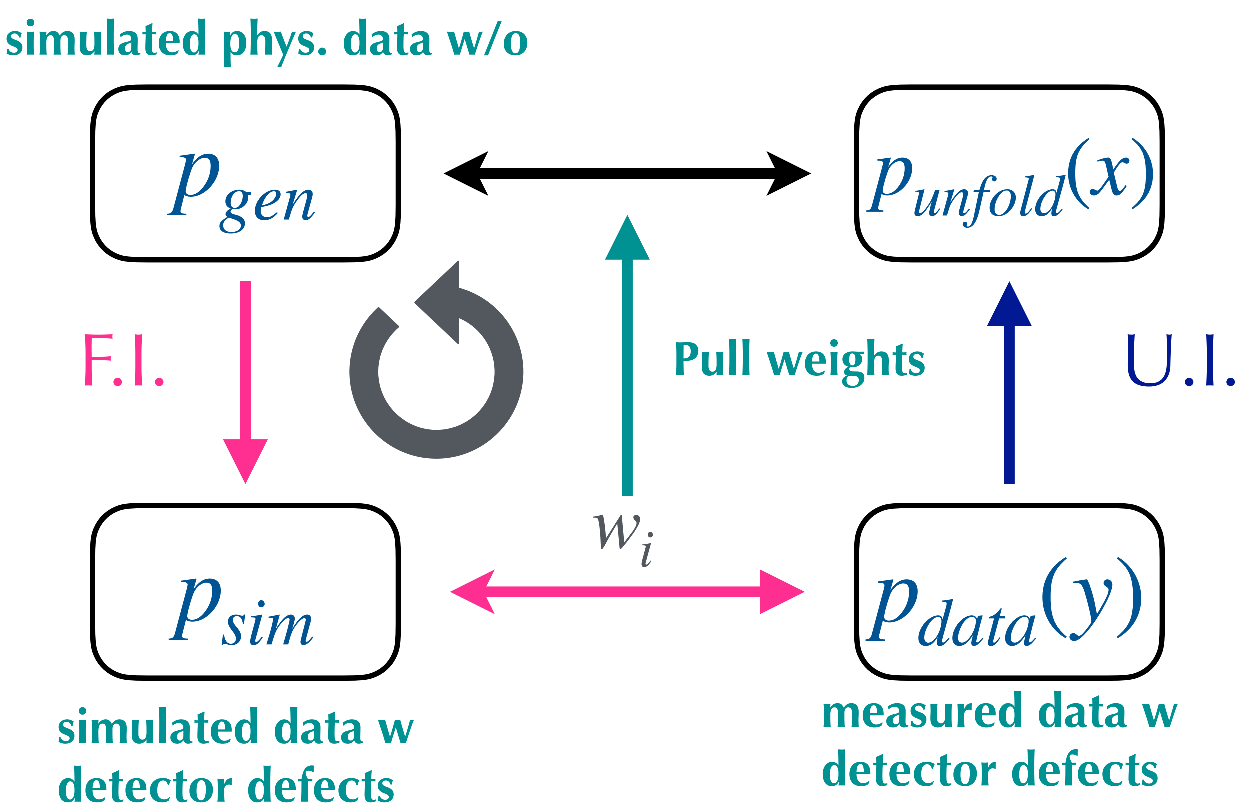
Machine Learning based Unfolding II

I. Re-weighting approach

PHYSICAL REVIEW LETTERS **124,** 182001 (2020)

OmniFold: A Method to Simultaneously Unfold All Observables

Anders Andreassen,[1,2,3,*] Patrick T. Komiske,[4,†] Eric M. Metodiev,[4,‡] Benjamin Nachman,[2,§] and Jesse Thaler[4,‖]



- The final goal :

$$p^{(n)}_{unfold}(x) = w_n \circledast p_{gen}(x)$$

II. Generative approach

- Two popular methods - Schrödinger bridges and Direct Diffusion

Improving Generative Model-based Unfolding with Schrödinger Bridges

Sascha Diefenbacher,[1,*] Guan-Horng Liu,[2,†] Vinicius Mikuni,[3,‡] Benjamin Nachman,[1,4,§] and Weili Nie[5,¶]

Kicking it Off(-shell) with Direct Diffusion

Anja Butter[1,3], Tomáš Ježo[2], Michael Klasen[2],
Mathias Kuschick[2], Sofia Palacios Schweitzer[1], and Tilman Plehn[1]

- Both of these methods aim to morph one non-trivial distribution p_{gen} to another p_{sim} via SDE and ODE respectively

- Apply the map to $p_{data}(y)$ to get $p_{unfold}(x)$

Lattice observables via trace estimation

- Observables in lattice QCD calculations can often be expressed as derivatives of $\ln Z$, e.g. quark number density : $\dfrac{\partial \ln \det M_f}{\partial \mu_f} = Tr\left(M_f^{-1} \dfrac{\partial M_f}{\partial \mu_f} \right)$

- Typical size of Fermion matrices : $N_\sigma^3 \times N_\tau \times N_c \times 4$, can go up to $\sim O(10^7 - 10^9)$

- *Estimating* the trace of the *inverse* of such large matrices is only accessible through "random noise method" that requires drawing random vectors η_i^n (i^{th} component of the n^{th} vector) which satisfy

$$\langle \eta_i \rangle = \lim_{L\to\infty} \frac{1}{L} \sum_{m=1}^{L} \eta_i^m = 0 \qquad \text{and} \qquad \langle \eta_i \eta_j \rangle = \lim_{L\to\infty} \frac{1}{L} \sum_{m=1}^{L} \eta_i^m \eta_j^m = \delta_{ij}$$

Noisy trace estimation with few sources

- The goal is to draw $L << N$ (linear dimension of M) vectors to estimate

$$\langle \eta^T M \eta \rangle_L \simeq TrM + \mathcal{O}\left(\frac{f(M)}{\sqrt{L}}\right)$$

! only get the true trace in $L \to \infty$!

