Teaching to extract spectral densities from lattice correlators to a broad audience of learning-machines

Alessandro De Santis

in collaboration with Michele Buzzicotti and Nazario Tantalo

40th International Symposium on Lattice Field Theory (Lattice 2023)

July 31, 2023 to August 4, 2023 - Fermilab, Chicago

TOR VERGATA
UNIVERSITY OF ROME
School of Mathematics, Physical and Natural Sciences

INFN
ROMA TOR VERGATA

2023 LATTICE

Based on arXiv:2307.00808:

Teaching to extract spectral densities from lattice correlators to a broad audience of learning-machines

Michele Buzzicotti,[1, *] Alessandro De Santis,[1, †] and Nazario Tantalo[1, ‡]

[1] *University and INFN of Roma Tor Vergata, Via della Ricerca Scientifica 1, I-00133, Rome, Italy*

(Dated: July 4, 2023)

This work

Introduction: hadronic spectral densities in Lattice QCD

Hadronic processes can be described in terms of spectral densities

- R-ratio: $e^+e^- \mapsto X$
  - ETMC Phys.Rev.Lett. 130, 241901 (2023)
- Inclusive hadronic τ decays: $\tau \mapsto \nu_\tau X$
  - **A. Evangelista's talk**
  - A. Evangelista et al. (next week)
- Inclusive decays of heavy mesons: $B \mapsto \ell\bar{\nu}X$
  - **A. Barone's talk**
  - S. Hashimoto PTEP, Volume 2017, Issue 5, (2017)
  - P. Gambino and S. Hashimoto PRL 125, 032001 (2020)
  - P. Gambino et al. JHEP volume 2022, Article n: 83 (2022)
  - A. Barone et al. arXiv:2305.14092 (2023)
- Deep Inelastic Scattering: $e^-P \mapsto e^-X$
  - M. T. Hansen et al. Phys.Rev.D 96, 094513 2017
- Radiative leptonic decays: $D_s \mapsto \ell\nu_\ell\gamma^*$
  - R. Frezzotti et al. arxiv:2306.07228
- Spectrum analysis
  - **A. Smecca's talk**
  - L. Del Debbio et al. EPJ C volume 83, Article number: 220 (2023)

Spectral densities are related to **Euclidean** correlation functions calculated on the **lattice**

$$C(t) = \int \mathrm{d}^3x\, e^{-i\boldsymbol{p}\cdot\boldsymbol{x}} \langle 0|\, \hat{O}_1 e^{-t\hat{H}+i\hat{\boldsymbol{P}}\cdot\boldsymbol{x}} \hat{O}_2 \,|0\rangle = \int_{E_0}^{\infty} \mathrm{d}E\, e^{-tE} \rho(E)$$

$$\rho(E) \equiv \langle 0|\, \hat{O}_1 \delta(\hat{H} - E)\delta^3(\hat{\boldsymbol{P}} - \boldsymbol{p})\hat{O}_2 \,|0\rangle$$

Spectral densities are related to **Euclidean** correlation functions calculated on the **lattice**

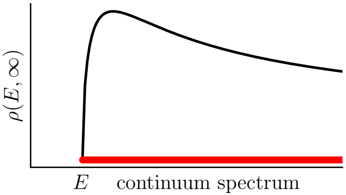
$$C(t) = \int \mathrm{d}^3x\, e^{-i\boldsymbol{p}\cdot\boldsymbol{x}} \langle 0| \hat{O}_1 e^{-t\hat{H}+i\hat{\boldsymbol{P}}\cdot\boldsymbol{x}} \hat{O}_2 |0\rangle = \int_{E_0}^{\infty} \mathrm{d}E\, e^{-tE} \rho(E)$$

$$\rho(E) \equiv \langle 0| \hat{O}_1 \delta(\hat{H} - E)\delta^3(\hat{\boldsymbol{P}} - \boldsymbol{p})\hat{O}_2 |0\rangle$$

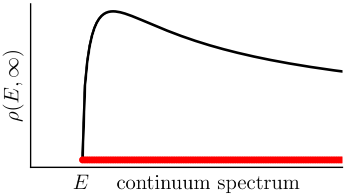
Extracting $\rho(E)$ from $C(t)$ is a **numerically ill-posed inverse problem**

- ▷ $t = a\tau \qquad \tau = 1, \cdots, T$ — finite amount of information
- ▷ $\bar{C}(t) \pm \Delta C(t)$ — imprecise data

Physics is associated with infinite volume spectral densities

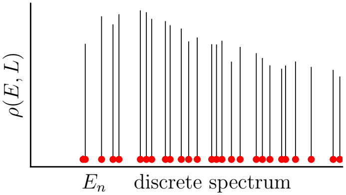


Physics is associated with infinite volume spectral densities

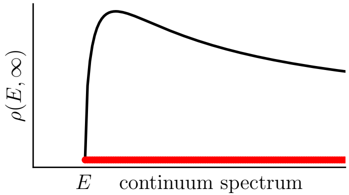


Finite volume spectral densities are **badly-behaving distributions**

$$\rho(E,L) = \sum_n \omega_n(L)\delta\big(E - E_n(L)\big)$$

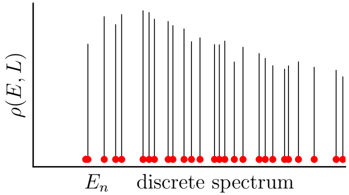


Physics is associated with infinite volume spectral densities



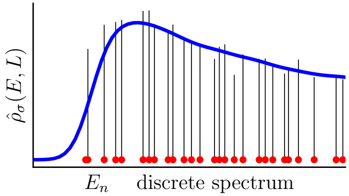
Finite volume spectral densities are **badly-behaving distributions**

