TOR VERGATA
UNIVERSITÀ DEGLI STUDI DI ROMA

nazario tantalo
nazario.tantalo@roma2.infn.it

Theory Institute LATTICE@TH 2024
CERN 10-07-2024

smeared R-ratio and applications to $g-2$

the R-ratio plays a fundamental role in particle physics since its introduction

LETTERE AL NUOVO CIMENTO VOL. IV, N. 1 4 Luglio 1970

Hadron Production in e^+e^- Collisions (*).

N. Cabibbo
Istituto di Fisica dell'Università - Roma
Istituto Nazionale di Fisica Nucleare - Sezione di Roma

G. Parisi and M. Testa
Istituto di Fisica dell'Università - Roma

(ricevuto il 30 Maggio 1970)

1. – The simple properties of deep inelastic electron-proton scattering has suggested models where these processes arise as interactions of virtual photons with an « elementary » component of the proton. These as yet unspecified elementary components of the proton have been given the name of « partons » by Feynman (1). The model has been studied by Bjorken and Paschos (2) and successively by Drell, Levy and Tung Mow Yan (3) who gave a field-theoretical treatment of the parton model, and were able to recover some of the experimentally observed properties of this process. In this letter we wish to extend the method of ref. (3) to the study of the total cross-section of electron-positron annihilation into hadrons.

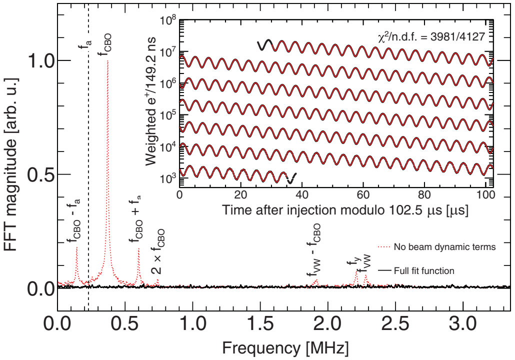
This treatment leads to an asymptotic (very high cross-section c.m. energy, $2E$) of the form

$$\sigma \to \frac{\pi\alpha^2}{12E^2}\left[\sum_{\text{spin }0}(Q_i)^2 + 4\sum_{\text{spin }\frac{1}{2}}(Q_i)^2\right], \qquad (1)$$

in the last few years, the importance of the R-ratio has been mainly associated with muon $g-2$ experiments

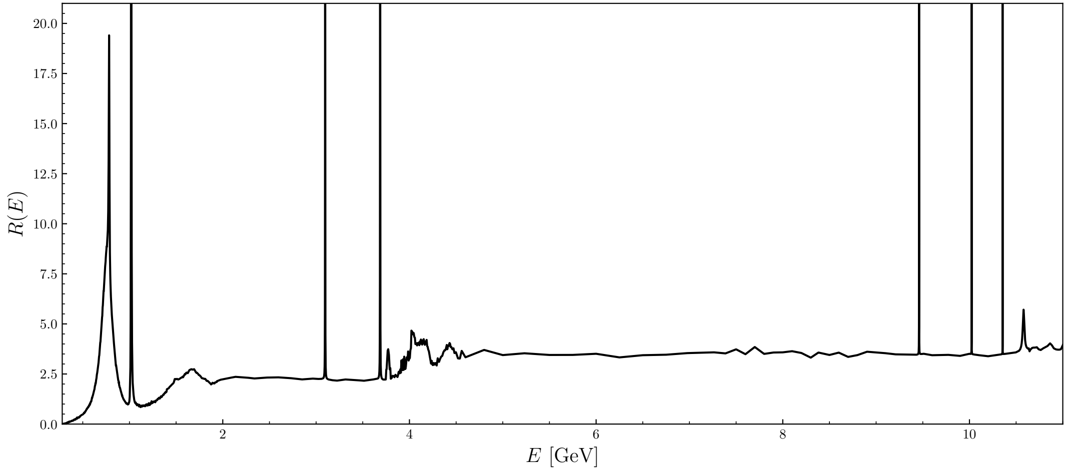
$$a_\mu^{\mathrm{HVP-LO}} = \int_0^{+\infty} d\omega \, f_{a_\mu}(\omega)\, R(\omega)$$

Munon $g-2$, PRL 131 (2023)



the R-ratio is much more than that. . .

KNT-19, PRD 101 (2020)



theoretically, $R(E)$ is a distribution

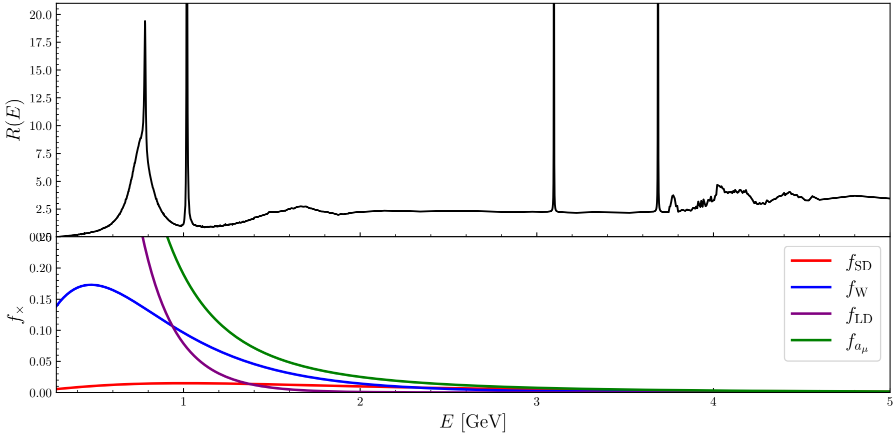
$$\langle 0|J^i_{\text{em}}(0)\delta(H-E)(2\pi)^3\delta^3(\boldsymbol{P})J^j_{\text{em}}(0)|0\rangle = -\frac{\delta^{ij}E^2}{12\pi^2}\,R(E)$$

theoretically, but also numerically (see later), it is convenient to define and study distributions as functionals

$$R(E) \quad \longmapsto \quad R[f] = \int_0^{+\infty} d\omega\, f(\omega)\, R(\omega)$$

and probe the energy dependence of $R(E)$ by changing $f(E)$

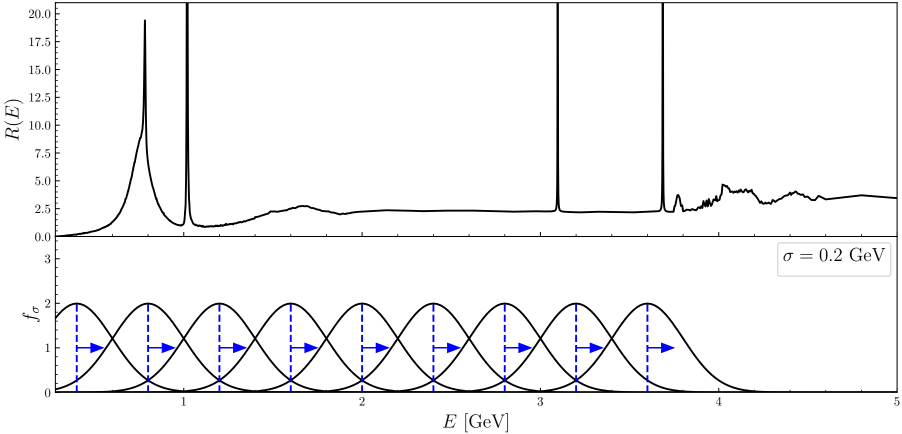
KNT-19, PRD 101 (2020)



$a_\mu^{\rm HVP-LO} \equiv R[f_{a_\mu}]$ can also be seen as a low-energy probe of $R(E)$

- $a_\mu^{\rm HVP-LO}$ is natively a low-energy observable, measured directly by letting muons wrap around in a magnetic field
- $R(E)$, as any other cross-section, is an energy-dependent probe of the theory and contains an infinite amount of information
- the experiments that measure $R(E)$ are totally different w.r.t. the ones that measure a_μ
- the comparison of the BNL+FNAL measurements, $a_\mu^{\rm HVP-LO,exp}$, with $\int_0^{+\infty} dE\, f_{a_\mu}(\omega)\, R^{\rm exp}(\omega)$ is a consistency check. . .
- to turn it into a first-principles test of the theory we must test $a_\mu^{\rm HVP-LO,exp}$ and $R^{\rm exp}(E)$ independently
- this can conveniently be done by using the unifying language

$$R[f] = \int_0^{+\infty} d\omega\, f(\omega)\, R(\omega)$$



a systematic study of $R(E)$ at different energies can be done by considering Gaussian energy bins

$$f(\omega) \equiv G_\sigma(E-\omega) = \frac{e^{-\frac{(E-\omega)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}\,, \qquad R_\sigma(E) \equiv R[f] = \int_0^{+\infty} d\omega\, G_\sigma(E-\omega)\, R(\omega)$$

PHYSICAL REVIEW D **99,** 094508 (2019)

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

- on the theoretical side, the required non-perturbative accuracy can be achieved by performing lattice simulations
- these give direct access to Euclidean correlators at finite L and a and necessarily affected by numerical and statistical errors
- the required information can be extracted by using the method that we developed to cope with this problem (that my friend j.bulava then called HLT)

- other methods are available on the market

a.rothkopf EPJ Web Conf. 274, 01004 (2022)

a.lupo and l.del.debbio talks

- and whenever i have a student named alessandro i devise a new one. . .

Eur. Phys. J. C (2024) 84:32
https://doi.org/10.1140/epjc/s10052-024-12399-0

THE EUROPEAN PHYSICAL JOURNAL C

Check for updates

Regular Article - Theoretical Physics

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

Michele Buzzicotti[a], **Alessandro De Santis**[b], **Nazario Tantalo**[c]

University and INFN of Roma Tor Vergata, Via della Ricerca Scientifica 1, 00133 Rome, Italy

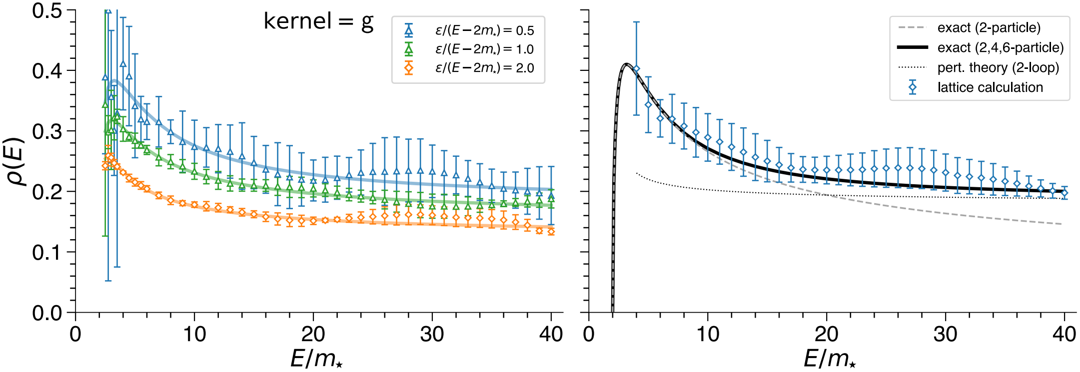
PHYSICAL REVIEW LETTERS **130,** 241901 (2023)

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))

- before addressing the phenomenologically relevant calculation of the smeared R-ratio on which i'm now going to focus. . .



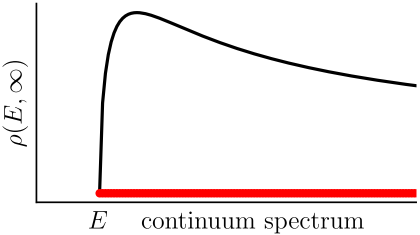
the HLT method has been stringently validated by performing the analogous calculation in the 2D non-linear $O(3)$ σ-model where the exact answer is known let's see how it works. . .

$$\rho(E) = \langle 0|\hat{O}_F\, \delta\left(\hat{H} - E\right)(2\pi)^3\delta^3\left(\hat{\boldsymbol{P}} - \boldsymbol{p}\right)\hat{O}_I|0\rangle$$

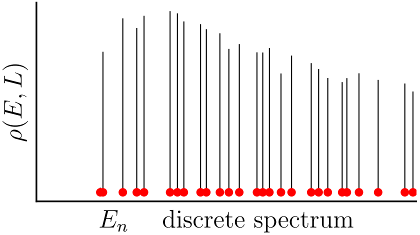
$$C(t) = \int \mathrm{d}^3x\, e^{-i\boldsymbol{p}\cdot\boldsymbol{x}}\langle 0|\hat{O}_F\, e^{-t\hat{H}+i\hat{\boldsymbol{P}}\cdot\boldsymbol{x}}\hat{O}_I|0\rangle = \int_{E_0}^{\infty}\mathrm{d}\omega\, e^{-t\omega}\,\rho(\omega)$$

$$t = a\tau\,, \qquad \tau = 1, \cdots, T$$

$$\rho(E) = \langle 0|\hat{O}_F\, \delta\left(\hat{H} - E\right)(2\pi)^3\delta^3\left(\hat{\boldsymbol{P}} - \boldsymbol{p}\right)\hat{O}_I|0\rangle$$



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



$$t = a\tau\,, \qquad \tau = 1,\cdots,T$$

$$f(\infty)=0\,,\qquad f(\omega)\in L^2_{\rm n}[E_0,\infty]$$

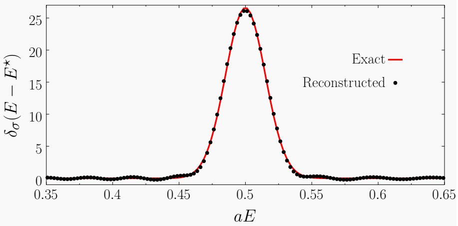
$$f(\omega)=\sum_{\tau=1}^{\infty} g_\tau\, e^{-\tau a\omega}\;,$$

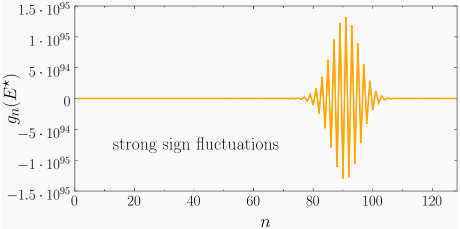
$$\rho[f]=\sum_{\tau=1}^{\infty} g_\tau\, C(a\tau)=\int_{E_0}^{\infty} d\omega\,\rho(\omega)\overbrace{\sum_{\tau=1}^{\infty} g_\tau\, e^{-\tau a\omega}}^{f(\omega)}$$

$$A_{\rm n}[\boldsymbol{g}]=\int_{E_0}^{\infty} d\omega\, w_{\rm n}(\omega)\left|f(\omega)-\sum_{\tau=1}^{T} g_\tau\, e^{-\tau a\omega}\right|^2$$

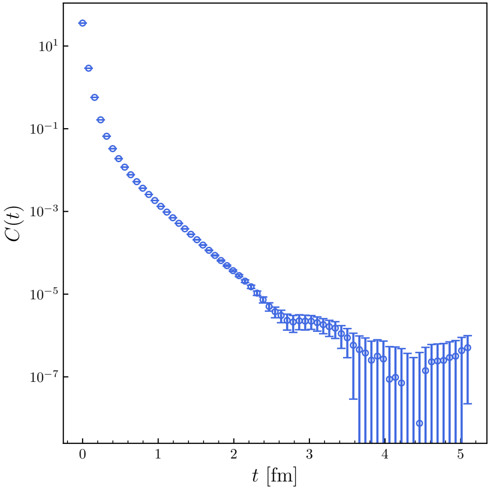
$$d_{\rm n}[\boldsymbol{g}]=\sqrt{\frac{A_{\rm n}[\boldsymbol{g}]}{A_{\rm n}[\boldsymbol{0}]}}$$

$aE_\star=0.5 \qquad \sigma=0.03E_\star \qquad T=128$





ID	$L^3 \times T$	a fm	aL fm	m_π GeV
B64	$64^3 \cdot 128$	0.07957(13)	5.09	0.1352(2)
B96	$96^3 \cdot 192$	0.07957(13)	7.64	0.1352(2)
C80	$80^3 \cdot 160$	0.06821(13)	5.46	0.1349(3)
D96	$96^3 \cdot 192$	0.05692(12)	5.46	0.1351(3)



in the case of the isoQCD R-ratio

$$C(t) = -\frac{1}{3}\sum_{i=1}^{3}\int d^3x\, T\langle 0|J_i(x)J_i(0)|0\rangle = \frac{1}{12\pi^2}\int_{2m_\pi}^{\infty} d\omega\, \omega^2 R(\omega)\, e^{-\omega t} = R\left[\frac{\omega^2\, e^{-\omega t}}{12\pi^2}\right]$$

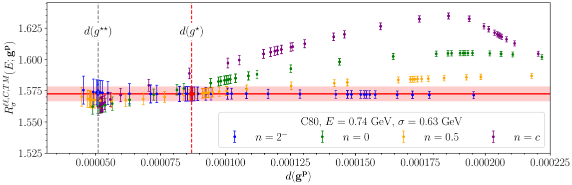
$$\Delta(\omega) = \frac{12\pi^2 G_\sigma(E-\omega)}{\omega^2}$$

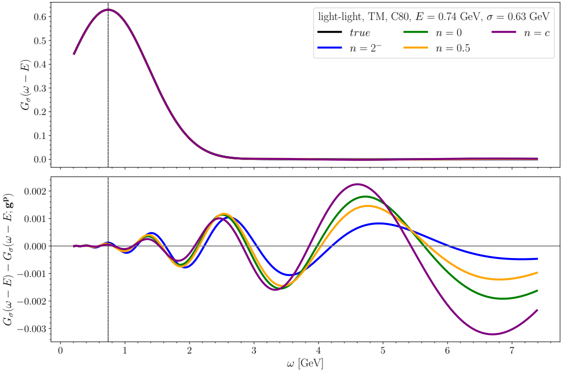
$$w_\alpha(\omega) = e^{\alpha a \omega} \ , \quad w_c(\omega) = \left(e^{a(\omega - 2m_\pi)} - 1\right)^{-\frac{1}{2}}$$

$$A_{\mathrm{n}}[\boldsymbol{g}] = \int_{2m_\pi}^{\infty} d\omega \, w_{\mathrm{n}}(\omega) \left| \Delta(\omega) - \sum_{\tau=1}^{T} g_\tau \, e^{-\tau a \omega} \right|^2$$

