Nuclear Science Computing Center at CCNU

Machine Learning Hadron Spectral Functions in Lattice QCD

Shi-Yang Chen[1]

H.-T. Ding[1], F.-Y. Liu[1,2], G. Papp[2] and C.-B.Yang[1]

[1] Central China Normal University

[2] Eötvös Loránd University

LATTICE21

JULY 26-30 2021, ZOOM/GATHER@MIT

Outline

1. **Hadron Spectral functions and Correlators**

2. **Application via neural net**
    - Construction of the neural net
    - Training and mock data tests
    - Application to lattice data

3. **Summary and outlook**

Hadron Spectral functions and Correlators

Hadron spectral functions: Carry all information about hadrons

1. **Quarkonia dissociation temperature** [T. Matsui & H. Satz, PLB 178, 416 (1986)]
2. **Heavy Quark diffusion coefficient** [P. Petreczky & D. Teaney, PRD 73,1649 (2006)]

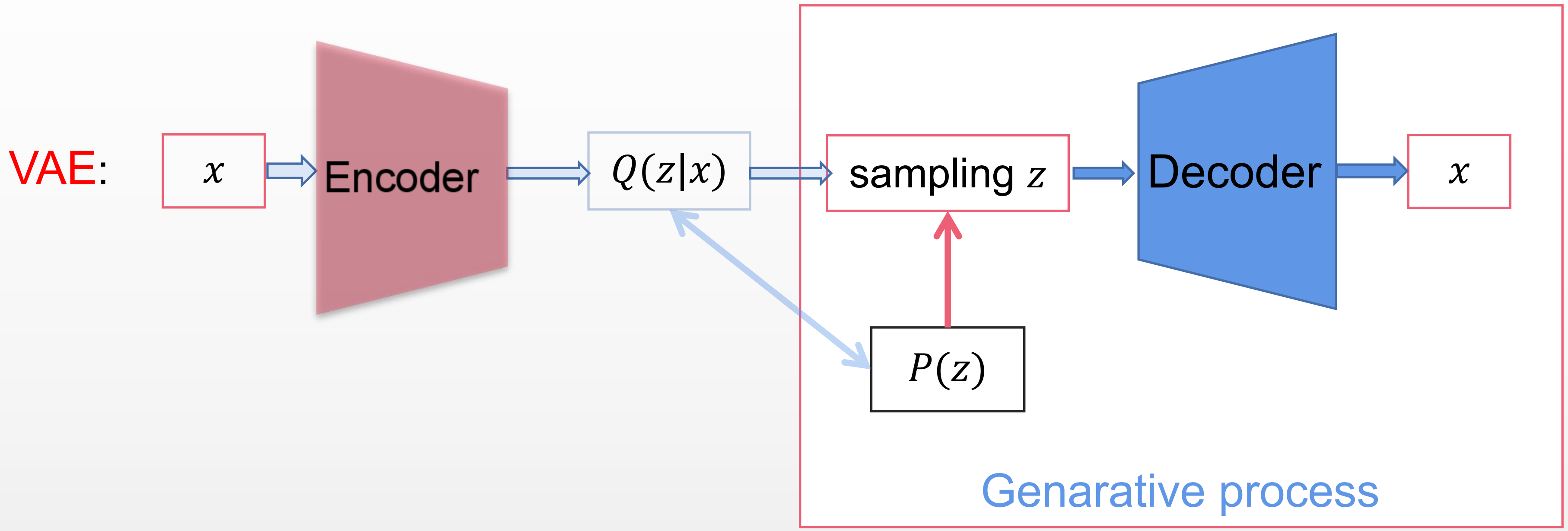
$$\underbrace{G(\tau,T)}_{\mathcal{O}(10)} = \sum_{x,y,z} \langle J_H(0,\vec{0}) J^+(\tau,\vec{x}) \rangle_T = \int_0^\infty \frac{d\omega}{2\pi} K(\tau,\omega,T) \underbrace{\rho(\omega)}_{\mathcal{O}(1000)}$$

reconstructing $\rho(\omega)$:inverse problem

Methods available in the literatures:

1. Maximum Entropy Method (MEM) [Asakawa, Hatsuda & Nakahara, '01]
2. New Bayesian Method (improved MEM) [Y.Burnier & A.Rothkopf, PRL 111,182003 (2013)]
3. Stochastic Approaches [H.-T. Ding et al., PRD 97, 094503 (2018)]
4. ...

Variational AutoEncoder (VAE) [C.Doersch, arXiv:1606.05908]



$P(z)$:Normal distirbution

Natural tool to tackle the inverse problem.

Our Goal:

1. find certain constraints on latent space Z.
2. build a modified VAE (mVAE) network.

Loss function

$$-\log P(\rho|G) = -\log\left[\int P(\rho, z|G)\ dz\right] = -\log\left[\int P(\rho|z,G)P(z|G)\ dz\right]$$

$$= -\log\left[\int P(\rho|z,G)\frac{P(z|G)}{Q(z|\rho_{gt},G_{gt})}Q(z|\rho_{gt},G_{gt})dz\right]$$

Jensen's inequality

$$\leq -\int Q(z|\rho_{gt},G_{gt})\log\left[P(\rho|z,G)\frac{P(z|G)}{Q(z|\rho_{gt},G_{gt})}\right]dz$$

$$= -E_{z\sim Q(z|\rho_{gt},G_{gt})}[\log P(\rho|z,G)\ ] + KL(Q(z|\rho_{gt},G_{gt})||P(z|G))$$

ρ_{gt}: ground truth value of the spectral function.
G_{gt}: correlator obtained from ρ_{gt}

$$E_{z\sim Q}[\log P\ ] = \int Q\log P\ dz$$

$$KL(Q||P) = \int Q\log\frac{Q}{P}dz$$

Loss function

$$-\log P(\rho|G) \leq -E_{z\sim Q(z|\rho_{gt},G_{gt})}[\log P(\rho|z,G)] + KL(Q(z|\rho_{gt},G_{gt})||P(z|G))$$

$$= -E_{z\sim Q(z|\rho_{gt},G_{gt})}\left[\log \frac{P(\rho|z)P(G|\rho,z)}{P(G|z)}\right] + KL(Q(z|\rho_{gt},G_{gt})||P(z|G))$$

Loss function

Prior information: $P(\rho|z) = Z_S e^{S}$, $\quad S = \int d\omega\, \rho_{gt} - \rho(z) - \rho(z) log \frac{\rho(z)}{\rho_{gt}}$.

Likelihood: $P(G|\rho,z) = Z_L e^{-L}$, $\quad L = \sum_\tau \frac{(G[\rho(z)] - G_{gt})^2}{\alpha^2(z)G^2}$.