$$\rho(E,L) = \sum_n \omega_n(L)\delta\big(E - E_n(L)\big)$$



Axiomatic: spectral densities must be smeared

$$\hat{\rho}_\sigma(E,L) = \int_0^\infty \mathrm{d}\omega K_\sigma(E,\omega)\rho(\omega,L)$$



$$K_\sigma(E,\omega) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(E-\omega)^2}{2\sigma^2}\right) \qquad \rho(E,\infty) = \lim_{\sigma\to 0}\lim_{L\to\infty} \hat{\rho}_\sigma(E,L)$$

We focus on the extracion of $\hat{\rho}_\sigma(E)$ smeared with a Gaussian of resolution $\sigma > 0$

$$K_\sigma(E,\omega) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(E-\omega)^2}{2\sigma^2}\right) \qquad \rho(E,\infty) = \lim_{\sigma\to 0} \lim_{L\to\infty} \hat{\rho}_\sigma(E,L)$$

We focus on the extracion of $\hat{\rho}_\sigma(E)$ smeared with a Gaussian of resolution $\sigma > 0$

A method based on linearity and Backus-Gilbert regularization already exists:

Extraction of spectral densities from lattice correlators

Martin Hansen,[1] Alessandro Lupo,[2] and Nazario Tantalo[3]

[1]*INFN Roma Tor Vergata, Via della Ricerca Scientifica 1, I-00133 Rome, Italy*
[2]*University of Rome Tor Vergata, Via della Ricerca Scientifica 1, I-00133 Rome, Italy*
[3]*University of Rome Tor Vergata and INFN Roma Tor Vergata, Via della Ricerca Scientifica 1, I-00133 Rome, Italy*

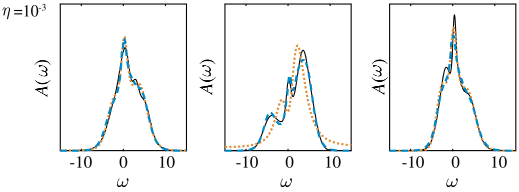
HLT

A. Lupo's talk for the connection between Bayesian and Backus-Gilbert methods

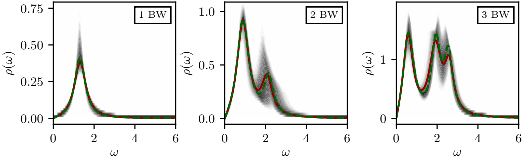
Machine Learning approach

The idea of using machine learning for spectral reconstruction is not original

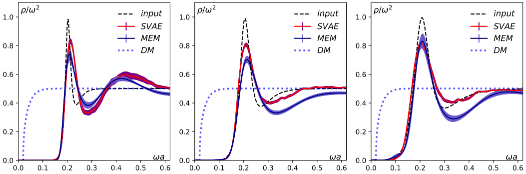
- ▷ Fourier et al. 2020



- ▷ Kades et al. 2021



- ▷ Chen et al. 2022



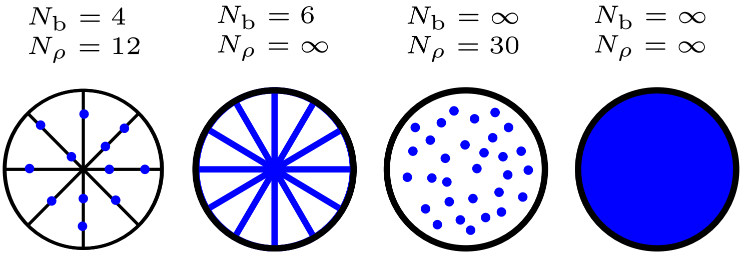
WHAT'S DIFFERENT?

1) Is it possible to devise a model independent training strategy ?

2) If such a strategy is found, is it then possible to quantify reliably , together with the statistical errors, also the unavoidable systematic uncertainties ?

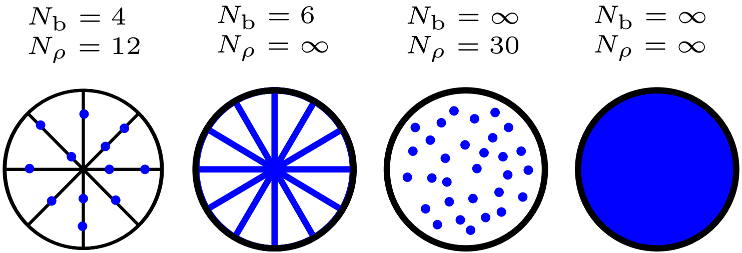
1) FUNCTIONAL-BASIS to parametrize the training set

N_{b} number of basis functions
N_{ρ} number of spectral functions in the space



1) FUNCTIONAL-BASIS to parametrize the training set

N_{b} number of basis functions
N_ρ number of spectral functions in the space



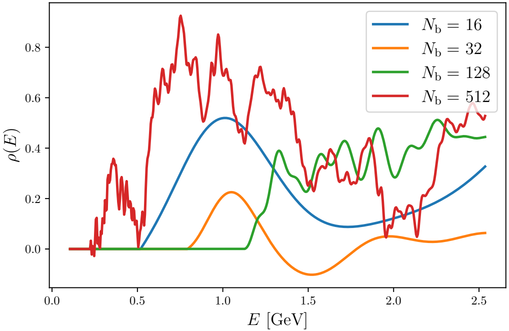
Limiting /either N_{b} and/or N_ρ means limiting the information to which the neural network is exposed.