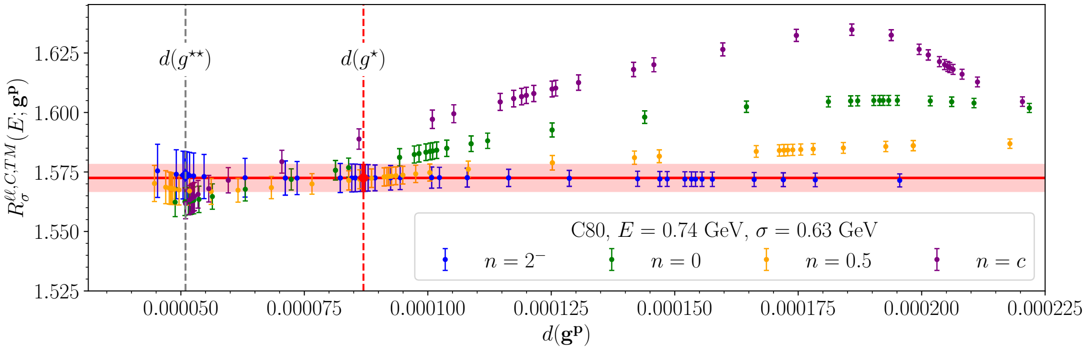
$$B[\boldsymbol{g}] = \sum_{\tau_{1,2}=1}^{T} g_{\tau_1} g_{\tau_2} \, \mathrm{Cov}_{\tau_1 \tau_2}$$

$$W_{\mathrm{n}}[\boldsymbol{g}] = A_{\mathrm{n}}[\boldsymbol{g}] + \lambda B[\boldsymbol{g}]$$

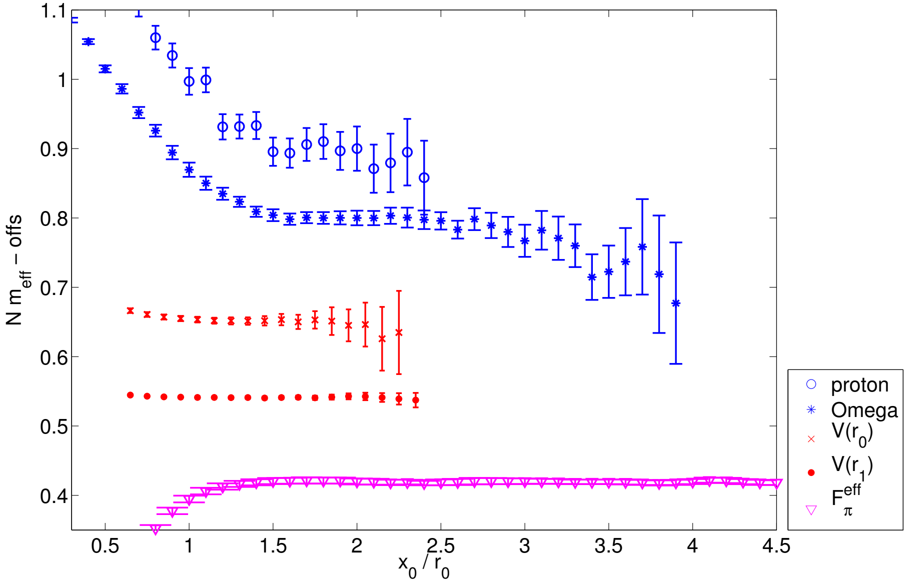




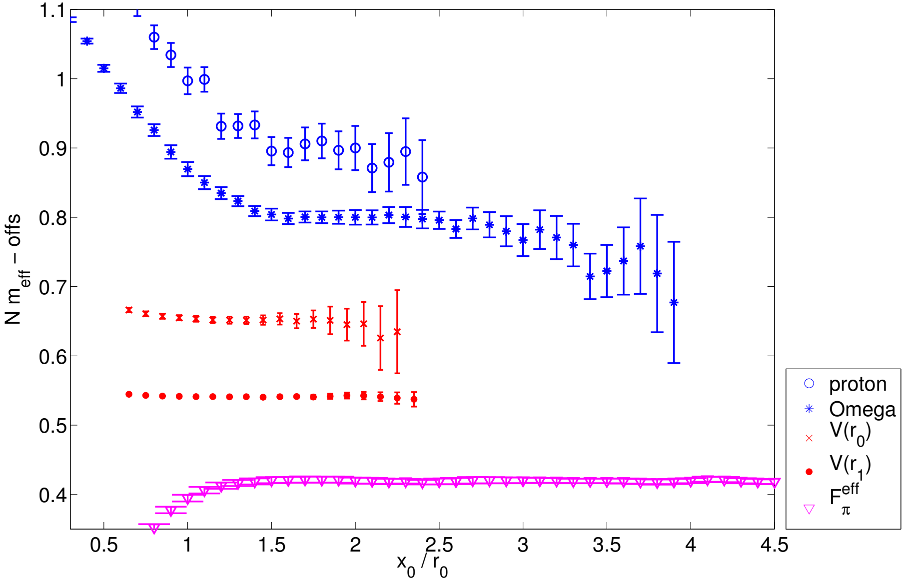
let’s have a closer look at the stability analysis. . .



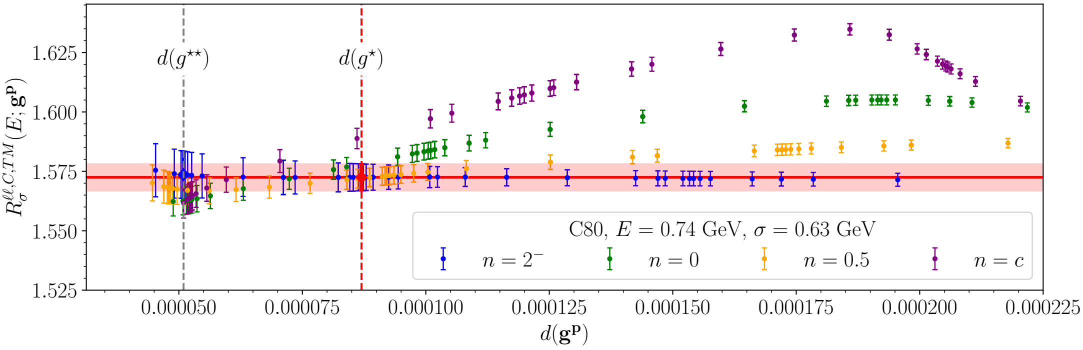
$d(g^{\star\star})$
$d(g^{\star})$
$R_{\sigma}^{\ell\ell,CTM}(E;\mathbf{g^p})$
$d(\mathbf{g^p})$
1.625
1.600
1.575
1.550
1.525
0.000050
0.000075
0.000100
0.000125
0.000150
0.000175
0.000200
0.000225
C80, $E = 0.74$ GeV, $\sigma = 0.63$ GeV
$n = 2^-$
$n = 0$
$n = 0.5$
$n = c$



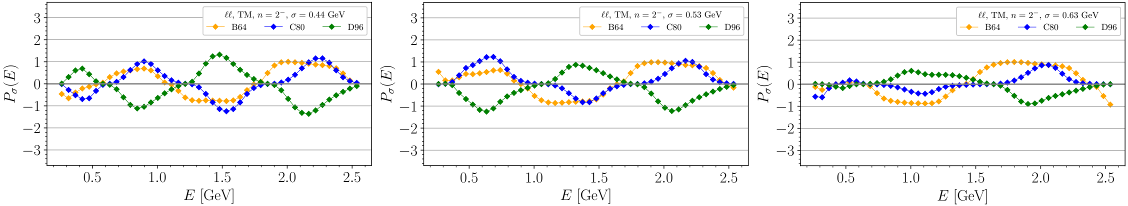
N m_{eff} – offs
x_0 / r_0
proton
Omega
$V(r_0)$
$V(r_1)$
F_π^{eff}
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.5
1
1.5
2
2.5
3
3.5
4
4.5



N m_eff − offs
x_0 / r_0
proton
Omega
V(r_0)
V(r_1)
F_π^eff

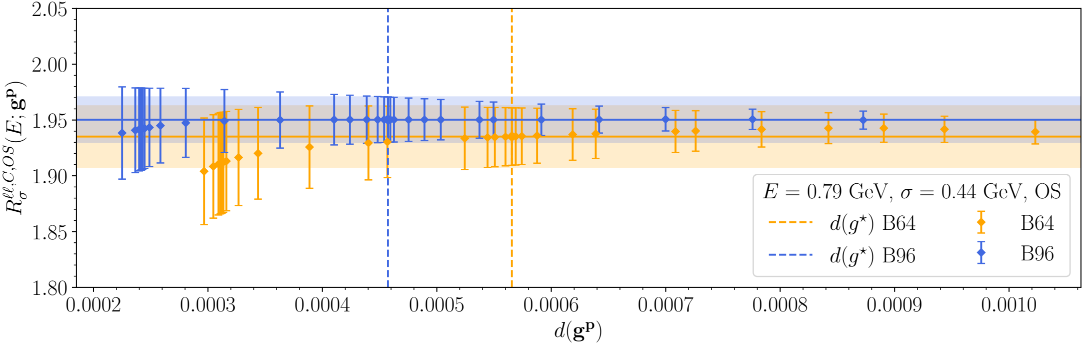


$$P_\sigma(E) = \frac{R_\sigma(E;\boldsymbol{g}^\star) - R_\sigma(E;\boldsymbol{g}^{\star\star})}{\Delta_\sigma^{\rm stat}(E;\boldsymbol{g}^{\star\star})}\,, \qquad \Delta_\sigma^{\rm rec}(E) = |R_\sigma(E;\boldsymbol{g}^\star) - R_\sigma(E;\boldsymbol{g}^{\star\star})|\,\mathrm{erf}\left(\frac{|P_\sigma(E)|}{\sqrt{2}}\right)$$



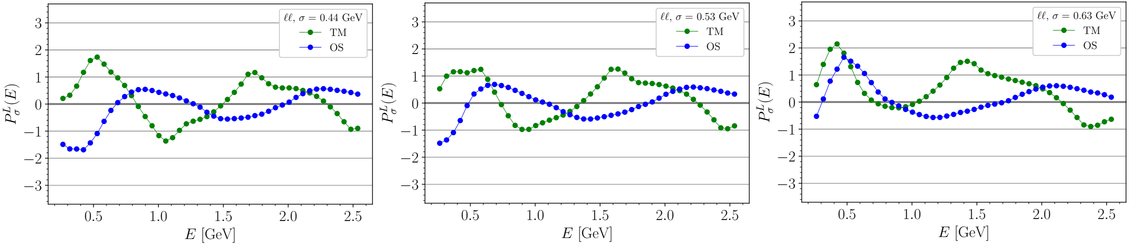
$$P_\sigma(E) = \frac{R_\sigma(E; \boldsymbol{g}^\star) - R_\sigma(E; \boldsymbol{g}^{\star\star})}{\Delta_\sigma^{\text{stat}}(E; \boldsymbol{g}^{\star\star})} \,, \qquad \Delta_\sigma^{\text{rec}}(E) = |R_\sigma(E; \boldsymbol{g}^\star) - R_\sigma(E; \boldsymbol{g}^{\star\star})| \, \text{erf}\left(\frac{|P_\sigma(E)|}{\sqrt{2}} \right)$$

let's look at finite volume effects. . .



$$P_\sigma^L(E) = \frac{R_\sigma(E;\frac{3L}{2}) - R_\sigma(E;L)}{\sqrt{\bar{\Delta}_\sigma(E;\frac{3L}{2})^2 + \bar{\Delta}_\sigma(E;L)^2}} ,$$

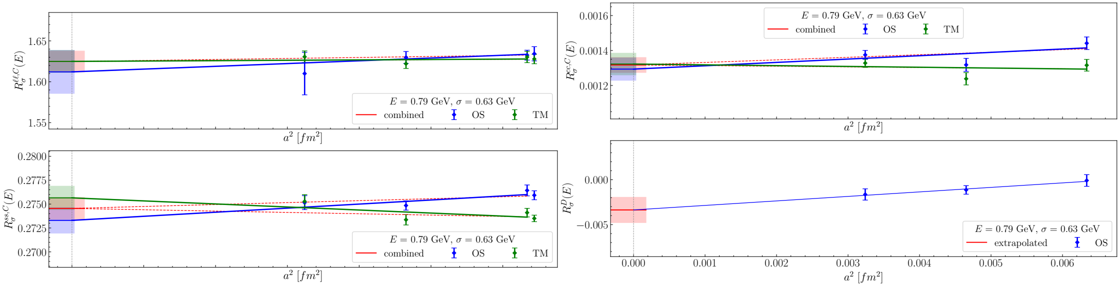
$$\Delta_\sigma^L(E) = \max_{OS,TM} \left\{ \left| R_\sigma(E;\frac{3L}{2}) - R_\sigma(E;L) \right| \mathrm{erf}\left(\frac{|P_\sigma^L(E)|}{\sqrt{2}} \right) \right\} , \qquad L \simeq 5.1 \text{ fm}$$



$$P^L_\sigma(E) = \frac{R_\sigma(E; \frac{3L}{2}) - R_\sigma(E; L)}{\sqrt{\bar{\Delta}_\sigma(E; \frac{3L}{2})^2 + \bar{\Delta}_\sigma(E; L)^2}} \ ,$$

$$\Delta^L_\sigma(E) = \max_{OS,TM} \left\{ \left| R_\sigma(E; \frac{3L}{2}) - R_\sigma(E; L) \right| \mathrm{erf}\left(\frac{\left| P^L_\sigma(E) \right|}{\sqrt{2}} \right) \right\} \ , \qquad L \simeq 5.1 \text{ fm}$$

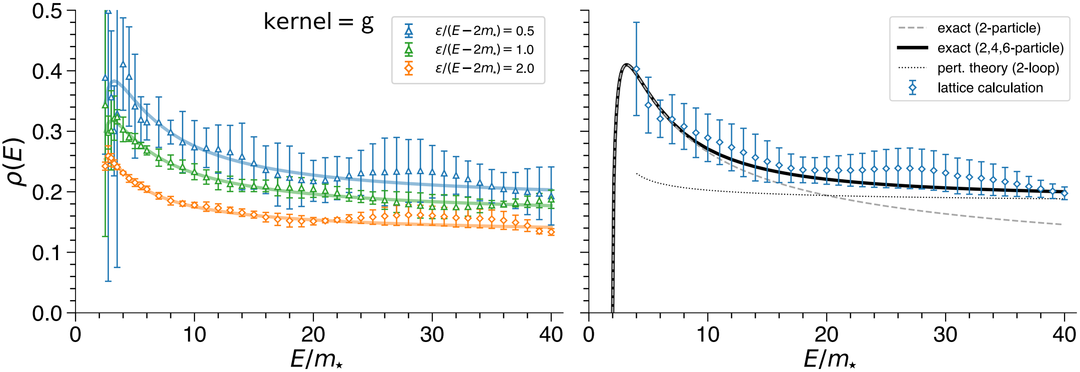
let's look at cutoff effects. . .



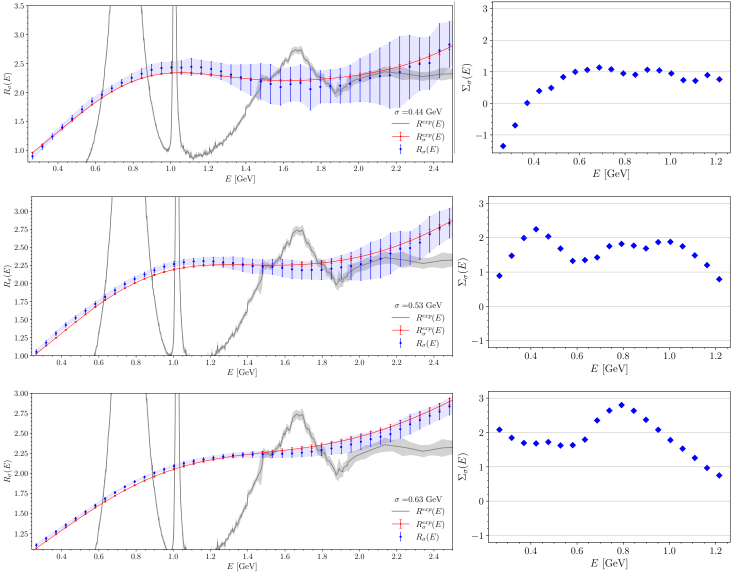
$E = 0.79$ GeV, $\sigma = 0.63$ GeV
combined
OS
TM
$R_\sigma^{\ell\ell,C}(E)$
1.65
1.60
1.55
$a^2\ [fm^2]$
$R_\sigma^{ss,C}(E)$
0.2800
0.2775
0.2750
0.2725
0.2700
$R_\sigma^{cc,C}(E)$
0.0016
0.0014
0.0012
$R_\sigma^{D}(E)$
0.000
−0.005
extrapolated
0.000
0.001
0.002
0.003
0.004
0.005
0.006

once you have all this. . .

and you have already seen this. . .



you can trust this:



m.davier et al., arXiv:2312.02053

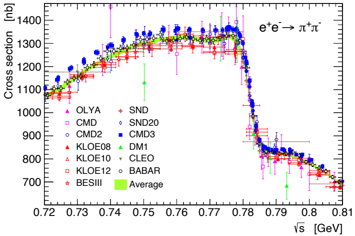
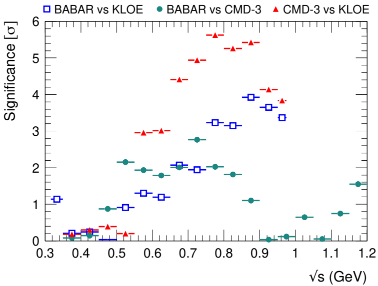
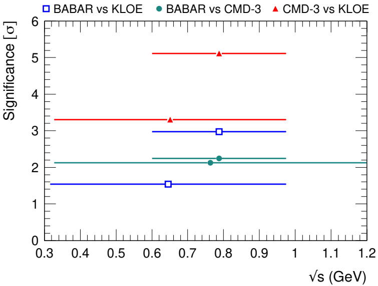


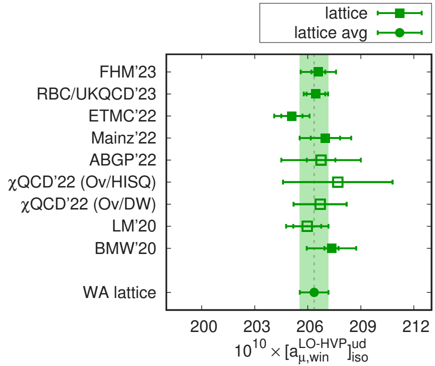
Fig. 1. Bare $e^+e^- \to \pi^+\pi^-$ cross section versus centre-of-mass energy in the ρ peak region. The error bars of the data points include statistical and systematic uncertainties added in quadrature. The green band shows the HVPTools combination within its 1σ uncertainty.

