smeared $R$-ratio and applications to $g - 2$
the $R$-ratio plays a fundamental role in particle physics since its introduction.
in the last few years, the importance of the $R$-ratio has been mainly associated with muon $g - 2$ experiments

$$a^\text{HVP-LO}_\mu = \int_0^{+\infty} d\omega f_{a\mu}(\omega) R(\omega)$$
the $R$-ratio is much more than that...
theoretically, $R(E)$ is a distribution

$$
\langle 0| j_{em}^i(0) \delta(H-E)(2\pi)^3 \delta^3(P) j_{em}^j(0)|0\rangle = -\frac{\delta^{ij} E^2}{12\pi^2} R(E)
$$

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

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

and probe the energy dependence of $R(E)$ by changing $f(E)$
$a_{\mu}^{\text{HVP-LO}} \equiv R[f_{a\mu}]$ can also be seen as a low-energy probe of $R(E)$.
\( a_{\mu}^{\text{HVP-LO}} \) 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 \( a_{\mu} \).

The comparison of the BNL+FNAL measurements, \( a_{\mu}^{\text{HVP-LO,exp}} \), with \( \int_{0}^{+\infty} dE f_{a_{\mu}}(\omega) R^{\text{exp}}(\omega) \) is a consistency check.

To turn it into a first-principles test of the theory we must test \( a_{\mu}^{\text{HVP-LO,exp}} \) and \( R^{\text{exp}}(E) \) 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

$$f(\omega) \equiv G_\sigma(E - \omega) = \frac{e^{-(E-\omega)^2/2\sigma^2}}{\sqrt{2\pi\sigma}}$$,

$$R_\sigma(E) \equiv R[f] = \int_0^{+\infty} d\omega \, G_\sigma(E - \omega) \, R(\omega)$$
Extraction of spectral densities from lattice correlators

Martin Hansen,1 Alessandro Lupo,2 and Nazario Tantalo3

1INFN Roma Tor Vergata, Via della Ricerca Scientifica 1, I-00133 Rome, Italy
2University of Rome Tor Vergata, Via della Ricerca Scientifica 1, I-00133 Rome, Italy
3University 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...
Regular Article - Theoretical Physics

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

Michele Buzzicotti\textsuperscript{a}, Alessandro De Santis\textsuperscript{b} \textsuperscript{c} and Nazario Tantalo\textsuperscript{c}

University and INFN of Roma Tor Vergata, Via della Ricerca Scientifica 1, 00133 Rome, Italy
before addressing the phenomenologically relevant calculation of the smeared $R$-ratio on which i’m now going to focus. . .
6 Conclusions

The aim of the preceding sections is to verify the procedure of ref. [1] for numerically computing smeared spectral densities (with an a priori specified smearing kernel) from lattice field theory correlation functions. In this regard the two-dimensional O(3) model usefully provides exact results against which the estimates can be checked. These checks, which are presented in figures 6 and 9, are satisfied and compare both $\rho_\epsilon(E)$ at finite $\epsilon$ and the results from $\epsilon \to 0$ extrapolations to determine $\rho(E)$ deep into the inelastic region where finite-volume methods have not yet been developed. The highest energy considered here is $E = 40 m^{\star}$, at which $\rho(E)$ is determined with a relative accuracy of 5% and differs significantly from the exact two-particle contribution $\rho(2)(E)$ given in eq. (2.6).

Apart from the 'usual' sources of systematic error due to the finite lattice spacing and finite-volume spacetime, we must also consider the imperfect reconstruction of the smearing kernel due to the finite number of input time slices and their associated statistical errors. All sources of systematic error have been estimated and included in figures 6 and 9 where the statistical and systematic errors are added in quadrature. Generally the errors due to the finite lattice extent are the largest source of systematic uncertainty, and are typically less than or comparable to the statistical errors.

The determination of $\rho_\epsilon(E)$ becomes increasingly difficult for smaller smearing widths $\epsilon$ at fixed energy $E$, and increasing $E$ with fixed $\epsilon$. As is evident from the right two panels of figure 6, it is difficult to achieve precise results outside of the elastic region for $\epsilon \lesssim m/2$ with the Gaussian smearing kernel. Better is to exploit the smoothness of $\rho(E)$ and scale $\epsilon \propto (E - 2m^{\star})$, so that an equal proportion of the smearing kernel 'leaks' down to the two particle threshold at each energy. This enables the determination of $\rho(E)$ in figure 9, which is the main result of this work.

the HLT method has been stringently validated by performing the analogous calculation in the 2D non-linear $O(3)$ $\sigma$-model where the exact answer is known.
let’s see how it works...
\[ \rho(E) = \langle 0 | \hat{O}_F \delta \left( \hat{H} - E \right) (2\pi)^3 \delta^3 \left( \hat{P} - \mathbf{p} \right) \hat{O}_I | 0 \rangle \]

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

\[ t = a\tau , \quad \tau = 1, \cdots, T \]
\[ \rho(E) = \langle 0| \hat{O}_F \, \delta \left( \hat{H} - E \right) (2\pi)^3 \delta^3 \left( \hat{P} - p \right) \hat{O}_I |0 \rangle \]

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

\[ t = a\tau , \quad \tau = 1, \cdots, T \]
\( f(\infty) = 0 \), \( f(\omega) \in L^2_\mathbb{R}[E_0, \infty] \)

\[
f(\omega) = \sum_{\tau=1}^{\infty} g_\tau e^{-\tau \alpha \omega},
\]

