



Nuclear Science  
Computing Center at CCNU



# Machine Learning Hadron Spectral Functions in Lattice QCD

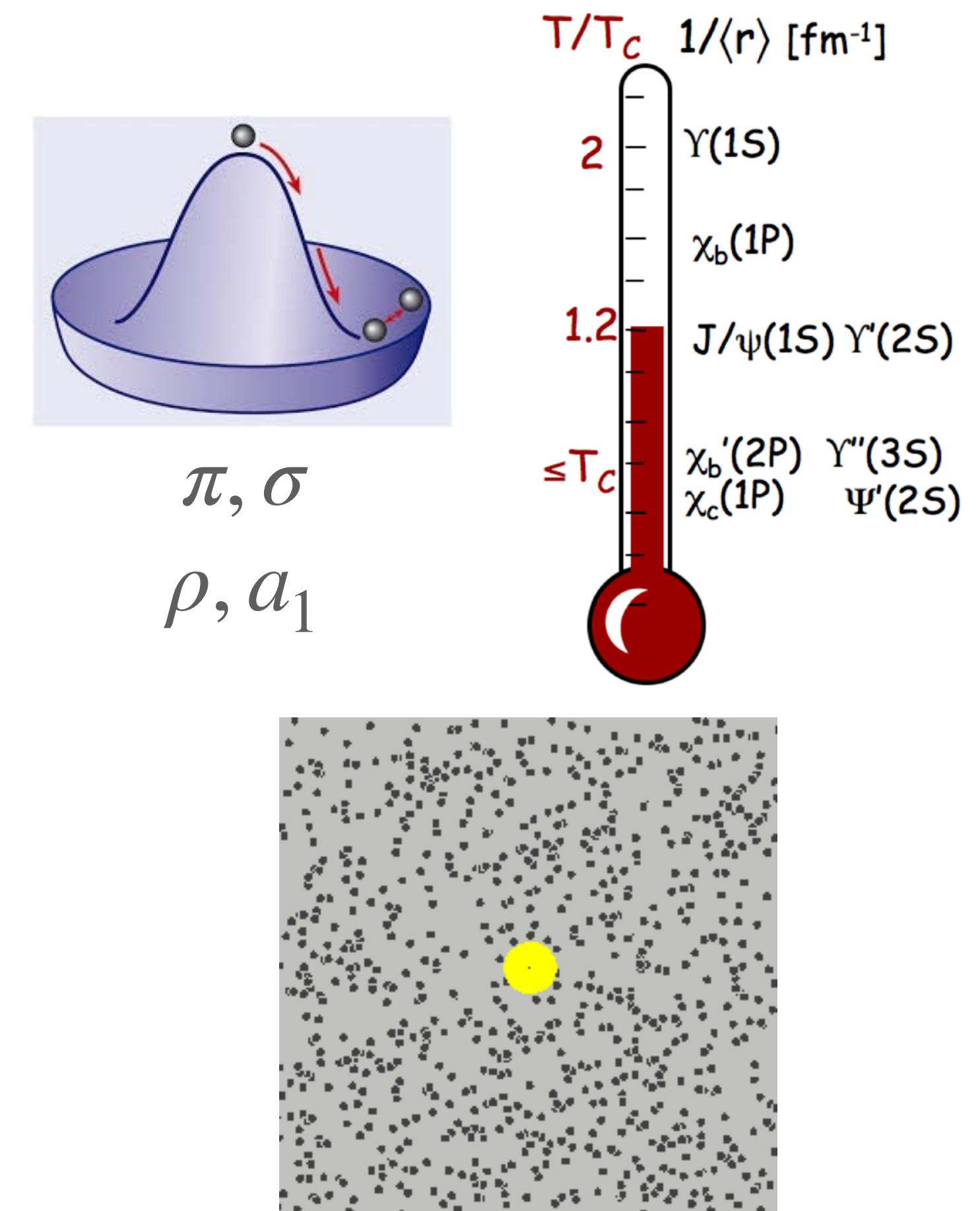
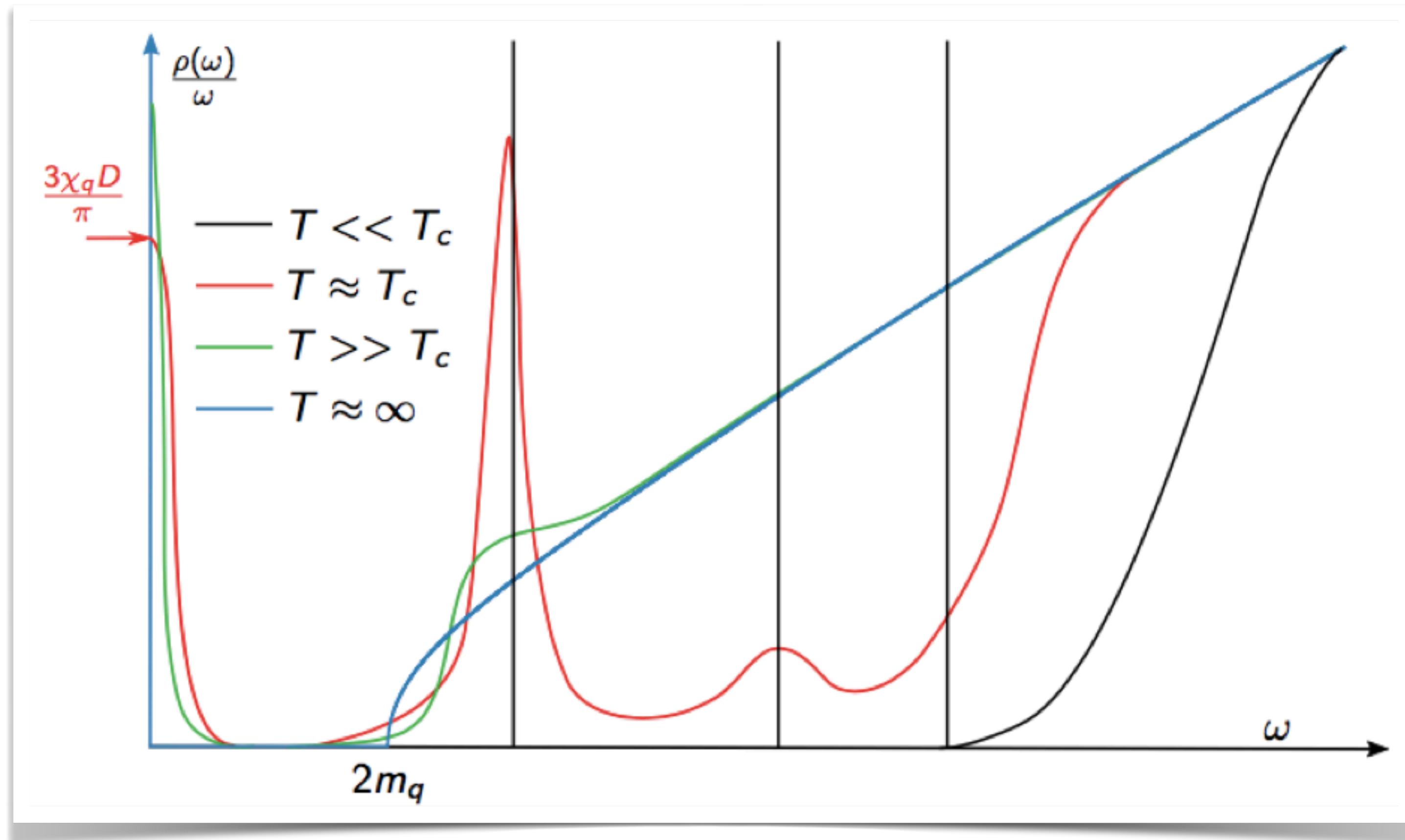
Heng-Tong Ding  
Central China Normal University

based on arXiv:2110.13521 & in collaboration with  
Shi-Yang Chen, Fei-Yi Liu, Gabor Papp, Chun-Bin Yang

workshop on “Efficient simulations on GPU hardware”  
Oct. 24-27, 2022 @ ETHZ

# Hadron Spectral Functions (SPF)

Dissociation of hadrons in the hot medium, diffusion of charges/quarks, symmetry restoration, electrical conductivity, thermal photon/dilepton rates...

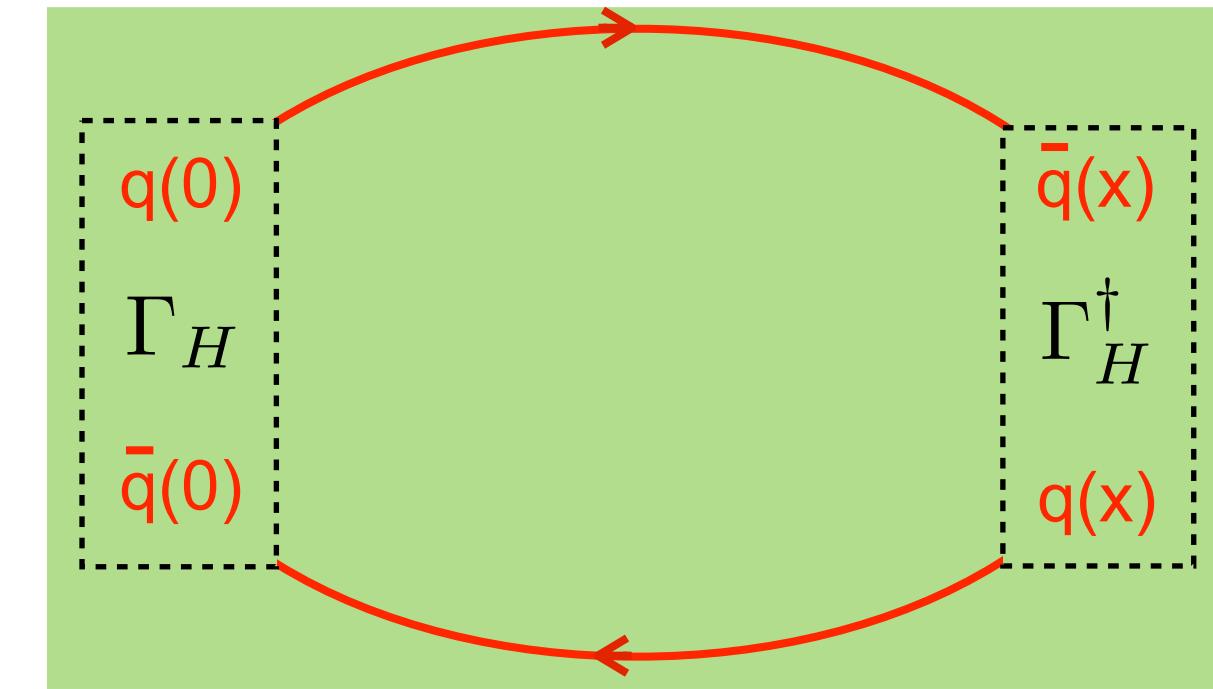


by courtesy of Hai-Tao Shu

# Temporal correlation and spectral functions

$$G(\tau, \vec{p}, T) = \sum_{\vec{x}} \exp(-i\vec{p} \cdot \vec{x}) \left\langle J_H(0, \vec{0}) J_H^\dagger(\tau, \vec{x}) \right\rangle \quad J_H(\tau, \vec{x}) = \bar{q}(\tau, \vec{x}) \Gamma_H q(\tau, \vec{x})$$

Channel	$\Gamma_H$	$^{2S+1}L_J$	$J^{PC}$	$c\bar{c}$	$M(c\bar{c})[\text{GeV}]$
PS	$\gamma_5$	$^1S_0$	$0^{-+}$	$\eta_c$	2.980(1)
VC	$\gamma_\mu$	$^3S_1$	$1^{--}$	$J/\psi$	3.097(1)
SC	1	$^3P_0$	$0^{++}$	$\chi_{c0}$	3.415(1)
AV	$\gamma_5 \gamma_\mu$	$^3P_1$	$1^{++}$	$\chi_{c1}$	3.510(1)



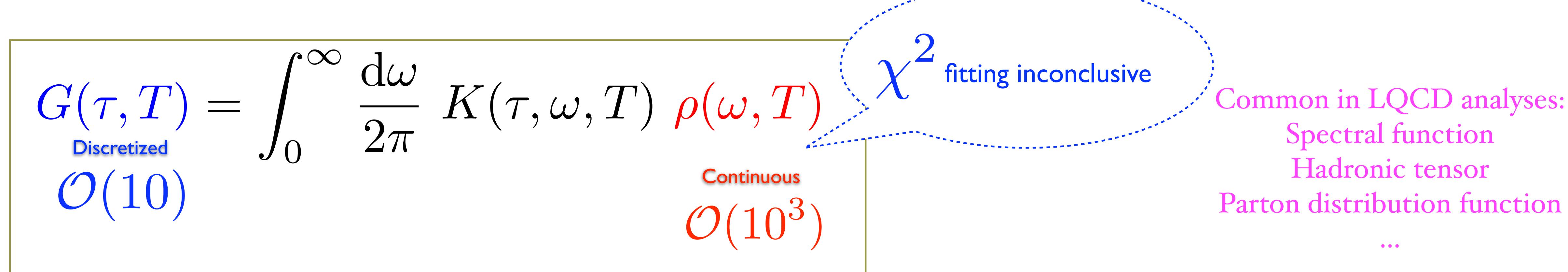
$$\rho(\omega) = 2\text{Im}D^R(\omega) = D^+(\omega) - D^-(\omega), \quad G(\tau, T) = D^+(-i\tau)$$

Spectral representation

$$G_H(\tau, \vec{p}, T) = \int_0^\infty \frac{d\omega}{2\pi} \rho_H(\omega, \vec{p}, T) \frac{\cosh(\omega(\tau - 1/2T))}{\sinh(\omega/2T)}$$

# Inverse problem

- A typical ill-posed inverse problem: infinite number of solutions exist!

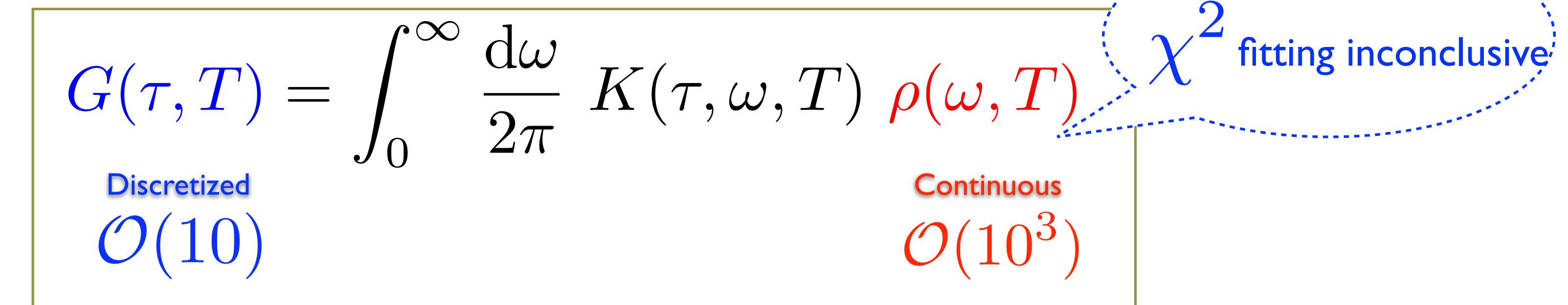


- Backus Gilbert method      G. Backus and F. Gilbert, Geophysical Journal International 16, 169 (1968)
- Tikhonov regularization      A. N. Tikhonov, Soviet Math. Dokl. 4, 1035 (1963)
- Maximum Entropy Method      M. Asakawa, T. Hatsuda & Y. Nakahara,  
Progress in Particle and Nuclear Physics 46 (2001) 459
- New Bayesian Method      Y. Burnier & A. Rothkopf, PRL 111, 182003 (2013)
- Stochastic optimization method      H.-T. Ding et al., PRD 97, 094503 (2018)
- ...