$\langle \eta_i M_{i,i} \eta_i \rangle \sim M_{i,i}$ $\qquad$ $\langle \eta_i M_{i,j} \eta_j \rangle_{i \neq j} \sim 0$

- Why is it important to try and use less random vectors?

- On the lattice - we have extra steps to reach the observable !

- Need to estimate traces of $M^{-1}, M^{-1}\frac{\partial M}{\partial \mu}$ ….

- First we need to solve $Mx = \eta^l$, as many times, using CG, as the sources to get x and construct the estimator $\langle \eta x \rangle$

Can Machine Learning help ?

- The idea to reduce the number noise vectors to estimate the trace using *machine learnt probe vectors* is not new

around for a while (2016) and not implemented in any real lattice calculation yet

and the talk on Monday by B. Choi @ 11.15

Estimation of matrix trace using machine learning

Boram Yoon

Los Alamos National Laboratory, CCS-7, Los Alamos, New Mexico 87545

Abstract

We present a new trace estimator of the matrix whose explicit form is not given but its matrix multiplication to a vector is available. The form of the estimator is similar to the Hutchison stochastic trace estimator, but instead of the random noise vectors in Hutchison estimator, we use small number of probing vectors determined by machine learning. Evaluation of the quality of estimates and bias correction are discussed. An unbiased estimator is proposed for the calculation of the expectation value of a function of traces. In the numerical experiments with random matrices, it is shown that the precision of trace estimates with $\mathcal{O}(10)$ probing vectors determined by the machine learning is similar to that with $\mathcal{O}(10000)$ random noise vectors.

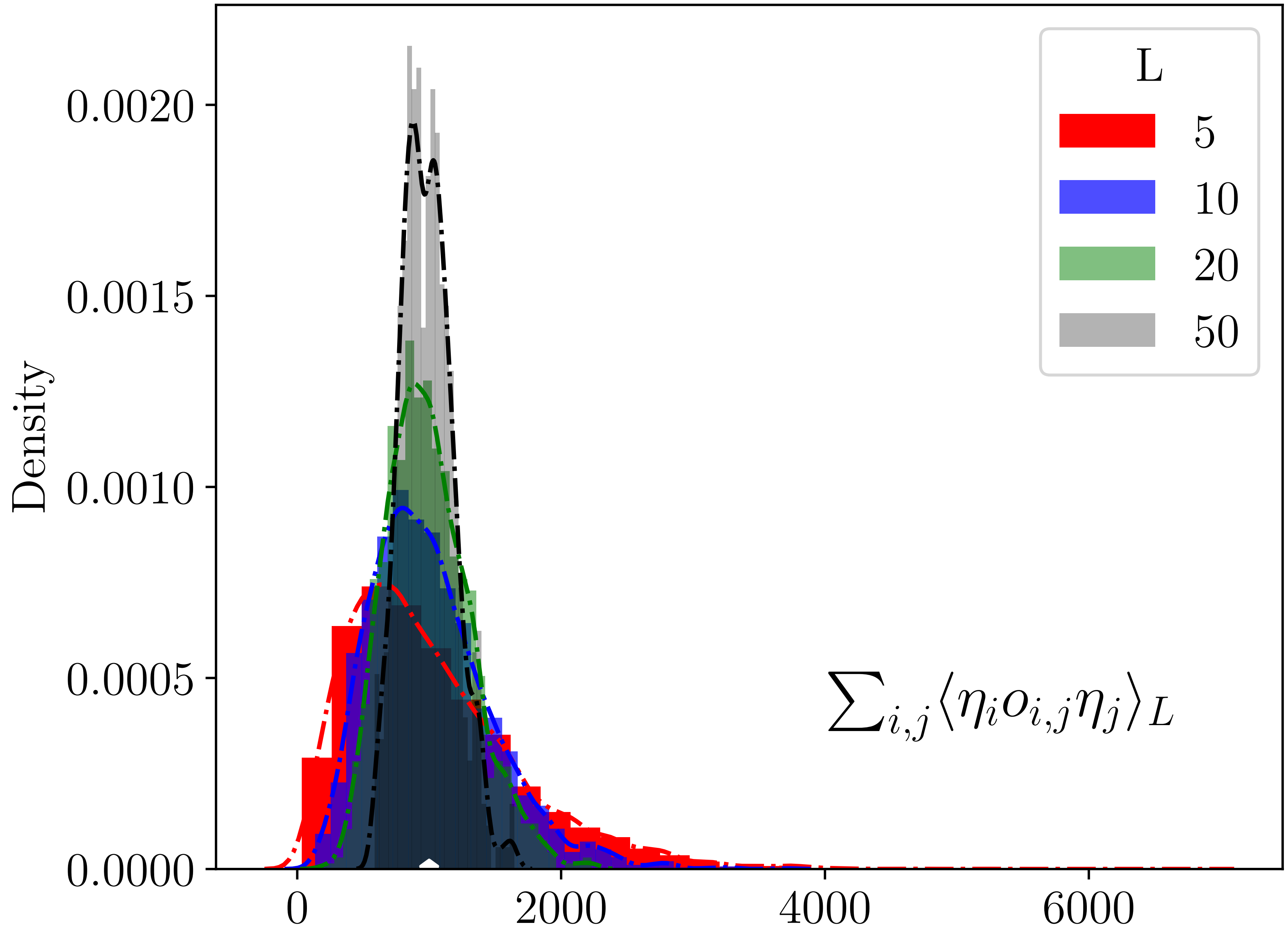
- The idea of this project was to look at the distributions of traces with varying number of sources - in the hope of training a NN to *unlearn* the effect of small number of random vectors

$$\langle \eta^T M \eta \rangle_{L_1} = f_{L_1,L_2} \circledast \langle \eta^T M \eta \rangle_{L_2} \quad \text{when} \quad L_1 < L_2$$