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

\[
A_n[g] = \int_{E_0}^{\infty} d\omega w_n(\omega) \left| f(\omega) - \sum_{\tau=1}^{T} g_\tau e^{-\tau \alpha \omega} \right|^2
\]

\[
d_n[g] = \sqrt{\frac{A_n[g]}{A_n[0]}}
\]

\[ aE_* = 0.5 \quad \sigma = 0.03E_* \quad T = 128 \]

Example with \( N = 128 \), \( aE_\dagger = 0.5 \) and \( \sigma = 0.03E_\dagger \):
in the case of the isoQCD $R$-ratio

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

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

\[ A_n[g] = \int_{2m_\pi}^\infty d\omega \, w_n(\omega) \left| \Delta(\omega) - \sum_{\tau=1}^T g_\tau e^{-\tau \alpha \omega} \right|^2 \]

\[ B[g] = \sum_{\tau_1, \tau_2=1}^T g_{\tau_1} g_{\tau_2} \text{Cov}_{\tau_1 \tau_2} \]

\[ W_n[g] = A_n[g] + \lambda B[g] \]
let's have a closer look at the stability analysis...
FIG. 8. Reconstructed kernels at dominated regime ($d(g^{**})$, $d(g^*)$) or on the C80 ensemble ($E = 0.74$ GeV, $\sigma = 0.63$ GeV). The blue and orange bands correspond to the results of the stability summarizing the comparison for all values of $d(gP)$. The plots on the left show the datasets, with the summary of the comparison of different ensembles for all values of $d(gP)$.

FIG. 9. Top-panel: Example of the comparison of $R_{e,C,TM}(E; gP)$ with the blue and orange bands for $d(g^{**})$ and $d(g^*)$. The red band corresponds to $E = 0.63$ GeV, and the green band corresponds to $E = 0.74$ GeV. Other panels (second panel, third panel) correspond to $E = 0.53$ GeV and $E = 0.44$ GeV, respectively. The red band corresponds to $E = 0.44$ GeV, and the green band corresponds to $E = 0.53$ GeV. The blue and orange bands correspond to the results of the stability summarizing the comparison for all values of $d(gP)$. The plots on the left show the datasets, with the summary of the comparison of different ensembles for all values of $d(gP)$.
the given flavour quantum numbers, their correlation functions have the least signal/noise problem in the Monte Carlo evaluation of the path integral.

Still restricting ourselves to isospin-symmetric QCD (isoQCD), we thus take it for granted that the choice $M_{i,2}$ is easy, and we do not need to discuss it in detail: the pseudoscalar meson masses are very good choices, and some variations for heavy quarks may provide further improvements.

The choice of $M_1$ is more difficult. From the point of view of physics, a natural choice is the nucleon mass, $M_{\text{nucl}}$. Unfortunately it has a rather bad signal/noise problem when quark masses are close to their physical values. The ratio of signal to noise of the correlation function at time $x_0$ from $N$ measurements behaves as

$$R_{\text{nucl}} \frac{S}{N} x_0 \propto \exp\left(\frac{(m_{\text{nucl}}^3 - m_{\pi}^3) x_0}{27 \text{ fm}}\right),$$

where the numerical value of $0.27 \text{ fm}$ uses the experimental masses. The behaviour in practice, but at still favourably large quark masses, is illustrated in Fig. 51. Because $0.5 < 1 < 1.5 < 2 < 2.5 < 3 < 3.5 < 4$, the situation for $M_{\text{proton}}$ becomes worse closer to the physical point, but may be changed by algorithmic improvements.

This property leads to large statistical errors and it is further difficult to control excited-state contaminations when statistical errors are large, it is useful to search for alternative physics scales. The community has gone this way, and we discuss some of them below.

For illustration, here we just give one example: the decay constants of leptonic $\pi$ decays have mass dimension one and can directly replace $M_1$ above. Figure 51 demonstrates their long and precise plateaux as a function of the Euclidean time. Advantages and disadvantages of this choice and others are discussed more systematically in Sec. 11.4.
the given flavour quantum numbers, their correlation functions have the least signal/noise problem in the Monte Carlo evaluation of the path integral.

Still restricting ourselves to isospin-symmetric QCD (isoQCD), we thus take it for granted that the choice $M_{i,j}$ is easy, and we do not need to discuss it in detail: the pseudoscalar meson masses are very good choices, and some variations for heavy quarks may provide further improvements.

The choice of $M_1$ is more difficult. From the point of view of physics, a natural choice is the nucleon mass, $M_1 = M_{\text{nucl}}$. Unfortunately it has a rather bad signal/noise problem when quark masses are close to their physical values. The ratio of signal to noise of the correlation function at time $x_0$ from $N$ measurements behaves as $R_{\text{nucl}} S/N_{x_0} \gtrsim p_N \exp\left(\frac{(m_{\text{nucl}}^2 m_\pi^2)}{x_0}\right) \simeq p_N \exp\left(\frac{x_0}{0.27 \text{ fm}}\right)$, \(468\).