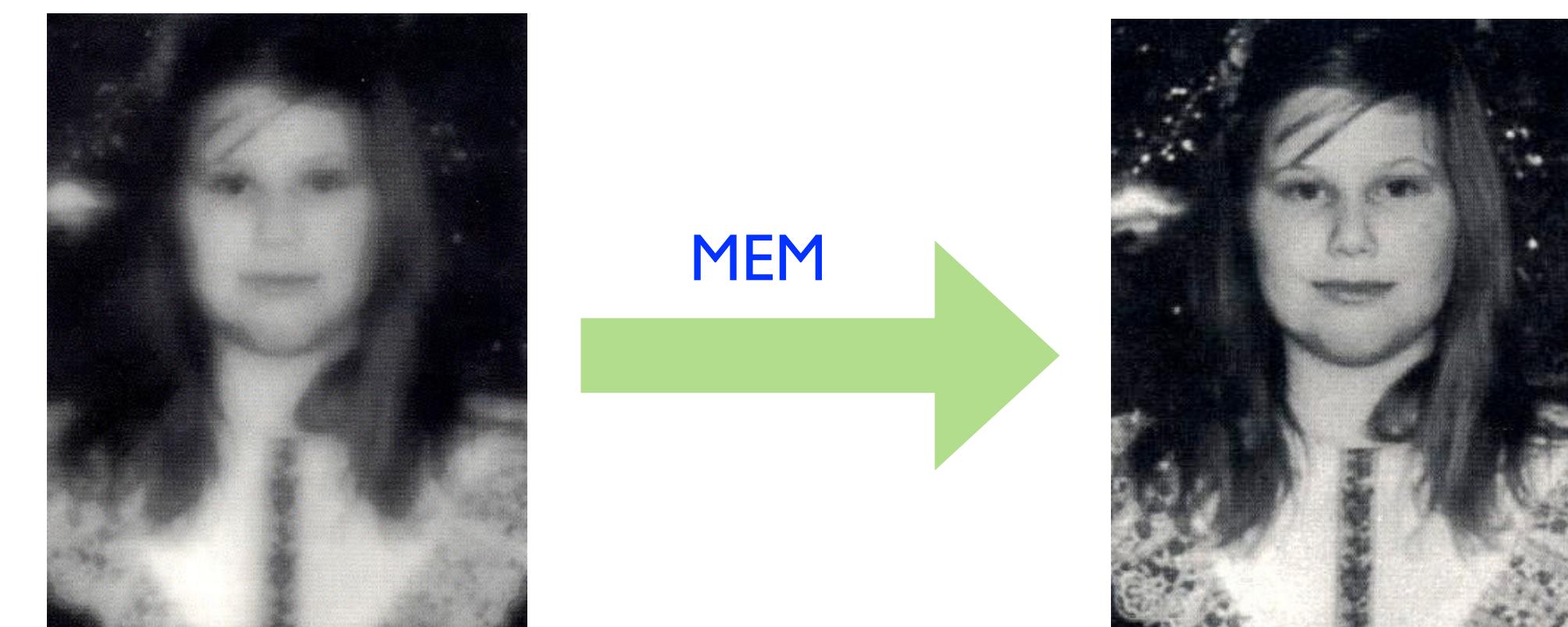
# Maximum Entropy Method (MEM)

[Asakawa, Hatsuda & Nakahara, '01]

- Extract spectral function (spf) without ansatz



- 📌 Successful in condensed matter physics, astrophysics, image processing...
- 📌 A method to obtain the most probable image from insufficient data
- 📌 A unique solution exists



# Introduction to MEM

[Asakawa, Hatsuda & Nakahara, '01]

- Based on the Bayesian theorem  $P[X|Y] = \frac{P[Y|X]P[X]}{P[Y]}$ ,  $P[X|Y]$ : Probability of X given Y

- Maximize the probability of  $P[\rho|GH] \propto P[G|\rho H] P[\rho|H]$

$P[G|\rho H] \propto \exp(-\chi^2/2)$  : likelihood function

$P[\rho|H] \propto \exp(\alpha S)$  : prior probability

$\rho$ : spectral function  
 $G$ : lattice data  
 $H$ : prior information on  $\rho$

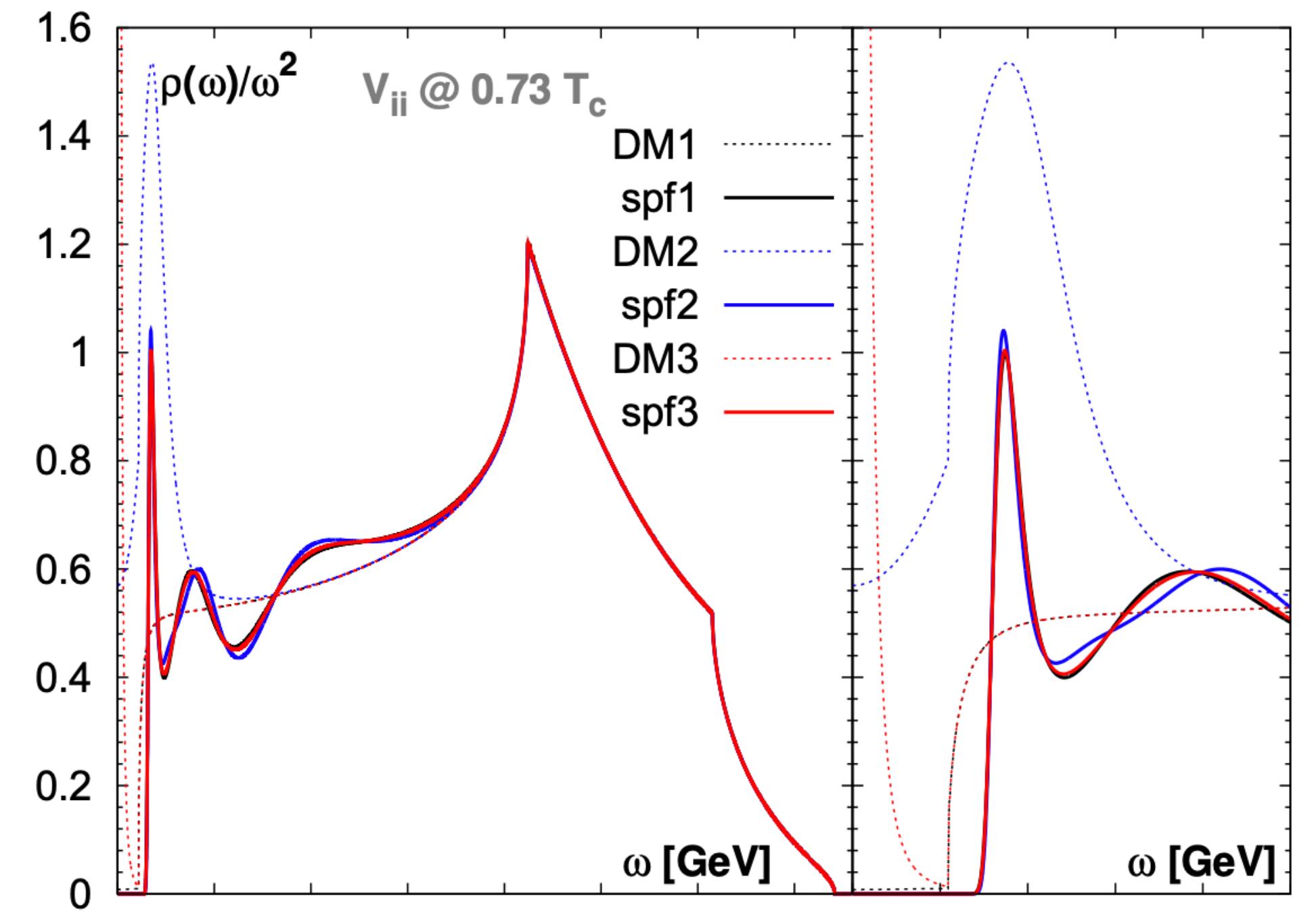
Information entropy:  $S = \int_0^\infty \frac{d\omega}{2\pi} \left[ \rho(\omega) - m(\omega) - \rho(\omega) \ln \left( \frac{\rho(\omega)}{m(\omega)} \right) \right]$

Default Model (DM):  $m(\omega)$ , includes the prior information on  $\rho$ , e.g.  $\rho$  is positive-definite  
only input parameter in the MEM analysis

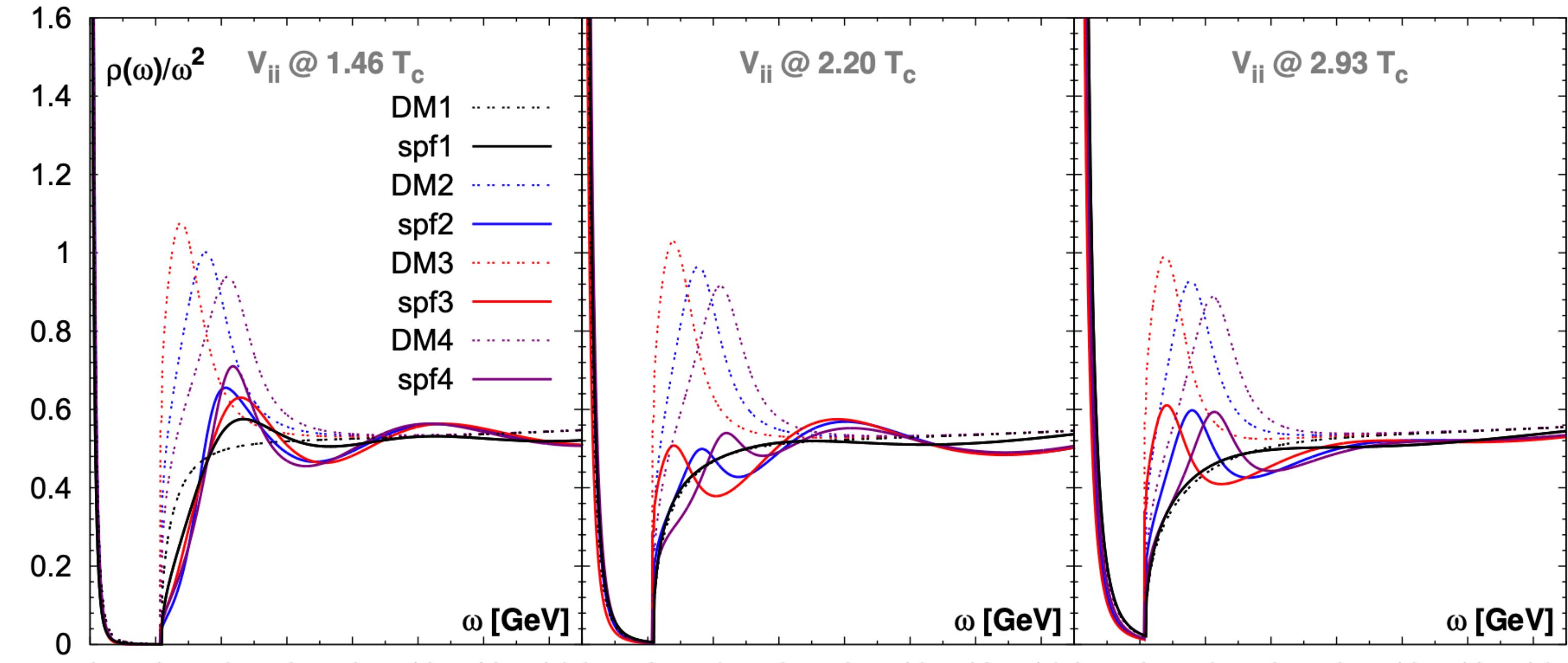
$\alpha$ : controls the relative weight between the likelihood and entropy



# Default model dependence in MEM analyses



$N_\tau = 96$



$N_\tau = 48$

$N_\tau = 32$

$N_\tau = 24$

# Inverse problem in machine learning

infinite number of solutions exist!

- A few studies have been carried out in reconstructing spectral functions using machine learning techniques
  - Kades et al., *s*, Phys. Rev. D 102, 096001 (2020),
  - Offler et al., PoS LATTICE2019, 076 (2019),
  - Zhou et al., Phys. Rev. D 104, 076011 (2021),
  - Horak et al., arXiv:2107.13464
- However, in most cases:
  - 1) the likelihood or simple difference in  $G$  and  $\rho$  are used in the loss function;
  - 2) the training inputs and wanted outputs have the same form
- Goal in our study:
  - Design a neural network that can learn to balance the prior & likelihood
  - Train the neural network using a general input