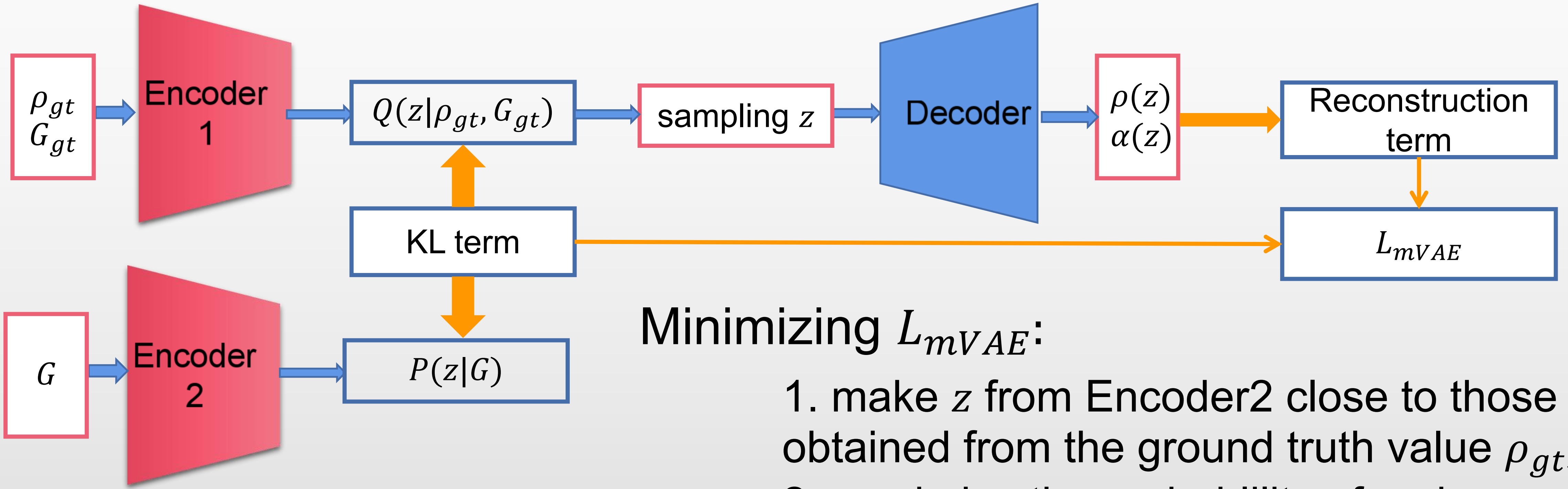
$$Q(z|\rho_{gt},G_{gt}) = \sum_i \prod_j \frac{\pi_i}{\sqrt{2\pi\sigma_{i,j}^2}} exp\left(-\frac{(z-\mu_{i,j})^2}{2\sigma_{i,j}^2}\right),$$

similar functional form for $P(z|G)$

Topology of Network: Training

Reconstruction term　　KL term

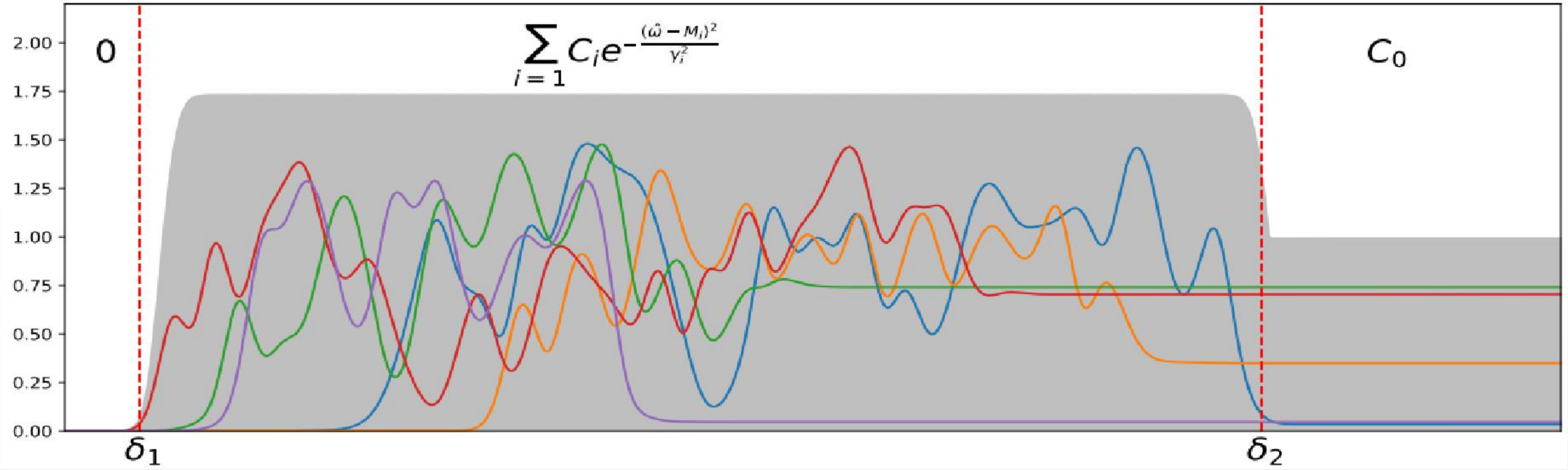
$$L_{mVAE} = -E_{z\sim Q(z|\rho_{gt},G_{gt})}\left[\log\frac{P(\rho|z)P(G|\rho,z)}{P(G|z)}\right] + KL(Q(z|\rho_{gt},G_{gt})||P(z|G))$$



Minimizing L_{mVAE}:

1. make z from Encoder2 close to those obtained from the ground truth value ρ_{gt}, G_{gt}
2. maximize the probablility of ρ given z, G