Model independence is achieved in the limit $(N_{\mathrm{b}}, N_\rho) \mapsto \infty$

We choose **Chebyshev polynomials** as basis functions

$$\rho(E) = \theta(E - E_0) \sum_{n=0}^{N_\mathrm{b}} c_n \left[T_n(E) - T_n(E_0)\right]$$

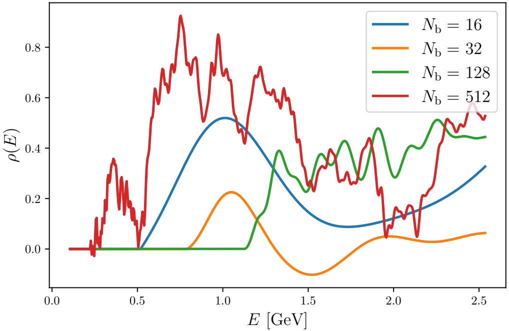
- c_n randomly generated
- E_0 mass gap



We choose **Chebyshev polynomials** as basis functions

$$\rho(E) = \theta(E - E_0) \sum_{n=0}^{N_\mathrm{b}} c_n \left[T_n(E) - T_n(E_0)\right]$$

- c_n randomly generated
- E_0 mass gap



and generate N_ρ unsmeared spectral densities

$$\rho(E) \quad \rightarrow \quad \begin{matrix} \text{INPUT} \quad \boldsymbol{C} = \{C(a), C(2a), \cdots\} \\ \text{OUTPUT} \quad \boldsymbol{\hat{\rho}_\sigma} = \{\hat{\rho}_\sigma(E_1), \hat{\rho}_\sigma(E_2), \cdots\} \end{matrix} \quad \rightarrow \quad \mathcal{T}_\sigma(N_\mathrm{b}, N_\rho) \quad \text{TRAINING SET}$$

2) ENSEMBLE OF MACHINES to quantify the error

We consider 3 architectures based on 1D Convolutional layers

Type	Maps	Size	Kernel size	Stride	Activation
Input		64			
Conv1D	32	32x32	3	2	LeakyReLu
Conv1D	64	16x64	3	2	LeakyReLu
Conv1D	128	8x128	3	2	LeakyReLu
Flatten		1024			
Fully conn.		256			LeakyReLu
Fully conn.		256			LeakyReLu
Output		47			
Parameters	371311				

ID	N_{n} number of neurons
arcS	94651
arcM	180871
arcL	371311

2) ENSEMBLE OF MACHINES to quantify the error

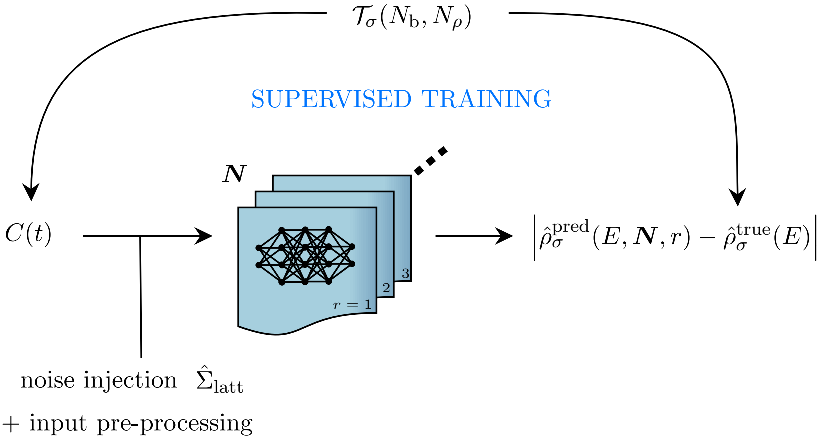
We consider 3 architectures based on 1D Convolutional layers

Type	Maps	Size	Kernel size	Stride	Activation
Input		64			
Conv1D	32	32x32	3	2	LeakyReLu
Conv1D	64	16x64	3	2	LeakyReLu
Conv1D	128	8x128	3	2	LeakyReLu
Flatten		1024			
Fully conn.		256			LeakyReLu
Fully conn.		256			LeakyReLu
Output		47			
Parameters	371311				

ID	$N_{\rm n}$ number of neurons
arcS	94651
arcM	180871
arcL	371311

- **The answer** of a machine with **finite** $N_{\rm n}$ neurons, **trained over a finite set** $\mathcal{T}_\sigma(N_{\rm b}, N_\rho)$ **cannot be exact**
- To quantify the network error we introduce $N_r = 20$ replica machines, the **ensemble of machines**, at fixed $\boldsymbol{N} = (N_{\rm n}, N_{\rm b}, N_\rho)$