Figure 51: Effective masses for $M_{\text{proton}}$, $\Omega_V(r_0)$, $\Omega_V(r_1)$, and $f_\pi$ on $N_f=2$ CLS ensemble $N6$ with $a=0.045 \text{ fm}$, $M_\pi=340 \text{ MeV}$ on a $48^3 \times 96$ lattice \(316\). All effective "masses" have been scaled such that the errors in the graph reflect directly the errors of the determined scales. They are shifted vertically by arbitrary amounts. Figure from Ref. 719. Note that this example is at still favourably large quark masses. The situation for $M_{\text{proton}}$ becomes worse closer to the physical point, but may be changed by algorithmic improvements.

For illustration, here we just give one example: the decay constants of leptonic $\pi$ or $K$ decays have mass dimension one and can directly replace $M_1$ above. Figure 51 demonstrates their long and precise plateaux as a function of the Euclidean time. Advantages and disadvantages of this choice and others are discussed more systematically in Sec. 11.4.
$P_\sigma (E) = \frac{R_\sigma (E; g^*) - R_\sigma (E; g^{**})}{\Delta_{\text{stat}} (E; g^{**})}$,

$\Delta_{\sigma}^{\text{rec}} (E) = |R_\sigma (E; g^*) - R_\sigma (E; g^{**})| \operatorname{erf} \left( \frac{|P_\sigma (E)|}{\sqrt{2}} \right)$
\[ P_\sigma(E) = \frac{R_\sigma(E; g^*) - R_\sigma(E; g^{**})}{\Delta_{\text{stat}}(E; g^{**})} \], \quad \Delta_{\text{rec}}(E) = |R_\sigma(E; g^*) - R_\sigma(E; g^{**})| \text{ erf} \left( \frac{|P_\sigma(E)|}{\sqrt{2}} \right)
let's look at finite volume effects...
\[ P^L_\sigma(E) = \frac{R_\sigma(E; \frac{3L}{2}) - R_\sigma(E; L)}{\sqrt{\Delta_{\sigma}(E; \frac{3L}{2})^2 + \Delta_{\sigma}(E; L)^2}}, \]

\[ \Delta^L_\sigma(E) = \max_{OS, TM} \left\{ \left| R_\sigma(E; \frac{3L}{2}) - R_\sigma(E; L) \right| \erf \left( \frac{|P^L_\sigma(E)|}{\sqrt{2}} \right) \right\}, \quad L \simeq 5.1 \text{ fm} \]
\[ P_\sigma^L(E) = \frac{R_\sigma(E; \frac{3L}{2}) - R_\sigma(E; L)}{\sqrt{\Delta_\sigma(E; \frac{3L}{2})^2 + \Delta_\sigma(E; L)^2}} , \]

\[ \Delta_\sigma^L(E) = \max_{OS, TM} \left\{ \left| R_\sigma(E; \frac{3L}{2}) - R_\sigma(E; L) \right| \text{erf} \left( \frac{|P_\sigma^L(E)|}{\sqrt{2}} \right) \right\} , \quad L \simeq 5.1 \text{ fm} \]
let's look at cutoff effects...
In the case of the connected contributions we performed both linear extrapolations in \( \exp \) and their combination with linear extrapolations in \( \exp \)
to Eq. (1). The smearing Gaussian corresponding to centering of \( a \) at Eq. (1), the connected strange-strange (\( \mu \)), the connected light-light (\( \alpha \)), and found them to be compatible within errors in all cases. The disconnected contribution has been obtained by generating bootstrap samples and here to compute the central values and covariance matrices. The lattice gauge ensembles used in this work. Generated by the ETMC, are listed in TABLE I.

In our lattice calculation we considered three values for \( \alpha = 0.79 \) GeV on the D96 ensemble, see Ref. [7] for more details. Here, in FIG. 2, we show an example (\( E = 0.79 \) GeV) of the di-electron spectrum, each of which simulating an independent measurement, from a multivariate Gaussian distribution. The results for each contribution, can be found in the supplementary material. For the di-electron spectrum, we performed both linear extrapolations in \( \exp \) and their combination with linear extrapolations in \( \exp \), each of which simulating an independent measurement, from a multivariate Gaussian distribution. The results for each contribution, can be found in the supplementary material. Here, in FIG. 2, we show an example (\( E = 0.79 \) GeV) of the di-electron spectrum, each of which simulating an independent measurement, from a multivariate Gaussian distribution. The results for each contribution, can be found in the supplementary material.
once you have all this...
and you have already seen this...
you can trust this:
and then you get this!
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?
\[ C(t) = C_{\text{no-LMA}}(t) + C_{\text{LMA}}(t) + C_{\text{blinding}}(t) \]
\begin{align*}
\sigma &= 500 \text{ MeV} \\
\text{no-LMA} \\
\text{LMA + blinding}
\end{align*}
<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>LMA+blinding</th>
<th>no-LMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0</td>
<td>250 MeV</td>
<td>250 MeV</td>
</tr>
</tbody>
</table>
\[ \sigma = 250 \text{ MeV} \]

- no-LMA
- LMA + blinding
\[ \sigma = 200 \text{ MeV} \]