# Variational auto-encoders (VAE)

Kingma, Welling,  
arXiv: 1312.6114

Encoder + Decoder + Loss function

Latent variable  $z$

Latent variable  $z$

Recognition/Inference  
model

Encoder  $q_\phi(z|x)$

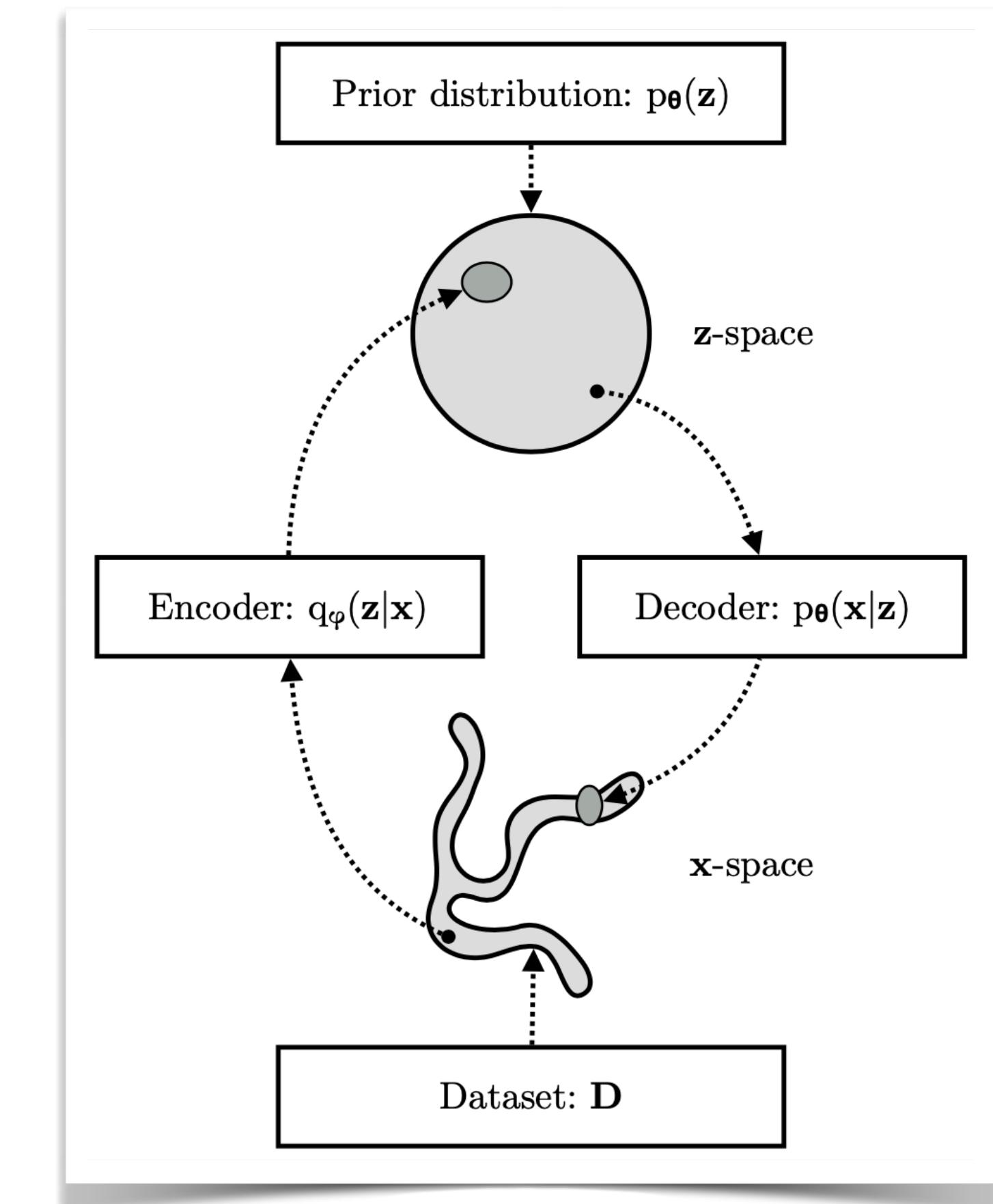
Decoder  $p_\theta(x|z)$

Data  $x$

Generative  
Model

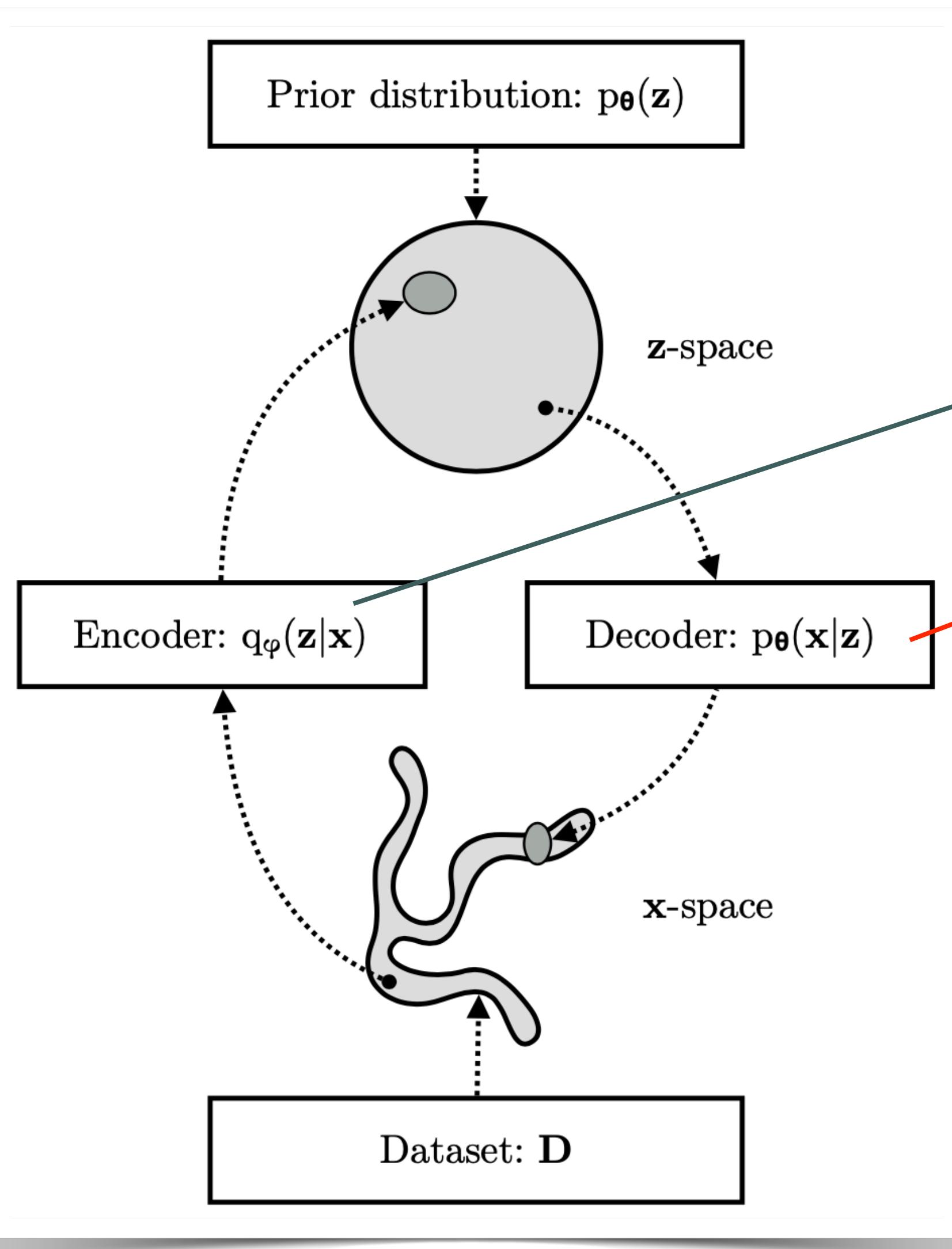
The Encoder compresses data into a latent space  $Z$

The Decoder reconstructs the data given the hidden representation



D. Kingma, M. Welling,  
arXiv: 1906.02691

# Loss function in VAE



$$\begin{aligned}
 \log p_\theta(x) &= \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(x)] \\
 &= \mathbb{E}_{q_\phi(z|x)} \left[ \log \frac{p_\theta(x, z)}{p_\theta(z|x)} \right] \\
 &= \mathbb{E}_{q_\phi(z|x)} \left[ \log \frac{p_\theta(x, z)}{q_\phi(z|x)} \frac{q_\phi(z|x)}{p_\theta(z|x)} \right] \\
 &= \underbrace{\mathbb{E}_{q_\phi(z|x)} \left[ \log \frac{p_\theta(x, z)}{q_\phi(z|x)} \right]}_{= \mathcal{L}_{\theta, \phi}(x) \text{ (ELBO)}} + \underbrace{\mathbb{E}_{q_\phi(z|x)} \left[ \log \frac{q_\phi(z|x)}{p_\theta(z|x)} \right]}_{= D_{KL}(q_\phi(z|x)||p_\theta(z|x)) \geq 0} \\
 &\quad \text{Kullback Leibler divergence between the model and data distributions}
 \end{aligned}$$

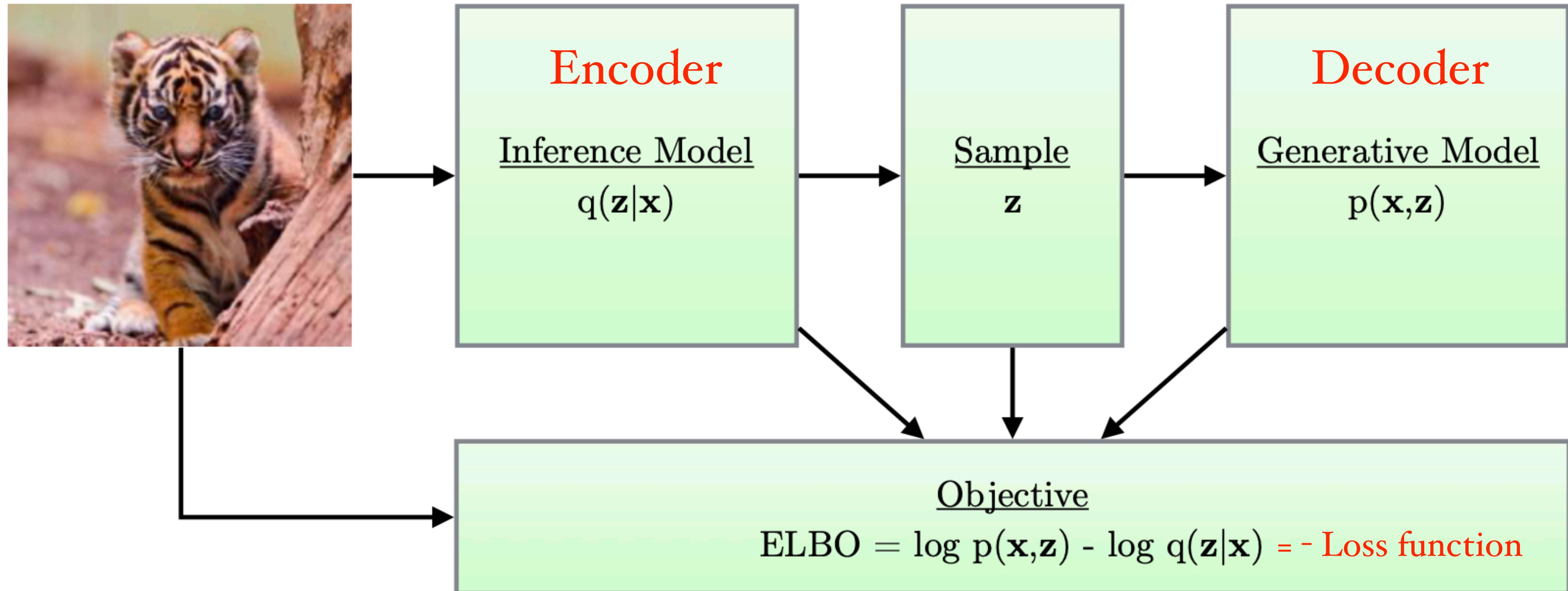
Maximize the ELBO  $\longleftrightarrow$  Minimize the KL divergence

Loss function = - ELBO

Minimize the loss function

# Computation flow in VAE

Datapoint



# A newly proposed SVAE

Chen, HTD, Liu, Papp & Yang,  
arXiv:2110.13521

Conditional  
Probability

$$\begin{aligned}
 \log P(\rho|G) &= \log \left( \int dz P(\rho|z, G)P(z|G) \right) && \xleftarrow{\quad} \\
 &\equiv \log \left( E_{Q(z|G_{gt}, \rho_{gt})} \left( P(\rho|z, G) \frac{P(z|G)}{Q(z|G_{gt}, \rho_{gt})} \right) \right) && \xleftarrow{\quad} \\
 &\geq E_{Q(z|G_{gt}, \rho_{gt})} \left( \log \left( P(\rho|z, G) \frac{P(z|G)}{Q(z|G_{gt}, \rho_{gt})} \right) \right) && \xleftarrow{\quad}
 \end{aligned}$$

**Evidence lower bound**

Jensen's inequality & the concave nature of the logarithm function

Define the loss function = - Evidence lower bound

$$\mathcal{L} = -E_{Q(z|G_{gt}, \rho_{gt})} \left( \log \left( P(\rho|z, G) \frac{P(z|G)}{Q(z|G_{gt}, \rho_{gt})} \right) \right)$$

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

$$\begin{aligned}
 &\equiv -E_{Q(z|G_{gt}, \rho_{gt})} \left[ \log P(\rho|z, G) \right] + KL \left( Q(z|G_{gt}, \rho_{gt}) \| P(z|G) \right) \\
 &\quad \text{Decoder} \qquad \text{Encoder 1} \qquad \text{Encoder 2}
 \end{aligned}$$

KL divergence:  
Distance between 2 probabilities

# Loss function in SVAE

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

$$P(\rho|z, G) = P(\rho|z) \cdot P(G|\rho, z) / P(G|z)$$

Prior  
information      
 Likelihood      
 Evidence

$$P(\rho|z) = \frac{1}{Z_S} e^S, \quad S = \int_0^\infty d\omega \left( \rho(\omega, z) - \rho_{gt}(\omega) - \rho(\omega, z) \log \left( \frac{\rho(\omega, z)}{\rho_{gt}(\omega)} \right) \right)$$

V.s. in MEM:  $S = \int_0^\infty d\omega \left( \rho(\omega) - m(\omega) - \rho(\omega) \ln \left( \frac{\rho(\omega)}{m(\omega)} \right) \right)$

$\rho_{gt}(\omega)$ : ground truth value of the spectral function, best prior info of  $\rho(\omega)$ !

Entropy term S + VAE  $\longrightarrow$  SVAE

# Loss function in SVAE

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

$$P(\rho|z, G) = P(\rho|z) \textcolor{red}{P(G|\rho, z)} / \textcolor{teal}{P(G|z)}$$

Prior  
information
Likelihood
Evidence

$$P(\rho|z) = \frac{1}{Z_S} e^S, \quad S = \int_0^\infty d\omega \left( \rho(\omega, z) - \boxed{\rho_{gt}(\omega)} - \rho(\omega, z) \log \left( \frac{\rho(\omega, z)}{\rho_{gt}(\omega)} \right) \right)$$

$$P(G|z, \rho) = \frac{1}{Z_L} e^{-L}, \quad L = \sum_{j=\tau_{min}}^{N_\tau/2} L_j = \sum_{j=\tau_{min}}^{N_\tau/2} \frac{\left( \hat{G}_j [\rho(z)] - G_j \right)^2}{2 \boxed{\alpha_j^2(z)} G_j^2}$$

G: correlator data  
\hat{G}: computed from \rho(z)

$\alpha_j(z)$ : controls the relative weight of the entropy (having  $\rho$  close to the prior) and the likelihood (having  $\rho$  close to the data G)

# Loss function in SVAE

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

$$P(\rho|z, G) = P(\rho|z) \textcolor{blue}{P(G|\rho, z)} / \textcolor{teal}{P(G|z)}$$