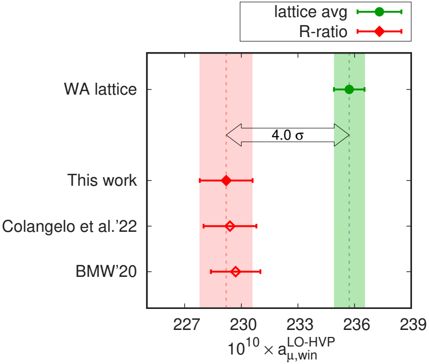




and then you get this!

m.davier et al., PRD 109 (2024), see the paper for the original lattice references



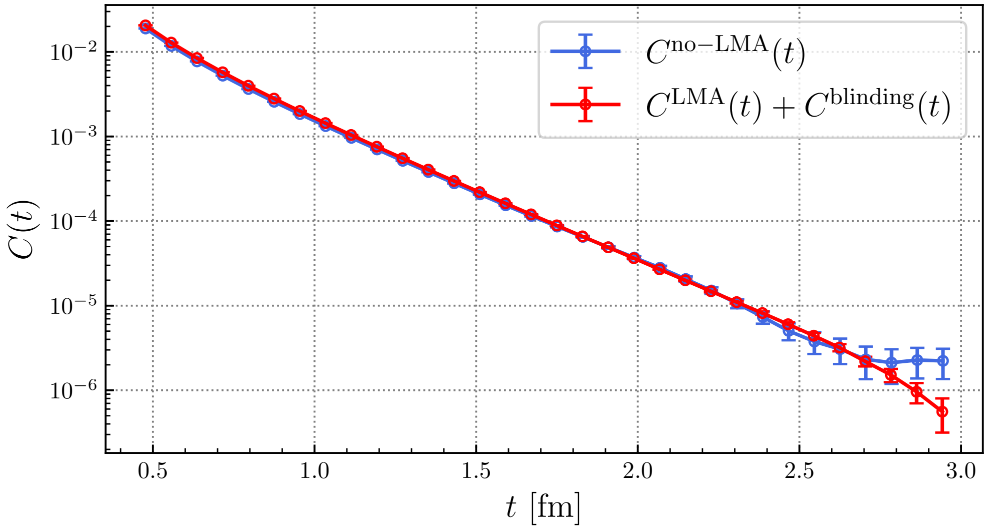


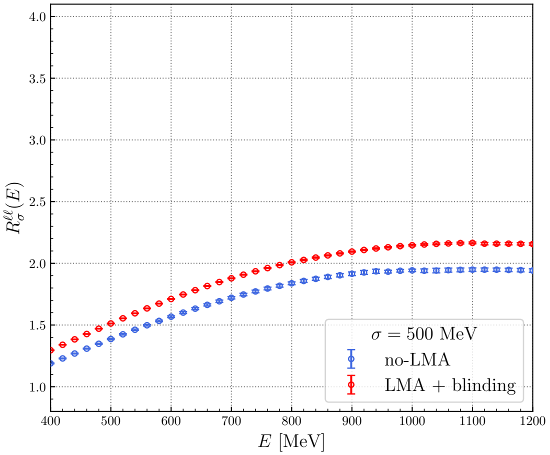
at present, since:

- the tension is at low-energy
- our lattice errors on $R_\sigma(E)$ are still rather large for $E > 1.5$ GeV

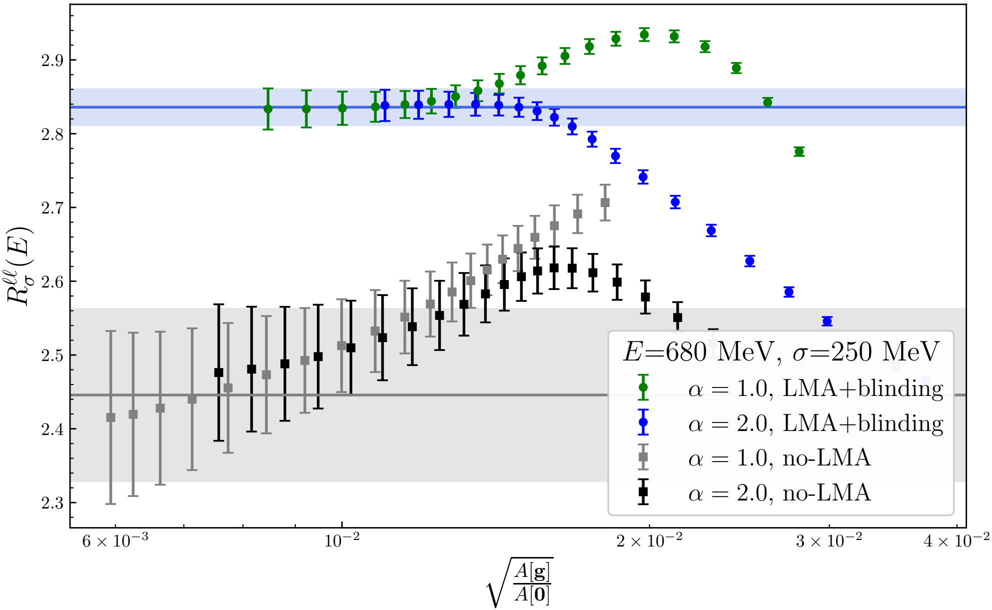
$R[f_{a_\mu}^W]$ makes the same job!

is there any chance to reduce the Gaussian bin size?

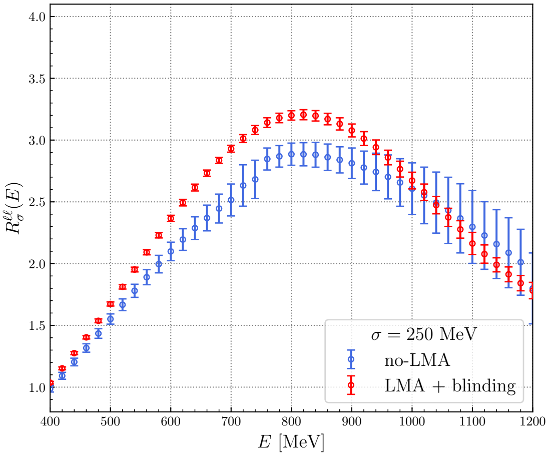

$C^{\text{no-LMA}}(t)$
$C^{\text{LMA}}(t) + C^{\text{blinding}}(t)$
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
$C(t)$
0.5
1.0
1.5
2.0
2.5
3.0
t [fm]



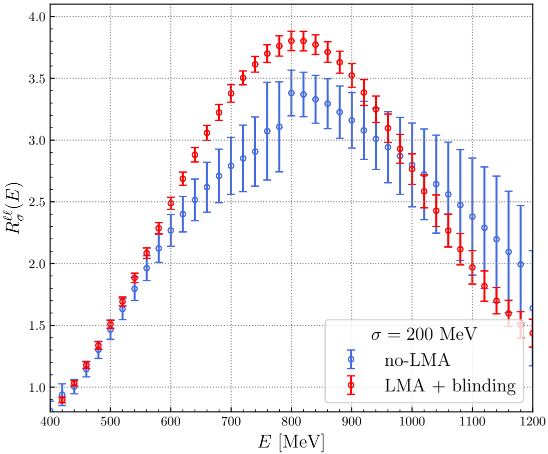
4.0
3.5
3.0
2.5
2.0
1.5
1.0
$R_\sigma^{\ell\ell}(E)$
400
500
600
700
800
900
1000
1100
1200
E [MeV]
$\sigma = 500$ MeV
no-LMA
LMA + blinding



2.9
2.8
2.7
2.6
2.5
2.4
2.3
$R_\sigma^{\ell\ell}(E)$
6×10^{-3}
10^{-2}
2×10^{-2}
3×10^{-2}
4×10^{-2}
$\sqrt{\frac{A[\mathbf{g}]}{A[\mathbf{0}]}}$
E=680 MeV, σ=250 MeV
$\alpha = 1.0$, LMA+blinding
$\alpha = 2.0$, LMA+blinding
$\alpha = 1.0$, no-LMA
$\alpha = 2.0$, no-LMA



$R^{\ell\ell}_{\sigma}(E)$
4.0
3.5
3.0
2.5
2.0
1.5
1.0
400
500
600
700
800
900
1000
1100
1200
E [MeV]
σ = 250 MeV
no-LMA
LMA + blinding



4.0
3.5
3.0
2.5
2.0
1.5
1.0
$R_\sigma^{\ell\ell}(E)$
400
500
600
700
800
900
1000
1100
1200
E [MeV]
σ = 200 MeV
no-LMA
LMA + blinding

let's go back to a_μ

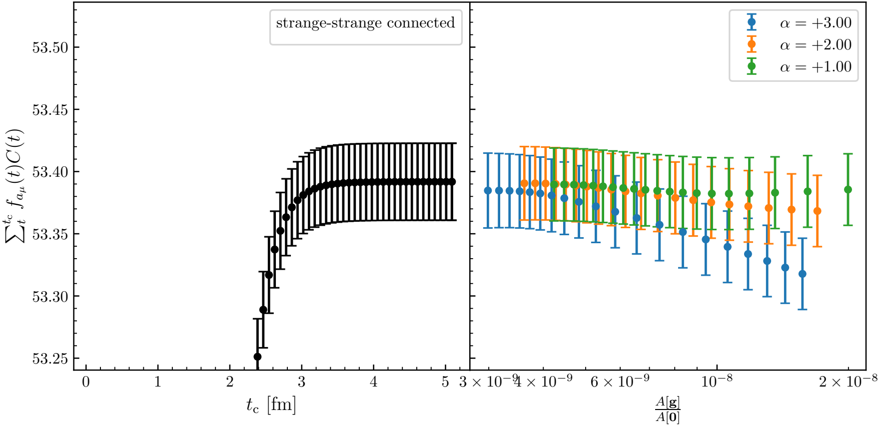
on the lattice this is estimated by

$$a_\mu^{\rm HVP-LO,standard} = \lim_{a\mapsto 0} \lim_{L\mapsto\infty} \lim_{t_c\mapsto\infty} a \sum_{t=0}^{t_c} \tilde{f}_{a_\mu}(at)\, C(at)$$

the big difference w.r.t. a Gaussian kernel is that the coefficients $\tilde{f}_{a_\mu}(at)$ do not require smoothing procedures

on the other hand, by using HLT one has a different estimator with different systematic errors

$$a_\mu^{\rm HVP-LO,HLT} = \lim_{a\mapsto 0} \lim_{L\mapsto\infty} \lim_{\lambda\mapsto 0} \int_0^\infty d\omega\, f_{a_\mu}^{HLT}(\omega)\, R^{a,L}(\omega)$$



strange-strange connected
53.50
53.45
53.40
53.35
53.30
53.25
$\sum_t^{t_c} f_{a_\mu}(t)C(t)$
0
1
2
3
4
5
t_c [fm]
$\alpha = +3.00$
$\alpha = +2.00$
$\alpha = +1.00$
3×10^{-9}
4×10^{-9}
6×10^{-9}
10^{-8}
2×10^{-8}
$\frac{A[\mathbf{g}]}{A[\mathbf{0}]}$

summarizing:

$$R[f] = \int_0^{+\infty} d\omega\, f(\omega)\, R(\omega)$$

- from the lattice perspective the best is $f(\omega) = e^{-\omega t}$
- from the phenomenological perspective one would like to have $f(\omega) = \delta(\omega - E)$
- let's do what we can with reasonable and, more importantly, trustable errors

before closing i'll take the chance to express my feelings concerning this spectral density business. . .

here (ALGT@CERN 2019 workshop) I presented, during an informal afternoon discussion, the HLT method that, at the time, was seen as a speculative idea. . .

that idea, together with many other important contributions, opened the way to the calculation of inclusive hadronic decay rates on the lattice

m.hansen et al. Phys.Rev.D 96 (2017)
s.hashimoto PTEP 2017 (2017)
p.gambino, s.hashimoto Phys.Rev.Lett. 125 (2020)
p.gambino et al. JHEP 07 (2022) 083
ETMC Phys.Rev.D 108 (2023)
ETMC Phys.Rev.Lett. 130 (2023)
a.barone et al. JHEP 07 (2023) 145
ETMC Phys.Rev.Lett. 132 (2024)

see the talks from
f.sanfilippo, p.gambino, a.barone, s.hashimoto

something that has been considered unfeasible for several years. . .

hic et nunc, agostino should be presenting now, hopefully will present in a couple of days, what I really think is an important theoretical step forward

a mathematically solid non-perturbative solution to the theoretical problem of extracting generic scattering amplitudes from lattice correlators

from the numerical perspective, this must still be seen as a speculative idea. . .

Scattering Amplitudes from Euclidean Correlators: Haag-Ruelle theory and approximation formulae

Agostino Patella[a,d], Nazario Tantalo[b,c]

[a] Humboldt Universität zu Berlin, Institut für Physik & IRIS Adlershof, Zum Grossen Windkanal 6, 12489 Berlin, Germany

[b] Università di Roma Tor Vergata, Dipartimento di Fisica, Via della Ricerca Scientifica 1, 00133 Rome, Italy

[c] INFN, Sezione di Tor Vergata, Via della Ricerca Scientifica 1, 00133 Rome, Italy

[d] DESY, Platanenallee 6, D-15738 Zeuthen, Germany

Abstract

In this work we provide a non-perturbative solution to the theoretical problem of extracting scattering amplitudes from Euclidean correlators in infinite volume. We work within the solid axiomatic framework of the Haag-Ruelle scattering theory and derive formulae which can be used to approximate scattering amplitudes arbitrarily well in terms of linear combinations of Euclidean correlators at discrete time separations. Our result generalizes and extends the range of applicability of a result previously obtained by Barata and Fredenhagen [1]. We provide a concrete procedure to construct such approximations, making our formulae ready to be used in numerical calculations of non-perturbative QCD scattering amplitudes. A detailed numerical investigation is needed to assess whether the proposed strategy can lead to the calculation of scattering amplitudes with phenomenologically satisfactory precision with presently available lattice QCD data. This will be the subject of future work. Nevertheless, the numerical accuracy and precision of lattice simulations is systematically improvable, and we have little doubts that our approach will become useful in the future.

i'm optimistic though. . .

and i'll take inspiration from what chis said at the end of his plenary talk in Villasimius where he presented the RBC-UKQCD results on $K \mapsto \pi\pi$

"if I were Italian I would be jumping for joy on stage!"

c.t.sachrajda @ lattice2010

who works with me knows that i'm all but a joyful man. . .

but i’m deeply italian!

backup slides

mathematically the problem is reduced to that of an inverse Laplace-transform mathematically the problem is reduced to that of an inverse Laplace-transform

to be performed numerically
by starting from a finite and noisy set of input data

$$\rho(E) = \langle 0|\hat{O}_F\, \delta\left(\hat{H} - E\right)(2\pi)^3\delta^3\left(\hat{\boldsymbol{P}} - \boldsymbol{p}\right)\hat{O}_I|0\rangle$$

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

$$t = a\tau\ , \qquad \tau = 1, \cdots, \frac{T}{a}$$

$\rho(E)$ contains an infinite amount of information

$\rho(E)$ contains an infinite amount of information

the problem, to be addressed numerically, has to be discretized

$$C(a\tau) = \int_{E_0}^{\infty} \mathrm{d}E \, e^{-\tau a E} \, \rho(E) \quad \mapsto \quad \sigma \sum_{m=0}^{N_E - 1} e^{-\tau a E_m} \, \hat{\rho}(E_m)$$

$$C(a\tau) = \int_{E_0}^{\infty} \mathrm{d}E \, e^{-\tau a E} \, \rho(E) \quad \mapsto \quad \sigma \sum_{m=0}^{N_E - 1} e^{-\tau a E_m} \, \hat{\rho}(E_m)$$

$$\hat{\mathcal{E}}_{\tau m} \equiv e^{-\tau a E_m} \; , \qquad \hat{G}_{nm} \equiv \left[\hat{\mathcal{E}}^T \hat{\mathcal{E}} \right]_{nm}$$

$$C(a\tau) = \int_{E_0}^{\infty} \mathrm{d}E \, e^{-\tau a E} \, \rho(E) \quad \mapsto \quad \sigma \sum_{m=0}^{N_E - 1} e^{-\tau a E_m} \, \hat{\rho}(E_m)$$

$$\hat{G} \, \hat{\boldsymbol{\rho}} = \frac{1}{\sigma} \hat{\mathcal{E}}^T \, \boldsymbol{C} \, , \qquad \hat{\boldsymbol{\rho}} = \frac{1}{\sigma} \hat{G}^{-1} \, \hat{\mathcal{E}}^T \, \boldsymbol{C}$$

$$C(a\tau) = \int_{E_0}^{\infty} \mathrm{d}E\, e^{-\tau a E}\, \rho(E) \quad \mapsto \quad \sigma \sum_{m=0}^{N_E - 1} e^{-\tau a E_m}\, \hat{\rho}(E_m)$$

$$\hat{G}\, \hat{\boldsymbol{\rho}} = \frac{1}{\sigma} \hat{\mathcal{E}}^T\, \boldsymbol{C}\ , \qquad \hat{\boldsymbol{\rho}} = \frac{1}{\sigma} \hat{G}^{-1}\, \hat{\mathcal{E}}^T\, \boldsymbol{C}$$

$$g_\tau(E_n) = \frac{1}{\sigma} \sum_{m=0}^{N_E - 1} \hat{G}^{-1}_{nm} \hat{\mathcal{E}}_{\tau m}\ , \qquad K(E_n, \omega) = \sum_{\tau=1}^{N_T - 1} g_\tau(E_n)\, e^{-a\tau\omega}$$

$$C(a\tau) = \int_{E_0}^{\infty} \mathrm{d}E\, e^{-\tau a E}\, \rho(E) \quad \mapsto \quad \sigma \sum_{m=0}^{N_E - 1} e^{-\tau a E_m}\, \hat{\rho}(E_m)$$

$$\hat{G}\, \hat{\boldsymbol{\rho}} = \frac{1}{\sigma} \hat{\mathcal{E}}^T\, \boldsymbol{C}\,, \qquad \hat{\boldsymbol{\rho}} = \frac{1}{\sigma} \hat{G}^{-1}\, \hat{\mathcal{E}}^T\, \boldsymbol{C}$$