- no-LMA
- LMA + blinding
let's go back to $a_\mu$

on the lattice this is estimated by

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

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

$$a_{\mu}^{\text{HVP-LO, HLT}} = \lim_{a \to 0} \lim_{L \to \infty} \lim_{\lambda \to 0} \int_0^\infty d\omega f_{a_\mu}^{\text{HLT}}(\omega) R^{a, L}(\omega)$$
\[ \sum_{t_c} f_a(t) C(t) \]

strange-strange connected

\[ A[g] \]

\[ A[0] \]

\[ \alpha = +3.00 \]

\[ \alpha = +2.00 \]

\[ \alpha = +1.00 \]
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


s.hashimoto PTEP 2017 (2017)
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. . .
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 $K \rightarrow \pi\pi$

“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
\[ \rho(E) = \langle 0 | \hat{O}_F \delta (\hat{H} - E) (2\pi)^3 \delta^3 (\hat{P} - \mathbf{p}) \hat{O}_I | 0 \rangle \]

\[ C(t) = \int d^3x \ e^{-i \mathbf{p} \cdot \mathbf{x}} \langle 0 | \hat{O}_F \ e^{-i \hat{H}t + i \hat{P} \cdot \mathbf{x}} \hat{O}_I | 0 \rangle = \int_{E_0}^{\infty} dE \ e^{-iE} \ \rho(E) \]

\[ t = a\tau, \quad \tau = 1, \cdots, \frac{T}{a} \]
$\rho(E)$ contains an infinite amount of information
\( \rho(E) \) contains an infinite amount of information

the problem, to be addressed numerically, has to be discretized
\[ C(a\tau) = \int_{E_0}^{\infty} dE \, e^{-\tau a E} \, \rho(E) \quad \leftrightarrow \quad \sigma \sum_{m=0}^{N_E - 1} e^{-\tau a E_m} \, \hat{\rho}(E_m) \]
\[
C(a\tau) = \int_{E_0}^{\infty} \, dE \, e^{-\tau a E} \rho(E) \quad \leftrightarrow \quad \sigma \sum_{m=0}^{N_E-1} e^{-\tau a E_m} \hat{\rho}(E_m)
\]

\[
\hat{E}_m \equiv e^{-\tau a E_m}, \quad \hat{G}_{nm} \equiv \left[ \hat{\mathcal{E}}^T \hat{\mathcal{E}} \right]_{nm}
\]
\[ C(a\tau) = \int_{E_0}^{\infty} dE \, e^{-aE \rho(E)} \quad \leftrightarrow \quad \sigma \sum_{m=0}^{N_E-1} e^{-aE_m} \dot{\rho}(E_m) \]

\[ \hat{G} \hat{\rho} = \frac{1}{\sigma} \hat{E}^T C \; , \quad \hat{\rho} = \frac{1}{\sigma} \hat{G}^{-1} \hat{E}^T C \]
\[ C(a\tau) = \int_{E_0}^{\infty} dE \, e^{-\tau aE} \rho(E) \quad \leftrightarrow \quad \sigma \sum_{m=0}^{N_E-1} e^{-\tau aE_m} \hat{\rho}(E_m) \]

\[ \hat{G} \hat{\rho} = \frac{1}{\sigma} \hat{\epsilon}^T C, \quad \hat{\rho} = \frac{1}{\sigma} \hat{G}^{-1} \hat{\epsilon}^T C \]

\[ g_\tau(E_n) = \frac{1}{\sigma} \sum_{m=0}^{N_E-1} \hat{G}_{nm}^{-1} \hat{\epsilon}_m, \quad K(E_n, \omega) = \sum_{\tau=1}^{N_T-1} g_\tau(E_n) e^{-\tau \omega} \]
\[
C(a\tau) = \int_{E_0}^{\infty} dE \ e^{-\tau aE} \rho(E) \quad \Leftrightarrow \quad \sigma \sum_{m=0}^{N_E-1} e^{-\tau aE_m} \hat{\rho}(E_m)
\]

\[
\hat{G} \hat{\rho} = \frac{1}{\sigma} \hat{\epsilon}^T \ C, \quad \hat{\rho} = \frac{1}{\sigma} \hat{G}^{-1} \hat{\epsilon}^T \ C
\]

\[
\hat{\rho}(E_n) = \sum_{\tau=1}^{N_T-1} g_{\tau}(E_n) \ C(a\tau) = \int_{E_0}^{\infty} d\omega \ K(E_n, \omega) \rho(\omega)
\]
\[
C(a\tau) = \int_{E_0}^{\infty} dE e^{-\tau a E} \rho(E) \quad \leftrightarrow \quad \sigma \sum_{m=0}^{N_E-1} e^{-\tau a E_m} \hat{\rho}(E_m)
\]

\[
\hat{G} \hat{\rho} = \frac{1}{\sigma} \hat{E}^T C, \quad \hat{\rho} = \frac{1}{\sigma} \hat{G}^{-1} \hat{E}^T C
\]

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

\[
K(E_n, E_m) = \frac{\delta_{nm}}{\sigma}, \quad K(E_n, \omega) \neq \delta(E_n - \omega)
\]
\[ C(t) = -\frac{1}{3} \sum_{i=1}^{3} \exp \left\{ -\frac{1}{2} \int_{0}^{\infty} E^2 R(E) e^{-Et} \mathrm{d}E \right\} \]
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 $O_1(f), \ldots, O_n(f)$ 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

$$\rho_L(E) = \sum_n w_n(L) \delta(E_n(L) - E), \quad \rho(E) \mapsto \rho[\Delta]$$
$$\rho_{\sigma,L}(E) = \int_0^\infty d\omega \, \Delta_\sigma(\omega, E) \rho_L(\omega)$$

$$\Delta_\sigma(\omega, E) = G_\sigma(\omega, E) = \frac{1}{\sqrt{2\pi} \sigma} \exp \left( -\frac{(E - \omega)^2}{2\sigma^2} \right)$$
\[ \rho(E) = \lim_{\sigma \to 0} \lim_{L \to \infty} \rho_{\sigma,L}(E) \]