Training samples



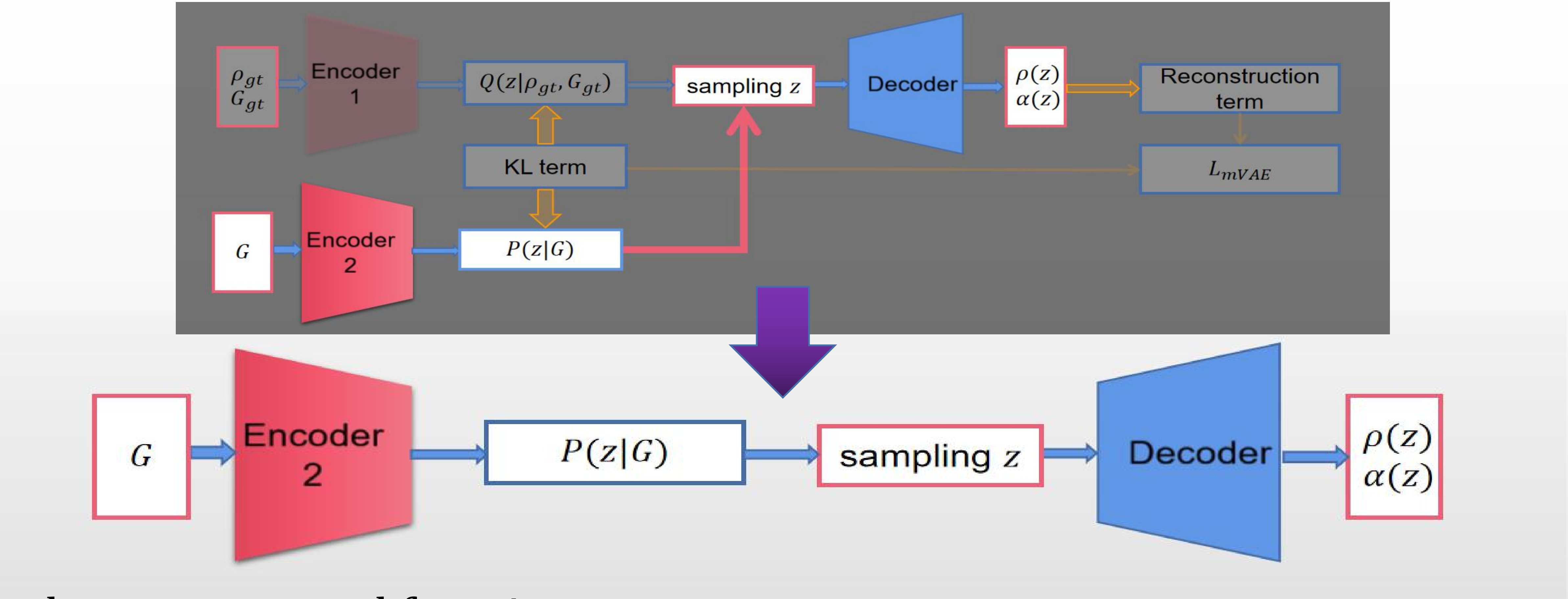
$$\rho_{gt} = \theta(\hat{\omega}, \delta_1, \zeta_1)\left[\sum_{i=1} C_i e^{-\left(\frac{\hat{\omega}-M_i}{\gamma_i}\right)^2}\left(1-\theta(\hat{\omega}, \delta_2, \zeta_2)\right) + C_0\theta(\hat{\omega}, \delta_2, \zeta_2)\right],$$

$$\theta(\hat{\omega}, \delta, \zeta) = \frac{1}{1+e^{-\frac{\hat{\omega}-\delta}{\zeta}}}$$

Parameter	Interval	Parameter	Interval
$\delta_1 \approx [M_c, 2M_c]$	[0.05,0.30]	$\delta_2 - \delta_1 \approx 2M_c$	[0.15,0.6]
$\zeta_{i=1,2} \approx 0.1M_c$	[0.005,0.015]	$C_{i=0,1,2,\ldots}$	[0,1]
$M_i \sim [M_c, 6M_c]$	[0.05, 0.08]	γ_i	0.75/N

M_c: charm quark mass in lattice spacing ($a^{-1} = 20GeV$); N: *the number of Gaussian peaks* ($N = 50$)