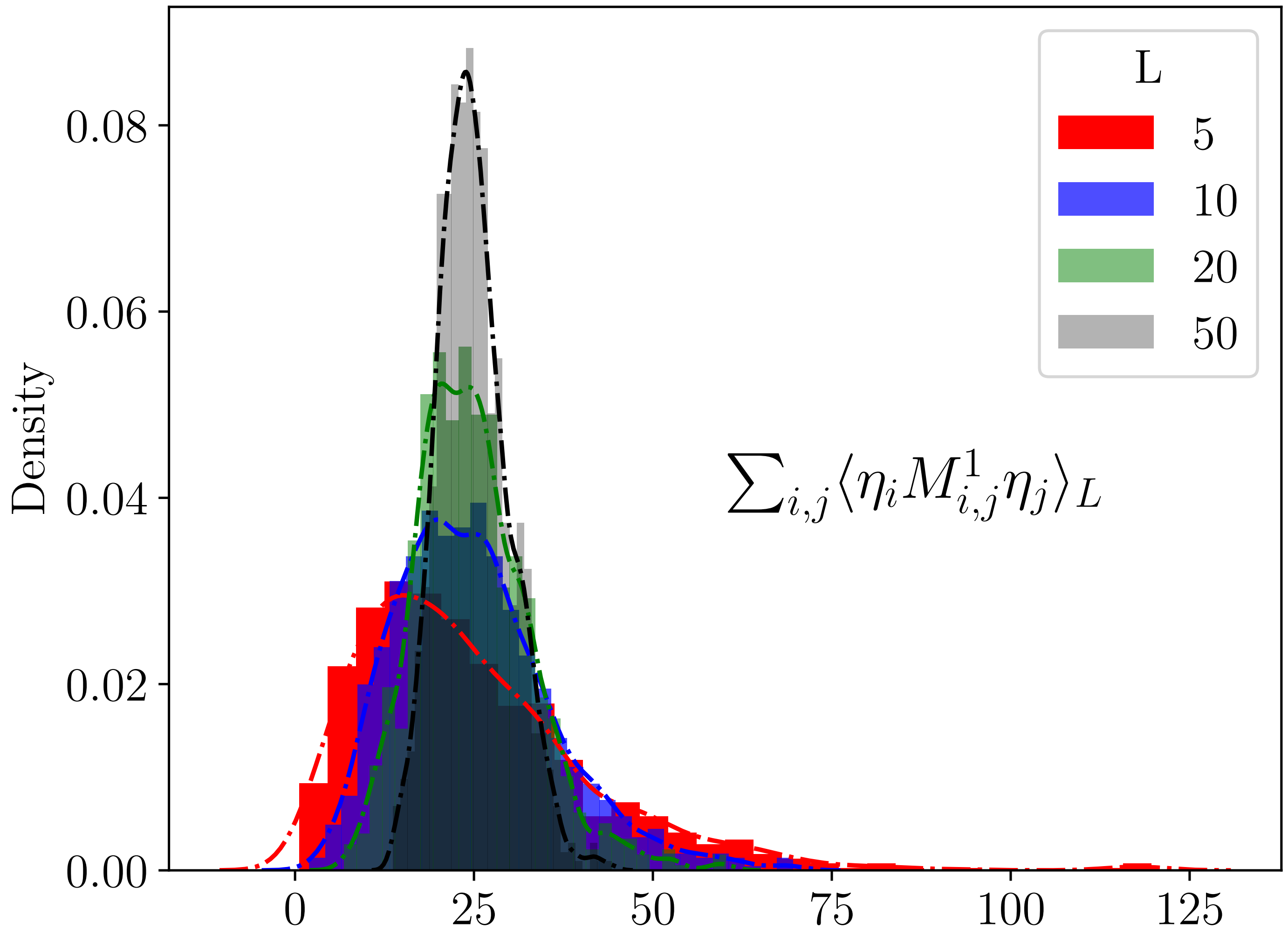
Distribution of random vectors & correlations

- **Observation I** : Distributions change with L *non-trivially* and depend on the random sources rather than matrix structure