**THE ORDER OF THE LIMITS IS IMPORTANT**
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

\[ \hat{\rho}_\sigma(E; L) = \int_{E_0}^\infty d\omega \Delta_\sigma(E, \omega) \rho_L(\omega) , \quad \rho(E) = \lim_{\sigma \to 0} \lim_{L \to \infty} \hat{\rho}_\sigma(E; L) \]
B64, $E = 0.79$ GeV, $\sigma = 0.44$ GeV

\[
d = 5.68 \times 10^{-4}
\]

\[
\frac{A_2[X^g]}{\Lambda_2[0]} = 10B[g]
\]

\[
d = 5.28 \times 10^{-3}
\]

\[
\frac{A_2[X^g]}{\Lambda_2[0]} = 10^5B[g]
\]

\[
d = 1.00 \times 10^{-2}
\]

\[
\frac{A_2[X^g]}{\Lambda_2[0]} = 10^6B[g]
\]
The effect of the noise regulator ($B[g]$ functional) manifests itself through large and oscillating coefficients. Extend arithmetic precision is required.

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

$\frac{d}{A_2[g]} = 3.58 \times 10^{-9}$

$B[g] = 0$

$\frac{A_2[g]}{A_2[0]} = 10B[g]$

$\frac{A_2[g]}{A_2[0]} = 10^4B[g]$

$\times 10^{\pm 5}$

$g(\tau)$

$\tau$
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


what we call the HLT method is the procedure to estimate reliably the errors
• 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

• what we call the HLT method is the procedure to estimate reliably the errors
Figure 9. Left: numerical results for $\rho(\epsilon)(E)$ for the Gaussian kernel and the spectral density including up to six-particle contributions (solid lines) smeared with the same kernel at different values of $\epsilon/(E-2m^*)$.

Right: results for $\rho(E)$ after extrapolation $\epsilon \to 0$, together with the exact two-particle contribution (light dashed line), the two-, four-, and six-particle contributions (dark solid line), and the 2-loop perturbative result (dark dotted line). Statistical and systematic errors due to the finite volume, continuum limit, and $\epsilon \to 0$ extrapolation are combined in quadrature as described in the text.

6 Conclusions

The aim of the preceding sections is to verify the procedure of ref. [1] for numerically computing smeared spectral densities (with an a priori specified smearing kernel) from lattice field theory correlation functions. In this regard the two-dimensional O(3) model usefully provides exact results against which the estimates can be checked. These checks, which are presented in figures 6 and 9, are satisfied and compare both $\rho(\epsilon)(E)$ at finite $\epsilon$ and the results from $\epsilon \to 0$ extrapolations to determine $\rho(E)$ deep into the inelastic region where finite-volume methods have not yet been developed. The highest energy considered here is $E = 40m^*$, at which $\rho(E)$ is determined with a relative accuracy of 5% and differs significantly from the exact two-particle contribution $\rho^{(2)}(E)$ given in eq. (2.6).

Apart from the 'usual' sources of systematic error due to the finite lattice spacing and finite-volume spacetime, we must also consider the imperfect reconstruction of the smearing kernel due to the finite number of input time slices and their associated statistical errors. All sources of systematic error have been estimated and included in figures 6 and 9 where the statistical and systematic errors are added in quadrature. Generally the errors due to the finite lattice extent are the largest source of systematic uncertainty, and are typically less than or comparable to the statistical errors.

The determination of $\rho(\epsilon)(E)$ becomes increasingly difficult for smaller smearing widths $\epsilon$ at fixed energy $E$, and increasing $E$ with fixed $\epsilon$. As is evident from the right two panels of figure 6, it is difficult to achieve precise results outside of the elastic region for $\epsilon \lesssim m/2$ with the Gaussian smearing kernel. Better is to exploit the smoothness of $\rho(E)$ and scale $\epsilon \propto (E-2m^*)$, so that an equal proportion of the smearing kernel 'leaks' down to the two particle threshold at each energy. This enables the determination of $\rho(E)$ in figure 9, which is the main result of this work.
Spacing and, therefore, and their correlation matrix is provided in the supplementary material.

Moreover, our theoretical results at different values of $R$ (red points) and $E$ (blue points) are shown in FIG. 3. This figure demonstrates the plots for central $R(E) = 0.44 \text{ GeV}$ (first row), $R(E) = 0.53 \text{ GeV}$ (second row), and $R(E) = 0.63 \text{ GeV}$ (third row).

Before ascribing this tension, of about three standard deviations, to new physics or to underestimated experimental errors, it is important to consider the contributions from the spectral reconstruction algorithm, continuum extrapolations, and finite volume effects. Although our theoretical errors, $\delta R$, are given by $\delta R = 0.63 \text{ GeV}$, those on the experimental values are smaller, $\delta R_{\text{exp}} = 0.44 \text{ GeV}$.