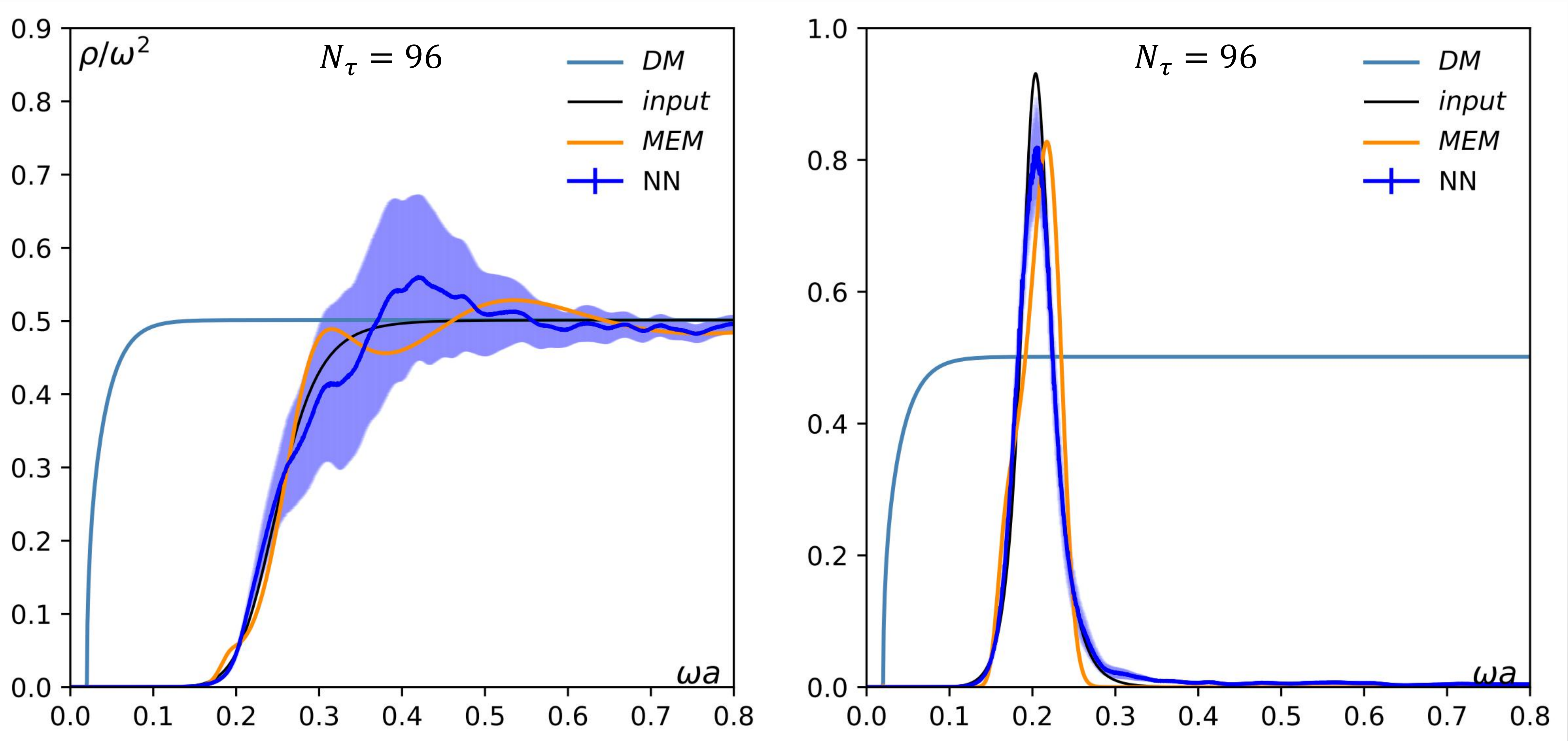
Topology of Network: Reconstruction



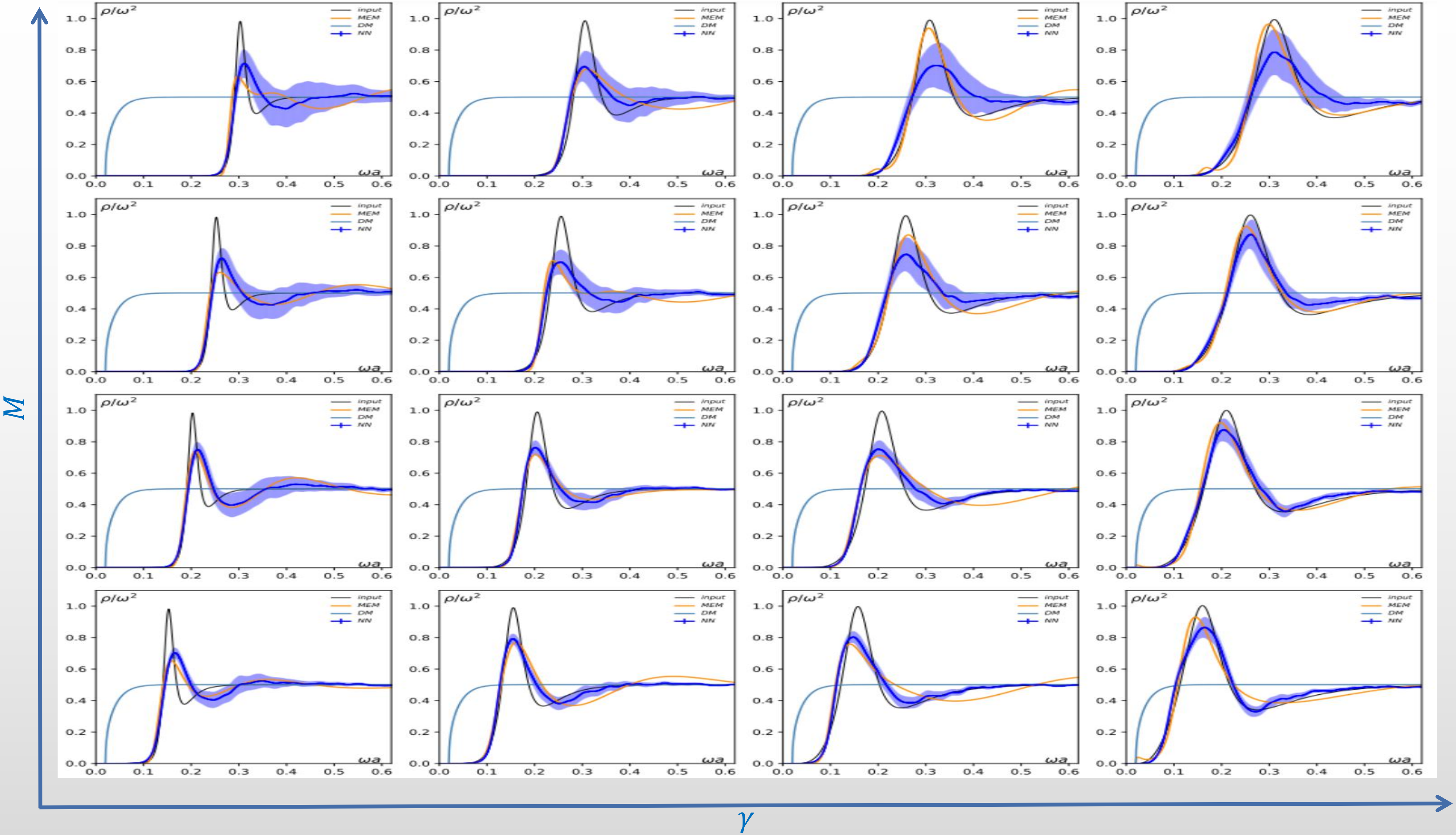
Final output spectral function:

$$\bar{\rho} = \int \rho(z)P(z|G)P(\rho|z,G)dz = \int \rho(z)P(z|G)\frac{P(\rho|z)P(G|z,\rho)}{\int P(\rho|z)P(G|z,\rho)dz}dz$$

Continuum and single peak mock data ($N_\tau = 96$)