distribution of random vectors as applied to a 1000 by 1000 matrix of all ones

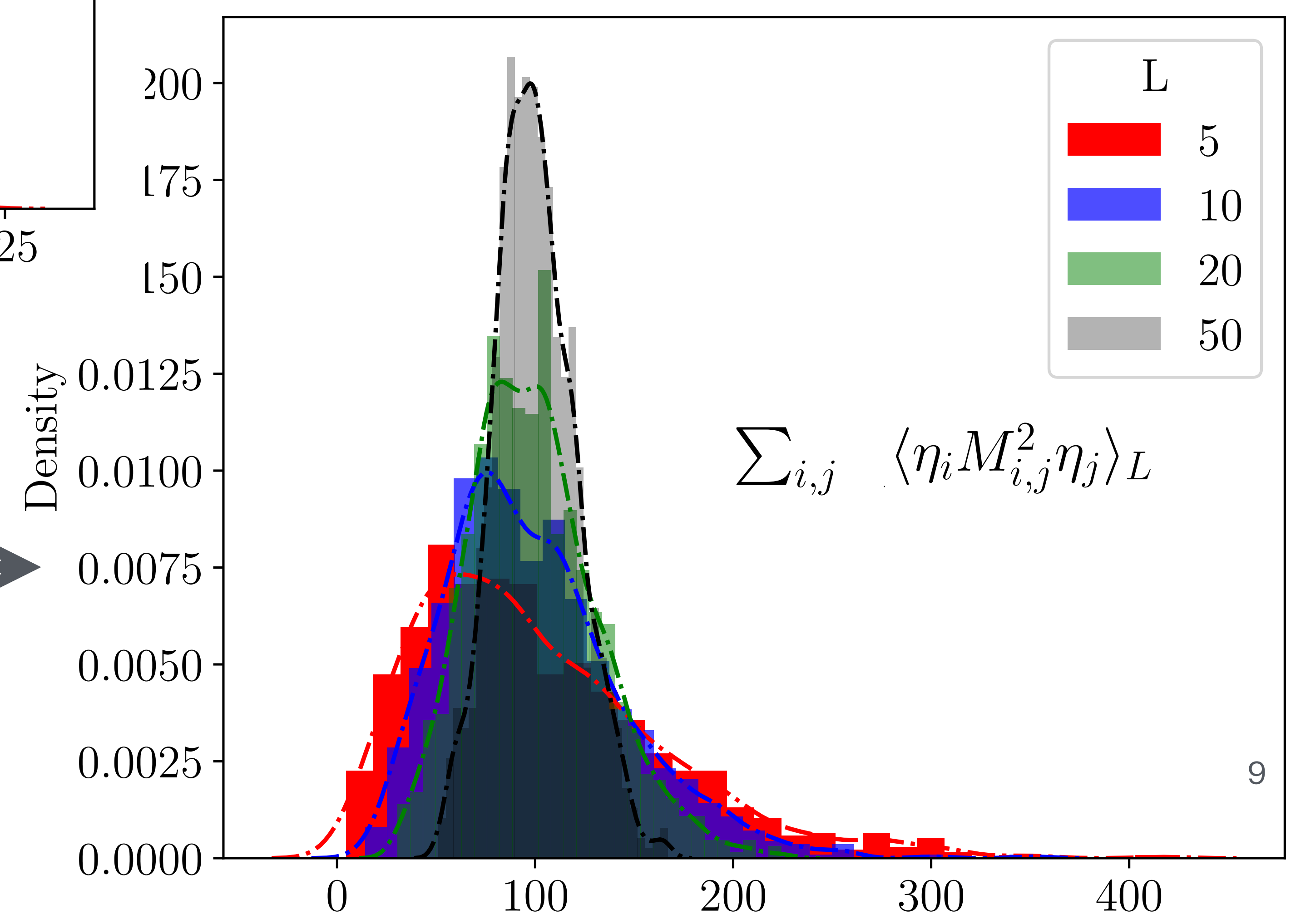


Distribution of random vectors & correlations



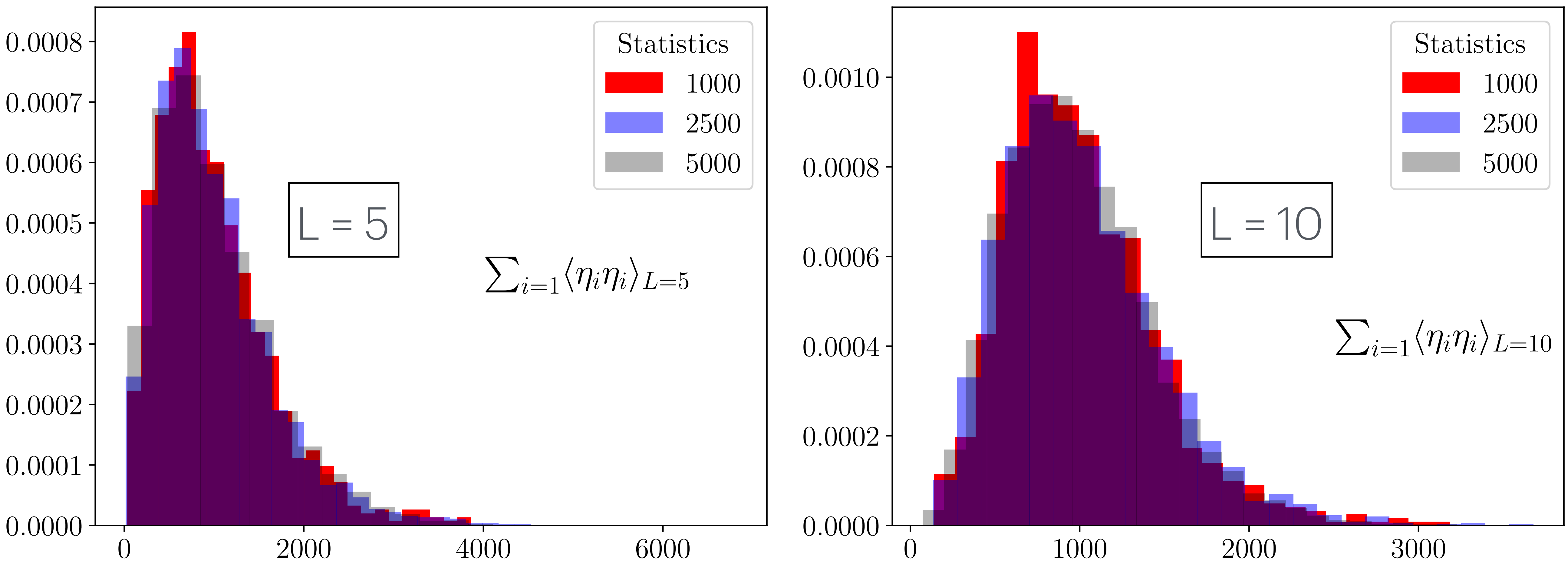
normal distr. , 1000 by 1000, dense matrix

poisson distr. , 1000 by 1000, sparse matrix



Distribution of random vectors & correlations

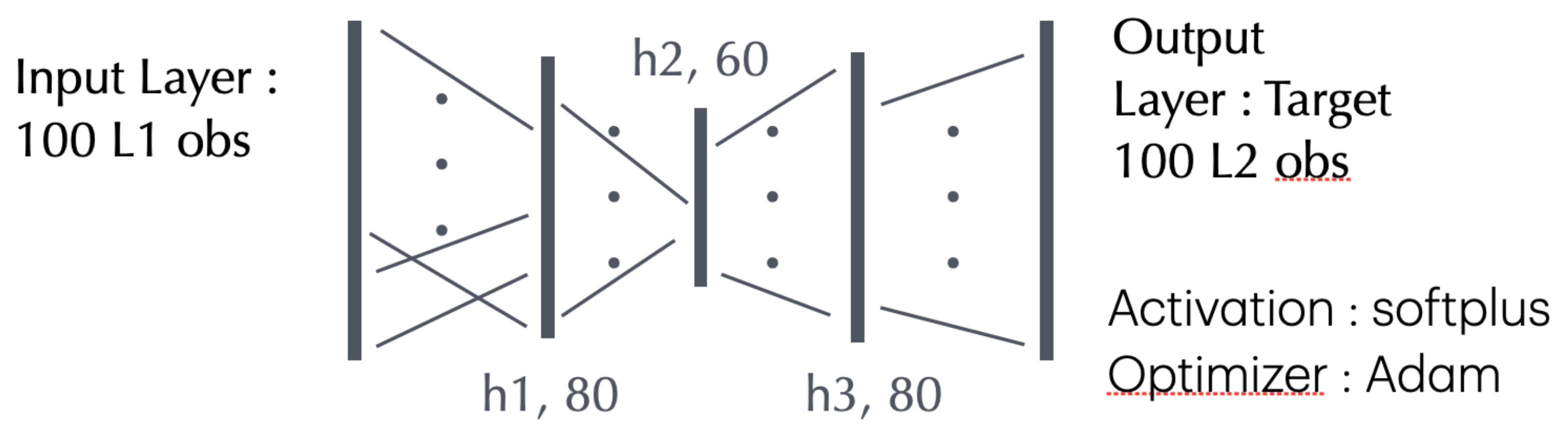
- **Observation II** : Distributions differ for different L - *not an artefact of low statistics !*