The error budget for $R_{\text{exp}}(E)$ is shown in FIG. 4. The uncertainties are calculated for each point and include contributions from theory and experiment.
Before ascribing this tension, of about three standard deviations, to new physics or to underestimated experimental errors, it is important to realize that the effects and also the ones coming from the spectral reconstruction algorithm, concentrate on the comparison of our first-principles results, obtained without using any input coming from hadron masses to fix the quark masses and the lattice parameters, that our lattice simulations have been calibrated by us-

\[
\Delta \left( E \right) / R \left( E \right) \\
\Delta^2 \left( E \right) / R \left( E \right) \\
\Delta^3 \left( E \right) / R \left( E \right) \\
\Delta^4 \left( E \right) / R \left( E \right) \\
\Delta^5 \left( E \right) / R \left( E \right)
\]

\( E \) [GeV]

ETMC Phys.Rev.Lett. 130 (2023)
TABLE I. ETMC gauge ensembles used in this work. The quoted pion masses have been obtained by a direct computation of the disconnected fermionic Wick contractions, together with a care-
ful study of the systematic uncertainties associated with continuum limit.

- **B64** 64 3 · 128 0.07957(13) 5.09 0.1352(2)
- **B96** 96 3 · 192 0.07957(13) 7.64 0.1352(2)
- **C80** 80 3 · 160 0.06821(13) 5.46 0.1349(3)
- **D96** 96 3 · 192 0.05692(12) 5.46 0.1351(3)

The lattice gauge ensembles used in this work are described in Ref. [7, 22] and analyze both the so-called correlated-constrained (red) and uncorrelated-unconstrained (blue) linear extrapolations in central values and covariance.

In the case of different experiments, no significant finite-volume effects were observed.

We considered three values for the central energies in the range $E = 0.44 \text{ GeV}$ and $E = 0.79 \text{ GeV}$ on the C80 ensemble and $E = 0.63 \text{ GeV}$ according to Eq. (1). The smearing Gaussian corresponding to center energy $E$ is $E \pm \sigma$.

In our lattice calculation we considered three values for the central energies in the range $E = 0.44 \text{ GeV}$ and $E = 0.79 \text{ GeV}$ on the C80 ensemble and $E = 0.63 \text{ GeV}$ according to Eq. (1). The smearing Gaussian corresponding to center energy $E$ is $E \pm \sigma$.

In order to compare our theoretical results with experimental values, from a multivariate Gaussian distribution using the KNT19 compilation [2] of the experimental data, we rely on the KNT19 compilation [2] of the experimental data.
Before ascribing this tension, of about three standard
precise to let us observe a tension of about three standard
right show the "pull"

\[\text{before ascribing this tension, of about three standard}
\]
\[\text{precision to let us observe a tension of about three standard}
\]
\[\text{before ascribing this tension, of about three standard}
\]
\[\text{precision to let us observe a tension of about three standard}
\]

Before ascribing this tension, of about three standard
precise to let us observe a tension of about three standard
right show the "pull"

Before ascribing this tension, of about three standard
precise to let us observe a tension of about three standard
right show the "pull"

Before ascribing this tension, of about three standard
precise to let us observe a tension of about three standard
right show the "pull"
\[ R_{ud}(\sigma) = 12\pi S_{EW} |V_{ud}|^2 \int_0^\infty \frac{dEE^2}{m_\tau^2} \left\{ K_T^\sigma \left( \frac{E}{m_\tau} \right) \rho_T(E^2) + K_L^\sigma \left( \frac{E}{m_\tau} \right) \rho_L(E^2) \right\} \]
The seminar is (mostly) based on:

1.6
1.7
1.8
1.9
2
2.1
0.01
R(⌧,TV)ud ()/|Vud|2
dT[gT]
↵=2, rmax=1↵=3, rmax=5↵=4, rmax=4↵=5, ...

FIG. 2: Stability-analysis plot for the three different contributions

A. Evangelista et al.
Phys. Rev. D 100, 074513 (2023)
\[ R_{ud}(\sigma) = 12\pi S E W |V_{ud}|^2 \int_0^\infty \frac{dE E^2}{m^3_{\tau}} \left\{ K_T^{\sigma} \left( \frac{E}{m_{\tau}} \right) \rho_T(E^2) + K_L^{\sigma} \left( \frac{E}{m_{\tau}} \right) \rho_L(E^2) \right\} \]

In light of the asymptotic-expansion formula of Eq. (48), we have carried out the extrapolation employing the light of the continuum extrapolated values of \( \rho_\tau \) and \( \rho_\tau \) and so on showing that their nonperturbative method can become a viable alternative to superallowed nuclear beta decays for obtaining \( V_{ud} \).

A. Evangelista et al.
\[ R_{ud}(\sigma) = 12\pi S_{EW} |V_{ud}|^2 \int_0^\infty \frac{dEE^2}{m_\tau^3} \left\{ K_T^\sigma \left( \frac{E}{m_\tau} \right) \rho_T(E^2) + K_L^\sigma \left( \frac{E}{m_\tau} \right) \rho_L(E^2) \right\} \]
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)
m.zhou et al. Phys.Rev.D 104 (2021)
s.chen et al. arXiv:2110.13521
t.lechien et al. SciPost Phys. 13 (2022)
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.

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 $\omega$ 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?