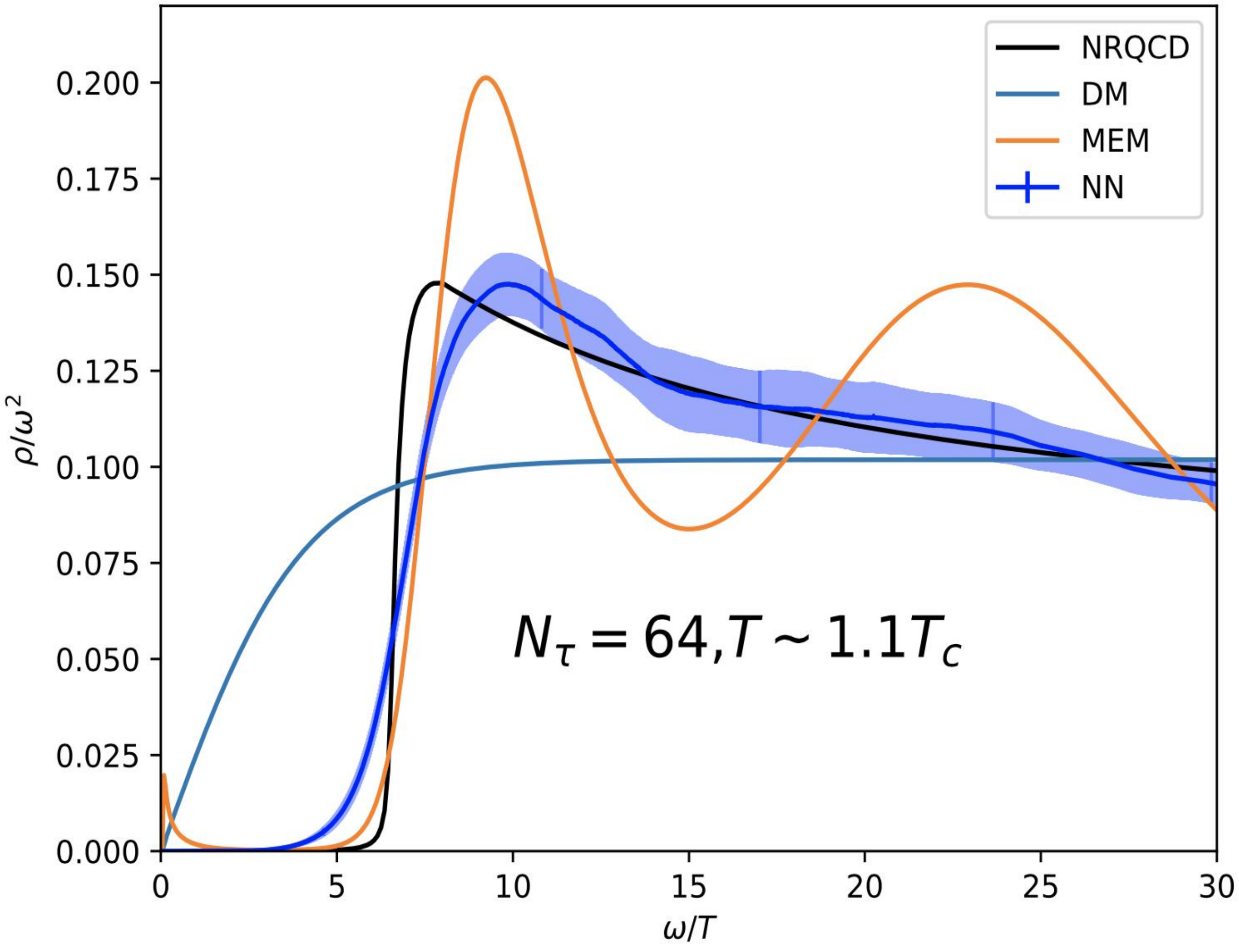


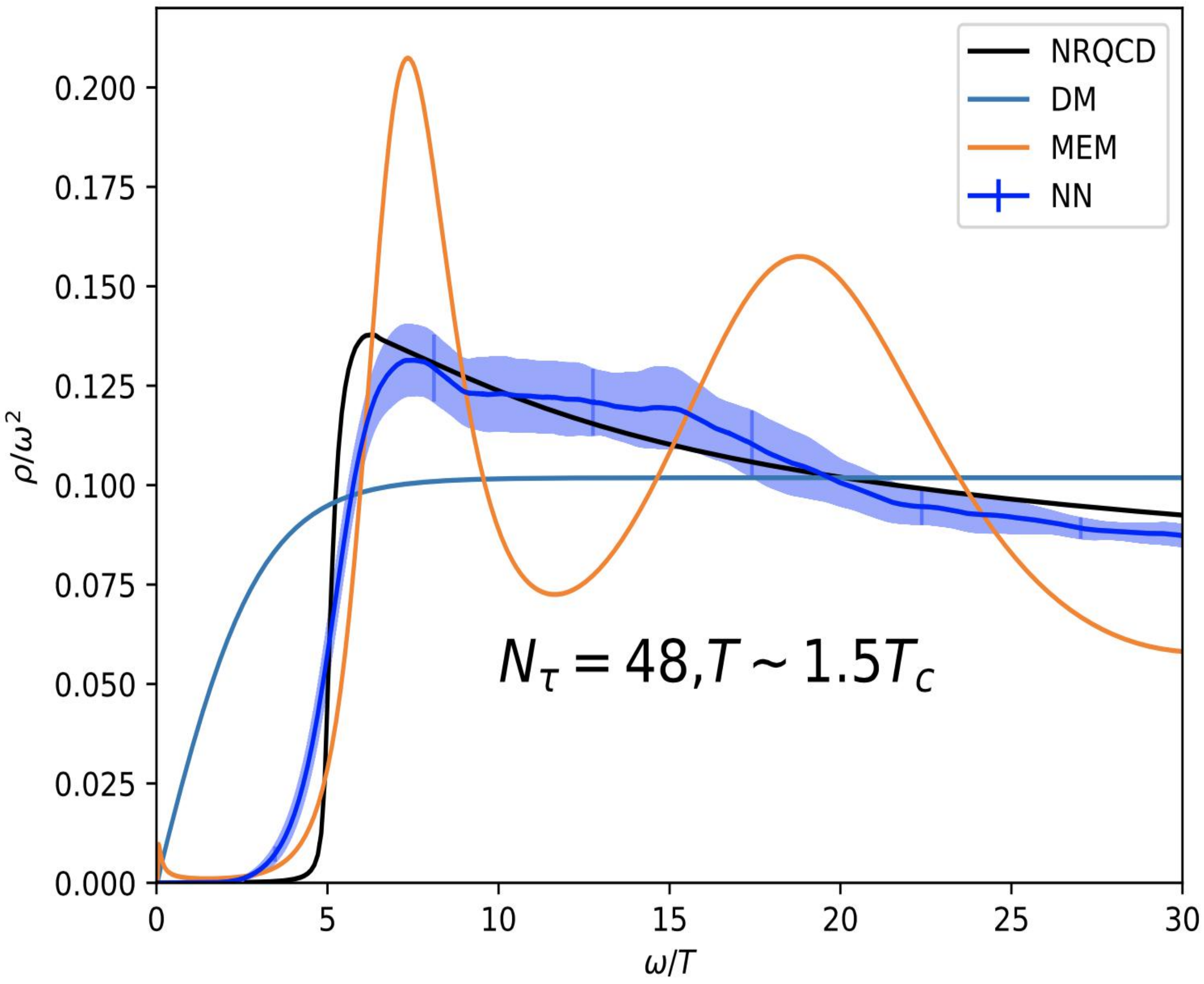
Single peak plus continuum mock data($N_\tau = 96$)



NRQCD motivated mock data

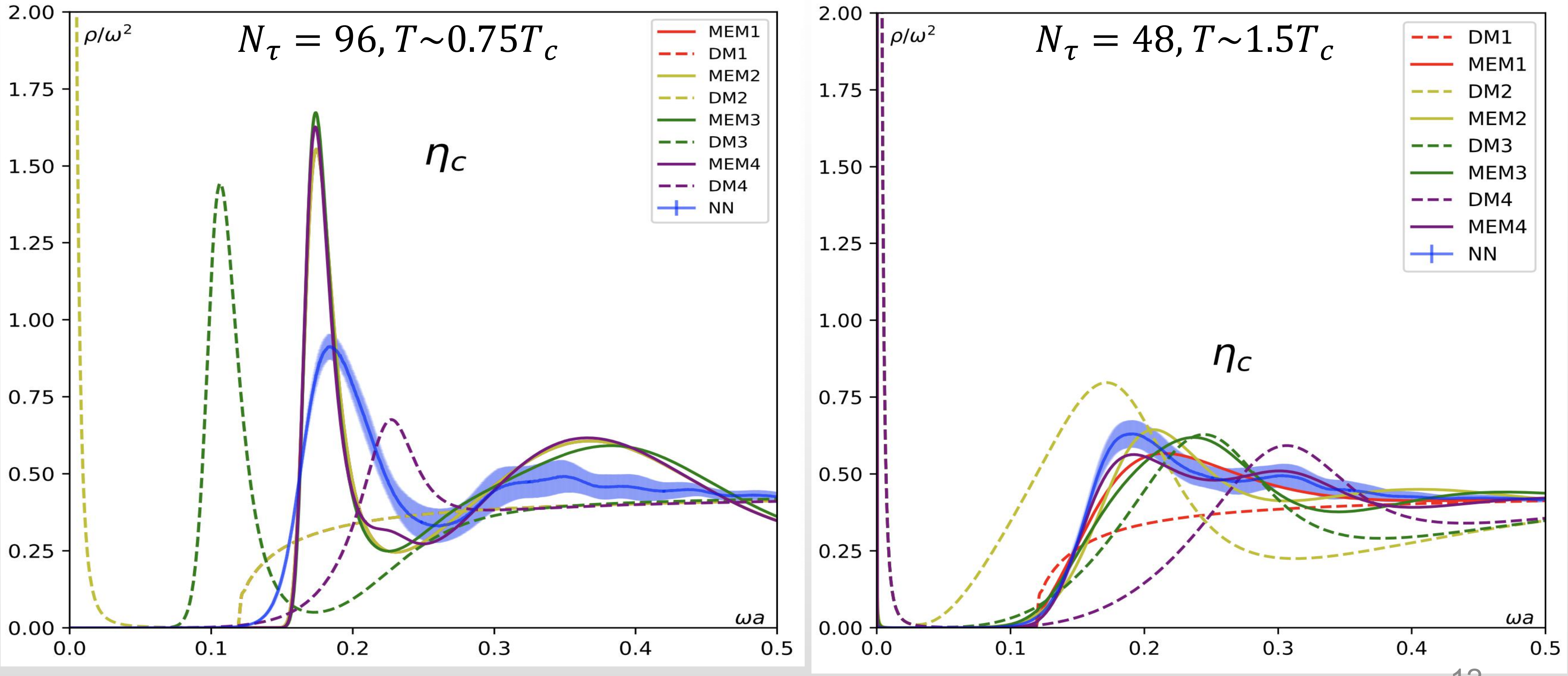
NRQCD motivated mock data: Y. Burnier et al., JHEP 1711 (2017) 206