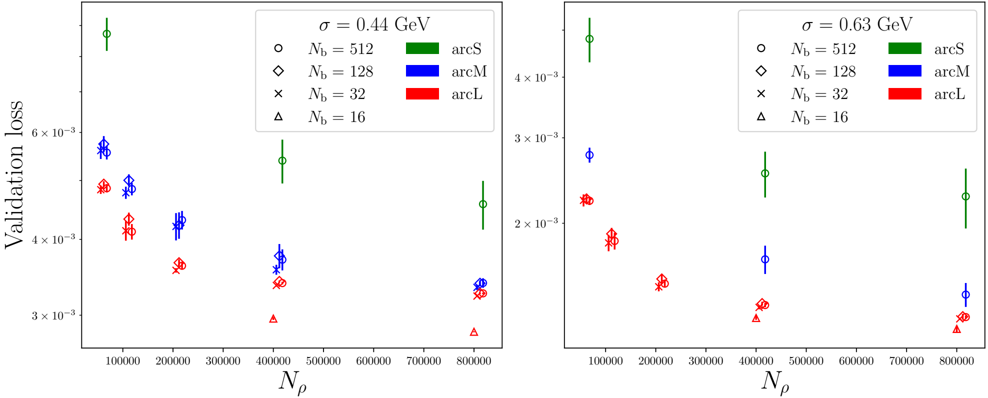
TEACHING the ensemble of machines



$\hat{\Sigma}_{\mathrm{latt}}$ is the covariance matrix of a true lattice correlator

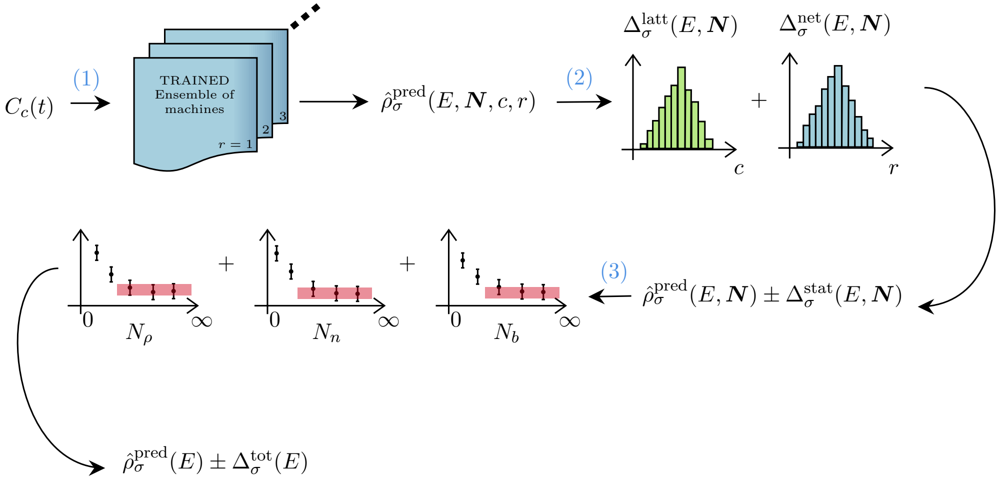
The broad audience of learning-machines

$$\text{MAE} \qquad \ell(\boldsymbol{w}) = \frac{\sum |\hat{\rho}_\sigma^{\text{pred}}(\boldsymbol{\omega}) - \hat{\rho}_\sigma^{\text{true}}|}{N_\rho}$$



$\mathcal{O}(3000)$ independent trainings

Quoting the final result with errors

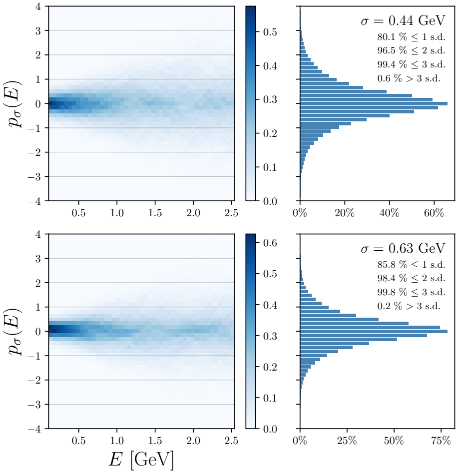


Validation on 2000-sample test sets

histograms of the significance

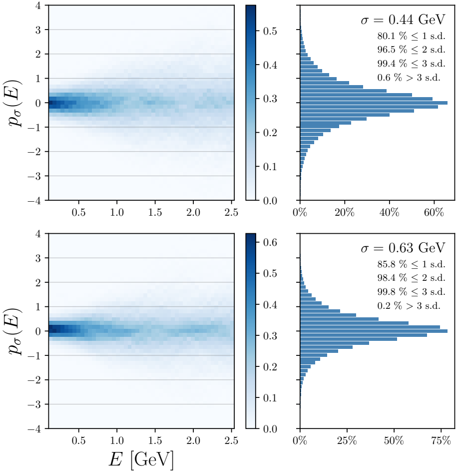
$$p_\sigma(E) = \frac{\hat{\rho}_\sigma^{\mathrm{pred}}(E) - \hat{\rho}_\sigma^{\mathrm{true}}(E)}{\Delta_{\mathrm{tot}}(E)}$$