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

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

Here, $M$ denotes the mass of the corresponding state, $\Gamma$ 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 $N_{BW} = 3$ 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 hardly 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

$N_b = 4$
$N_\rho = 12$

$N_b = 6$
$N_\rho = \infty$

$N_b = \infty$
$N_\rho = 30$

$N_b = \infty$
$N_\rho = \infty$
we have chosen Chebyshev polynomials as basis functions

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

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

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

\[
\rho_\sigma(E) = \int_{E_0}^{\infty} d\omega K_\sigma(E, \omega) \rho(\omega),
\]

\[
K_\sigma(E, \omega) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(E-\omega)^2}{2\sigma^2}}, \quad \sigma = \{0.44, 0.63\} \text{ GeV}
\]
the ensemble of machines

- **The answer** of a machine with finite $N_n$ neurons, trained over a finite set $T_\sigma (N_b, N_\rho )$ cannot be exact

- To quantify the network error we therefore introduced $N_r$ replica machines at fixed $N = (N_n, N_b, N_\rho )$
training the ensemble of machines

\[ \mathcal{T}_\sigma(N_b, N_\rho) \]

SUPERVISED TRAINING

\[ C(t) \rightarrow \left| \hat{\rho}_\sigma^{\text{pred}}(E, N, r) - \hat{\rho}_\sigma^{\text{true}}(E) \right| \]
quoting predictions
let’s now consider a new $\rho$, again extracted on the Chebyshev basis but never seen during the trainings and this time with $N_b = 2 \times N_b^{\text{max}} = 1024$
let's now consider a new \( \rho \), again extracted on the Chebyshev basis but never seen during the trainings and this time with

\[
N_b = 2 \times N_{b}^{\text{max}} = 1024
\]
let’s now consider a new $\rho$, again extracted on the Chebyshev basis but never seen during the trainings and this time with

$$N_b = 2 \times N_b^{\text{max}} = 1024$$
let's now repeat the previous experiment 2000 times, with random $N_b$

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

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

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

going 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)
m.zhou et al. Phys.Rev.D 104 (2021)
s.chen et al. arXiv:2110.13521
t.lechien et al. SciPost Phys. 13 (2022)
• 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

---

**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 $\omega$ 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?

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

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

Here, $M$ denotes the mass of the corresponding state, $\Gamma$ 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 $N_{BW} = 3$ 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 hardly 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

\[ N_b = 4 \quad N_b = 6 \quad N_b = \infty \quad N_b = \infty \]

\[ N_\rho = 12 \quad N_\rho = \infty \quad N_\rho = 30 \quad N_\rho = \infty \]
we have chosen Chebyshev polynomials as basis functions

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

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

$N_b = 16$
$N_b = 32$
$N_b = 128$
$N_b = 512$
\[ \hat{\rho}_\sigma(E) = \int_{E_0}^\infty d\omega \, G_\sigma(E, \omega) \rho(\omega) \]

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

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

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

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

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

\[ \rho^i(E; N_b) \rightarrow (C, \hat{\rho}_\sigma)^i \]

\[ \mathbb{C} \left[ C^i, \left( \frac{C^i(a)}{C_{\text{latt}}(a)} \right)^2 \sigma_{\text{latt}} \right] \rightarrow C^i_{\text{noisy}} \]

\[ (C_{\text{noisy}}, \hat{\rho}_\sigma)^i \in \mathcal{T}_\sigma(N_b, N_\rho), \quad i = 1, \cdots, N_\rho \]
we considered 3 architectures with sizes in the proportion

\[ N_n^{arcS} : N_n^{arcM} : N_n^{arcL} = 1 : 2 : 3 \]

<table>
<thead>
<tr>
<th>Type</th>
<th>Maps</th>
<th>Size</th>
<th>Kernel size</th>
<th>Stride</th>
<th>Activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>64</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conv1D</td>
<td>2</td>
<td>32x2</td>
<td>3</td>
<td>2</td>
<td>LeakyReLu</td>
</tr>
<tr>
<td>Conv1D</td>
<td>4</td>
<td>16x4</td>
<td>3</td>
<td>2</td>
<td>LeakyReLu</td>
</tr>
<tr>
<td>Conv1D</td>
<td>8</td>
<td>8x8</td>
<td>3</td>
<td>2</td>
<td>LeakyReLu</td>
</tr>
<tr>
<td>Flatten</td>
<td>384</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fully conn.</td>
<td>256</td>
<td></td>
<td></td>
<td></td>
<td>LeakyReLu</td>
</tr>
<tr>
<td>Fully conn.</td>
<td>256</td>
<td></td>
<td></td>
<td></td>
<td>LeakyReLu</td>
</tr>
<tr>
<td>Output</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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 $\mathcal{T}_\sigma(N_b, N_\rho)$ cannot be exact

- to quantify the network error we therefore introduced $N_r$ replica machines at fixed $\mathcal{N} = (N_n, N_b, N_\rho)$
training the ensemble of machines

\[ \mathcal{T}_\sigma(N_b, N_\rho) \]

SUPERVISED TRAINING

\[ C(t) \rightarrow N \rightarrow N_\rho \]

\[ \hat{\rho}_\sigma^{\text{pred}}(E, N, r) - \hat{\rho}_\sigma^{\text{true}}(E) \]
\[ \ell(w) = \frac{1}{N_\rho} \sum_{i=1}^{N_\rho} |\hat{\rho}_{\sigma}^{\text{pred},i}(w) - \hat{\rho}_{\sigma}^i| \]