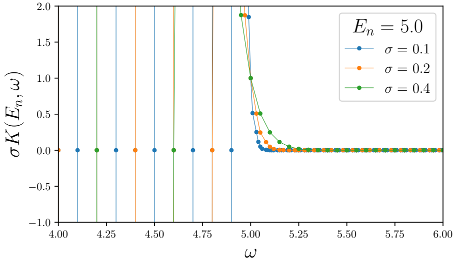
$$\hat{\rho}(E_n) = \sum_{\tau=1}^{N_T - 1} g_\tau(E_n)\, C(a\tau) = \int_{E_0}^{\infty} \mathrm{d}\omega\, K(E_n, \omega)\, \rho(\omega)$$

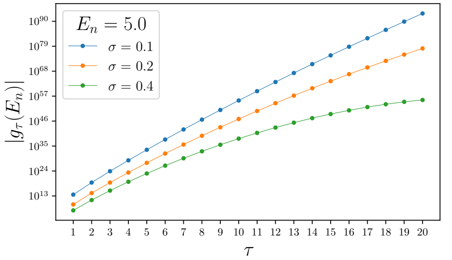
$$C(a\tau) = \int_{E_0}^{\infty} \mathrm{d}E\, e^{-\tau a E}\, \rho(E) \quad \mapsto \quad \sigma \sum_{m=0}^{N_E-1} e^{-\tau a E_m}\, \hat{\rho}(E_m)$$

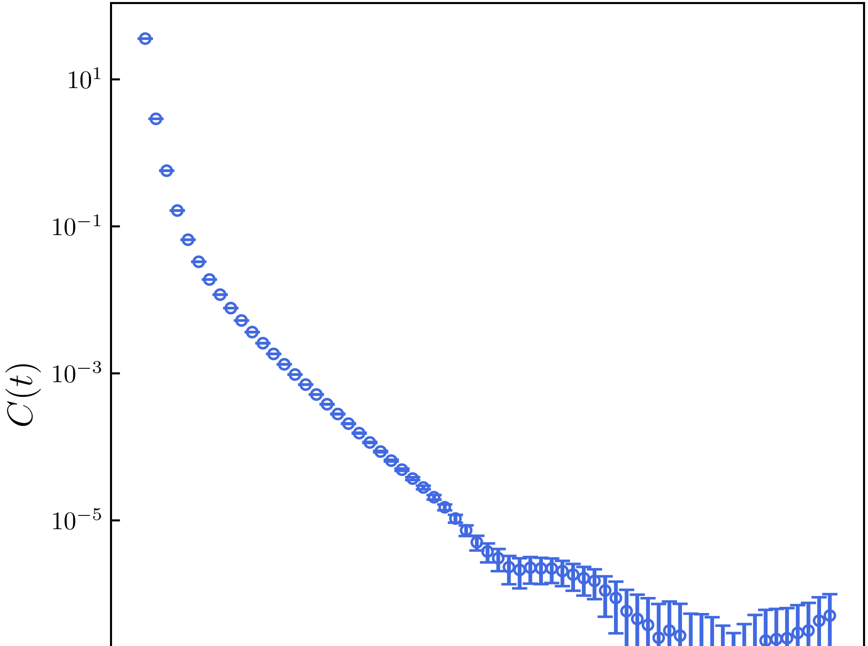
$$\hat{G}\,\hat{\boldsymbol{\rho}} = \frac{1}{\sigma}\hat{\mathcal{E}}^T \boldsymbol{C}\,, \qquad \hat{\boldsymbol{\rho}} = \frac{1}{\sigma}\hat{G}^{-1}\,\hat{\mathcal{E}}^T\,\boldsymbol{C}$$

$$\hat{\rho}(E_n) = \sum_{\tau=1}^{N_T-1} g_\tau(E_n)\, C(a\tau) = \int_{E_0}^{\infty} \mathrm{d}\omega\, K(E_n,\omega)\,\rho(\omega)$$

$$K(E_n, E_m) = \frac{\delta_{nm}}{\sigma}\,, \qquad K(E_n,\omega) \neq \delta(E_n - \omega)$$







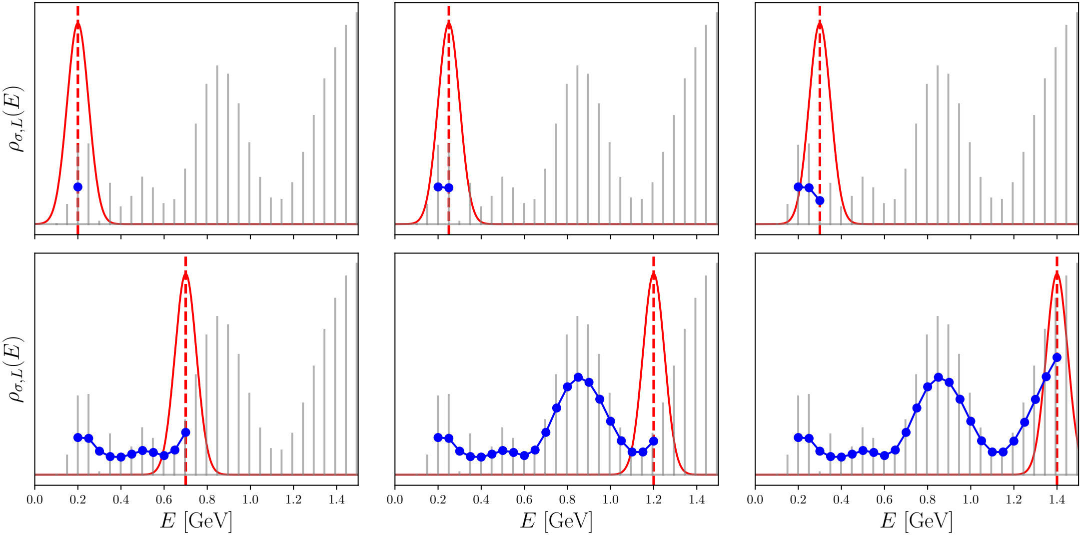
$C(t)$
10^{1}
10^{-1}
10^{-3}
10^{-5}

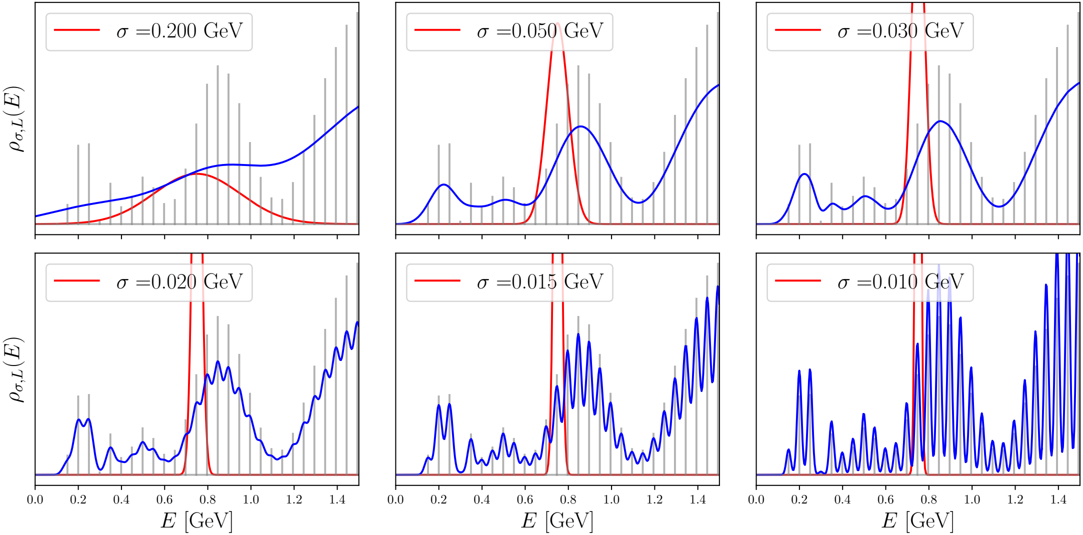
Axiom W1: For each test function f, i.e. for a function with a compact support and continuous derivatives of any order, there exists a set of operators $O_1(f), \cdots, O_n(f)$ which, together with their adjoints, are defined on a dense subset of the Hilbert state space, containing the vacuum. **The fields O are operator-valued tempered distributions**. The Hilbert state space is spanned by the field polynomials acting on the vacuum (cyclicity condition).

spectral densities **must be smeared**, in particular on finite volumes where

$$\rho_L(E) = \sum_n w_n(L)\, \delta\left(E_n(L) - E\right) \ , \qquad\qquad \rho(E) \mapsto \rho[\Delta]$$

$$\rho_{\sigma,L}(E) = \int_0^\infty \mathrm{d}\omega\ \Delta_\sigma(\omega, E)\rho_L(\omega) \qquad \Delta_\sigma(\omega, E) = G_\sigma(\omega, E) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(E-\omega)^2}{2\sigma^2}\right)$$





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

THE ORDER OF THE LIMITS IS IMPORTANT

PHYSICAL REVIEW D **99,** 094508 (2019)

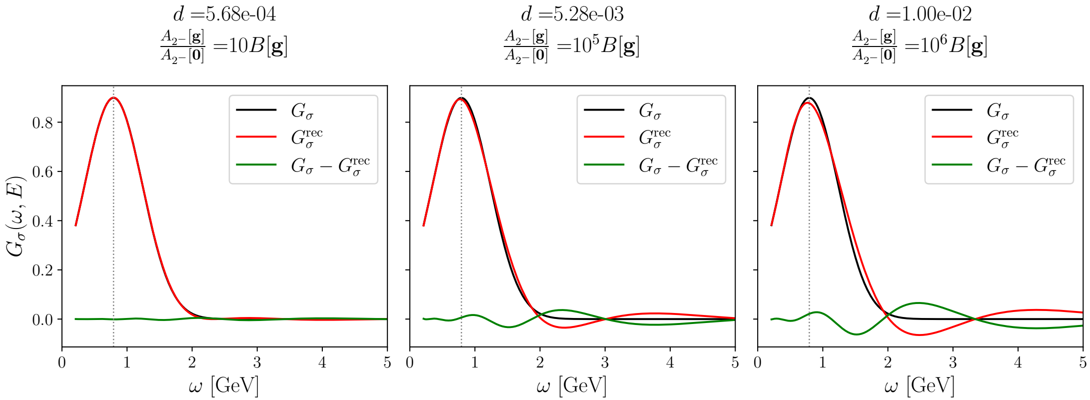
Extraction of spectral densities from lattice correlators

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*

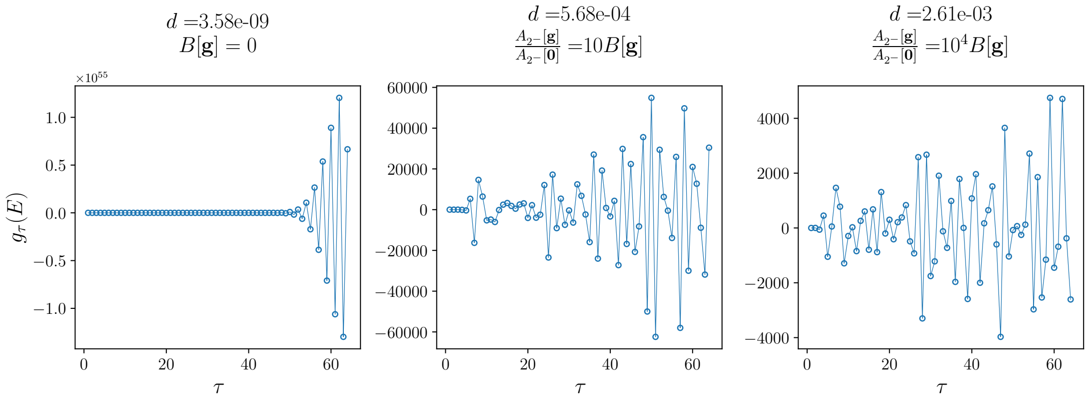
having this in mind, we developed a method (that my friend j.bulava then called HLT) that allows to extract **smeared spectral densities** from lattice correlators

$$\hat{\rho}_\sigma(E;L) = \int_{E_0}^{\infty} \mathrm{d}\omega\, \Delta_\sigma(E,\omega)\rho_L(\omega)\ , \qquad \rho(E) \stackrel{?}{=} \lim_{\sigma\mapsto 0}\lim_{L\mapsto\infty} \hat{\rho}_\sigma(E;L)$$

B64, $E = 0.79$ GeV, $\sigma = 0.44$ GeV



B64, $E = 0.79$ GeV, $\sigma = 0.44$ GeV



./picse/B64stability

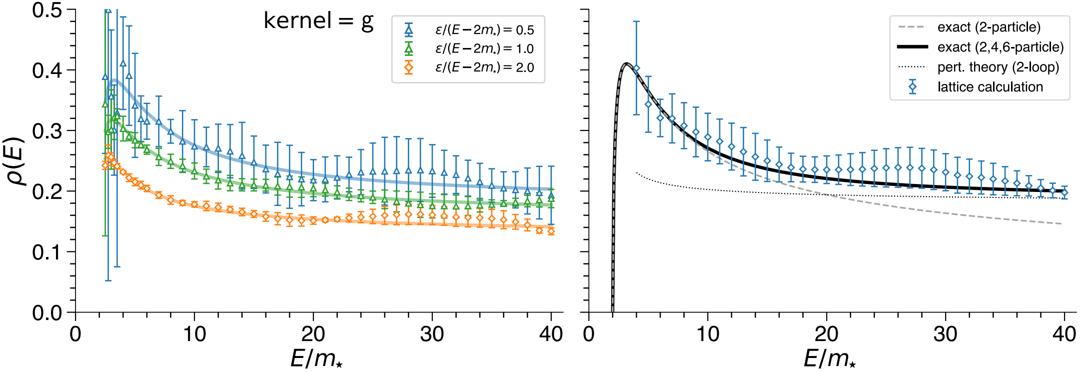
./picse/B64stability

- we didn't know it (again john) but the mathematics of the HLT method was already known
  f.pijpers, m.thompson Astron.Astrophys. 262 (1992)

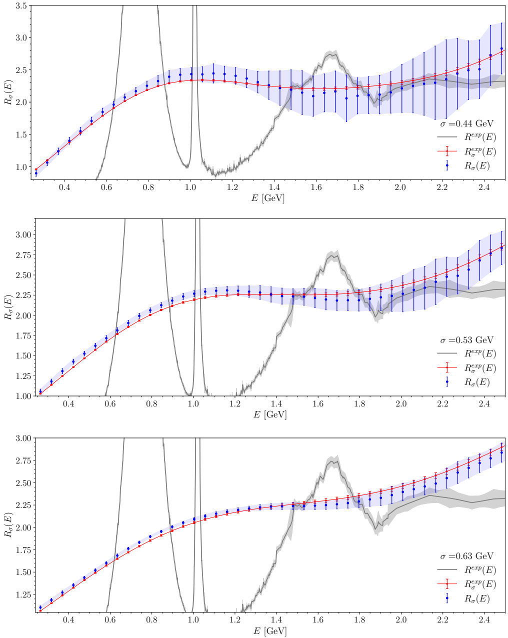
- it is a generalization of the Backus-Gilbert method that I learnt by reading
  m.t.hansen, h.b.meyer, d.robaina Phys.Rev.D 96 (2017) 9
  g.backus, f.gilbert Geophys.J.Int. 16 (1968) 169

- moreover, the HLT method can also be interpreted within the Bayesian language of the Gaussian Processes
  a.valentine, m.sambridge, Geophys.J.Int. 220 (2020), ETMC Phys.Rev.Lett. 130 (2023)

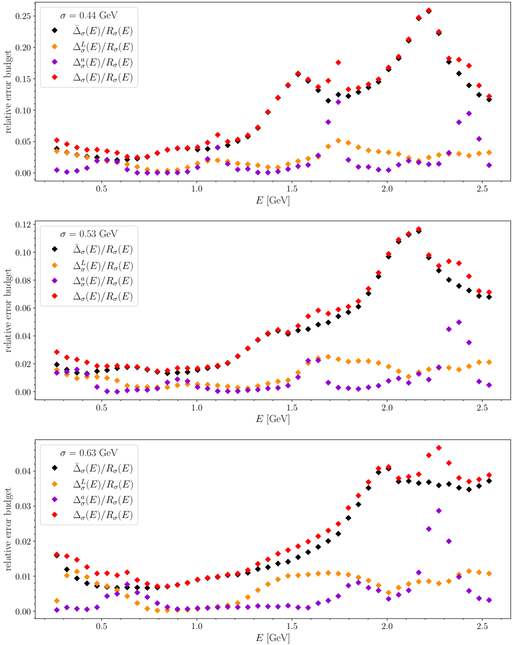
- what we call **the HLT method is the procedure to estimate reliably the errors** j.bulava et al. JHEP 07 (2022) 034