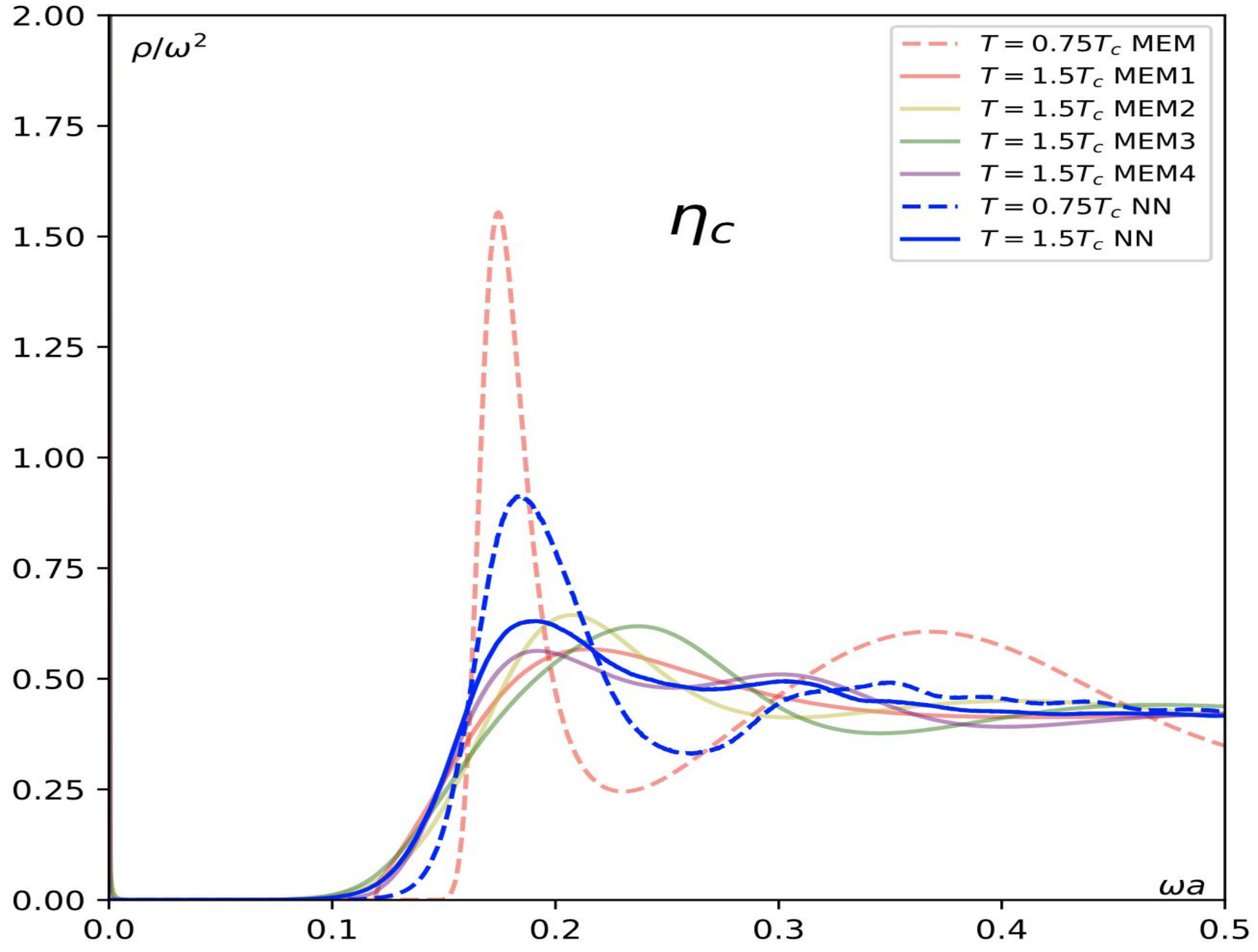
Hadron spectral functions from lattice data

Lattice data: [H.-T. Ding, et al., RRD 86, 014509, 07 2012]
MEM results and DMs: [H.-T. Ding, et al., PRD 97, 094503 (2018)]



Hadron spectral functions from lattice data

Lattice data: [H.-T. Ding, et al., RRD 86, 014509, 07 2012]
MEM results and DMs: [H.T. Ding, et al., PRD 97, 094503 (2018)]



Summary

- To extract spectral functions ρ, we designed a neural network (NN) based on the Variational autoEncoder and Bayes' theorem including an entropy term with ρ_{gt} included
- In the training process, we use a general ρ made of Gaussian peaks
- In the mock data tests, we found that results from the NN are comparable to those from the MEM
- Output spectral functions from both the NN and MEM suggest the thermal modification of η_c from 0.75 Tc to 1.5 Tc

Outlook

- Extend the study on the ρ including the transport peak
- Study correlators at T=0 where $K(\tau, \omega, T) = e^{-\omega\tau}$ is much simpler