Prior  
information
Likelihood
Evidence

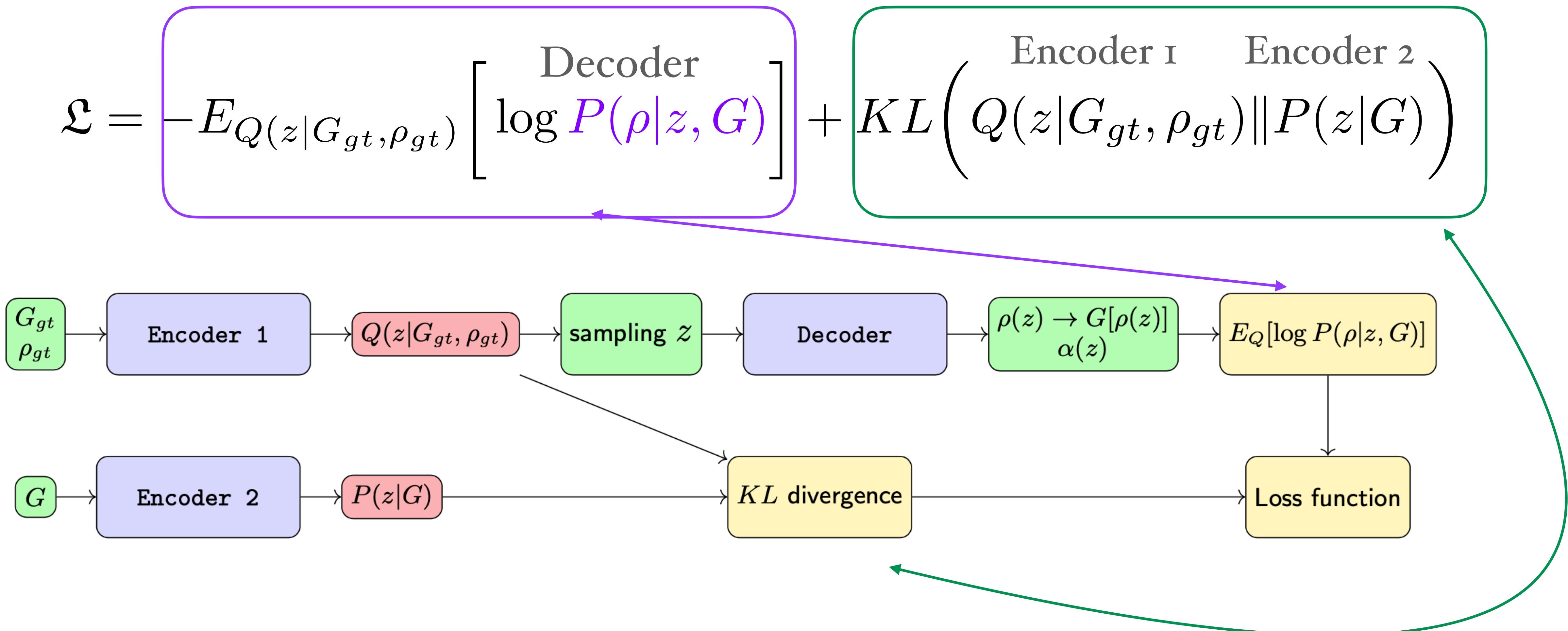
$$P(\rho|z) = \frac{1}{Z_S} e^S, \quad S = \int_0^\infty d\omega \left( \rho(\omega, z) - \boxed{\rho_{gt}(\omega)} - \rho(\omega, z) \log \left( \frac{\rho(\omega, z)}{\rho_{gt}(\omega)} \right) \right)$$

$$P(G|z, \rho) = \frac{1}{Z_L} e^{-L}, \quad L = \sum_{j=\tau_{min}}^{N_\tau/2} L_j = \sum_{j=\tau_{min}}^{N_\tau/2} \frac{\left( \hat{G}_j [\rho(z)] - G_j \right)^2}{2 \boxed{\alpha_j^2(z)} G_j^2}$$

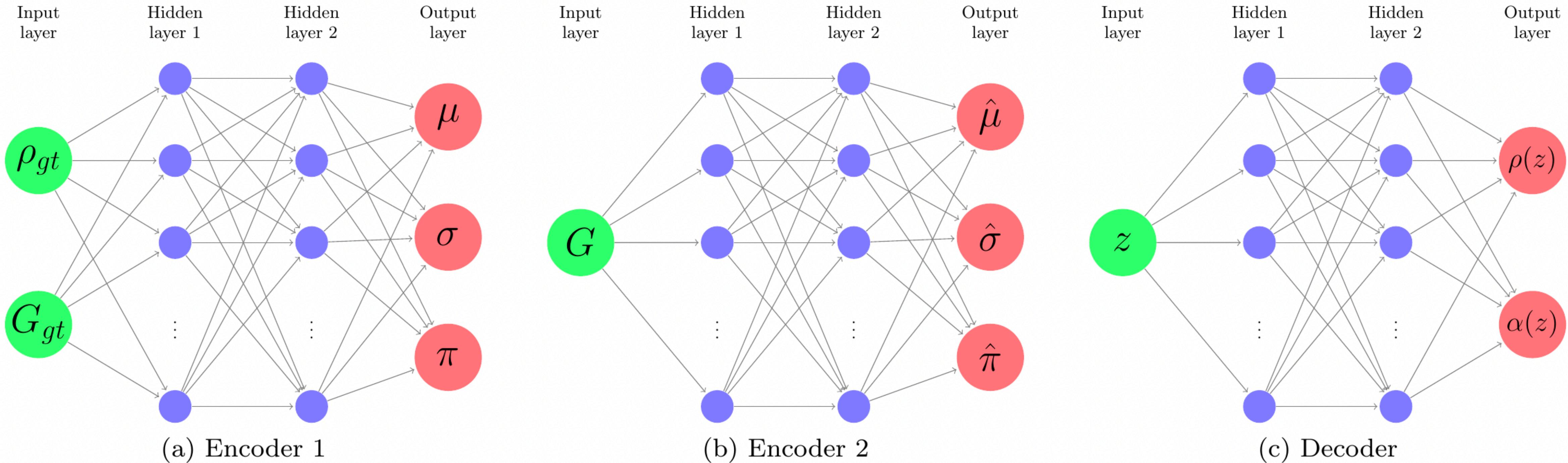
G: correlator data  
 $\hat{G}$ : computed from  $\rho(z)$

$$P(G|z) = \int \mathcal{D}\rho(z) P(G|z, \rho) P(\rho|z) = \int \mathcal{D}\rho(z) \frac{1}{Z_S Z_L} e^{-L+S}$$

# topology of the SVAE for training



# Encoders and decoder



$$Q(z|\rho_{gt}, G_{gt}) = \sum_i^{N_g} \pi_i \prod_k^{N_z} \mathcal{N}(\mu_{i,k}, \sigma_{i,k}).$$

$$P(z|G) = \sum_i^{N_g} \hat{\pi}_i \prod_k^{N_z} \mathcal{N}(\hat{\mu}_{i,k}, \hat{\sigma}_{i,k}).$$

$$\rho(z), \alpha(z)$$

$Q$  and  $P(z|G)$ : parameterized based on the Gauss mixture model  
 with  $\mathcal{N}$  the Gauss function, outputs from Encoder 1 and 2

$\rho(z), \alpha(z)$ : outputs from Decoder

# Loss function of the SVAE

Prior

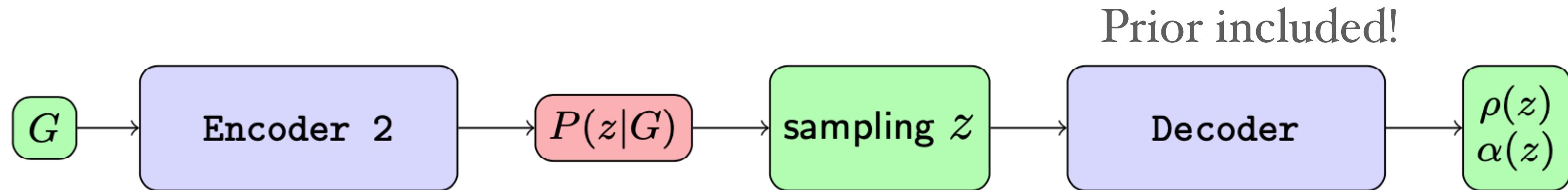
$$\mathcal{L}_{\text{SVAE}} = \int_{z \sim Q(z|\rho_{gt}, G_{gt})} \left\{ \sum_l^{N_\omega} \left[ \rho_{gt,l} - \rho_l(z) + \rho_l(z) \log \left( \frac{\rho_l(z)}{\rho_{gt,l}} \right) \right] + \sum_{j=\tau_{min}}^{N_\tau/2} \frac{1}{2} \left[ \left( \frac{\hat{G}_j[\rho(z)] - G_j}{\alpha_j(z) G_j} \right)^2 + \frac{2\pi - 1}{2\pi} \log (\alpha_j^2(z) G_j^2) \right] \right\} dz$$

KL divergence

$$+ \sum_{i=1}^{N_g} \left[ \frac{\pi_i}{2} \sum_{k=1}^{N_z} \left( \frac{(\mu_{i,k} - \hat{\mu}_{i,k})^2}{\hat{\sigma}_{i,k}^2} + \log \frac{\hat{\sigma}_{i,k}^2}{\sigma_{i,k}^2} \frac{\sigma_{i,k}^2}{\hat{\sigma}_{i,k}^2} - 1 \right) + \pi_i \log \left( \pi_i / \hat{\pi}_i \right) \right]$$

Minimize  $\mathcal{L}_{\text{SVAE}}$  to obtain  $\alpha, \pi, \mu, \hat{\pi}, \hat{\mu}, \sigma$

# Reconstruction of SPF in SVAE



$$\tilde{\rho}_\Theta = \frac{1}{N_c} \sum_{m=1}^{N_c} \int dz \rho(z) P(G_m | \rho, z) P(z | G_m) \rightarrow \frac{1}{N_c} \sum_{m=1}^{N_c} \sum_{n=1}^{N_s} \frac{\rho(z_n) P(G_m | \rho, z_n)}{\sum_{k=1}^{N_s} P(G_m | \rho, z_k)} \equiv \frac{1}{N_c} \sum_{m=1}^{N_c} \tilde{\rho}_{\Theta,m}$$

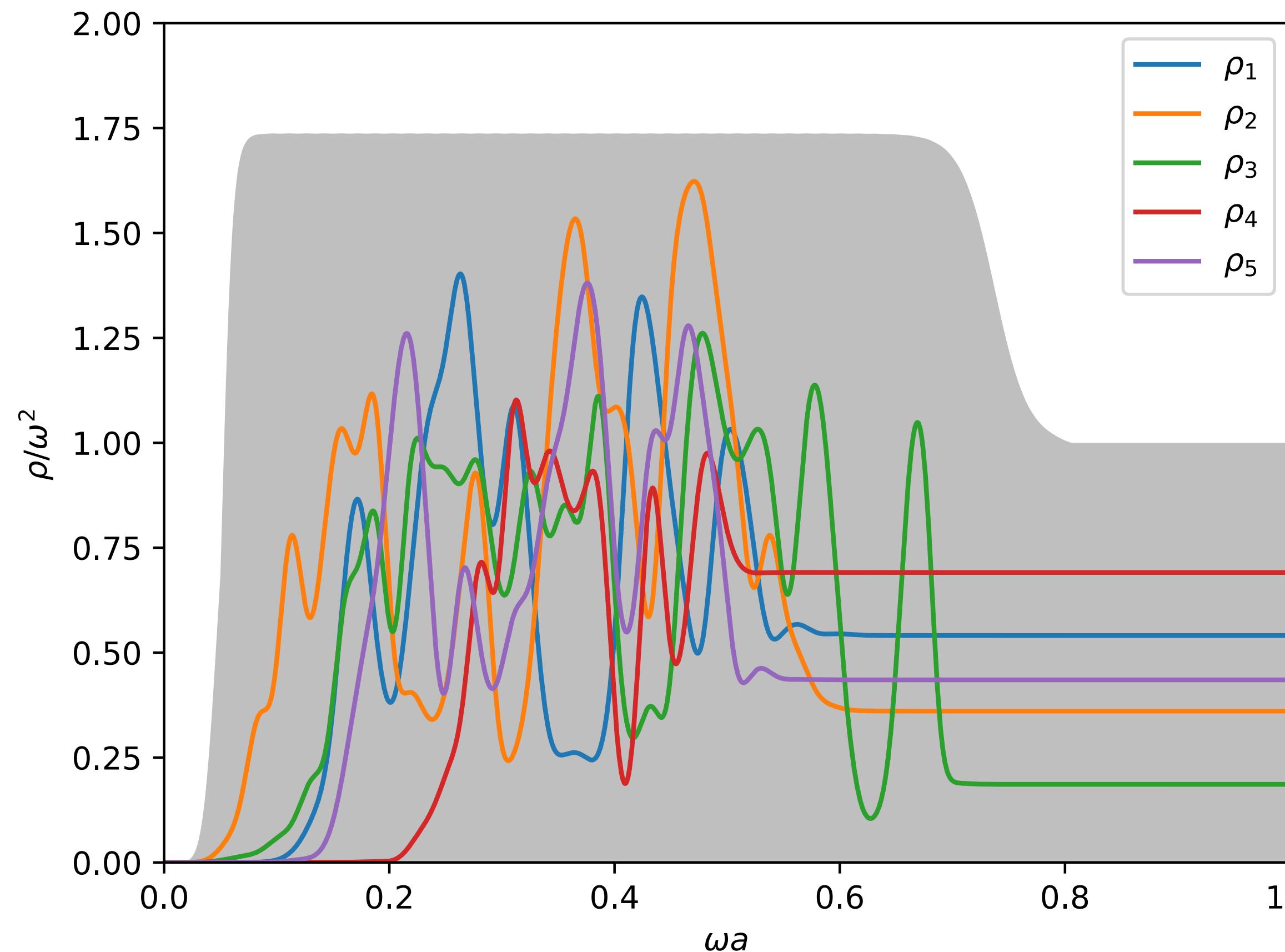
$\Theta$ : a certain epoch,  $N_c$ : # of configurations,  $N_s$ : # of samples in  $z$

Two dimensional bootstrap sampling (in configuration and epoch spaces) to obtain final results of SPF and its uncertainty

$$\bar{\rho} = \frac{1}{N_{btp}} \sum_{i=1}^{N_{btp}} \tilde{\rho}_i , \quad \sigma_\rho = \sqrt{\frac{1}{N_{btp}} \sum_{i=1}^{N_{btp}} (\tilde{\rho}_i - \bar{\rho})^2}$$

# A general spectral function used for the training

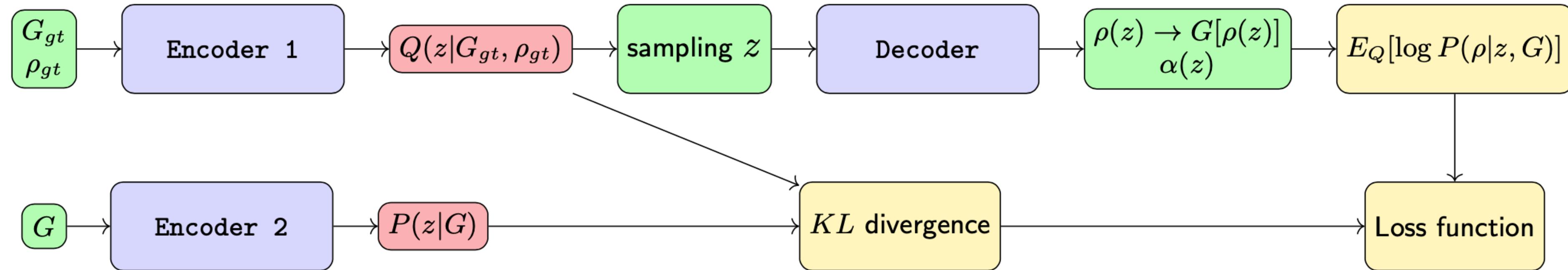
$$\frac{\rho_{train}}{\omega^2} = \hat{\theta}(\omega, \delta_1, \zeta_1) \left( \sum_{i=1}^{\hat{N}_g} C_i e^{-\left(\frac{\omega - M_i}{\gamma}\right)^2} \left( 1 - \hat{\theta}(\omega, \delta_2, \zeta_2) \right) + C_{i=0} \hat{\theta}(\omega, \delta_2, \zeta_2) \right)$$



$\hat{\theta}$ : smeared step functions  
 $\hat{N}_g=50$ : # of Gauss peaks  
 $M_i$ : fixed uniformly  
 $\gamma$ : fixed

Parameters	interval	Parameters	interval
$M_i$	$[0.05, 0.8]$	$\zeta_2 (\sim [0.1, 0.2]m_c)$	$[0.005, 0.02]$
$C_i$	$[0, 1]$	$\delta_1 (\sim [1, 3]m_c)$	$[0.05, 0.3]$
$\zeta_1 (\sim [0.1, 0.2]m_c)$	$[0.005, 0.02]$	$\delta_2 (\sim 8m_c)$	$[\delta_1 + 0.1, \delta_1 + 0.6]$

# Training strategy in details



$$\frac{\rho_{train}}{\omega^2} = \hat{\theta}(\omega, \delta_1, \zeta_1) \left( \sum_{i=1}^{\hat{N}_g} C_i e^{-(\frac{\omega - M_i}{\gamma})^2} \left( 1 - \hat{\theta}(\omega, \delta_2, \zeta_2) \right) + C_{i=0} \hat{\theta}(\omega, \delta_2, \zeta_2) \right)$$

1. feed  $\rho_{gt}$  and  $G_{gt}$  obtained from  $\rho_{train}$  and corresponding correlator to Encoder 1
2. feed  $G$  generated stochastically to Encoder2; Gaussian noise for  $G$  sample from  $\mathcal{N}(G_{gt}, \sigma)$  with  $\sigma = b(\tau) \times G_{gt}(\tau)$  and  $b(\tau) = \sigma(\tau)/G_{lat}(\tau)$
3. Perform the feeding in steps 1 and 2 simultaneously
4. For each  $N_\tau$  repeat steps 1, 2 & 3.