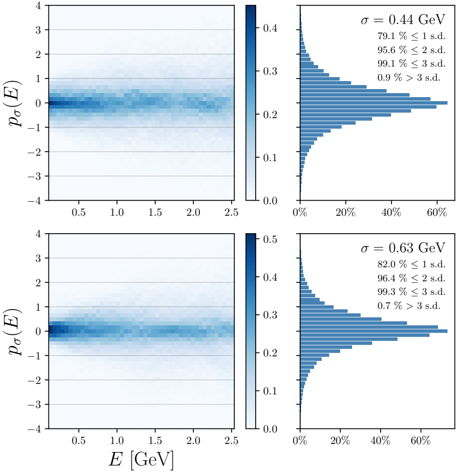
$$\rho(E) = \sum_{n=0}^{N_{\mathrm{b}} \leq 1024} c_n T_n(E)$$



$$\rho(E) = \sum_{n=0}^{N_b \leq 1024} c_n T_n(E)$$

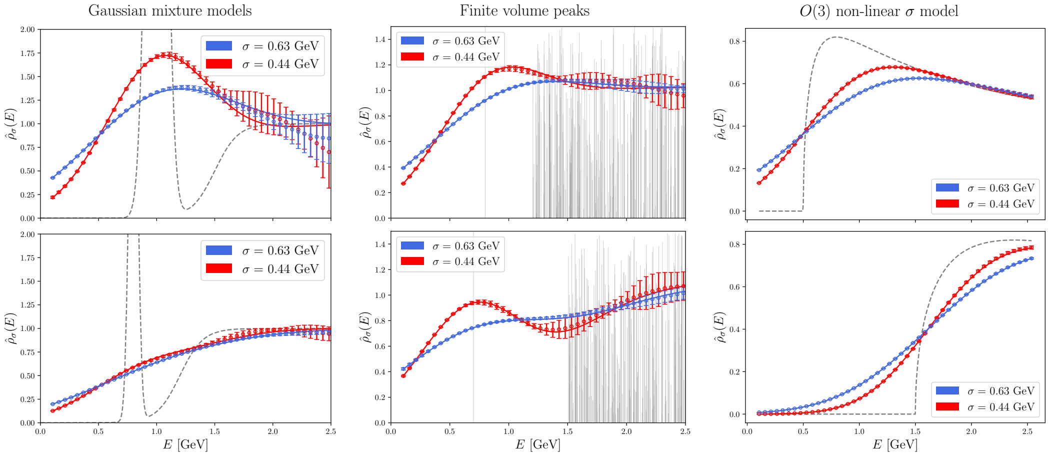
$$\rho(E) = \sum_{n=0}^{N_{\text{peaks}}} \omega_n \delta(E - E_n)$$





We observe deviations less than 1:2:3 standard deviations in about 80% : 95% : 99% of the cases

Mock data inspired by physical models



None of these spectral densities are included in the training set

True lattice data: the R-ratio

$$R(E) = \frac{\sigma(e^+e^- \to \mathrm{hadrons})}{\sigma(e^+e^- \to \mu^+\mu^-)}$$

$$C(t) = \int_0^\infty \mathrm{d}\omega \, \frac{\omega^2}{12\pi^2} e^{-t\omega} R(\omega) \qquad\qquad C(t) = -\frac{1}{3}\sum_{i=1}^{3} \int \mathrm{d}^3x \, \hat{T} \, \langle 0 | \, \hat{J}_i^{\mathrm{em,had}}(x) \hat{J}_i^{\mathrm{em,had}}(0) \, | 0 \rangle$$

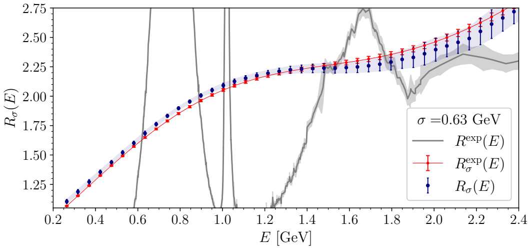
PHYSICAL REVIEW LETTERS **130**, 241901 (2023)

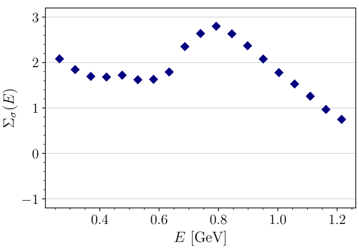
Probing the Energy-Smeared *R* Ratio Using Lattice QCD

Constantia Alexandrou,[1,2] Simone Bacchio,[2] Alessandro De Santis,[3] Petros Dimopoulos,[4] Jacob Finkenrath,[2] Roberto Frezzotti,[3] Giuseppe Gagliardi,[5] Marco Garofalo,[6] Kyriakos Hadjiyiannakou,[1,2] Bartosz Kostrzewa,[7] Karl Jansen,[8] Vittorio Lubicz,[9] Marcus Petschlies,[6] Francesco Sanfilippo,[5] Silvano Simula,[5] Nazario Tantalo,[3,*] Carsten Urbach,[6] and Urs Wenger[10]

(Extended Twisted Mass Collaboration (ETMC))

R-ratio

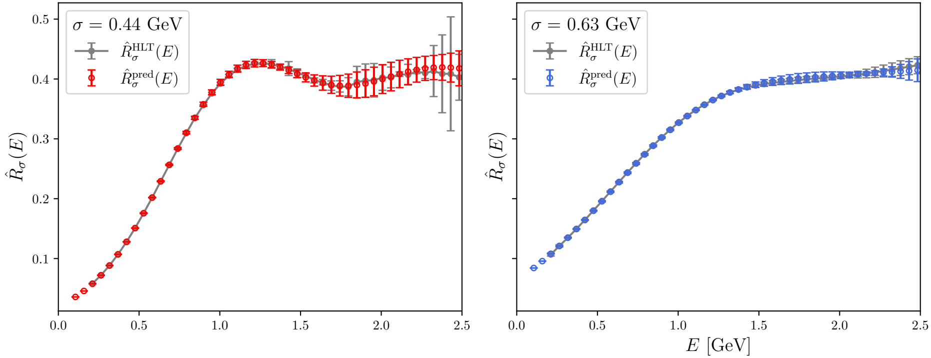




ID	$L^3 \times T$	a fm	aL fm	m_π GeV
B64	$64^3 \times 128$	0.07957(13)	5.09	0.1352(2)