- As L *increases*, skewness *decreases* and binder cumulant *increases*

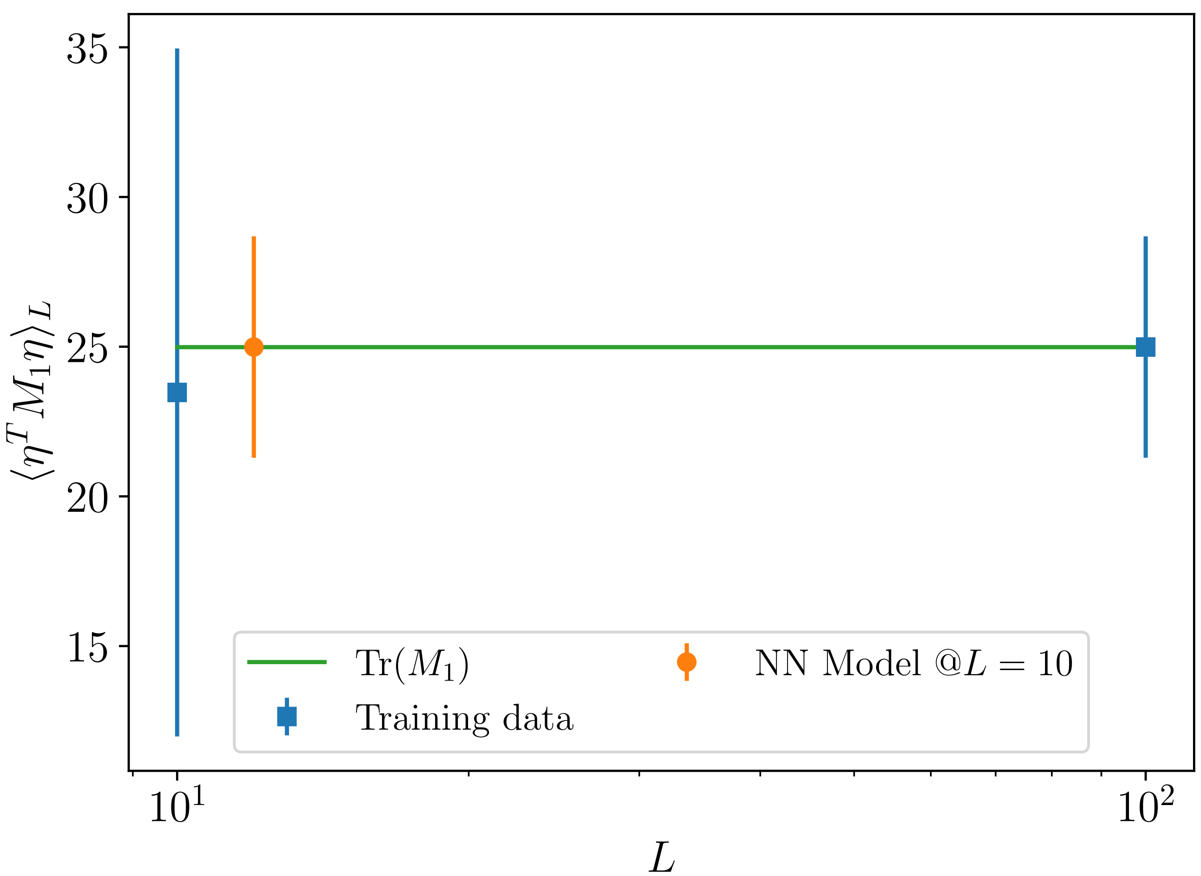
Results for mock matrices I

- A simple fully connected, network with 3 hidden layers



- Training : Iteratively update the network parameters to minimize the difference || A L1 - L2 || to learn A