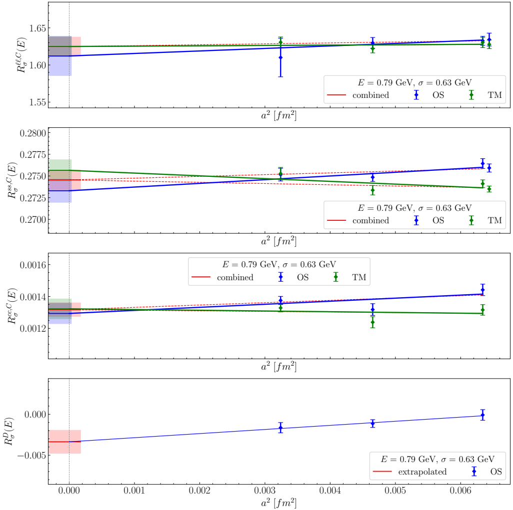
ETMC Phys.Rev.Lett. 130 (2023)



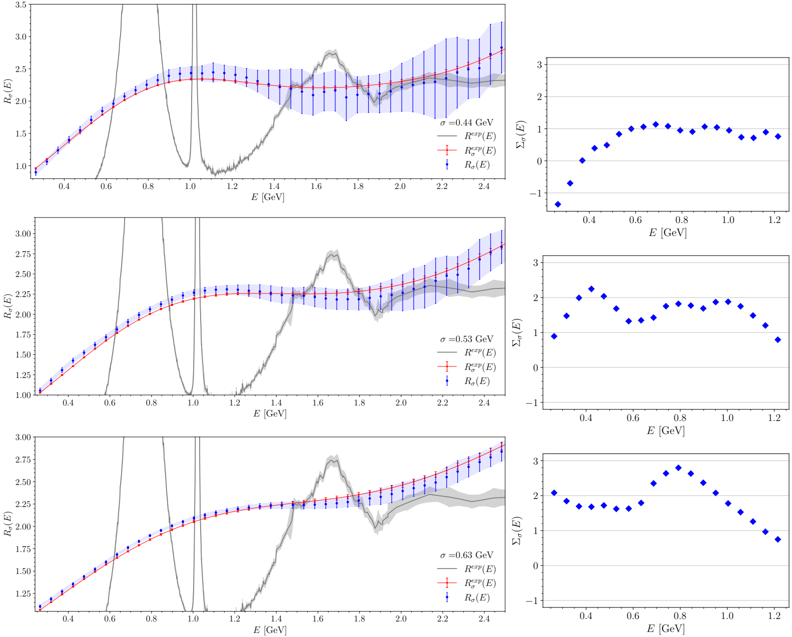
ETMC Phys.Rev.Lett. 130 (2023)



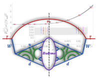
ETMC Phys.Rev.Lett. 130 (2023)



ETMC Phys.Rev.Lett. 130 (2023)



$$R_{ud}(\sigma) = 12\pi S_{EW} |V_{ud}|^2 \int_0^\infty \frac{dE E^2}{m_\tau^3} \left\{ K_T^\sigma\left(\frac{E}{m_\tau}\right) \rho_T(E^2) + K_L^\sigma\left(\frac{E}{m_\tau}\right) \rho_L(E^2) \right\}$$



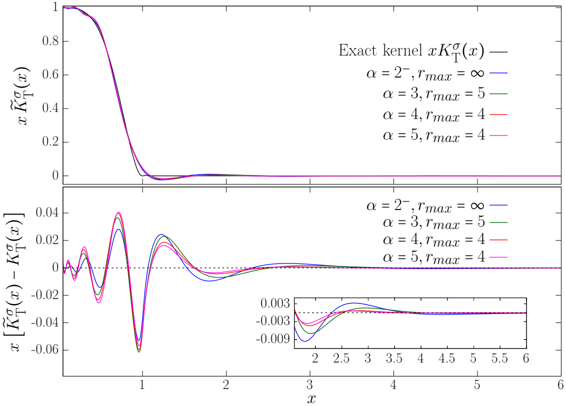
EDITORS' SUGGESTION

Inclusive hadronic decay rate of the τ lepton from lattice QCD

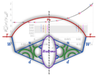
The authors express the inclusive hadronic decay rate of the tau lepton as an integral over the spectral density of the two-point correlator of the weak $V - A$ hadronic current which they compute fully nonperturbatively in lattice QCD. In a lattice QCD computation with all systematic errors except for isospin breaking effects under control, they then obtain the CKM matrix element V_{ud} with subpercent errors showing that their nonperturbative method can become a viable alternative to superallowed nuclear beta decays for obtaining V_{ud}.

A. Evangelista *et al.*

Phys. Rev. D **108**, 074513 (2023)



$$R_{ud}(\sigma) = 12\pi S_{EW} |V_{ud}|^2 \int_0^\infty \frac{dE E^2}{m_\tau^3} \left\{ K_T^\sigma \left(\frac{E}{m_\tau} \right) \rho_T(E^2) + K_L^\sigma \left(\frac{E}{m_\tau} \right) \rho_L(E^2) \right\}$$



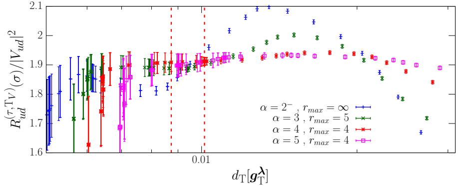
EDITORS' SUGGESTION

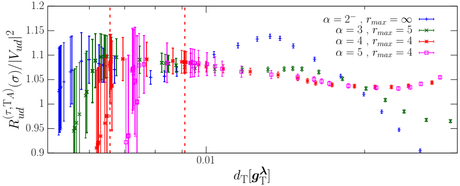
Inclusive hadronic decay rate of the τ lepton from lattice QCD

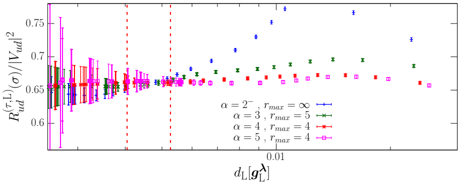
The authors express the inclusive hadronic decay rate of the tau lepton as an integral over the spectral density of the two-point correlator of the weak $V - A$ hadronic current which they compute fully nonperturbatively in lattice QCD. In a lattice QCD computation with all systematic errors except for isospin breaking effects under control, they then obtain the CKM matrix element V_{ud} with subpercent errors showing that their nonperturbative method can become a viable alternative to superallowed nuclear beta decays for obtaining V_{ud}.

A. Evangelista *et al.*

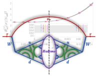
Phys. Rev. D **108**, 074513 (2023)







$$R_{ud}(\sigma) = 12\pi S_{EW} |V_{ud}|^2 \int_0^\infty \frac{dE E^2}{m_\tau^3} \left\{ K_T^\sigma \left(\frac{E}{m_\tau} \right) \rho_T(E^2) + K_L^\sigma \left(\frac{E}{m_\tau} \right) \rho_L(E^2) \right\}$$

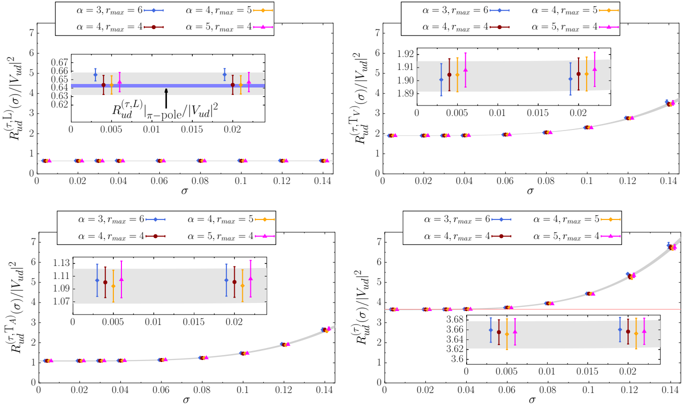


EDITORS' SUGGESTION

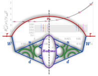
Inclusive hadronic decay rate of the τ lepton from lattice QCD

The authors express the inclusive hadronic decay rate of the tau lepton as an integral over the spectral density of the two-point correlator of the weak $V - A$ hadronic current which they compute fully nonperturbatively in lattice QCD. In a lattice QCD computation with all systematic errors except for isospin breaking effects under control, they then obtain the CKM matrix element V_{ud} with subpercent errors showing that their nonperturbative method can become a viable alternative to superallowed nuclear beta decays for obtaining V_{ud}.

A. Evangelista *et al.*
Phys. Rev. D **108**, 074513 (2023)



$$R_{ud}(\sigma) = 12\pi S_{EW} |V_{ud}|^2 \int_0^\infty \frac{dE E^2}{m_\tau^3} \left\{ K_T^\sigma \left(\frac{E}{m_\tau} \right) \rho_T(E^2) + K_L^\sigma \left(\frac{E}{m_\tau} \right) \rho_L(E^2) \right\}$$



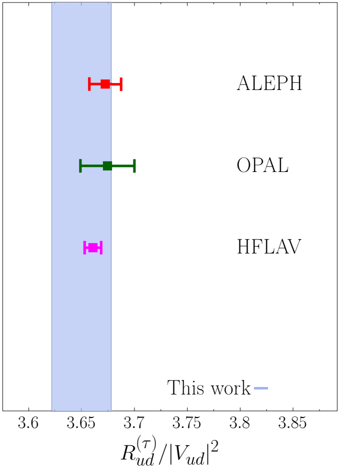
EDITORS' SUGGESTION

Inclusive hadronic decay rate of the τ lepton from lattice QCD

The authors express the inclusive hadronic decay rate of the tau lepton as an integral over the spectral density of the two-point correlator of the weak $V - A$ hadronic current which they compute fully nonperturbatively in lattice QCD. In a lattice QCD computation with all systematic errors except for isospin breaking effects under control, they then obtain the CKM matrix element V_{ud} with subpercent errors showing that their nonperturbative method can become a viable alternative to superallowed nuclear beta decays for obtaining V_{ud}.

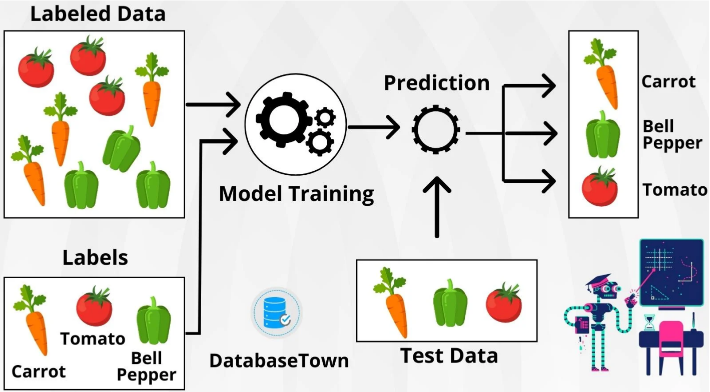
A. Evangelista *et al.*

Phys. Rev. D **108**, 074513 (2023)



do we still have time?

- supervised deep-learning techniques are powerful tools to address classification but also regression problems, particularly in the case of noisy input data



- supervised deep-learning techniques are powerful tools to address classification but also regression problems, particularly in the case of noisy input data

- the idea of using them to extract spectral densities is certainly not original

j.karpie et al. JHEP 04, 057 (2019)
l.kades et al. Phys.Rev.D 102 (2020)
m.zhou et al. Phys.Rev.D 104 (2021)
s.chen et al. arXiv:2110.13521
l.wang et al. Phys.Rev.D 106 (2022)
t.lechien et al. SciPost Phys. 13 (2022)
s.shi et al. Comput.Phys.Commun. 282 (2023)

- supervised deep-learning techniques are powerful tools to address classification but also regression problems, particularly in the case of noisy input data

- the idea of using them to extract spectral densities is certainly not original

- these are non-linear tools that could perform better than linear algorithms such as the HLT method

I.kades et al. Phys.Rev.D 102 (2020)

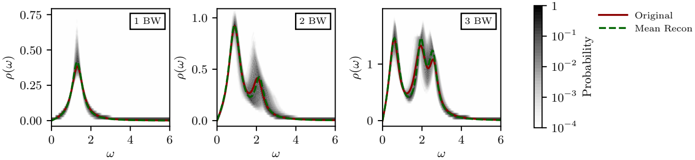
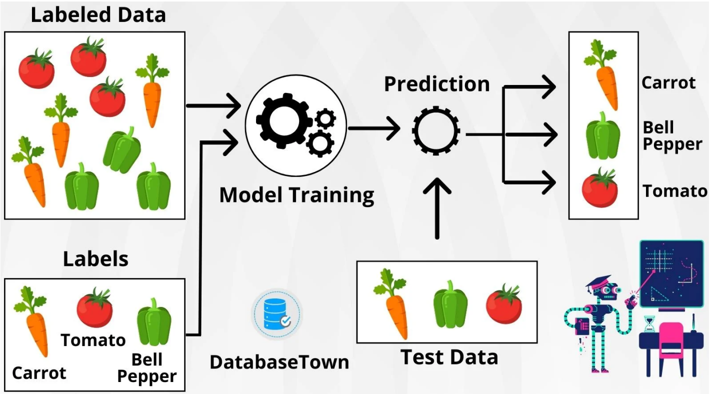


FIG. 1. Examples of mock spectral functions reconstructed via our neural network approach for the cases of one, two and three Breit-Wigner peaks. The chosen functions mirror the desired locality of suggested reconstructions around the original function (red line). Additive, Gaussian noise of width 10^{-3} is added to the discretised analytic form of the associated propagator of the same original spectral function multiple times. The shaded area depicts for each frequency ω the distribution of resulting outcomes, while the dashed green line corresponds to the mean. The results are obtained from the FC parameter network optimised with the parameter loss. The network is trained on the largest defined parameter space which corresponds to the volume Vol O. The uncertainty for reconstructions decreases for smaller volumes as illustrated in Figure 4. A detailed discussion on the properties and problems of a neural network based reconstruction is given in Section IV A.

- supervised deep-learning techniques are powerful tools to address classification but also regression problems, particularly in the case of noisy input data
- the idea of using them to extract spectral densities is certainly not original
- these are non-linear tools that could perform better than linear algorithms such as the HLT method
- what about the dependence of the results upon the training set?



- supervised deep-learning techniques are powerful tools to address classification but also regression problems, particularly in the case of noisy input data
- the idea of using them to extract spectral densities is certainly not original
- these are non-linear tools that could perform better than linear algorithms such as the HLT method
- what about the dependence of the results upon the training set?

I.kades et al. Phys.Rev.D 102 (2020)

Mock spectral functions are constructed using a superposed collection of Breit-Wigner peaks based on a parametrisation obtained directly from one-loop perturbative quantum field theory. Each individual Breit-Wigner is given by

$$\rho^{(BW)}(\omega) = \frac{4A\Gamma\omega}{(M^2+\Gamma^2-\omega^2)^2+4\Gamma^2\omega^2} \, . \qquad (7)$$

Here, M denotes the mass of the corresponding state, Γ its width and A amounts to a positive normalisation constant.

Spectral functions for the training and test set are constructed from a combination of at most $N_{\mathrm{BW}} = 3$ different Breit-Wigner peaks. Depending on which type of

in designing a new supervised deep-learning algorithm, together with **m.buzzicotti** and **a.de santis**, we wanted to address the following two pivotal questions:

in designing a new supervised deep-learning algorithm, together with **m.buzzicotti** and **a.de santis**, we wanted to address the following two pivotal questions:

- is it possible to devise a **model independent** training strategy?

in designing a new supervised deep-learning algorithm, together with **m.buzzicotti** and **a.de santis**, we wanted to address the following two pivotal questions:

- is it possible to devise a **model independent** training strategy?