(C. Alexandrou et al. 2022)

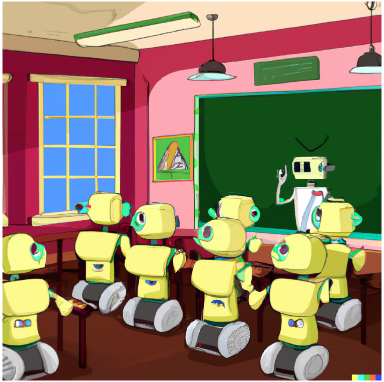
Strange-strange connected contribution:



The two **totally unrelated** methods are in exceptional agreement

Conclusions

- ▷ **Supervised deep learning techniques** can be used to extract smeared hadronic spectral densities from lattice correlators in a **model-independent way**
- ▷ The **systematic errors can be reliably quantified** and the predictions can be used in phenomenological analyses
- ▷ Admittedly, the procedure that we propose to do might end up to be numerically demanding and can possibly be simplified, but there is no free-lunch in physics!
- ▷ Here **we taught a lesson to a broad audience of learning-machines** but the subject of the lesson is just a particular topic
- ▷ **The idea of teaching systematically to a broad audience of machines is much more general** and can be used to estimate reliably the systematic errors in many other applications



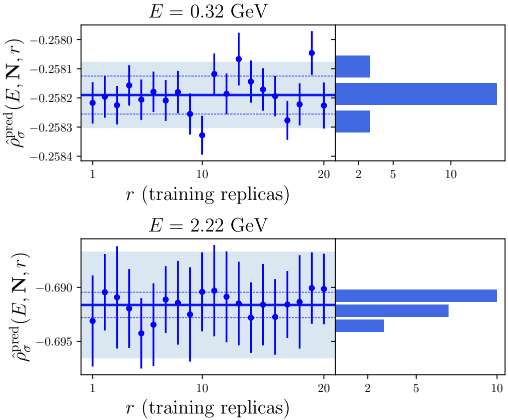
A classroom where the students are replaced by robots (DALL·E)

THANK YOU FOR THE ATTENTION!

Backup slides

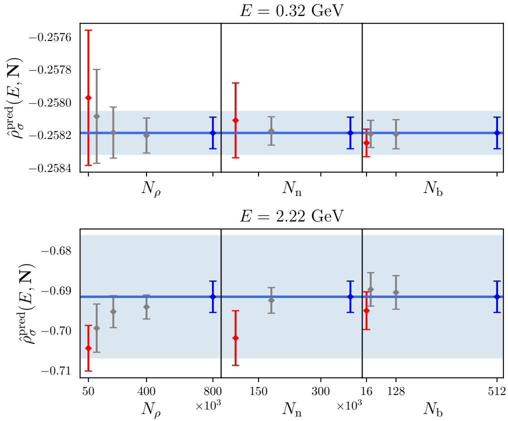
New unsmeared $\rho(E)$ generated with $N_{\mathrm{b}} = 1024$

Collect and combine the different responses from the ensembles of machines



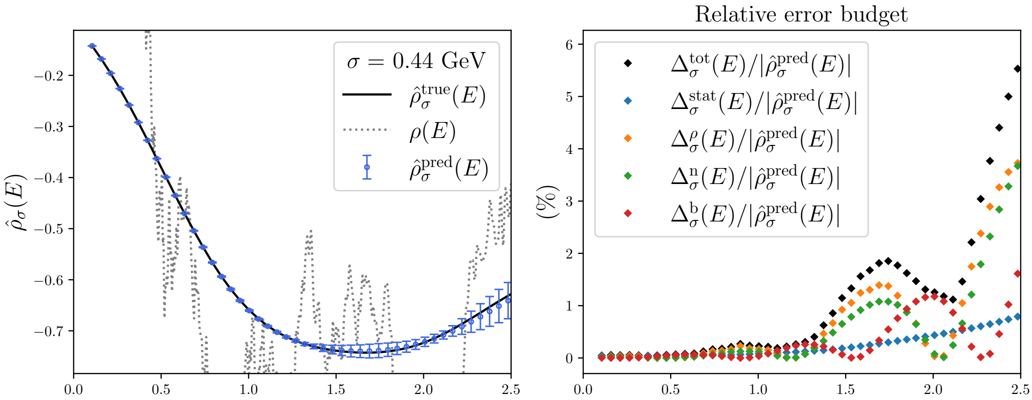
$$\Delta_{\sigma}^{\mathrm{stat}}(E, \boldsymbol{N}) = \sqrt{\left[\Delta_{\sigma}^{\mathrm{latt}}(E, \boldsymbol{N})\right]^2 + \left[\Delta_{\sigma}^{\mathrm{net}}(E, \boldsymbol{N})\right]^2}$$

Perform numerically the $\boldsymbol{N} \to \infty$ limit in a plateau sense and quote the total error



$$\Delta_{\sigma}^{\mathrm{tot}}(E) = \sqrt{\left[\Delta_{\sigma}^{\mathrm{stat}}\right]^2 + \left[\Delta_{\sigma}^{\rho}\right]^2 + \left[\Delta_{\sigma}^{\mathrm{b}}\right]^2 + \left[\Delta_{\sigma}^{\mathrm{n}}\right]^2}$$