Thanks

Appendix1:loss function

$$L_{mVAE} =- E_{z\sim P(z|\rho_{gt},G_{gt})}\left[log\ \frac{P(\rho|z)P(G|\rho,z)}{P(G|z)}\right]- \ kl(P(\rho|z.G)||P(z|G))$$

$$\leq \int\left\{\left[\frac{\alpha^2(z)}{p}\right]^{\frac{1}{p}}+\sum_\tau\frac{(G[\rho(z)]-G)^2}{\alpha^2(z)G^2}+\ \sum_j\rho_{gt,j}-\rho_j(z)-\rho_j(z)log\frac{\rho_j(z)}{\rho_{gt,j}}\right\}P(z|\rho_{gt},G_{gt})dz$$

$$+\sum_i\left\{\pi_i\sum_k\left[\frac{(\mu_k-\hat{\mu}_k)^2}{2\hat{\sigma}_k^2}+\frac{1}{2}log\frac{\hat{\sigma}_k^2}{\sigma_k^2}+\frac{\sigma_k^2}{2\hat{\sigma}_k^2}+\frac{1}{2}\right]\right\}$$

where $P(G|z)$ is normalization function and,

$$P(G|z)\ =\ \int d\rho(z)\frac{Z_S}{\sqrt{2\pi\alpha^2(z)}}e^{S-L}\leq\left[\int d\rho(z)\,Z_Se^{pS}\right]^{\frac{1}{p}}\left[\int d\rho(z)\ \frac{e^{-qL}}{\sqrt{2\pi\alpha^2(z)G^2}}\right]^{\frac{1}{q}}\propto\left[\frac{\alpha^2(z)}{p}\right]^{\frac{1}{p}}.$$

and $p\geq 1$ is a hyperparameter, and $\frac{1}{p}+\frac{1}{q}=1.$

$kl(P(\rho|z.G)||P(z|G)$ term can be simplified as ,

$$kl(P(\rho|z.G)||P(z|G))\leq\sum_i\left\{\pi_i\sum_k\left[\frac{(\mu_k-\hat{\mu}_k)^2}{2\hat{\sigma}_k^2}+\frac{1}{2}log\frac{\hat{\sigma}_k^2}{\sigma_k^2}+\frac{\sigma_k^2}{2\hat{\sigma}_k^2}+\frac{1}{2}\right]+log\frac{\pi_i}{\pi_i}\right\}=\sum_i\left\{\pi_i\sum_k\left[\frac{(\mu_k-\hat{\mu}_k)^2}{2\hat{\sigma}_k^2}+\frac{1}{2}log\frac{\hat{\sigma}_k^2}{\sigma_k^2}+\frac{\sigma_k^2}{2\hat{\sigma}_k^2}+\frac{1}{2}\right]\right\}$$

Appendix2:The uniqueness of final result

Recalling the definition of $P(G|z)$ *and* $P(G|\rho,z)$*, thus,* $P(G|z)$ *is the function of* $\alpha(z)$*.*

$$P(G|z)=\int d\rho(z)P(\rho|z)P(G|\rho,z)=\int d\rho(z)\frac{Z_s e^{S-L}}{\sqrt{2\pi\alpha^2(z)}};\qquad P(G|\rho,z)=\frac{e^{-\sum_\tau\frac{(G[\rho(z)]-G)^2}{\alpha^2(z)G^2}}}{\sqrt{2\pi\alpha^2(z)G^2}}$$

And the loss function has only one stationary points for $\alpha(z)$.

$$\frac{\sqrt{p}|G[\rho(z)]-G|}{G}=arg\,min_{\alpha(z)}\,L_{mVAE}$$

It means any optimal soultions for L_{mVAE}, $P(G|z)=P'(G|z)$ and $P(G|\rho,z)=P'(G|\rho,z)$.

$$P(z|G)=\frac{P(G|z)P(z)}{P(G)}=\frac{P'(G|z)P(z)}{P(G)}=P'(z|G)$$

L_{mVAE} *has an unique Encoder2*

$$L_{mVAE}=E_{z\sim P(z|\rho_{gt},G_{gt})}[log\,P'(\rho|z.G)]+kl(P'(z|\rho.G)||P'(z|G))$$
$$=E_{z\sim P(z|\rho_{gt},G_{gt})}[log\,P'(\rho|z.G)]+kl(P'(z|\rho.G)||P(z|G))$$

$kl(P'(\rho|z.G)||P(z|G))$ is a convex function, thus, $P'(z|\rho.G)=P(z|\rho.G)$, and

L_{mVAE} *has an unique Encoder1*

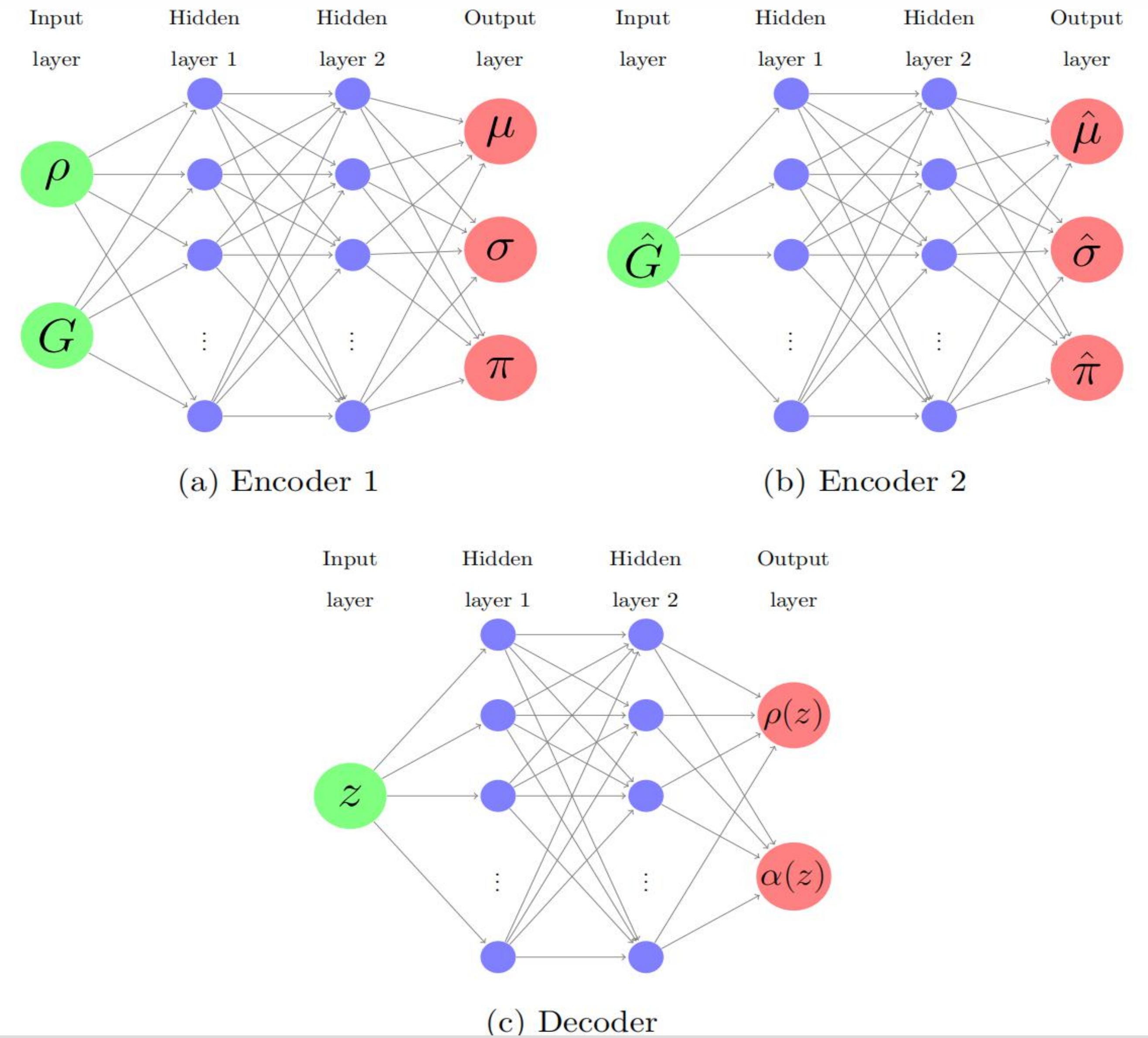
$$P(\rho|z,G)=\frac{P(\rho|z)P(G|\rho,z)}{P(G|z)}$$