- if such a strategy is found, is it then possible to quantify reliably, together with the statistical errors, also the unavoidable **systematic uncertainties**?

the importance of these questions can hadrdly be underestimated

- under the **working assumption that a sufficiently large neural network can perform any task**, limiting either the size of the network or the information to which it is exposed during the training process means limiting its ability to solve the problem in full generality

- addressing the second question makes the difference between providing a possibly efficient but qualitative solution to the problem and providing a scientific numerical tool to be used in order to derive theoretical predictions for phenomenological analyses

we faced the challenge and devised a strategy to extract **smeared spectral densities** built on the following two pillars:

we faced the challenge and devised a strategy to extract **smeared spectral densities** built on the following two pillars:

- the introduction of a **functional-basis** to parametrize the correlators and the smeared spectral densities of the **training** sets in a **model independent** way

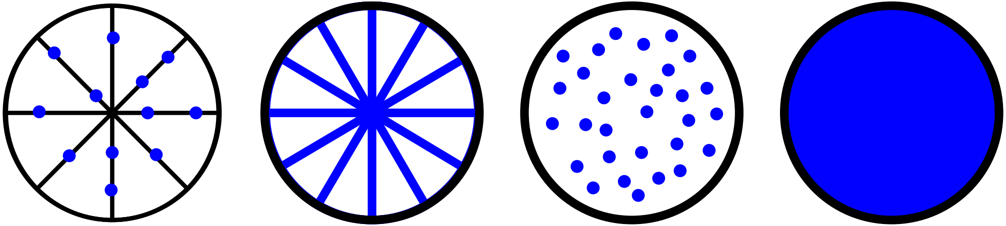
we faced the challenge and devised a strategy to extract **smeared spectral densities** built on the following two pillars:

- the introduction of a **functional-basis** to parametrize the correlators and the smeared spectral densities of the **training** sets in a **model independent** way

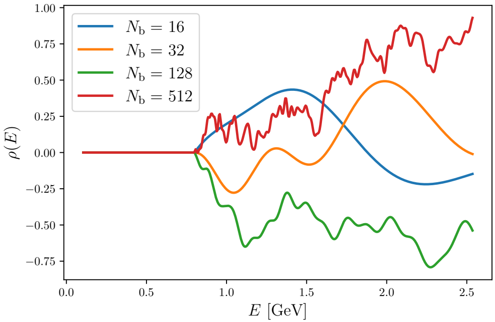
- the introduction of the **ensemble of machines**, the broad audience mentioned in the title, to estimate the **systematic errors**

parametrizing the space of possible unsmeared spectral densities

$N_{\mathrm{b}} = 4$, $N_\rho = 12$ | $N_{\mathrm{b}} = 6$, $N_\rho = \infty$ | $N_{\mathrm{b}} = \infty$, $N_\rho = 30$ | $N_{\mathrm{b}} = \infty$, $N_\rho = \infty$



- we have chosen **Chebyshev polynomials as basis functions**



$$\rho(E; N_b) = \theta(E - E_0) \sum_{n=0}^{N_b} c_n \left[T_n\left(x(E)\right) - T_n\left(x(E_0)\right)\right] , \qquad x(E) = 1 - 2e^{-E}$$

$$c_0 = r_0 \; ; \qquad c_n = \frac{r_n}{n^{1+\varepsilon}} , \quad n > 0 , \qquad r_n \in [-1, 1] , \qquad E_0 \in [0.2, 1.3] \text{ GeV}$$

building the training sets

- we wanted to analyze a lattice correlator already used in **ETMC Phys.Rev.Lett. 130 (2023)** to extract the R-ratio with the HLT method

- therefore, also in the case of mock data we measured energies in GeV and set

$$C(t) = \int_{E_0}^{\infty} \mathrm{d}\omega \, \frac{\omega^2}{12\pi^2} \left[e^{-t\omega} + e^{-(T-t)\omega} \right] \rho(\omega) \; , \qquad T = 64a$$

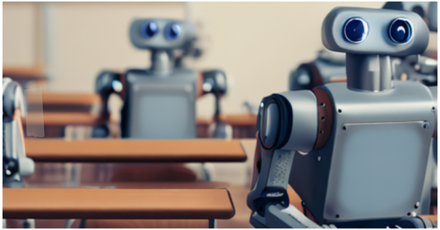
$$\hat{\rho}_\sigma(E) = \int_{E_0}^{\infty} \mathrm{d}\omega \, K_\sigma(E,\omega) \rho(\omega) \; ,$$

$$K_\sigma(E,\omega) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(E-\omega)^2}{2\sigma^2}} \; , \qquad \sigma = \{0.44, 0.63\} \text{ GeV}$$

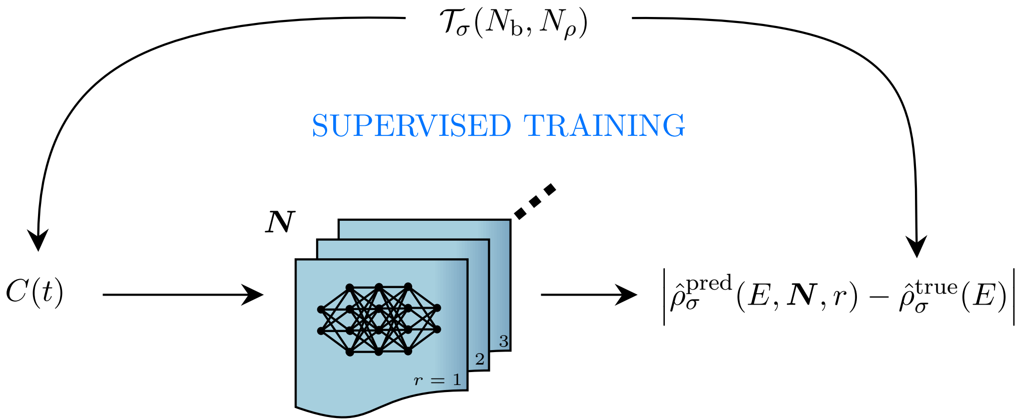
the ensemble of machines

- **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**

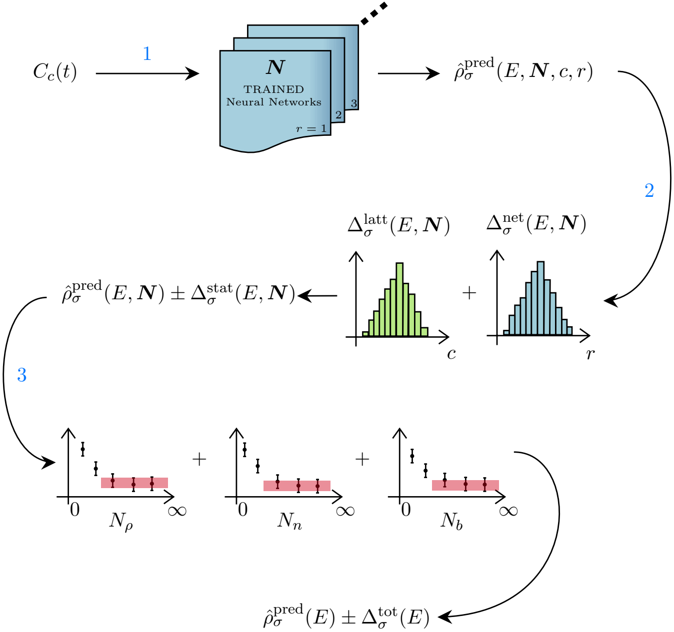
- to quantify the network error we therefore introduced $N_{\rm r}$ replica machines at fixed $\boldsymbol{N} = (N_{\rm n}, N_{\rm b}, N_\rho)$

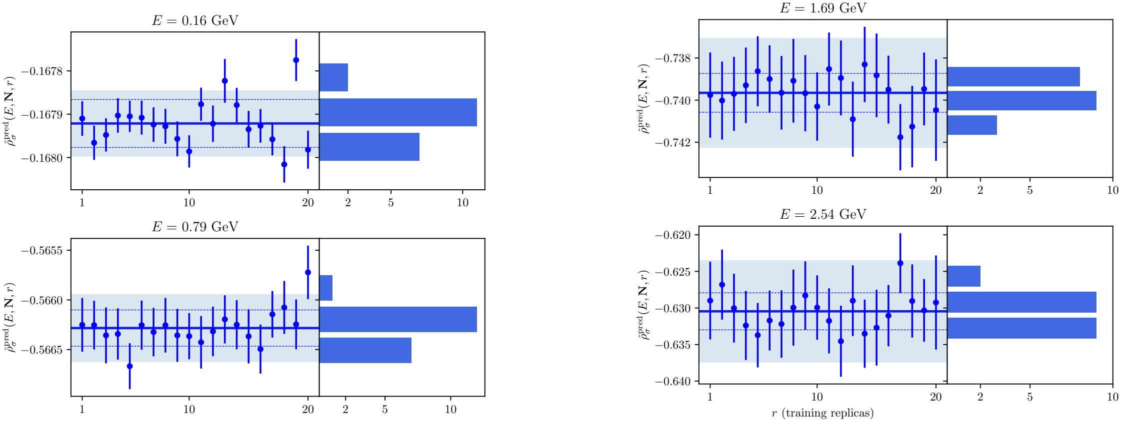


training the ensemble of machines



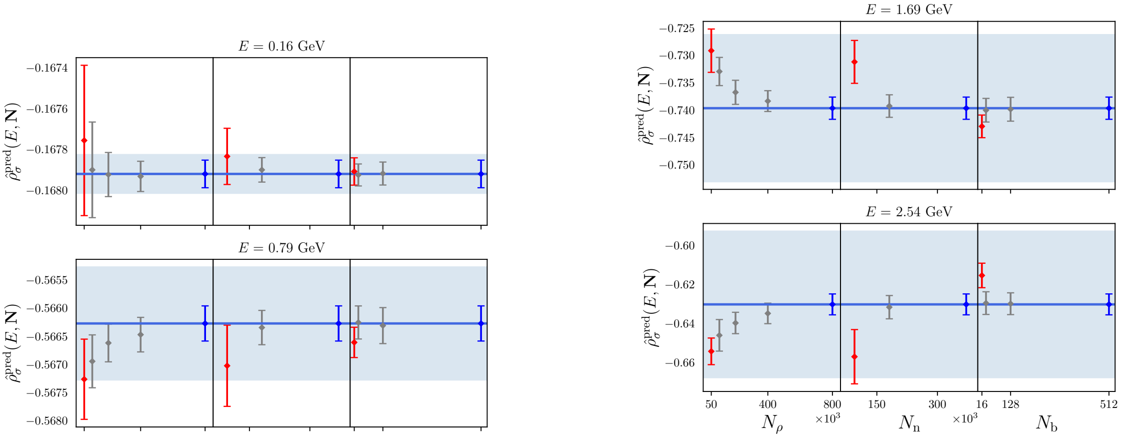
quoting predictions





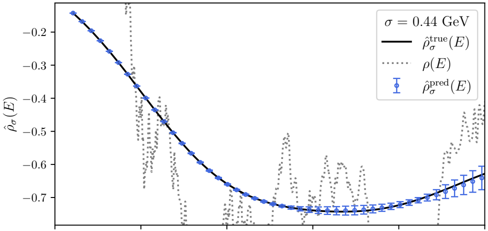
let's now consider a new ρ, again extracted on the Chebyshev basis but never seen during the trainings and this time with

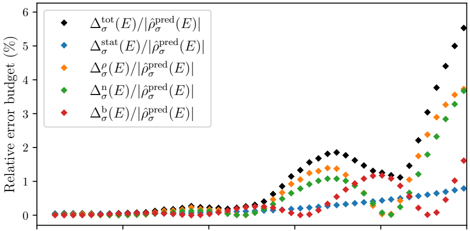
$$N_b = 2 \times N_b^{max} = 1024$$



let's now consider a new ρ, again extracted on the Chebyshev basis but never seen during the trainings and this time with

$$N_{\rm b} = 2 \times N_{\rm b}^{\rm max} = 1024$$



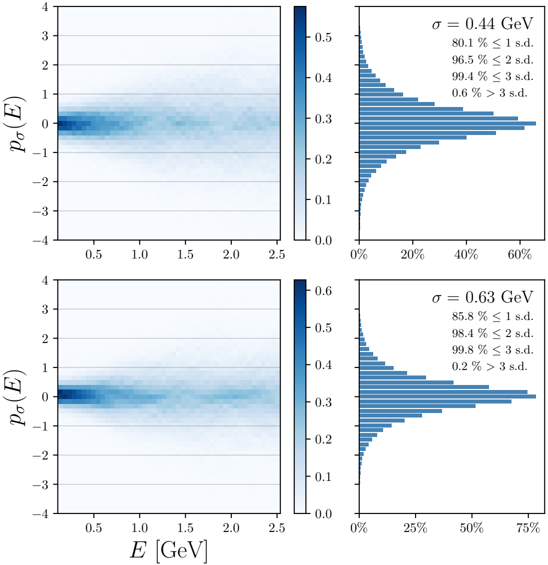


let's now consider a new ρ, again extracted on the Chebyshev basis but never seen during the trainings and this time with

$$N_{\mathrm{b}} = 2 \times N_{\mathrm{b}}^{\max} = 1024$$

let's now repeat the previous experiment 2000 times, with random N_b

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

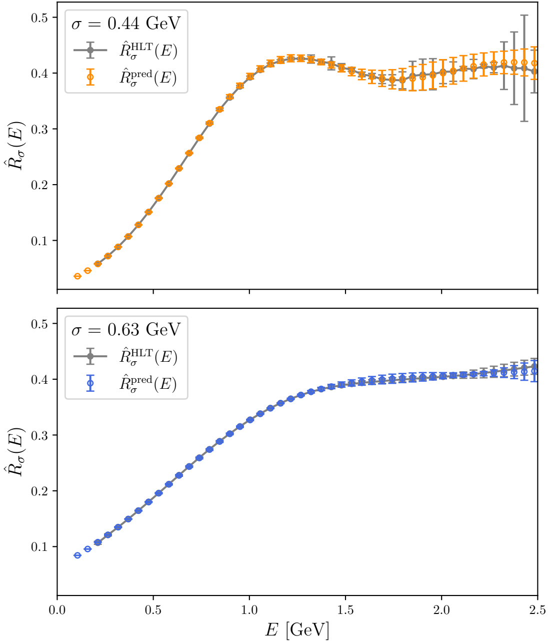


ID	$L^3 \times T$	a fm	aL fm	m_π GeV
B64	$64^3 \cdot 128$	0.07957(13)	5.09	0.1352(2)

let's now look at **true lattice data**, the connected strange-strange contribution to the R-ratio, and at the **comparison with the HLT method**

$$C_{\rm latt}(t) = -\frac{1}{3}\sum_{i=1}^{3}\int d^3x\, \mathrm{T}\langle 0|J_i(x)J_i(0)|0\rangle$$

$$= \int_0^\infty \mathrm{d}\omega\, \frac{\omega^2}{12\pi^2} e^{-t\omega} R(\omega)$$



it is now possible to extract smeared spectral densities from lattice correlators

these can be used to compute (smeared) inclusive hadronic decay rates from first-principles

getting unsmeared spectral densities, with a precision relevant for phenomenology, is much more challenging, but not impossible

the next step are exclusive hadronic decays. . .

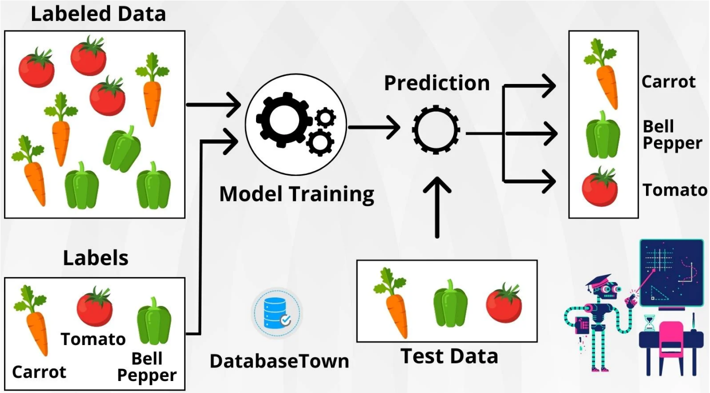
backup slides

so, why another method?

so, why another method?

well, why not. . .

- supervised deep-learning techniques are powerful tools to address classification but also regression problems, particularly in the case of noisy input data



- supervised deep-learning techniques are powerful tools to address classification but also regression problems, particularly in the case of noisy input data

- the idea of using them to extract spectral densities is certainly not original

j.karpie et al. JHEP 04, 057 (2019)
l.kades et al. Phys.Rev.D 102 (2020)
m.zhou et al. Phys.Rev.D 104 (2021)
s.chen et al. arXiv:2110.13521
l.wang et al. Phys.Rev.D 106 (2022)
t.lechien et al. SciPost Phys. 13 (2022)
s.shi et al. Comput.Phys.Commun. 282 (2023)

- supervised deep-learning techniques are powerful tools to address classification but also regression problems, particularly in the case of noisy input data

- the idea of using them to extract spectral densities is certainly not original

- these are non-linear tools that could perform better than linear algorithms such as the HLT method

I.kades et al. Phys.Rev.D 102 (2020)

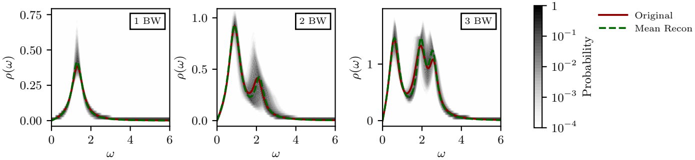


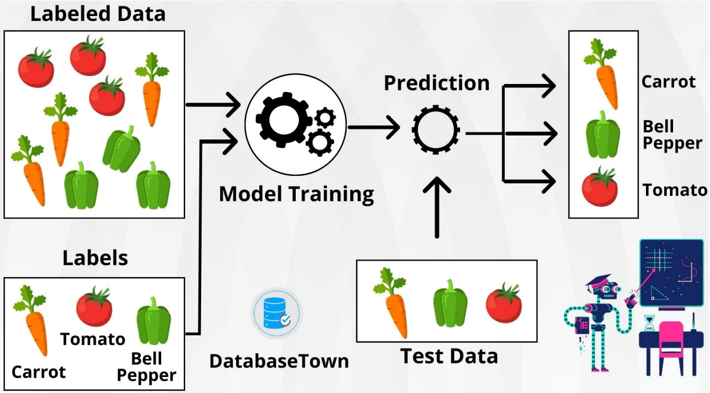
FIG. 1. Examples of mock spectral functions reconstructed via our neural network approach for the cases of one, two and three Breit-Wigner peaks. The chosen functions mirror the desired locality of suggested reconstructions around the original function (red line). Additive, Gaussian noise of width 10^{-3} is added to the discretised analytic form of the associated propagator of the same original spectral function multiple times. The shaded area depicts for each frequency ω the distribution of resulting outcomes, while the dashed green line corresponds to the mean. The results are obtained from the FC parameter network optimised with the parameter loss. The network is trained on the largest defined parameter space which corresponds to the volume Vol O. The uncertainty for reconstructions decreases for smaller volumes as illustrated in Figure 4. A detailed discussion on the properties and problems of a neural network based reconstruction is given in Section IV A.

- supervised deep-learning techniques are powerful tools to address classification but also regression problems, particularly in the case of noisy input data

- the idea of using them to extract spectral densities is certainly not original

- these are non-linear tools that could perform better than linear algorithms such as the HLT method

- what about the dependence of the results upon the training set?



- supervised deep-learning techniques are powerful tools to address classification but also regression problems, particularly in the case of noisy input data
- the idea of using them to extract spectral densities is certainly not original
- these are non-linear tools that could perform better than linear algorithms such as the HLT method
- what about the dependence of the results upon the training set?