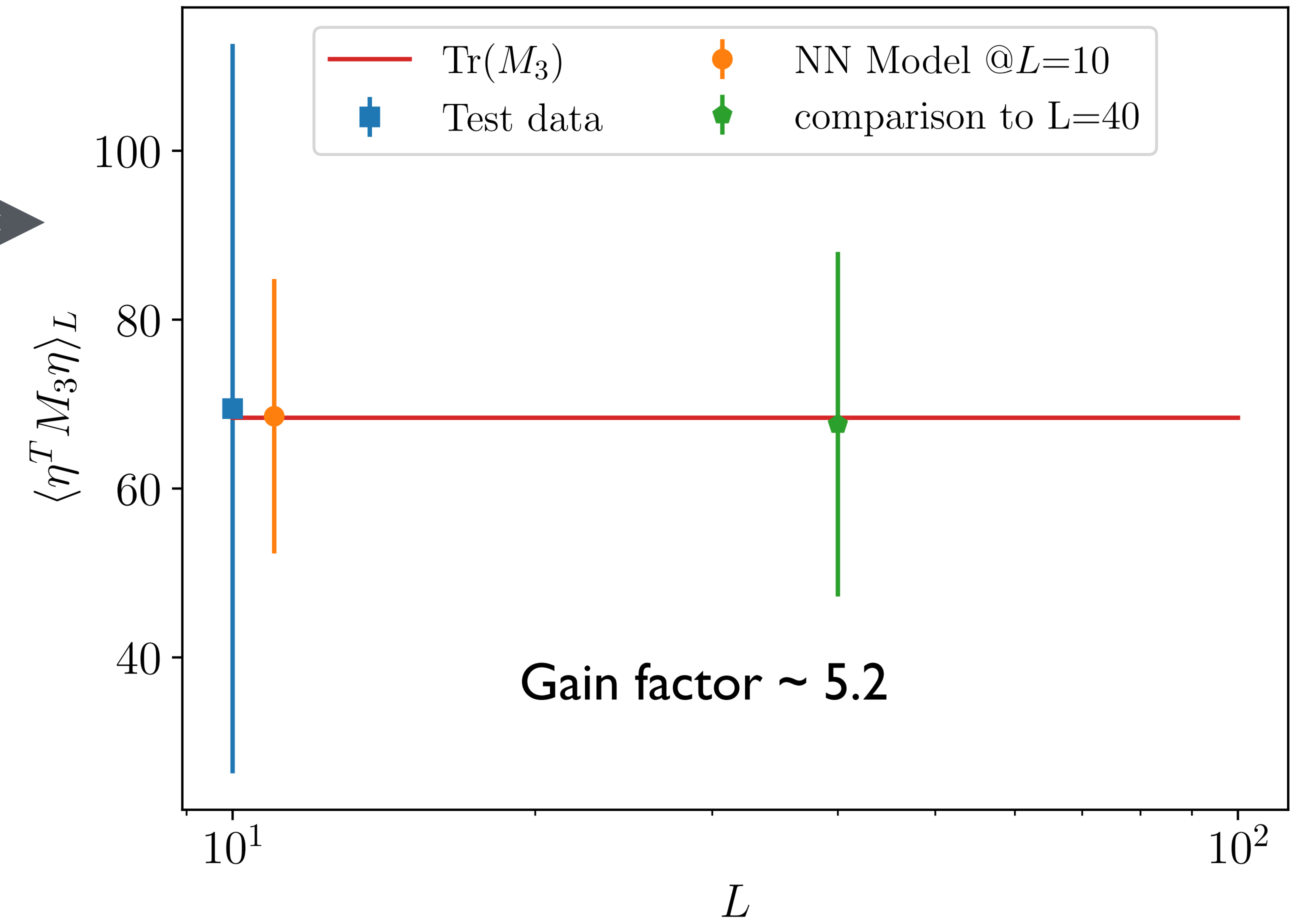
M_1 : A dense 1K by 1K matrix with elements drawn from a gaussian

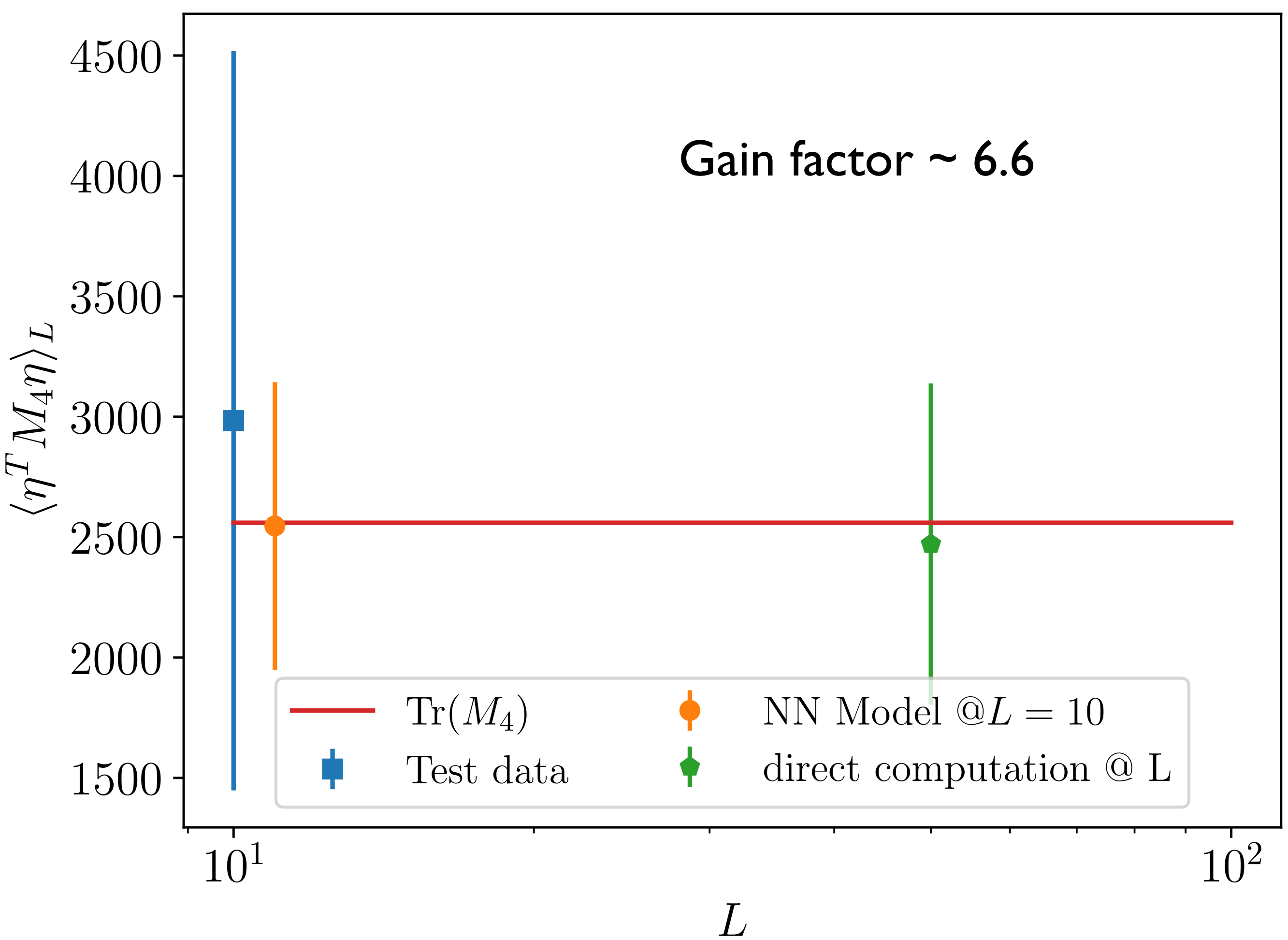


Results for mock matrices II

- The same model applied to other matrices or different structures and sizes

M_3 : A sparse 1000 by 1000 matrix with elements drawn from a poisson distribution