# Test strategy

• SPF used for tests

I:

$$\begin{cases} \rho_{res} = C_{res} \frac{\omega^2}{\left(\frac{\omega^2}{M_{res}\Gamma} - \frac{M_{res}}{\Gamma}\right)^2 + 1}, \\ \rho_{cont} = C_{cont} \frac{3\omega^2}{8\pi} \theta(\omega^2 - 4M_{cont}^2) \tanh\left(\frac{\omega}{4T}\right) \sqrt{1 - \left(\frac{2M_{cont}}{\omega}\right)^2} \left(2 + \left(\frac{2M_{cont}}{\omega}\right)^2\right) \end{cases}$$

II:  $\rho_{pert}(\omega) = A^{match} \rho^{\text{pNRQCD}}(\omega) \theta(\omega^{match} - \omega) + \rho^{vac}(\omega) \theta(\omega - \omega^{match})$ .

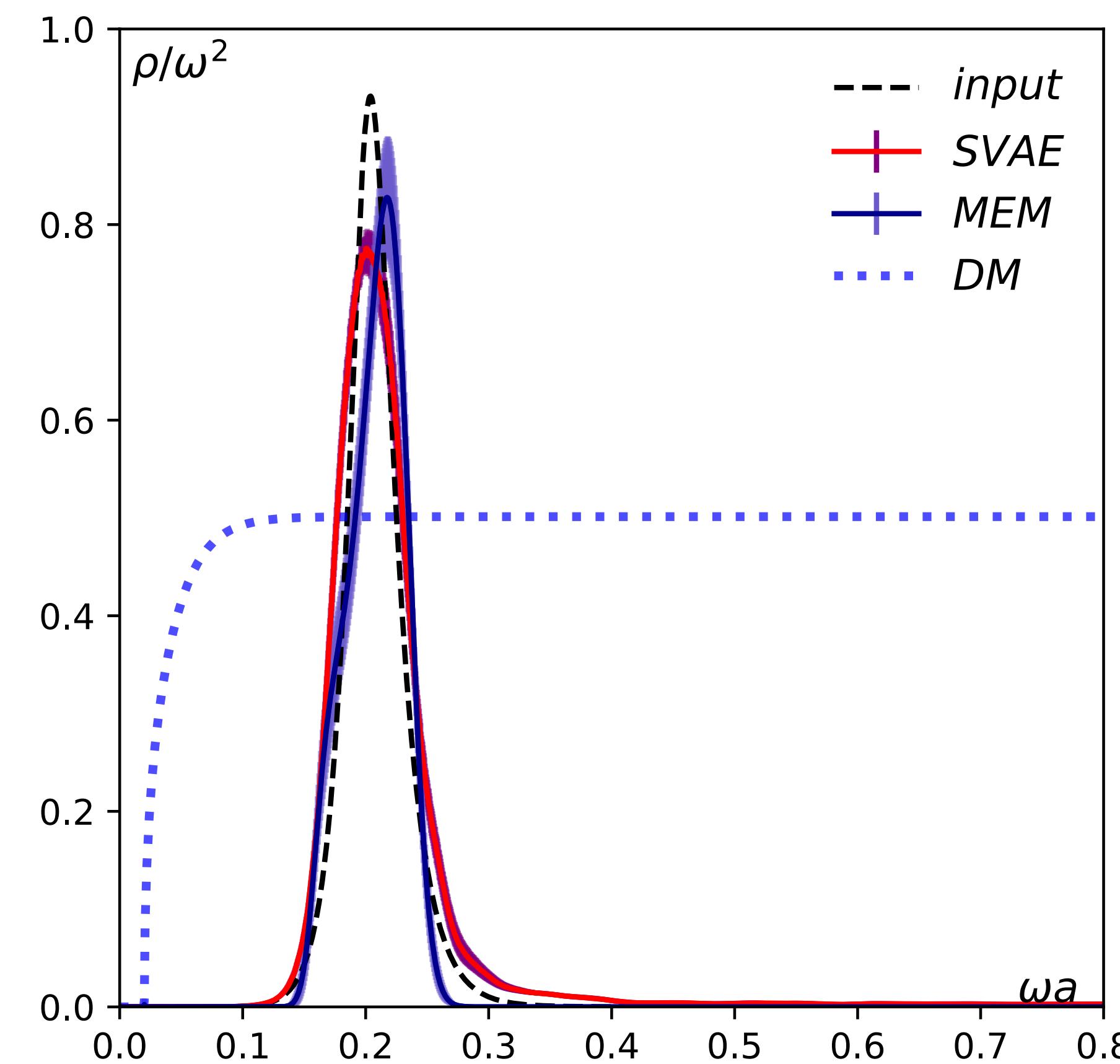
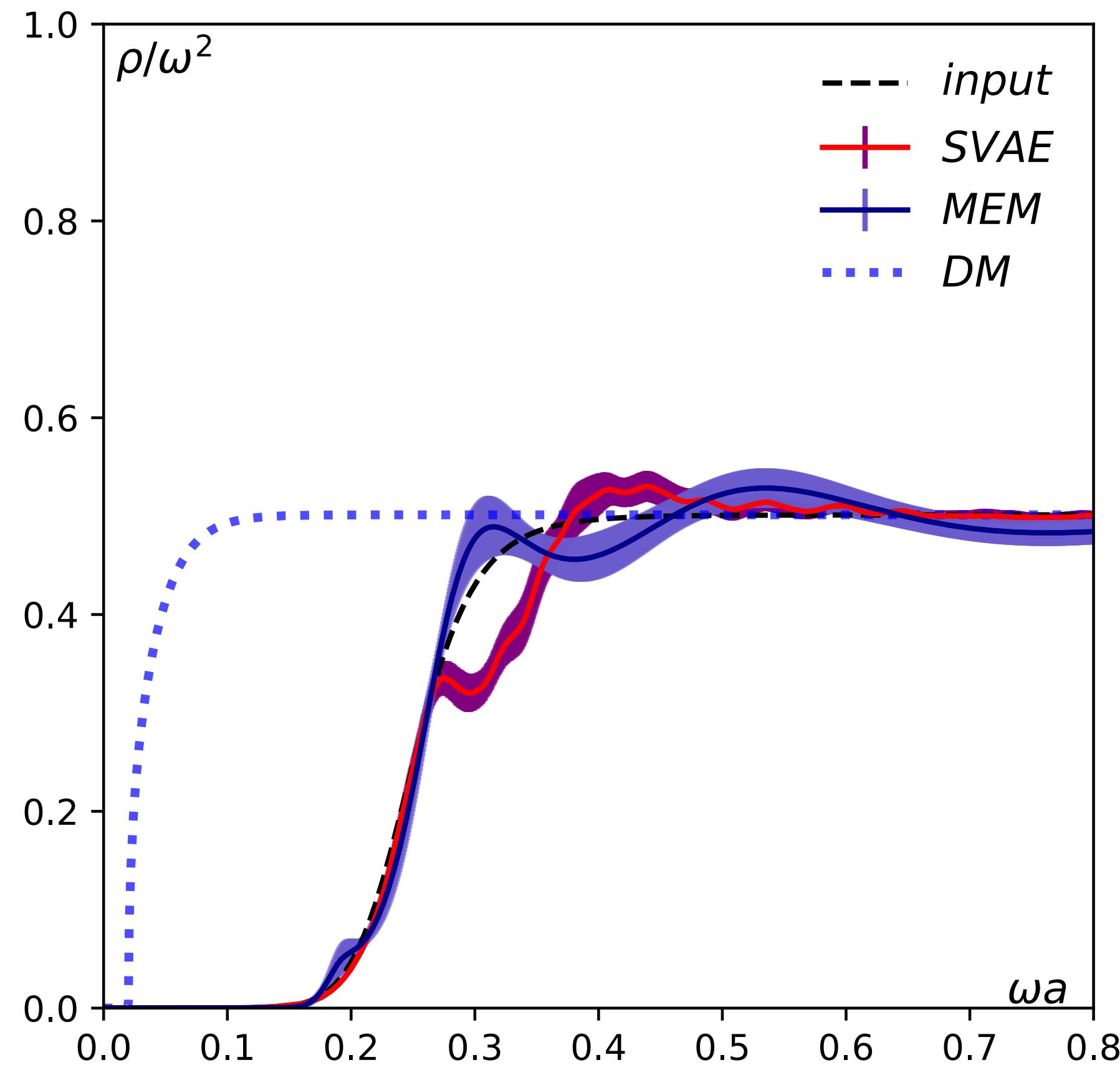
III:  $\rho_{res+cont} = \zeta(\omega, M_{res}, \Gamma) \rho_{res}(\omega, C_{res}, M_{res}, \Gamma) (1 - \zeta(\omega, M_{res} + \Gamma, \Gamma)) + \zeta(\omega, M_{res} + \Gamma, \Gamma) \rho_{cont}(C_{cont}, M_{cont})$ ,

• Construct mock data, i.e. noisy correlators using the above test spectral functions

• Noise models?

# Mock data tests: I

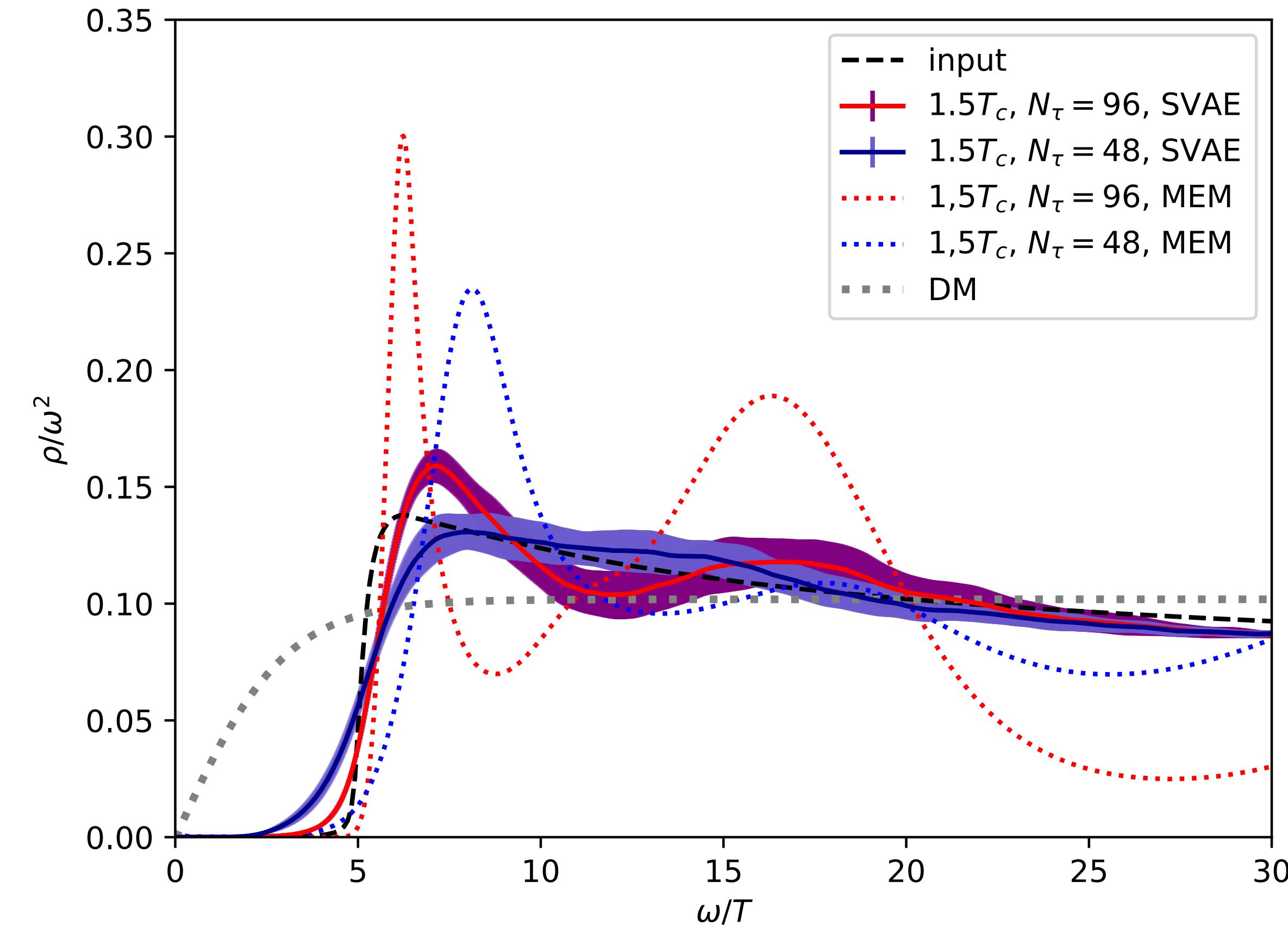
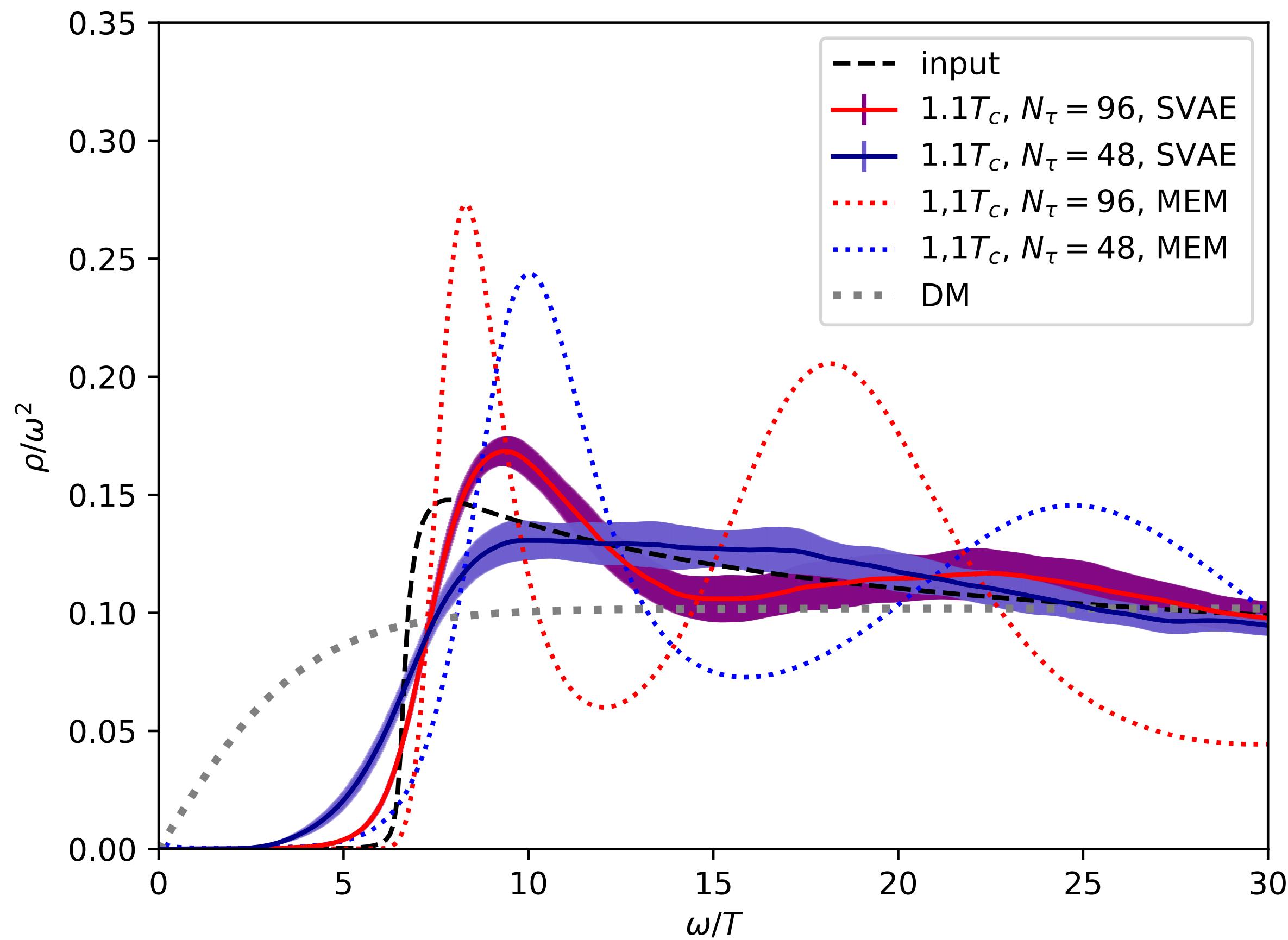
Two extreme cases with  $N_\tau = 96$