I.kades et al. Phys.Rev.D 102 (2020)

Mock spectral functions are constructed using a superposed collection of Breit-Wigner peaks based on a parametrisation obtained directly from one-loop perturbative quantum field theory. Each individual Breit-Wigner is given by

$$\rho^{(BW)}(\omega) = \frac{4A\Gamma\omega}{(M^2+\Gamma^2-\omega^2)^2+4\Gamma^2\omega^2} \,. \qquad (7)$$

Here, M denotes the mass of the corresponding state, Γ its width and A amounts to a positive normalisation constant.

Spectral functions for the training and test set are constructed from a combination of at most $N_{\mathrm{BW}} = 3$ different Breit-Wigner peaks. Depending on which type of

in designing a new supervised deep-learning algorithm, together with **m.buzzicotti** and **a.de santis**, we wanted to address the following two pivotal questions:

in designing a new supervised deep-learning algorithm, together with **m.buzzicotti** and **a.de santis**, we wanted to address the following two pivotal questions:

- is it possible to devise a **model independent** training strategy?

in designing a new supervised deep-learning algorithm, together with **m.buzzicotti** and **a.de santis**, we wanted to address the following two pivotal questions:

- is it possible to devise a **model independent** training strategy?
- if such a strategy is found, is it then possible to quantify reliably, together with the statistical errors, also the unavoidable **systematic uncertainties**?

the importance of these questions can hadrdly be underestimated

- under the **working assumption that a sufficiently large neural network can perform any task**, limiting either the size of the network or the information to which it is exposed during the training process means limiting its ability to solve the problem in full generality

- addressing the second question makes the difference between providing a possibly efficient but qualitative solution to the problem and providing a scientific numerical tool to be used in order to derive theoretical predictions for phenomenological analyses

we faced the challenge and devised a strategy to extract **smeared spectral densities** built on the following two pillars:

we faced the challenge and devised a strategy to extract **smeared spectral densities** built on the following two pillars:

- the introduction of a **functional-basis** to parametrize the correlators and the smeared spectral densities of the **training** sets in a **model independent** way

we faced the challenge and devised a strategy to extract **smeared spectral densities** built on the following two pillars:

- the introduction of a **functional-basis** to parametrize the correlators and the smeared spectral densities of the **training** sets in a **model independent** way

- the introduction of the **ensemble of machines**, the broad audience mentioned in the title, to estimate the **systematic errors**

parametrizing the space of possible unsmeared spectral densities

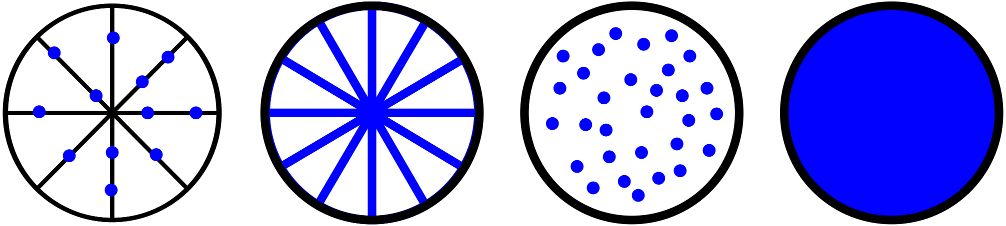
$N_\mathrm{b} = 4$, $N_\rho = 12$ | $N_\mathrm{b} = 6$, $N_\rho = \infty$ | $N_\mathrm{b} = \infty$, $N_\rho = 30$ | $N_\mathrm{b} = \infty$, $N_\rho = \infty$

$$N_\mathrm{b} = 4 \quad N_\rho = 12$$

$$N_\mathrm{b} = 6 \quad N_\rho = \infty$$

$$N_\mathrm{b} = \infty \quad N_\rho = 30$$

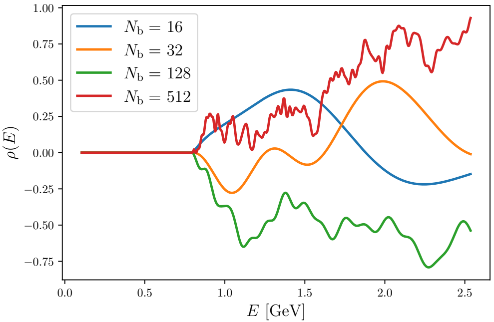
$$N_\mathrm{b} = \infty \quad N_\rho = \infty$$



- we have chosen **Chebyshev polynomials as basis functions**

$$\rho(E;N_{\mathrm{b}}) = \theta(E-E_0)\sum_{n=0}^{N_{\mathrm{b}}} c_n \left[T_n\left(x(E)\right) - T_n\left(x(E_0)\right)\right] \ , \qquad x(E) = 1 - 2e^{-E}$$

$$c_0 = r_0 \ ; \qquad c_n = \frac{r_n}{n^{1+\varepsilon}} \ , \quad n > 0 \ , \qquad r_n \in [-1,1] \ , \qquad E_0 \in [0.2, 1.3] \ \text{GeV}$$

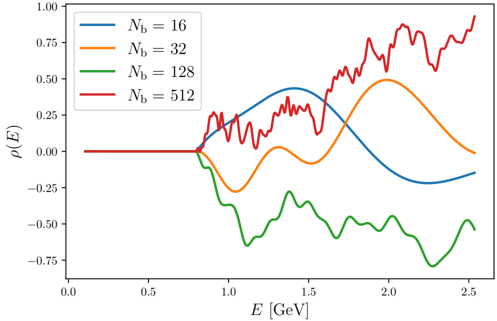


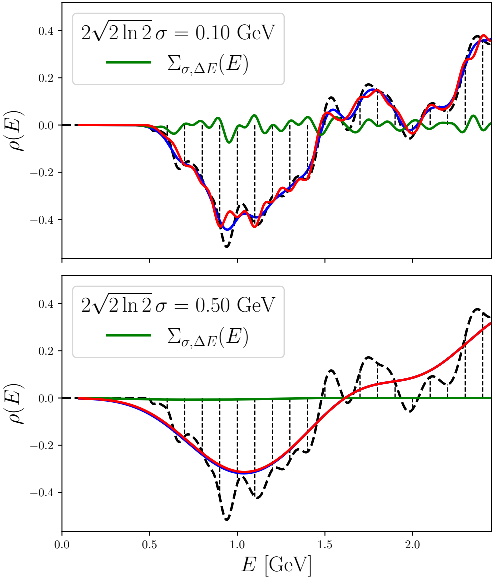
$N_b = 16$
$N_b = 32$
$N_b = 128$
$N_b = 512$
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
$\rho(E)$
0.0
0.5
1.0
1.5
2.0
2.5
E [GeV]

$$\hat{\rho}_\sigma(E) = \int_{E_0}^{\infty} \mathrm{d}\omega\, G_\sigma(E,\omega)\rho(\omega)$$

$$= \Delta E \sum_{n=0}^{\infty} G_\sigma(E,\omega_n)\rho(\omega_n) + \Sigma_{\sigma,\Delta E}(E)$$

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





building the training sets

- we wanted to analyze a lattice correlator already used in **ETMC Phys.Rev.Lett. 130 (2023)** to extract the R-ratio with the HLT method

- therefore, also in the case of mock data we measured energies in GeV and set

$$C(t) = \int_{E_0}^{\infty} \mathrm{d}\omega \, \frac{\omega^2}{12\pi^2} \left[e^{-t\omega} + e^{-(T-t)\omega} \right] \rho(\omega) \, , \qquad T = 64a$$

$$\hat{\rho}_\sigma(E) = \int_{E_0}^{\infty} \mathrm{d}\omega \, K_\sigma(E,\omega)\rho(\omega) \, ,$$

$$K_\sigma(E,\omega) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(E-\omega)^2}{2\sigma^2}} \, , \qquad \sigma = \{0.44, 0.63\} \text{ GeV}$$

building the training sets

- we taught to the networks to distinguish the physical information from the noise by injecting the **noise of the lattice correlator** in our training sets as follows

$$\rho^i(E; N_\mathrm{b}) \quad \mapsto \quad (\boldsymbol{C}, \hat{\boldsymbol{\rho}}_{\boldsymbol{\sigma}})^i$$

$$\mathbb{G}\left[\boldsymbol{C}^i, \left(\frac{C^i(a)}{C_\mathrm{latt}(a)}\right)^2 \hat{\Sigma}_\mathrm{latt}\right] \quad \mapsto \quad \boldsymbol{C}^i_\mathrm{noisy}$$

$$(\boldsymbol{C}_\mathrm{noisy}, \hat{\boldsymbol{\rho}}_{\boldsymbol{\sigma}})^i \in \mathcal{T}_\sigma(N_\mathrm{b}, N_\rho) , \qquad i = 1, \cdots, N_\rho$$

- we considered 3 architectures with sizes in the proportion

$$N_{\rm n}^{\rm arcS} \ : \ N_{\rm n}^{\rm arcM} \ : \ N_{\rm n}^{\rm arcL} = 1 : 2 : 3$$

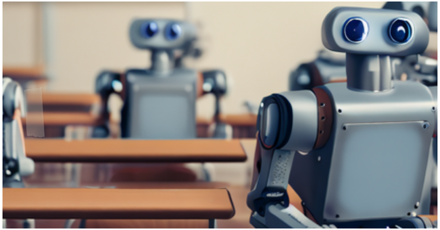
Type	Maps	Size	Kernel size	Stride	Activation
Input		64			
Conv1D	2	32x2	3	2	LeakyReLu
Conv1D	4	16x4	3	2	LeakyReLu
Conv1D	8	8x8	3	2	LeakyReLu
Flatten		384			
Fully conn.		256			LeakyReLu
Fully conn.		256			LeakyReLu
Output		47			
Parameters	94651				

TABLE II. *arcS*: the smallest neural network architecture used in this work. The architecture is of the type feedforward and the structure can be read from top to bottom of the table. It consist of three 1D convolutional layers with an increasing number of maps followed by two fully connected layers. The two blocks are intermediated by one flatten layer. The column denoted by "Size" reports the shape of the signal produced by the corresponding layer. The stride of the filters is set to 2 in such a way that the dimension of the signal is halved at 1D convolutional layer thus favouring the neural network to learn a more abstract representation of the input data. As activation functions we use the LeakyReLu with negative slope coefficient set to -0.2. The neurons with activation functions are also provided with biases. The output is devoid of activation function in order not to limit the output range. The bottom line reports the total number of trainable parameters.

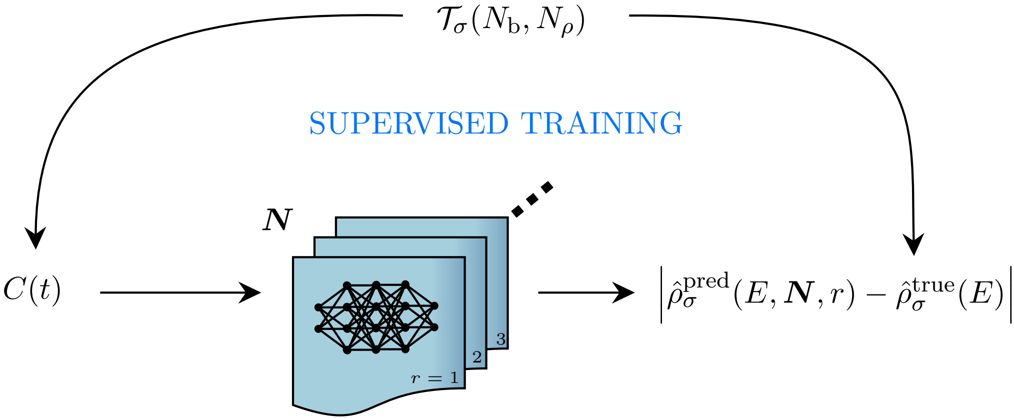
the ensemble of machines

- **the answer** of a machine with **finite** N_n neurons, trained over a **finite set** $\mathcal{T}_\sigma(N_\mathrm{b}, N_\rho)$ **cannot be exact**

- to quantify the network error we therefore introduced N_r replica machines at fixed $\boldsymbol{N} = (N_\mathrm{n}, N_\mathrm{b}, N_\rho)$



training the ensemble of machines

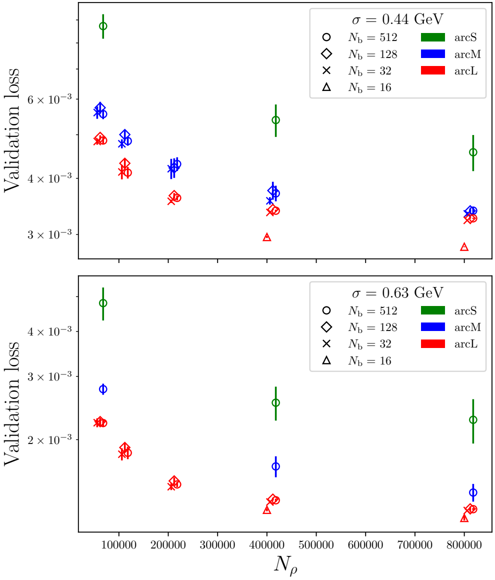


$$\ell(\boldsymbol{w}) = \frac{1}{N_\rho} \sum_{i=1}^{N_\rho} \left| \hat{\boldsymbol{\rho}}_\sigma^{\mathrm{pred},i}(\boldsymbol{w}) - \hat{\boldsymbol{\rho}}_\sigma^i \right|$$

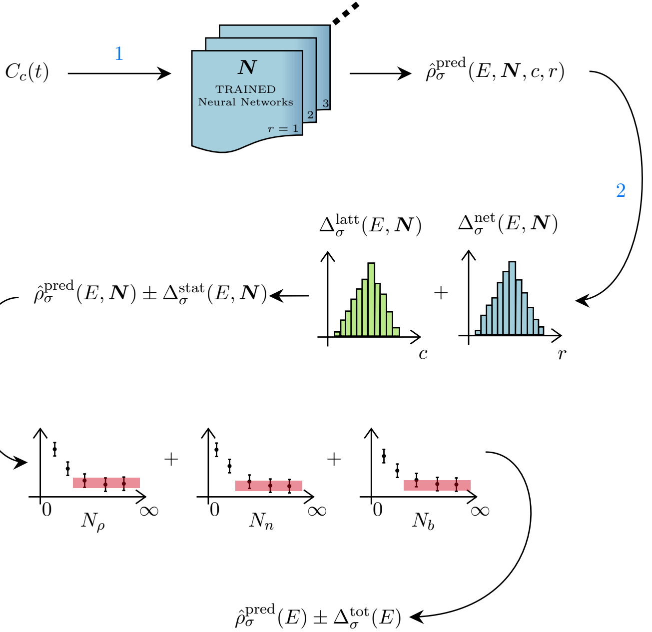
$N_{\mathrm{r}} = 20$

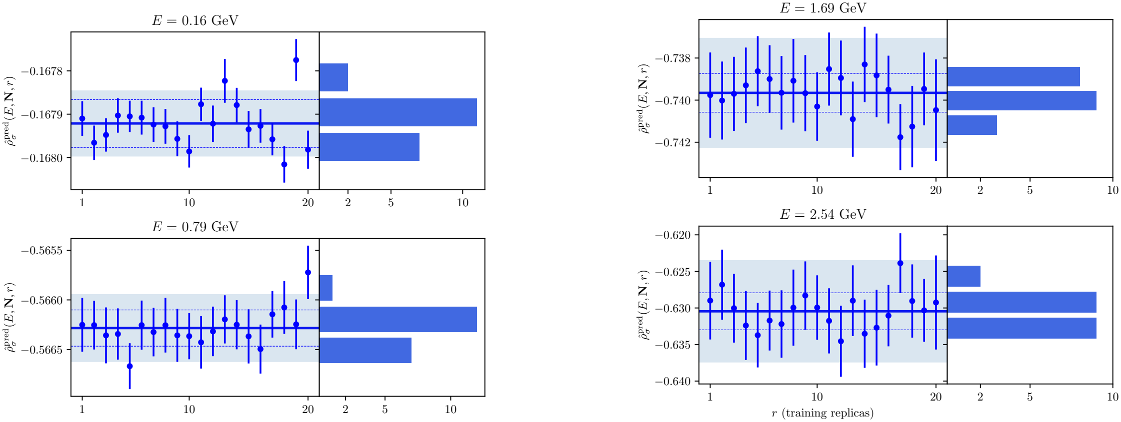
$N_{\mathrm{b}} = \{16, 32, 128, 512\}$

$N_\rho = \{50, 100, 200, 400, 800\} \times 10^3$



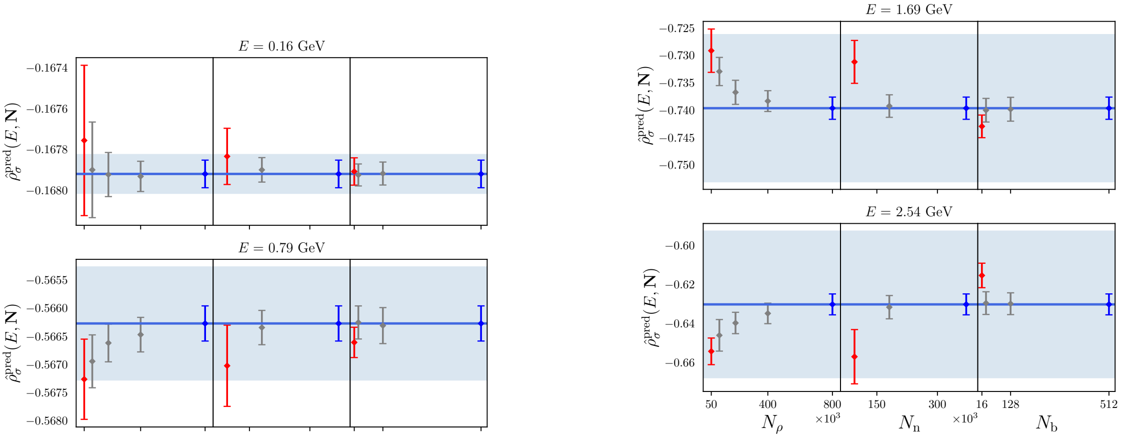
quoting predictions





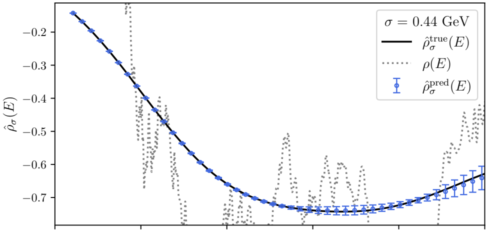
let's now consider a new ρ, again extracted on the Chebyshev basis but never seen during the trainings and this time with