Final result



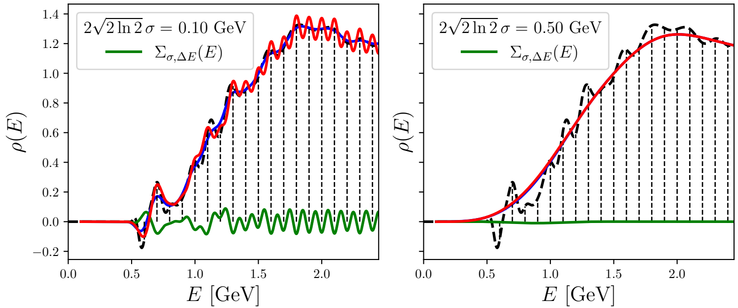
The systematic uncertainties associated with $N < \infty$ are **not negligible**

Wait, wait, wait ... distributions cannot be represented by the Chebyshev basis

Right! But we are interested in **smeared** spectral densities

$$\hat{\rho}_\sigma(E) = \int_{E_0}^{\infty} \mathrm{d}\omega\, K_\sigma(E,\omega)\rho(\omega) = \Delta E \sum_{n=0}^{\infty} K_\sigma(E,\omega_n)\rho(\omega_n) + \Sigma_{\sigma,\Delta E}(E)$$

$$= \int_{E_0}^{\infty} \mathrm{d}\omega\, K_\sigma(E,\omega)\rho_\delta(\omega) + \Sigma_{\sigma,\Delta E}(E)$$



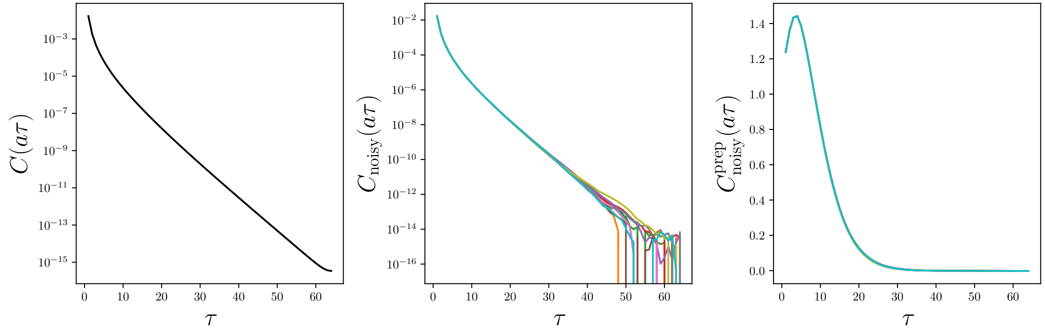
Everything should work as long as $\sigma \gg$ peaks separation (validated!)

Noise injection via true lattice data covariance matrix

$$\boldsymbol{C}_{\text{noisy}} \in \mathbb{G}\left[\boldsymbol{C}, \left(\frac{C(a)}{C_{\text{latt}}(a)}\right)^2 \hat{\Sigma}_{\text{latt}}\right]$$

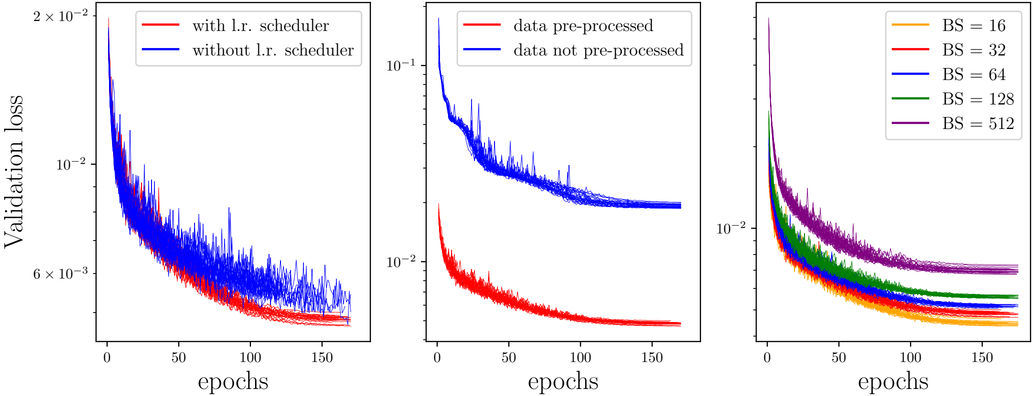
Standardization of input data

$$C^{\text{prep}}_{\text{noisy}}(a\tau) = \frac{C_{\text{noisy}}(a\tau) - \mu(\tau)}{\gamma(\tau)} \qquad \mu(\tau) = \frac{1}{N_\rho}\sum_{i=1}^{N_\rho} C_i(a\tau) \qquad \gamma(\tau) = \sqrt{\frac{\sum_{i=1}^{N_\rho}\left(C_i(a\tau) - \mu(\tau)\right)^2}{N_\rho}}$$