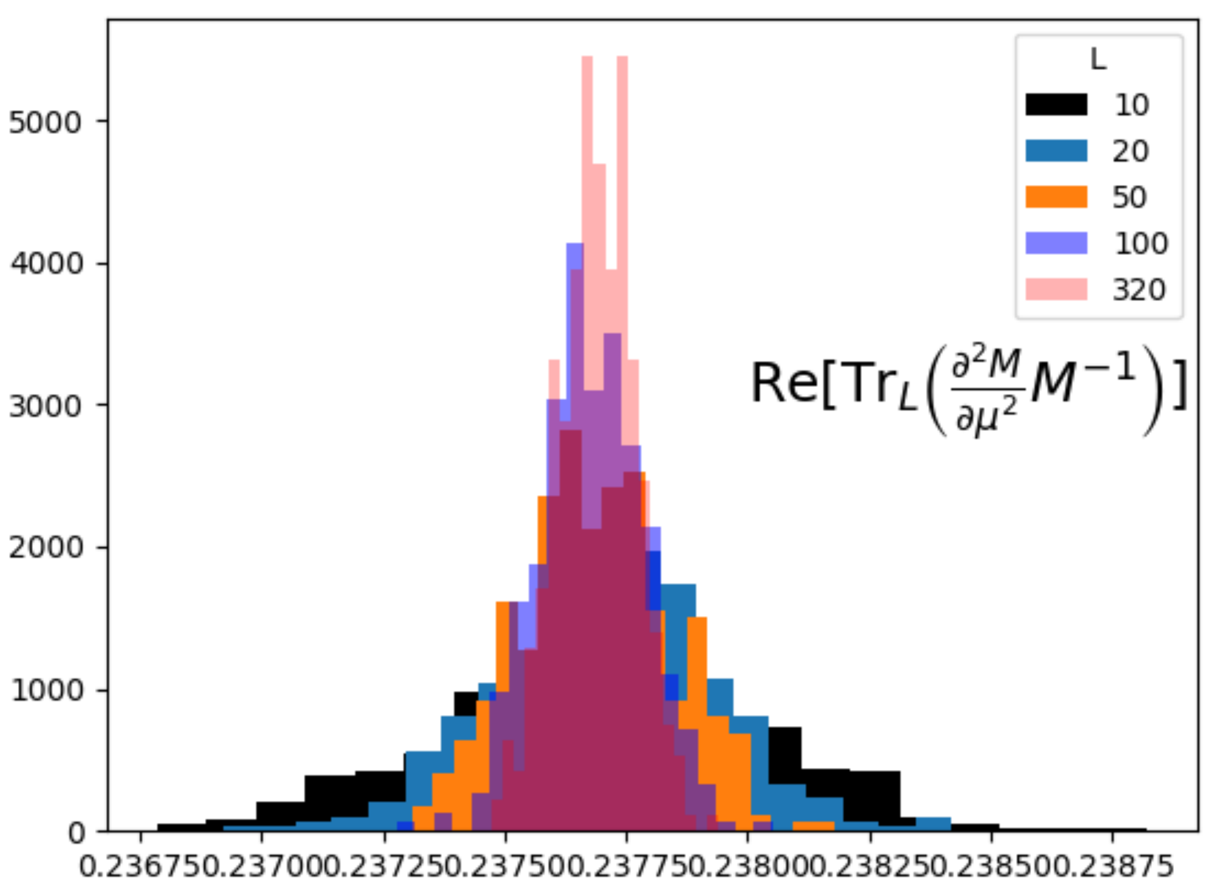
M_4 : A sparse 10K by 10K matrix with elements drawn from a gaussian

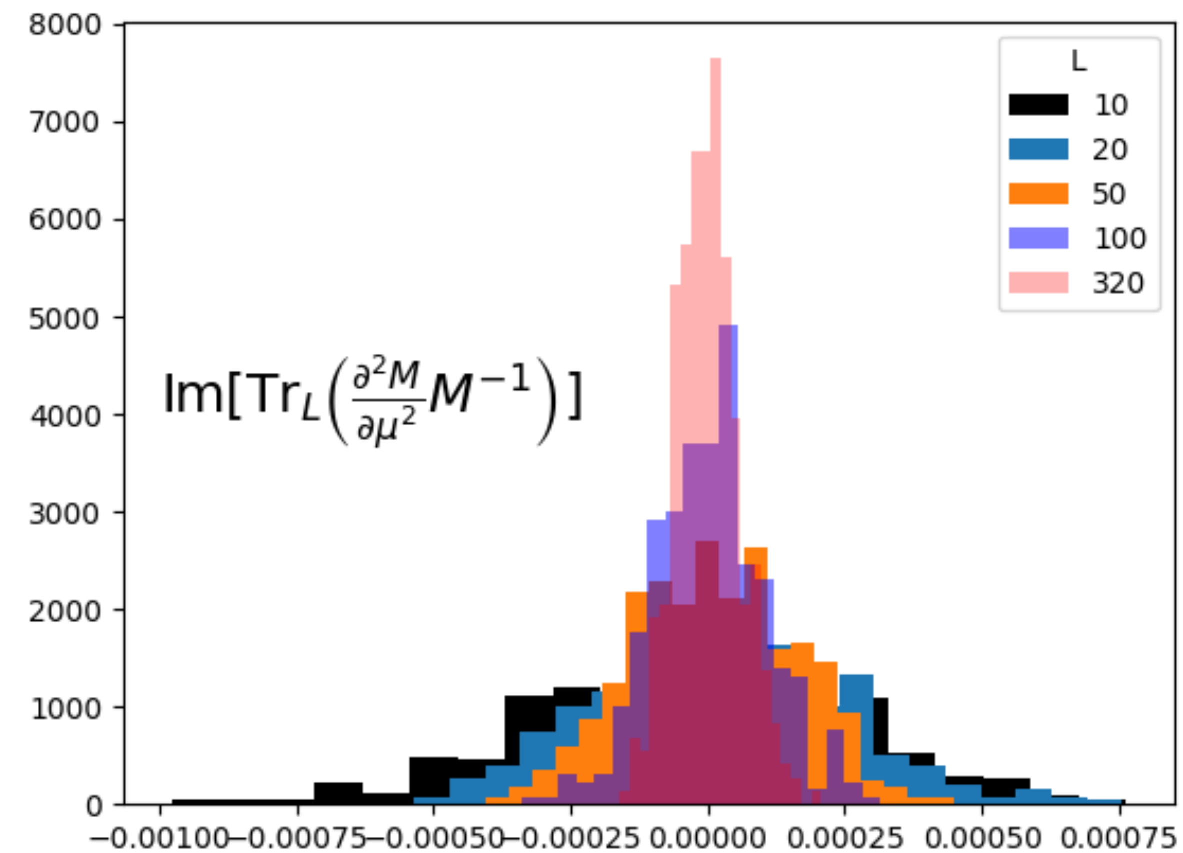
Lattice data : What happens here?