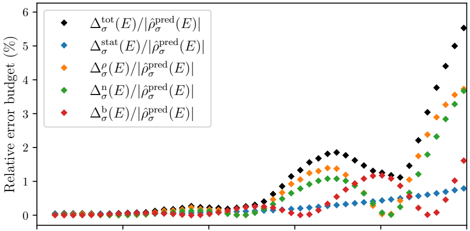
$$N_{\rm b} = 2 \times N_{\rm b}^{\rm max} = 1024$$



let's now consider a new ρ, again extracted on the Chebyshev basis but never seen during the trainings and this time with

$$N_{\rm b} = 2 \times N_{\rm b}^{\rm max} = 1024$$



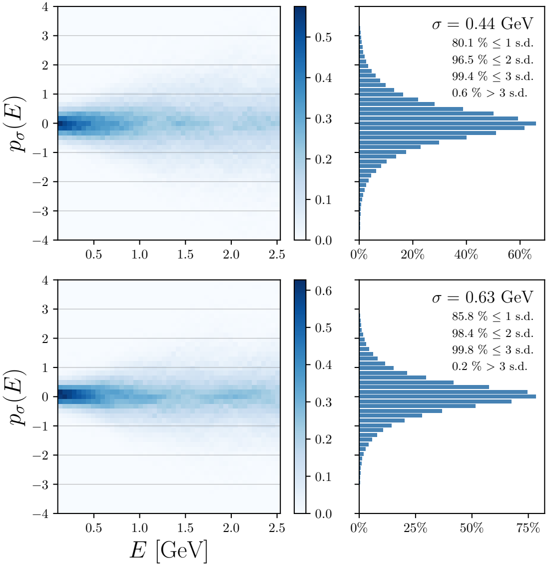


let's now consider a new ρ, again extracted on the Chebyshev basis but never seen during the trainings and this time with

$$N_{\rm b} = 2 \times N_{\rm b}^{\rm max} = 1024$$

let's now repeat the previous experiment 2000 times, with random $N_{\rm b}$

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

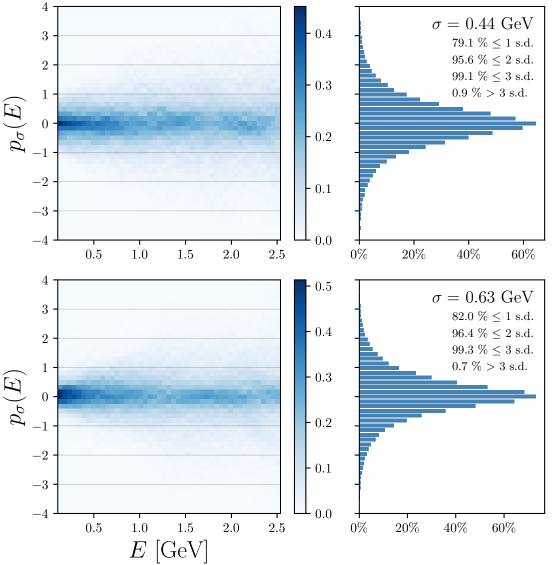


let's now consider another 2000 random unsmeared spectral densities mimicking what we can get on finite volumes

$$\rho(E) = \sum_{n=1}^{N_{\text{peaks}}} c_n \delta(E - E_n)$$

$$N_{\text{peaks}} = 5000 \,, \qquad E_0 \in [0.3, 1.3] \text{ GeV}$$

$$E_n \in [E_0, 15] \text{ GeV} \,, \quad c_n \in [-0.01, 0.01]$$



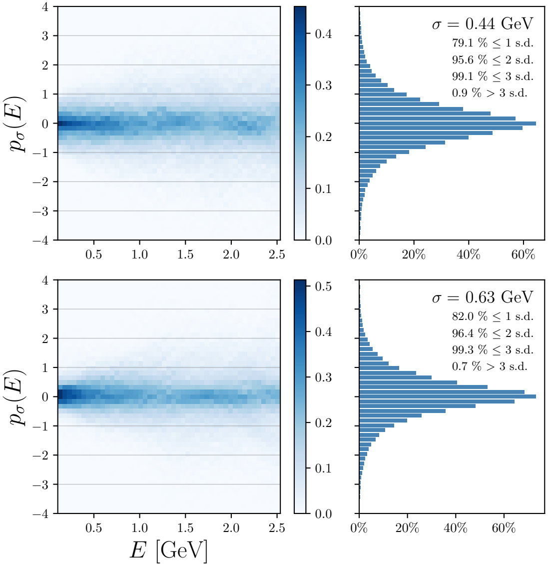
let's now consider another 2000 random unsmeared spectral densities mimicking what we can get on finite volumes

$$\rho(E) = \sum_{n=1}^{N_{\text{peaks}}} c_n \delta(E - E_n)$$

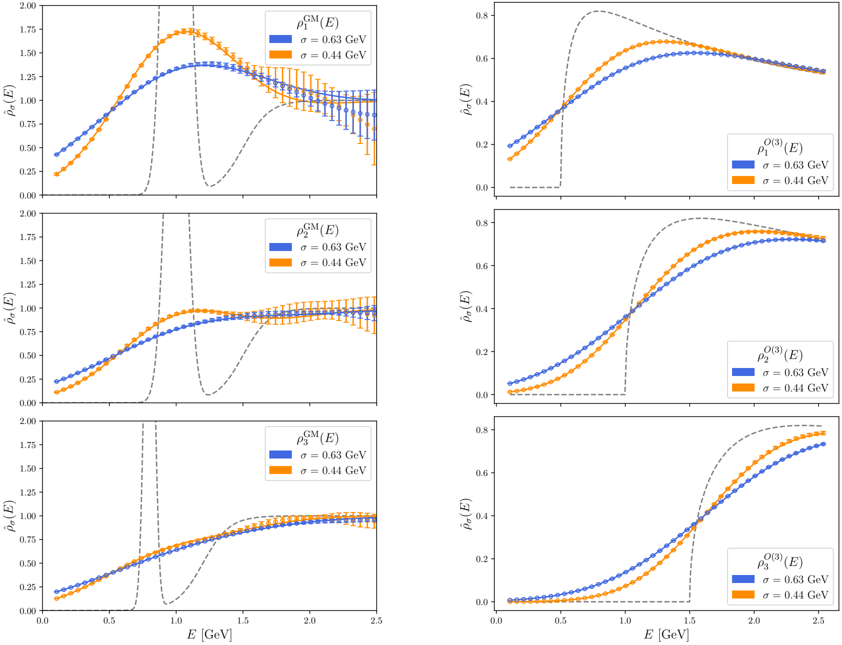
$$N_{\text{peaks}} = 5000\ , \qquad E_0 \in [0.3, 1.3]\ \text{GeV}$$

$$E_n \in [E_0, 15]\ \text{GeV}\ , \quad c_n \in [-0.01, 0.01]$$

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



let's now consider **mock data inspired by physics models**

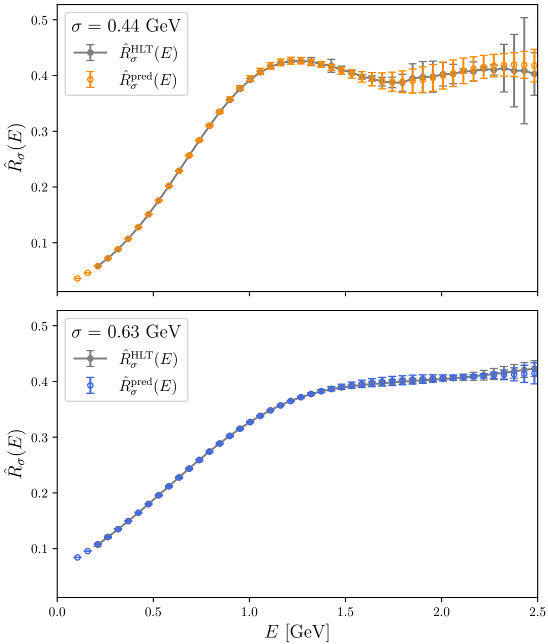


ID	$L^3 \times T$	a fm	aL fm	m_π GeV
B64	$64^3 \cdot 128$	0.07957(13)	5.09	0.1352(2)

let's now look at **true lattice data**, the connected strange-strange contribution to the R-ratio, and at the **comparison with the HLT method**

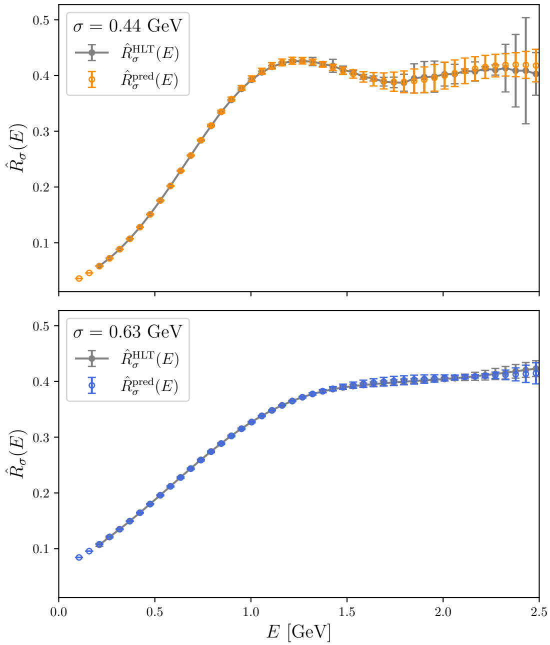
$$C_{\text{latt}}(t) = -\frac{1}{3}\sum_{i=1}^{3}\int d^3x\, \mathrm{T}\langle 0|J_i(x)J_i(0)|0\rangle$$

$$= \int_0^\infty \mathrm{d}\omega\, \frac{\omega^2}{12\pi^2} e^{-t\omega} R(\omega)$$



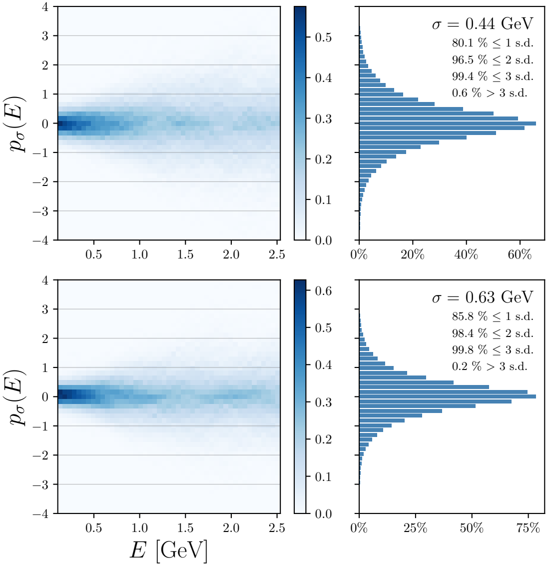
what **we** have learned?

- supervised deep learning techniques can be used to extract smeared hadronic spectral densities from lattice correlators in a model-independent way

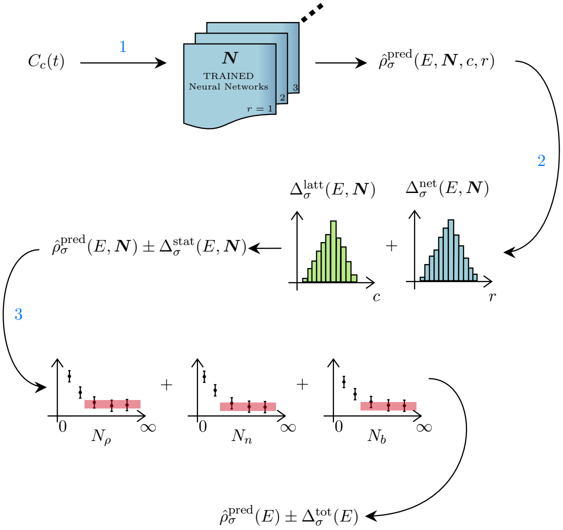


- supervised deep learning techniques can be used to extract smeared hadronic spectral densities from lattice correlators in a model-independent way

- the systematic errors can be reliably quantified and the predictions can be used in phenomenological analyses



- 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 that might end up to be numerically demanding and can possibly be simplified, but there is no free-lunch in physics!

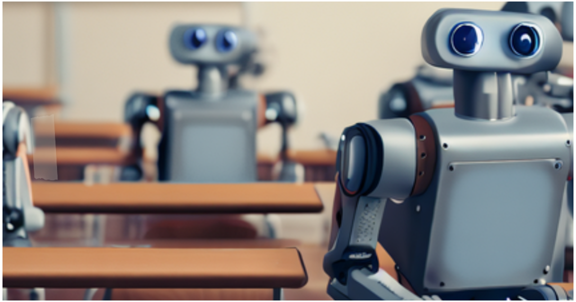


- supervised deep learning techniques can be used to extract smeared hadronic spectral densities from lattice correlators in a model-independent way

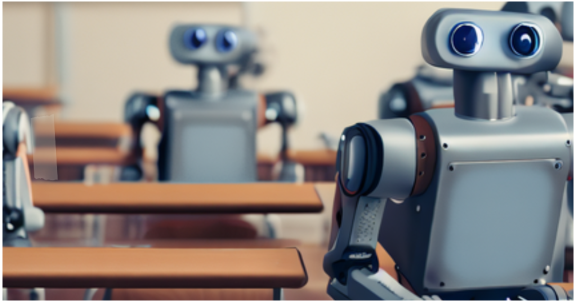
- the systematic errors can be reliably quantified and the predictions can be used in phenomenological analyses

- admittedly, the procedure that we propose to do that might end up to be numerically demanding and can possibly be simplified, but there is no free-lunch in physics!

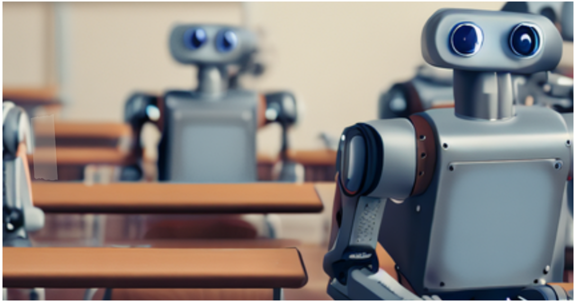
- here we taught a lesson to a broad audience of learning machines



- 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 that 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
- the subject of the lesson is just a particular topic. . .



- 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 that 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
- 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



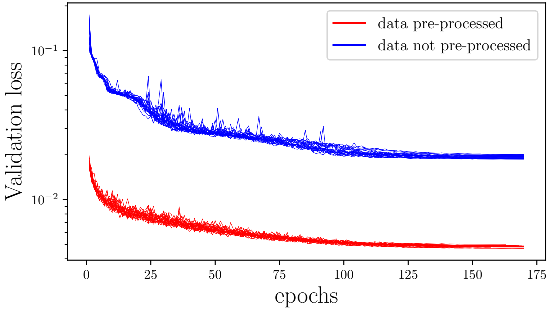
a major impact in the machine-learning performance is played by the way the data are presented to the neural network

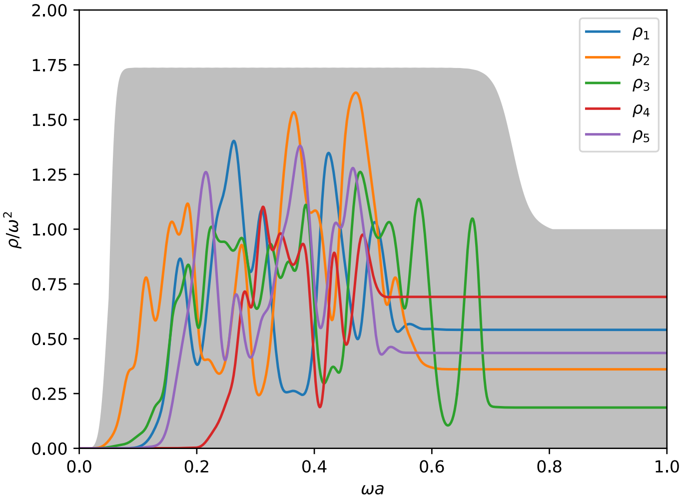
we standardized input data at fixed training set $\mathcal{T}_\sigma(N_{\rm b}, N_\rho)$ as follows

$$\mu(t) = \frac{1}{N_\rho} \sum_{i=1}^{N_\rho} C_i(t)$$

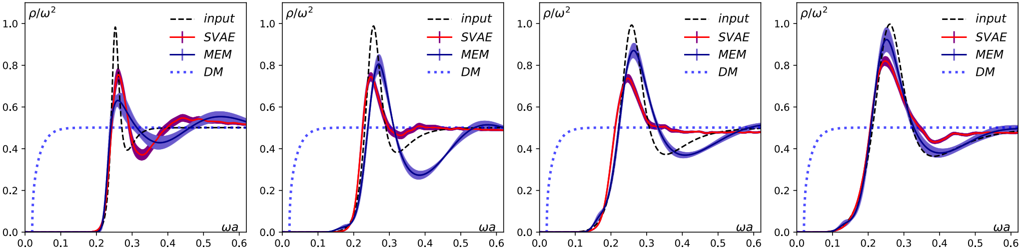
$$\gamma(t) = \sqrt{\frac{\sum_{i=1}^{N_\rho} \left(C_i(t) - \mu(t)\right)^2}{N_\rho}}$$

$$C_{\rm noisy}(t) \mapsto C'_{\rm noisy}(t) = \frac{C_{\rm noisy}(t) - \mu(t)}{\gamma(t)}$$





the approach closest in spirit to the one that we propose is that of **s.chen et al. arXiv:2110.13521**



the approach closest in spirit to the one that we propose is that of **s.chen et al. arXiv:2110.13521**