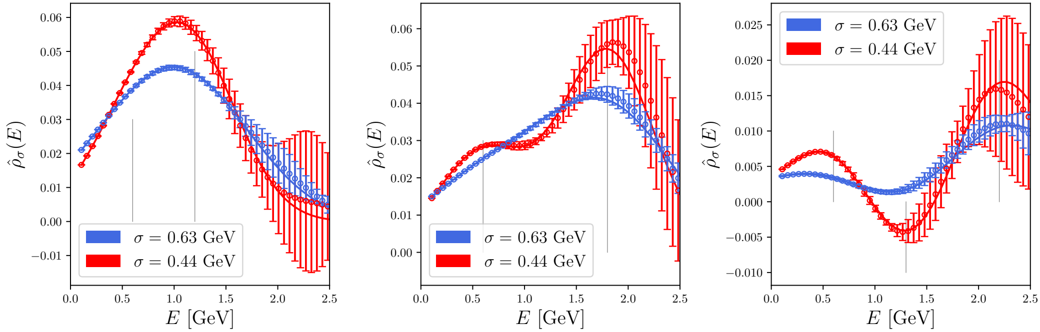


Hyperparameters tuning

- Learning rate scheduler $\eta(e) = \frac{\theta(e-25)\eta(e-1)}{1+e\cdot 4\times 10^{-4}}$
- Mini-Batch Gradient Descent algorithm. Batch Size = 32 + Adam optimizer

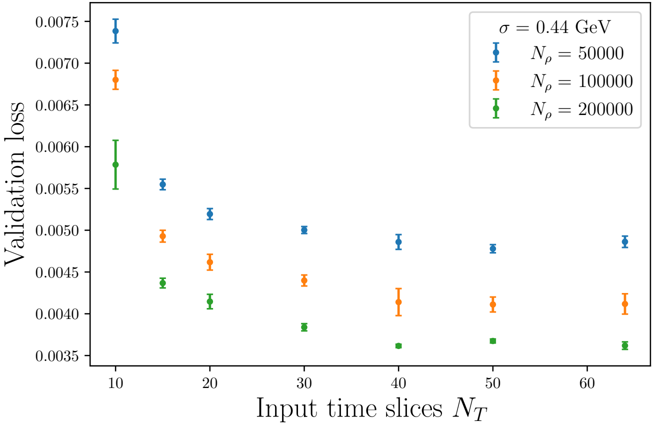


Performance in extreme cases



The error increases according to the increase of the severity of the inverse problem

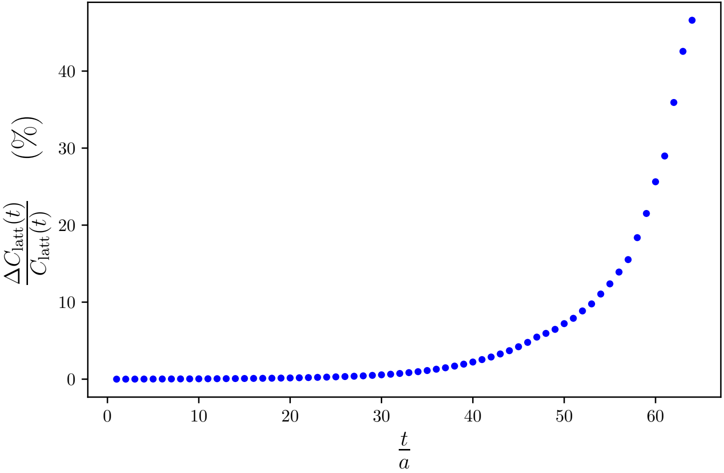
Impact of the number of input time slices

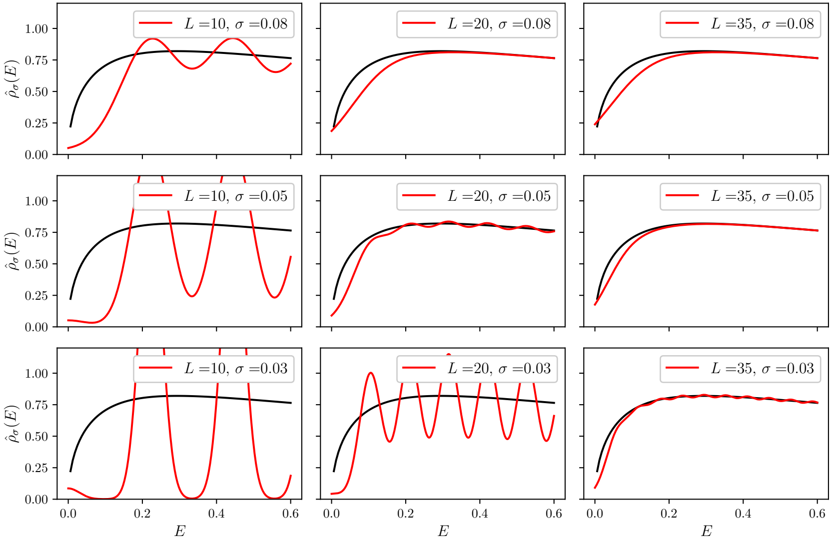


The performance of the neural network gets to a saturation point around $N_T \sim 40$

Noise level

We model the noise using the covariance matrix of the two-point vector strange-strange correlation function produced by ETM Collaboration on B64 ensemble





$$\rho(E) = \lim_{\sigma \to 0} \lim_{L \to \infty} \hat{\rho}_\sigma(E, L)$$

the order of the limits is important

We choose **Chebyshev polynomials** as basis functions

$$\rho(E) = \theta(E - E_0) \sum_{n=0}^{N_b} c_n \left[T_n(x(E)) - T_n(x(E_0))\right]$$

- c_n randomly generated
- E_0 mass gap
- $x(E) = 1 - 2e^{-E}$
- $c_n \in \frac{r_n}{n^{1+\varepsilon}} \qquad r_n \in [-1, 1]$