Thus unique Encoder2 + unique Encoder1 → unique Decoder or $P(\rho|z)=P'(\rho|z)$*,*
neglecting internal symmetry $\gamma: z\rightarrow\gamma(z)$*, s.t.* $P'(z|G)=P(\gamma(z)|G)=P(z|G)$*.*

Thus, network has an uniquely final output:

$$\bar{\rho}'=\int dz\;\rho(z)P(\rho|G)=\int dz\,\rho(z)P'(z|G)\frac{P'(\rho|z)P'(G|\rho,z)}{P'(G|z)}=\int dz\,\rho(z)P(z|G)\frac{P(\rho|z)P(G|\rho,z)}{P(G|z)}=\bar{\rho}$$

Appendix3:Subnetwork stucture



(a) Encoder 1 (b) Encoder 2

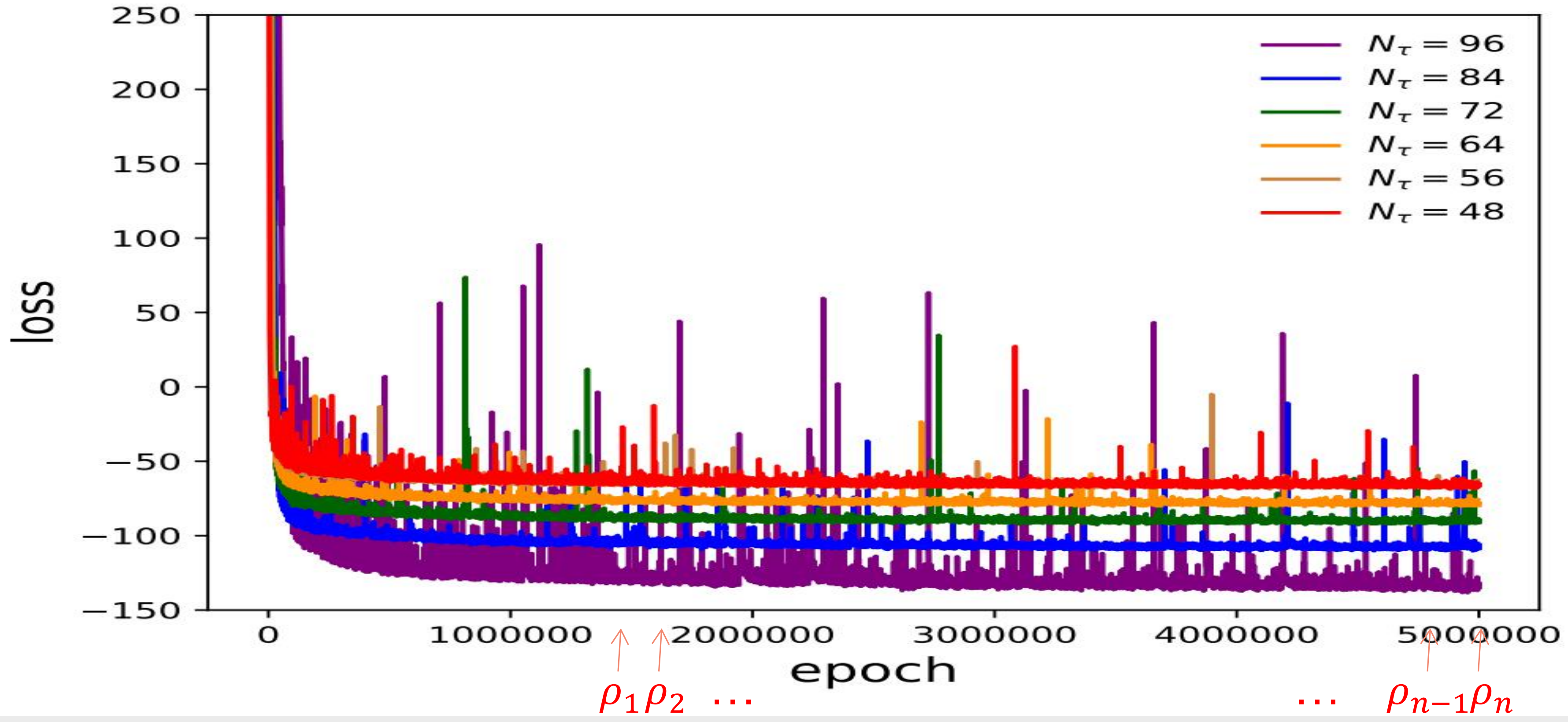
(c) Decoder

Appendix3:Hyperparameters

Network \ Layer	Input Layer	Hidden Layer 1	Hidden Layer 2	Output Layer
Encoder 1	3000: ρ_i	1500	750	350
	$43(T<T_c)/29(T>T_c)$: G_k			350
Encoder 2	$43(T<T_c)$ $/29(T>T_c)$: $\hat{G}_k$	150	200	350
				350
Decoder	350: z_j	750	1500	3000
				3000

Subnet \ Layer	Layer 1	Layer 2	Output Layer
Encoder1	Prelu(x)	Prelu(x)	None
			$\exp(x/2)$
Encoder2	Prelu(x)	Prelu(x)	None
			$\exp(x/2)$
Decoder	Prelu(x)	Prelu(x)	$\mathrm{Max}(0, x^2 - 0.0001)$
			$\exp(x/2)$

Appendix4:Time history and error



Systematic error: the standard deviation of $\rho_1, \rho_2, ..., \rho_{n-1}, \rho_n$