# Mock data tests: II

Input spf obtained from the Non-Relativistic QCD

Y. Burnier et al.,  
JHEP 11 (2017) 206



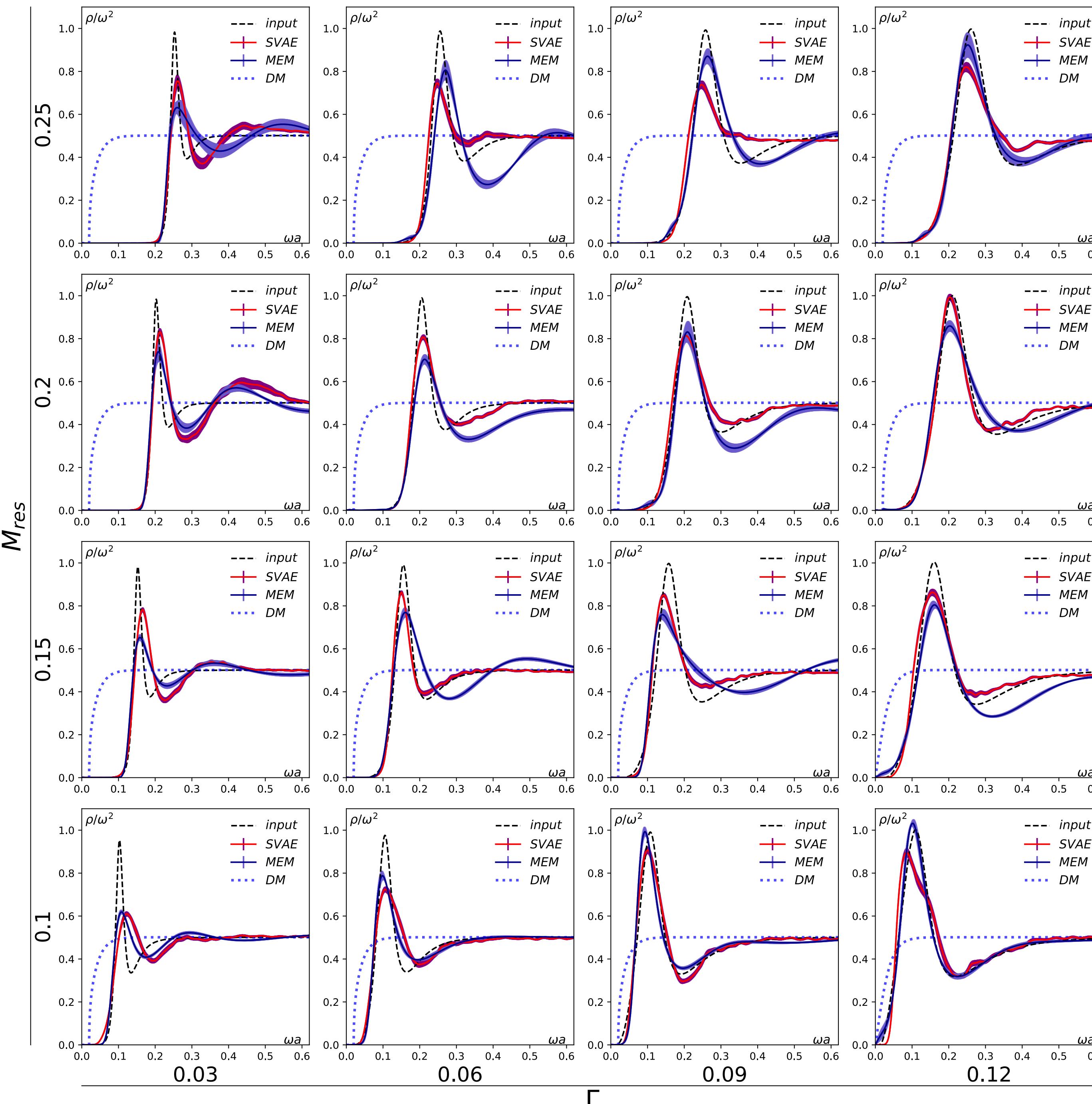
# Mock data tests: III

One resonance peak + continuum

$$N_\tau = 96$$

Dependences on the peak location  $M_{res}$  and width  $\Gamma$  of input spectral functions

Reconstruction quantity depends more on  $\Gamma$  rather than  $M_{res}$



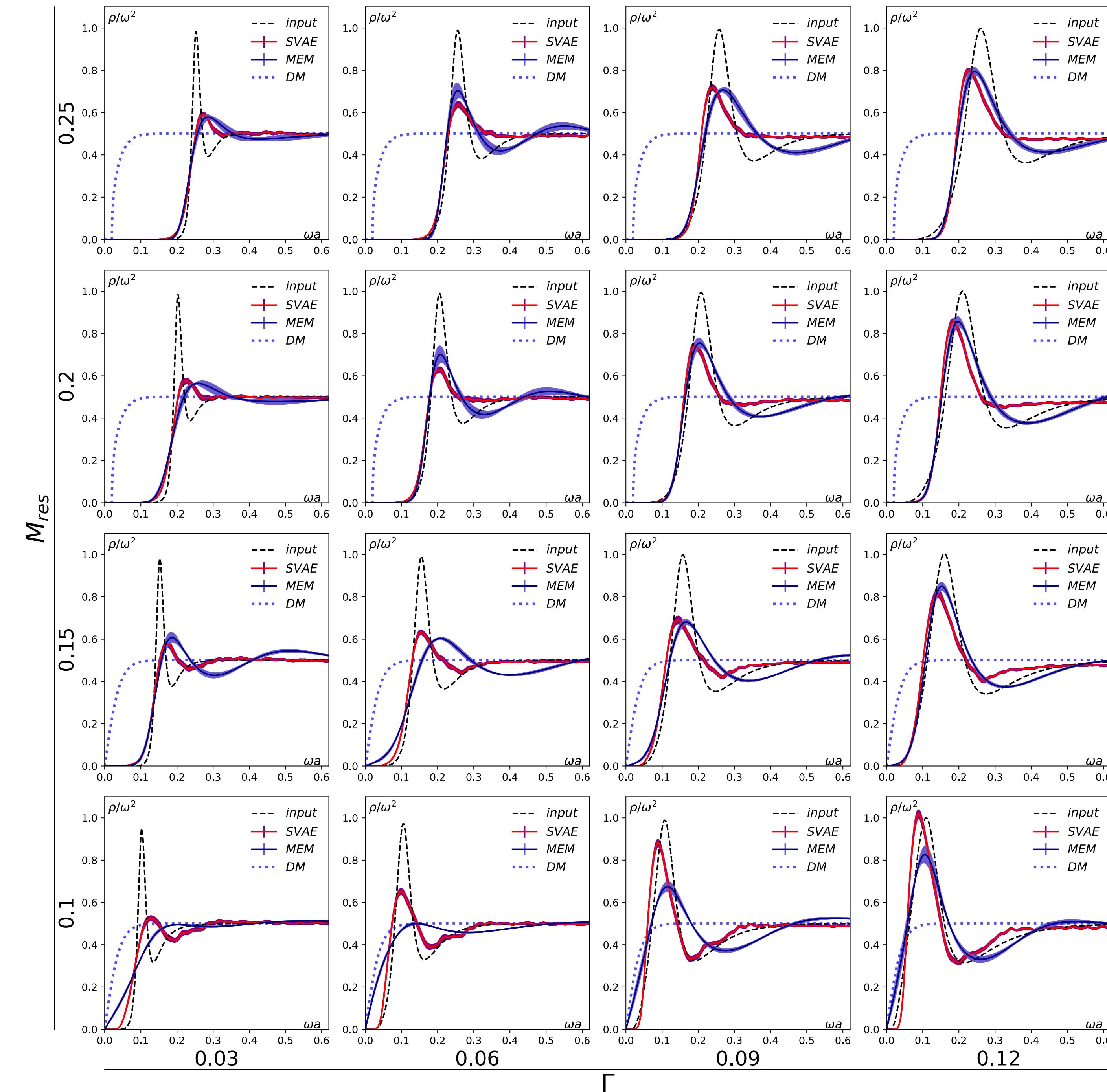
# Mock data tests: III

One resonance peak + continuum

$$N_\tau = 48$$

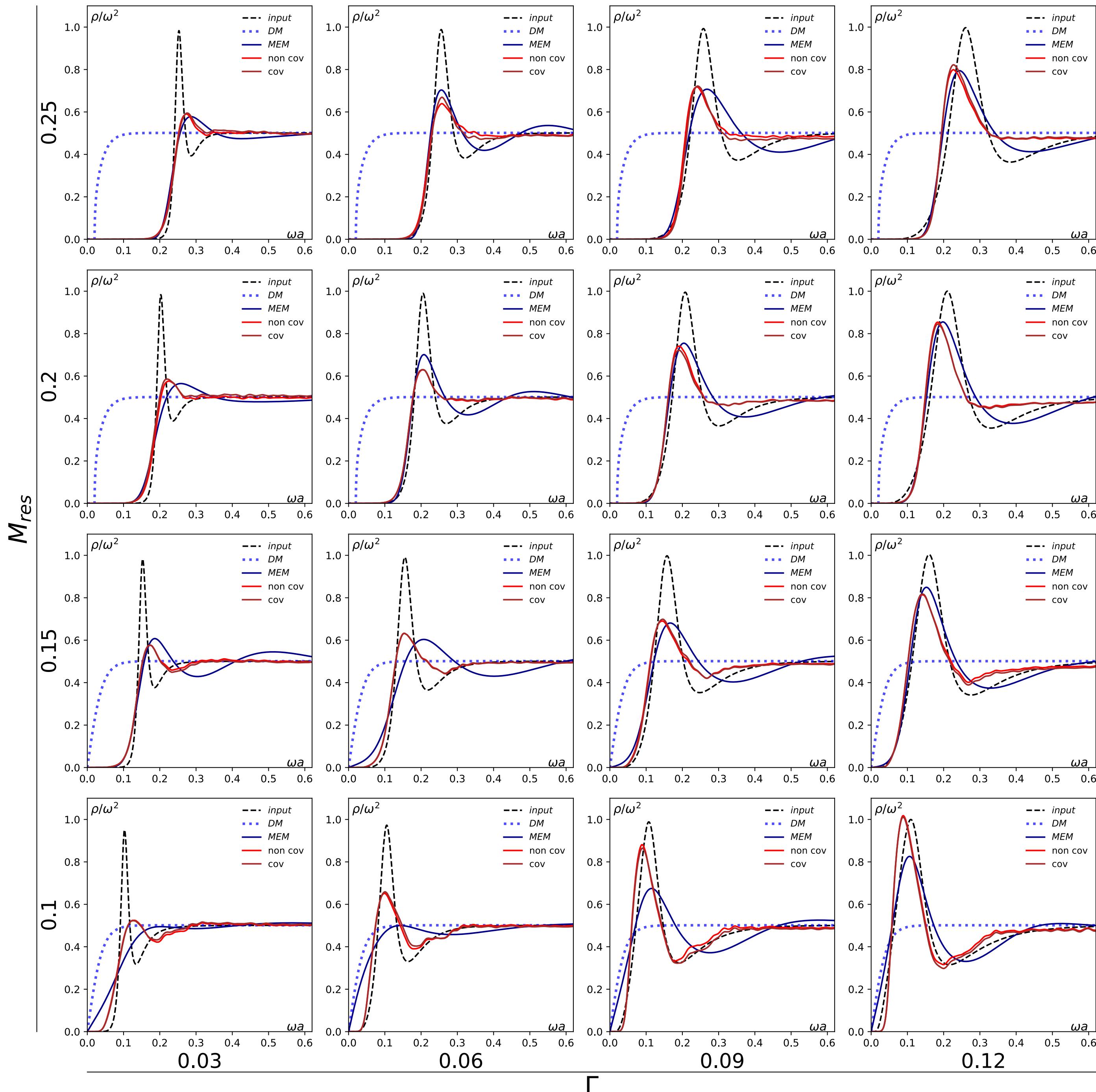
Dependences on the peak location  $M_{res}$  and width  $\Gamma$  of input spectral functions