- To apply this to lattice we need data in the form of various independent sets of measurements for different number of random sources

- Using data from our (Bielefeld - Parma collaboration) recent imaginary μ simulations [arXiv:2405.10196] Complex $\mu \rightarrow$ complex traces !

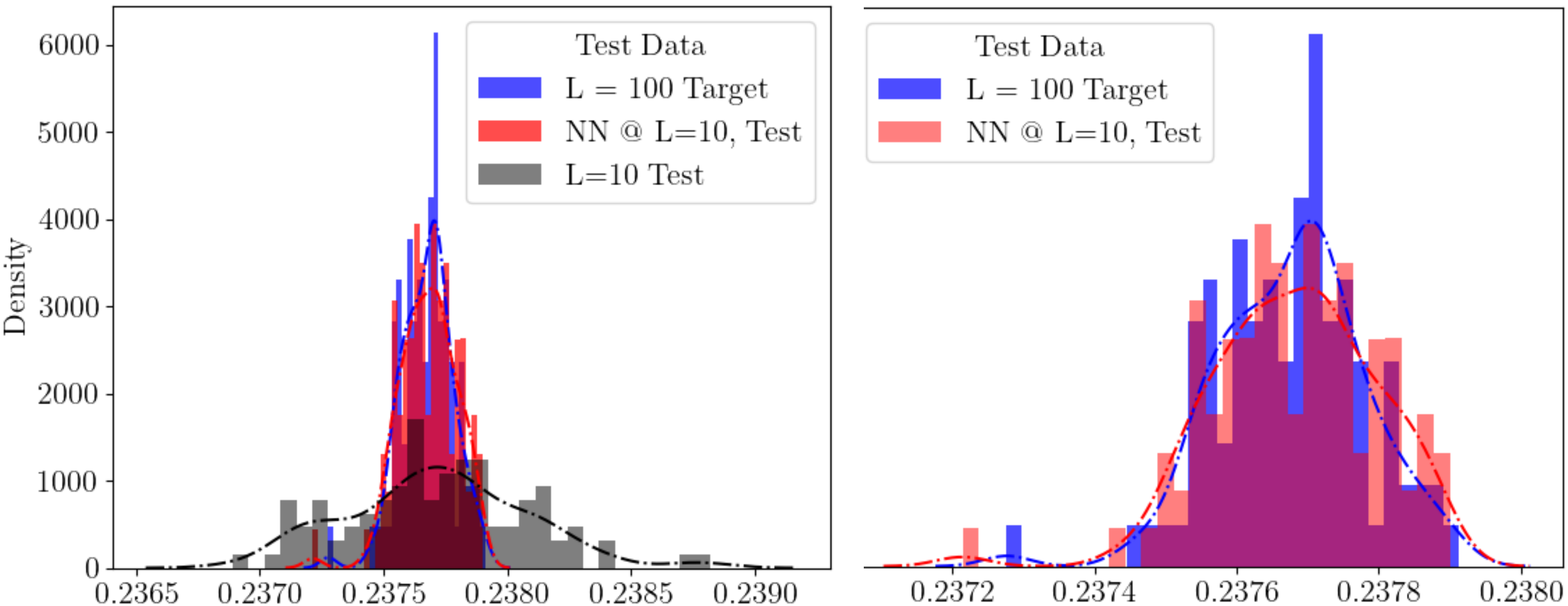
- Can we spilt the analysis into real and imaginary parts of observables - since Trace and $\mathbb{E}$ are linear operation?





Lattice data : What happens here?

- Although distributions not like mock data - low statistics and/or no fixed matrix?
- What does the model learn?



Summary

- Invitation to consider the well-developed Unfolding algorithms developed by the experimental community to lattice problems like inversion

- Motivate the problem of trace estimation as a “detector defect” problem that can be unfolded

- Some success on mock matrix data

- Lattice : There is no fixed matrix - as each gauge configuration has statistical fluctuations. This adds a dimension of complexity not present in the mock data

- Can one think of other ways to improve the Hutchinson trace estimator?

A BOUND FOR THE ERROR IN THE NORMAL APPROXIMATION TO THE DISTRIBUTION OF A SUM OF DEPENDENT RANDOM VARIABLES

CHARLES STEIN
STANFORD UNIVERSITY

Bernoulli Society for Mathematical Statistics and Probability

On Stein's method for products of normal random variables and zero bias couplings
Author(s): ROBERT E. GAUNT

- Data from other projects very welcome