\( N_r = 20 \)

\( N_b = \{16, 32, 128, 512\} \)

\( N_\rho = \{50, 100, 200, 400, 800\} \times 10^3 \)
quoting predictions
let's now consider a new $\rho$, again extracted on the Chebyshev basis but never seen during the trainings and this time with

$$N_b = 2 \times N_b^{\text{max}} = 1024$$
let’s now consider a new $\rho$, again extracted on the Chebyshev basis but never seen during the trainings and this time with 

$$N_b = 2 \times N_b^{\text{max}} = 1024$$
let's now consider a new $\rho$, again extracted on the Chebyshev basis but never seen during the trainings and this time with

$$N_b = 2 \times N_b^{\text{max}} = 1024$$
let's now repeat the previous experiment 2000 times, with random $N_b$

$$p_\sigma(E) = \frac{\hat{p}_\sigma^{\text{pred}}(E) - \hat{p}_\sigma^{\text{true}}(E)}{\Delta_\sigma^{\text{tot}}(E)}$$
let's now consider another 2000 random unsmeared spectral densities mimicking what we can get on finite volumes

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

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

$$E_n \in [E_0, 15] \ \text{GeV} , \quad c_n \in [-0.01, 0.01]$$
let’s now consider another 2000 random unsmeared spectral densities mimicking what we can get on finite volumes

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

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

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

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.
<table>
<thead>
<tr>
<th>ID</th>
<th>$L^3 \times T$</th>
<th>$a$ fm</th>
<th>$aL$ fm</th>
<th>$m_\pi$ GeV</th>
</tr>
</thead>
<tbody>
<tr>
<td>B64</td>
<td>$64^3 \cdot 128$</td>
<td>0.07957(13)</td>
<td>5.09</td>
<td>0.1352(2)</td>
</tr>
</tbody>
</table>

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

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

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

![Graph showing \( \hat{R}_\sigma(E) \) for different values of \( \sigma \)]
• 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!

\[
\begin{align*}
C_c(t) & \rightarrow N \rightarrow \hat{\rho}_{\text{pred}}^{\sigma}(E, N, c, r) \\
\hat{\rho}_{\text{pred}}^{\sigma}(E, N, c, r) & \rightarrow \Delta_{\text{latt}}^{\sigma}(E, N) + \Delta_{\text{net}}^{\sigma}(E, N) \rightarrow \hat{\rho}_{\text{pred}}^{\sigma}(E, N) \pm \Delta_{\text{stat}}^{\sigma}(E, N) \\
\hat{\rho}_{\text{pred}}^{\sigma}(E, N) \pm \Delta_{\text{stat}}^{\sigma}(E, N) & \rightarrow \hat{\rho}_{\text{true}}^{\sigma}(E) \pm \Delta_{\text{tot}}^{\sigma}(E) \\
\end{align*}
\]
• 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 $\mathcal{T}_\sigma(N_b, N_\rho)$ as follows

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

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

$$C_{\text{noisy}}(t) \mapsto C'_{\text{noisy}}(t) = \frac{C_{\text{noisy}}(t) - \mu(t)}{\gamma(t)}$$
For each $G_{gt}(\tau)$ we add a Gaussian noise constructing the stochastic data $G(\tau)$ to be fed to the "Encoder 2" (cf. Fig. 2). Thus $G(\tau)$ is sampled from a Gaussian distribution with its mean value $G_{gt}(\tau)$ and its variance having the following form.

Here $b(\tau) = lat(\tau)/G_{lat}(\tau)$ and the variance is chosen to mimic the noise level of the lattice QCD data, i.e. $b(\tau)$ becomes larger as $|\tau|N/2$ becomes smaller and the largest relative error is $b(N\tau/2) = 1.5\%$.

Note the values of the noise level $b(\tau)$ in the training data are the same as used in the mock data tests for consistency.

The feeding process in each training epoch described in procedures 1 and 2 is performed simultaneously with one mini-batch containing 100 training samples.

For each value of $N\tau$ we repeat the training process described in procedures 1, 2 and 3.

The approach closest in spirit to the one that we propose is that of s.chen et al. arXiv:2110.13521
Figure 6. Mock data test with input spectral function containing a resonance peak and a continuum part (cf. Eq. (27) with $N = 96$. The black dashed line denotes the input spectral function. From left to right the width of the resonance peak in the input spectral function, $\sigma$, is increased with the peak location of the resonance peak, $M_{\text{res}}$, fixed in each row, while from bottom to top $M_{\text{res}}$ is increased with $\sigma$ fixed in each column. The red solid line and purple band represent the mean values and uncertainties of spectral functions reconstructed from the SVAE, respectively. The black solid line and blue band denote the mean values and uncertainties of spectral functions reconstructed from the MEM with the blue dotted line the default model.

different combinations: either Gaussian noise or log-normal is used in the training or test process. We found that results of spectral functions obtained in following cases have minor difference compared to our current results with Gaussian noise used in both training and tests. These cases are: 1) the log-normal noise is used in both the training and tests; 2) a mismatch in the noise model: Gaussian noise is used in the training while log-normal noise is used in the tests and vice versa and 3) the multivariate Gaussian noise, i.e. including correlations among different time slices, the approach closest in spirit to the one that we propose is that of s.chen et al. arXiv:2110.13521