Reconstruction quantity depends more on  $\Gamma$  rather than  $M_{res}$

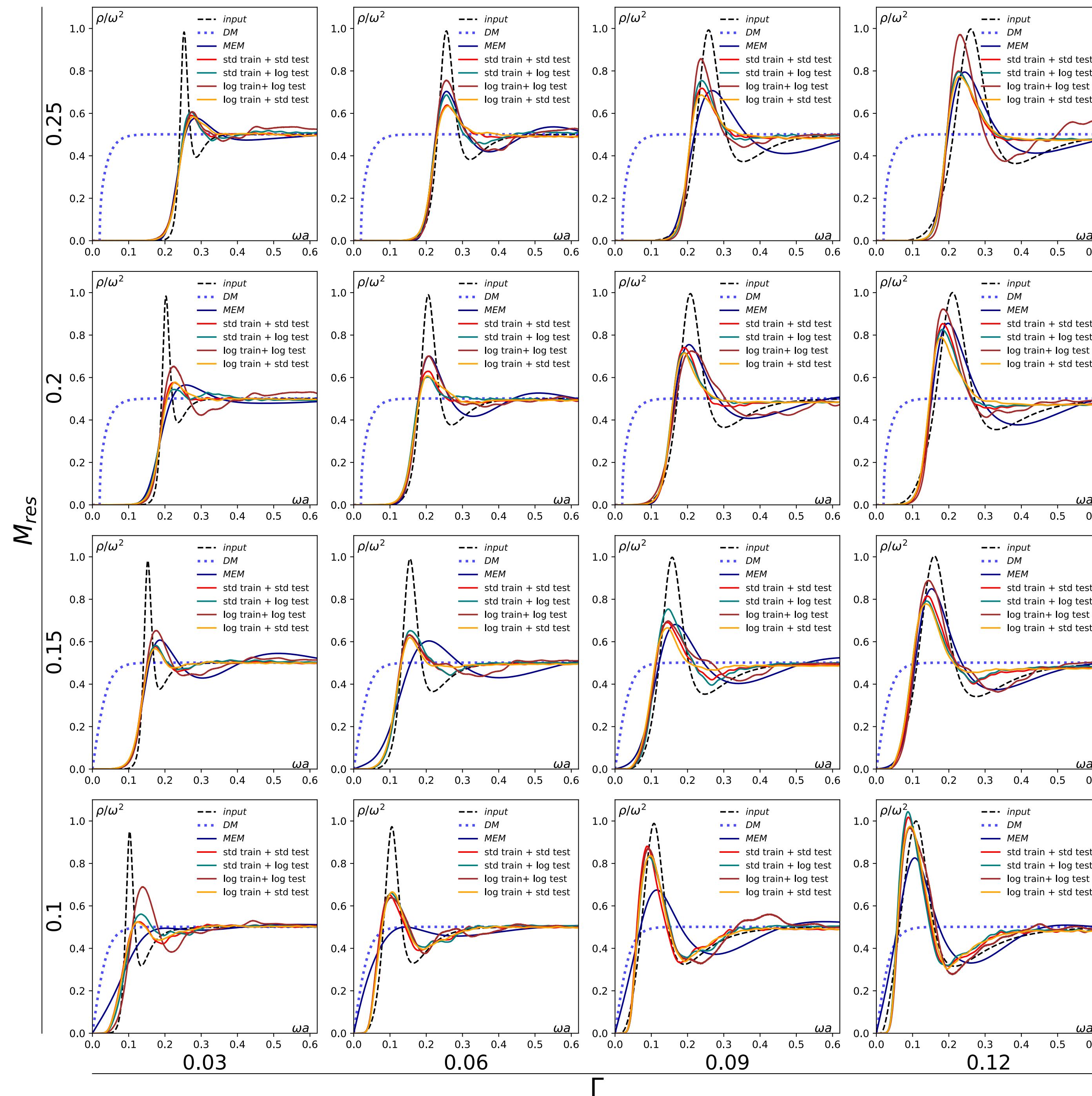


# Marginal dependence on noise models

wo and with covariance matrix (cov)  
in the noise model of test correlators



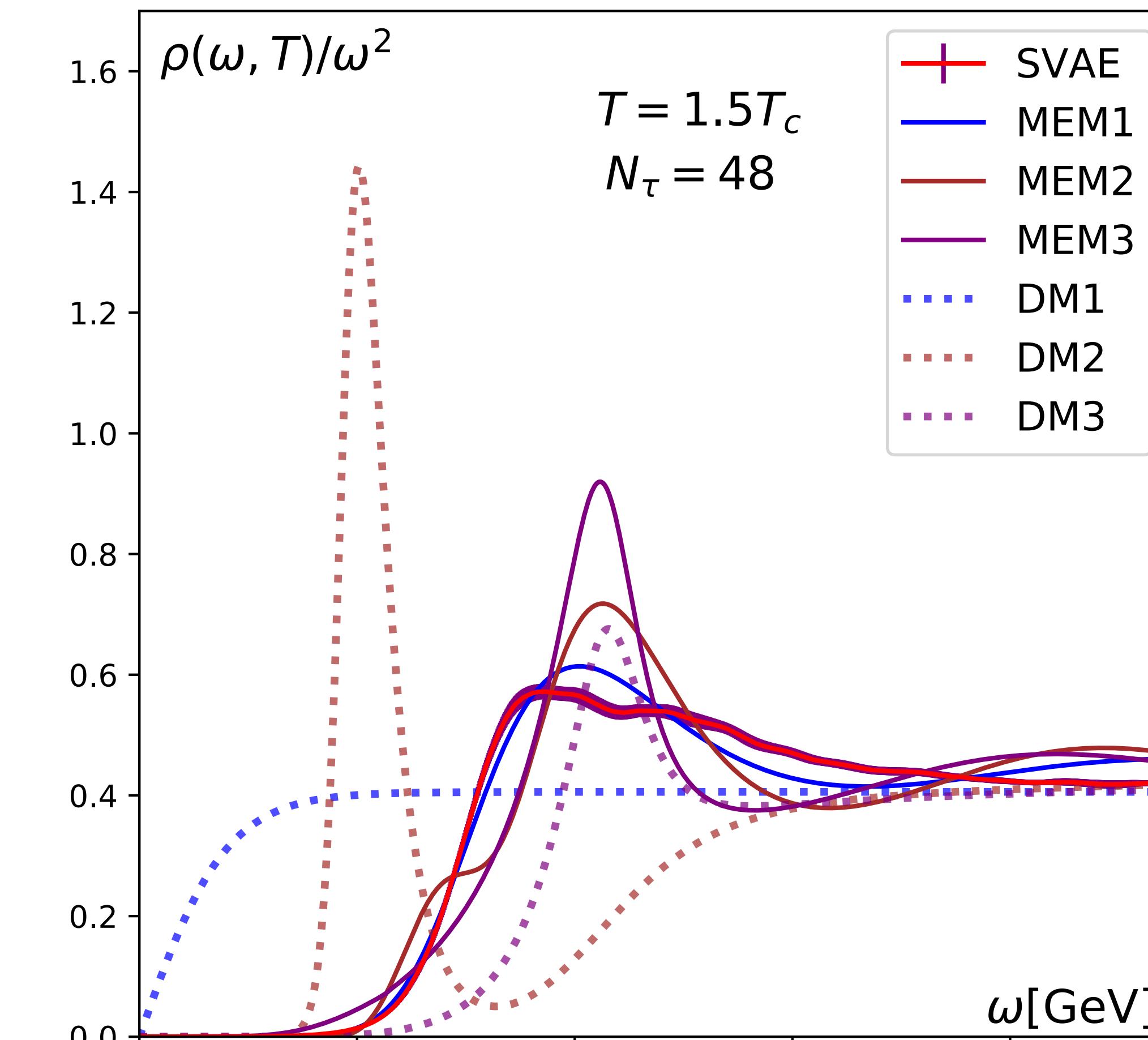
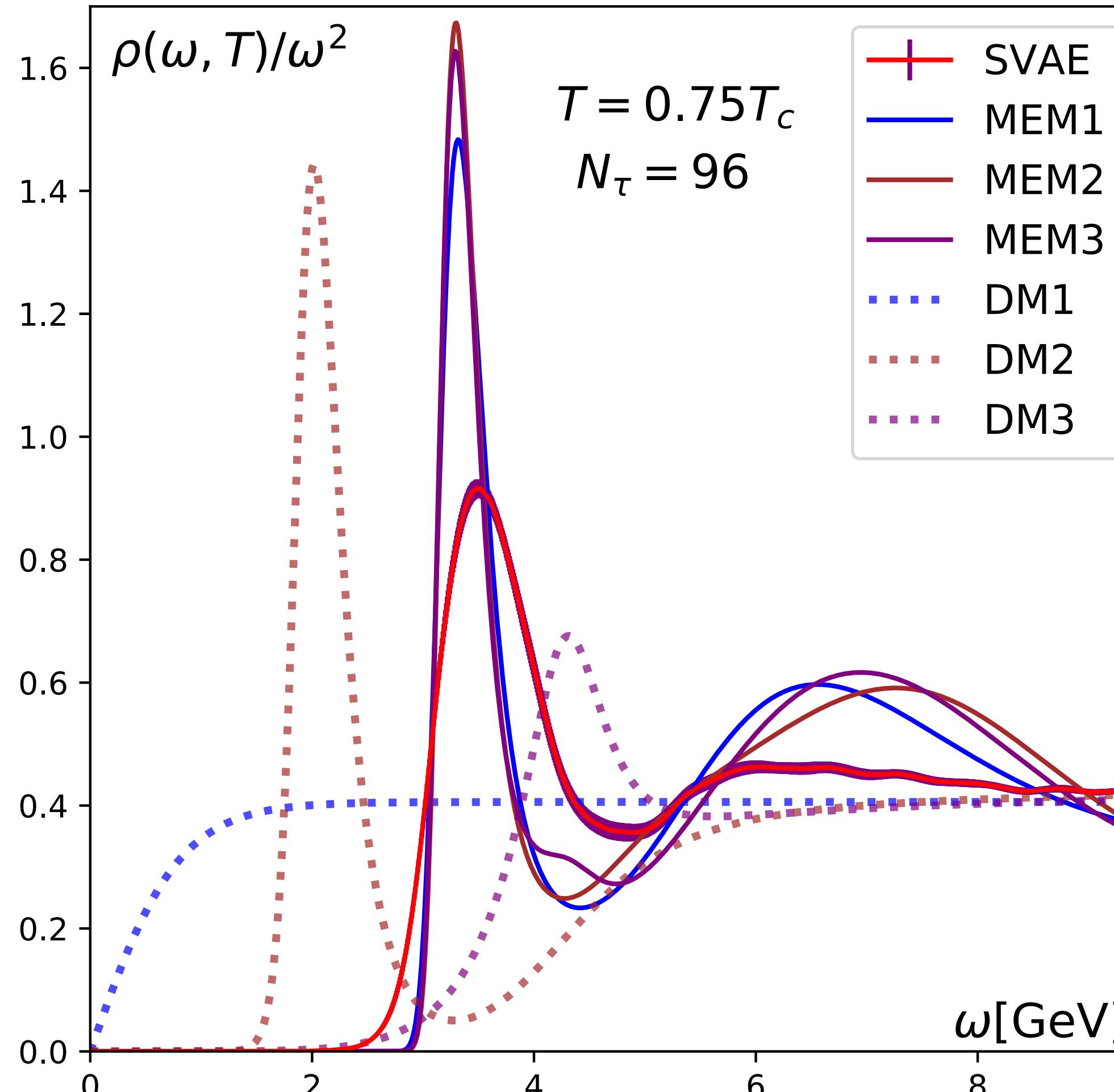
# Marginal dependence on noise models



	Training	Test
Gaussian	Gaussian	Gaussian
Log normal	Log normal	Log normal
Gaussian	Log normal	Gaussian
Log normal	Log normal	Log normal

# Application to Lattice QCD data: charmonium correlator in the pseudo-scalar channel

Clover-improved Wilson fermions on quenched lattices  
 $128^3 \times 96$  ( $0.75T_c$ ) and  $128^3 \times 48$  ( $1.5T_c$ ) with a fixed scale approach

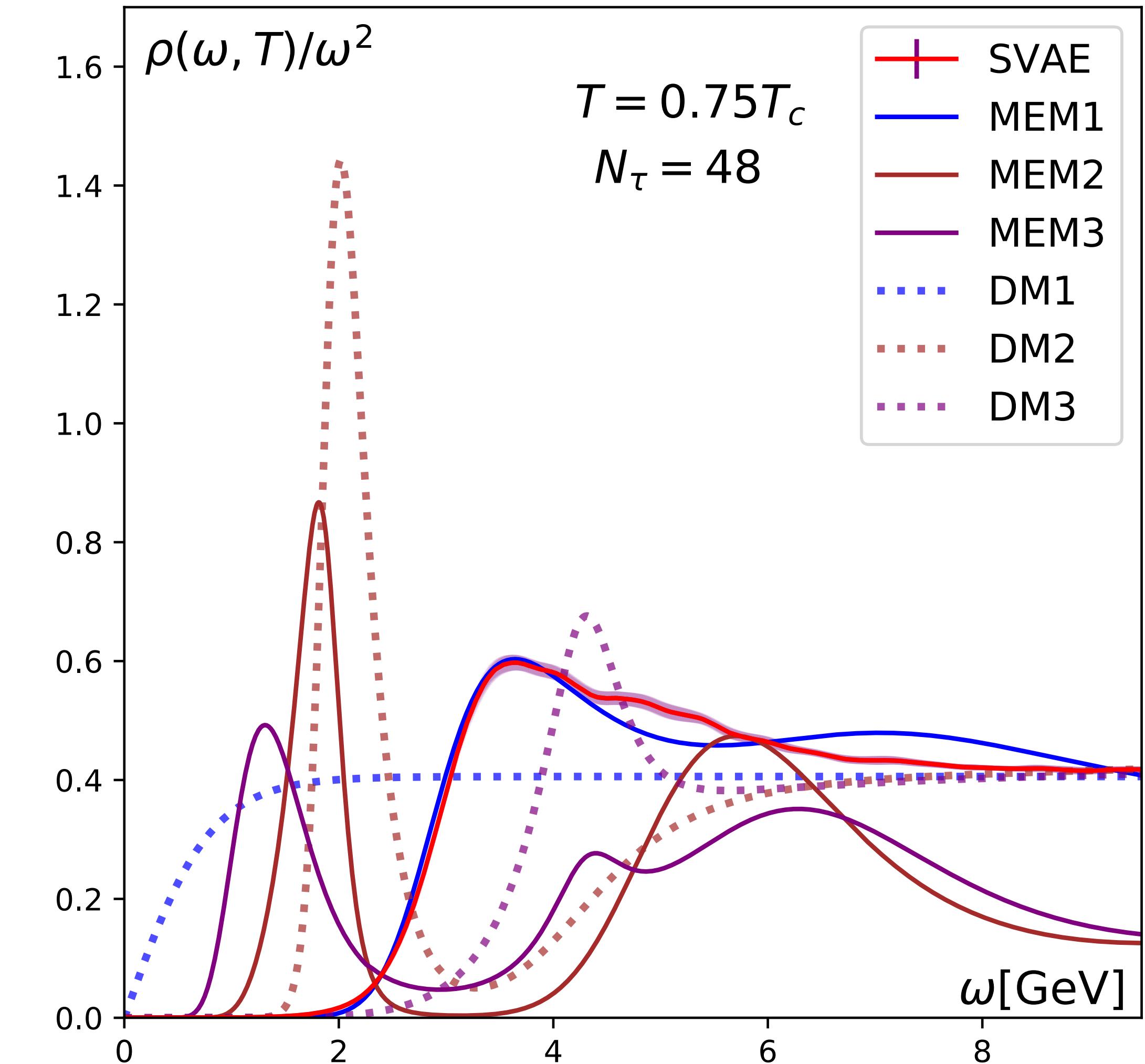


# Caution: $N_\tau$ dependence

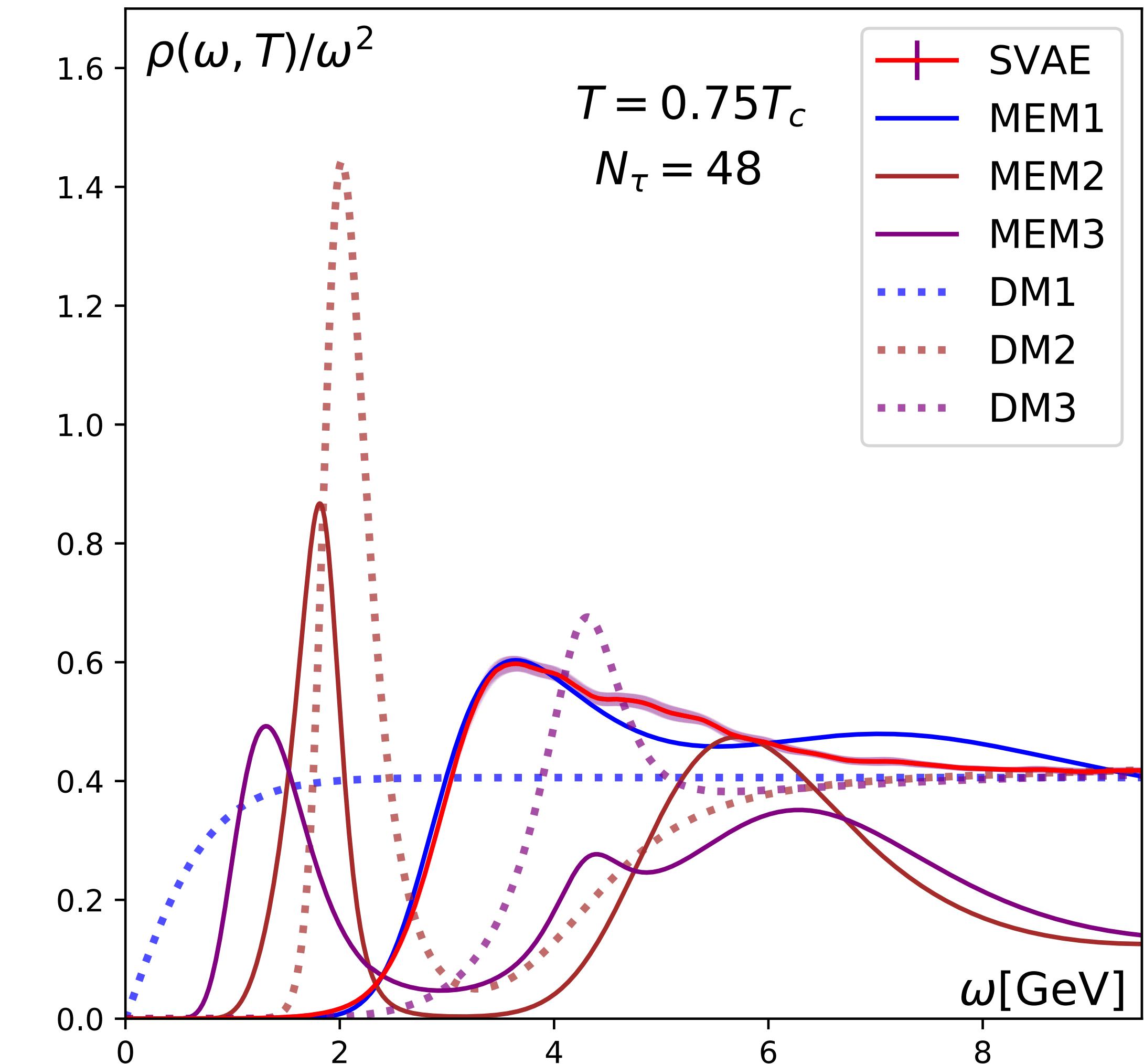
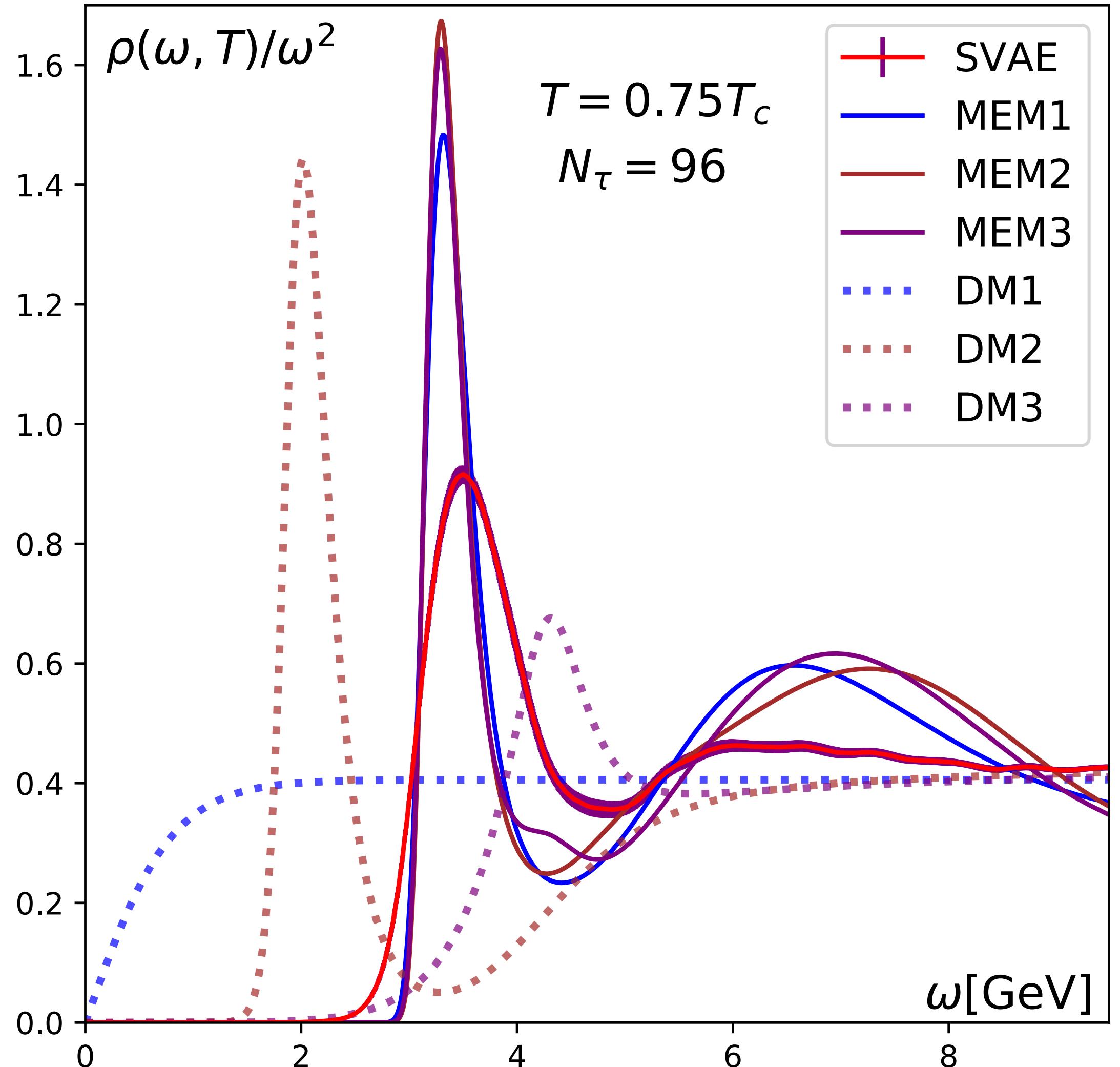
Reconstruct correlator with spf fixed at  $T' = 0.75T_c$ :  $N'_\tau = 96$  to  $N_\tau = 48$

$$G_{rec}(\tau, T; T') = \int_0^\infty d\omega K(\tau, T, \omega) \rho(\omega, T')$$

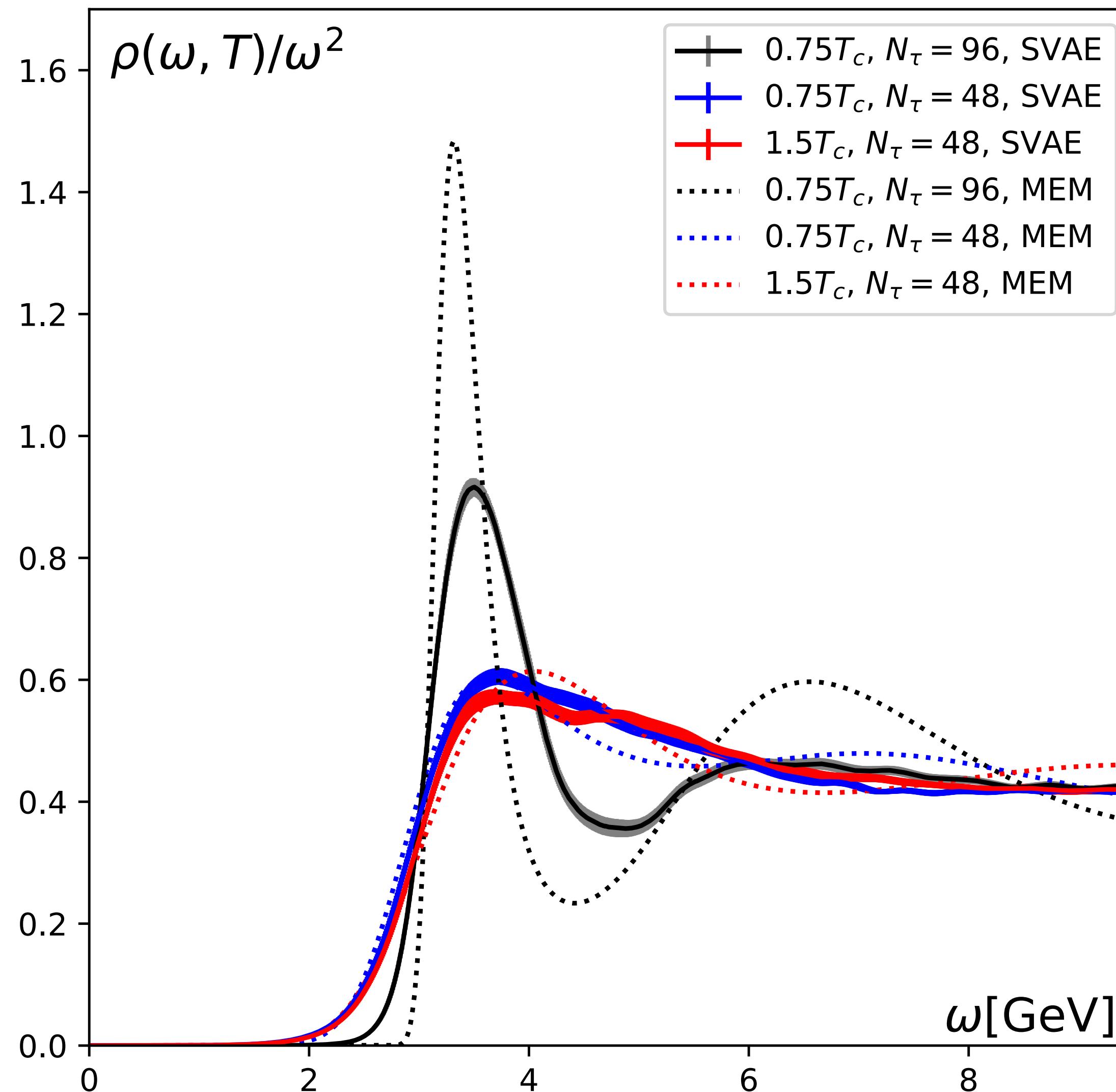
$$G_{rec}(\tau, T; T') = \sum_{\tau'=\tau; \Delta\tau'=N_\tau}^{N'_\tau-N_\tau+\tau} G(\tau', T')$$



# Caution: $N_\tau$ dependence



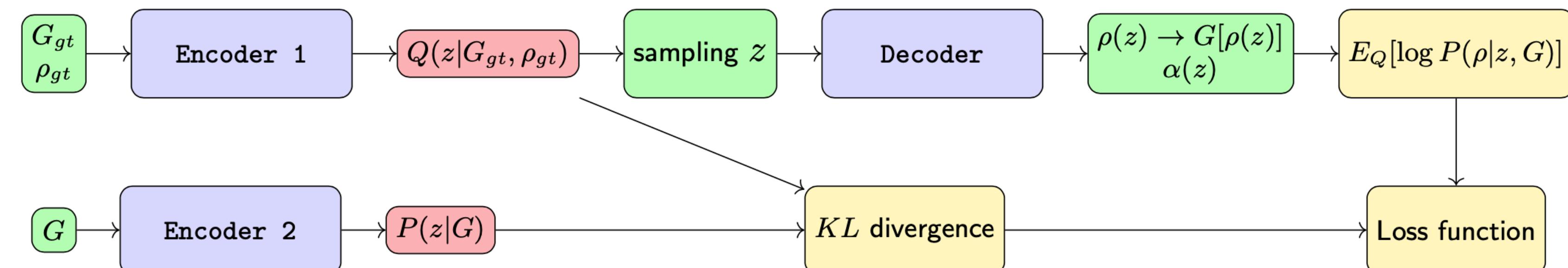
# Summary plot on the SPF



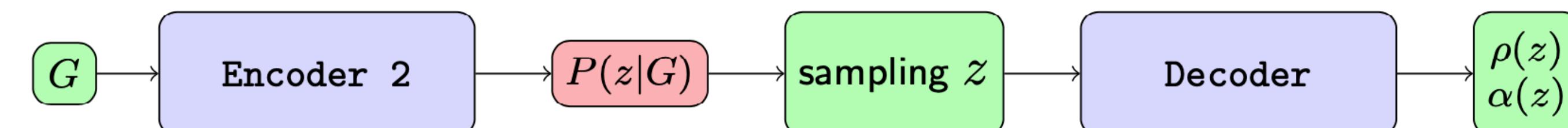
- $N_\tau=98$ : peak location agree within error from sVAE and MEM
- $N_\tau=48$ : SVAE is consistent with MEM using a featureless DM at both  $0.75 T_c$  and  $1.5 T_c$
- Hard to tell the fate of  $\eta_c$  at  $1.5 T_c$  with  $N_t=48$  data

# Summary & Outlook

- We proposed a novel neural network, SVAE, which can be trained to obtain the most probable image of the spectral function
- The loss function of SVAE includes an entropy and a likelihood term which are balanced by a weight  $\alpha(z)$
- Training: A general  $\rho$  is used with corresponding correlator having the error of correlators mimic to LQCD



• Reconstruction:



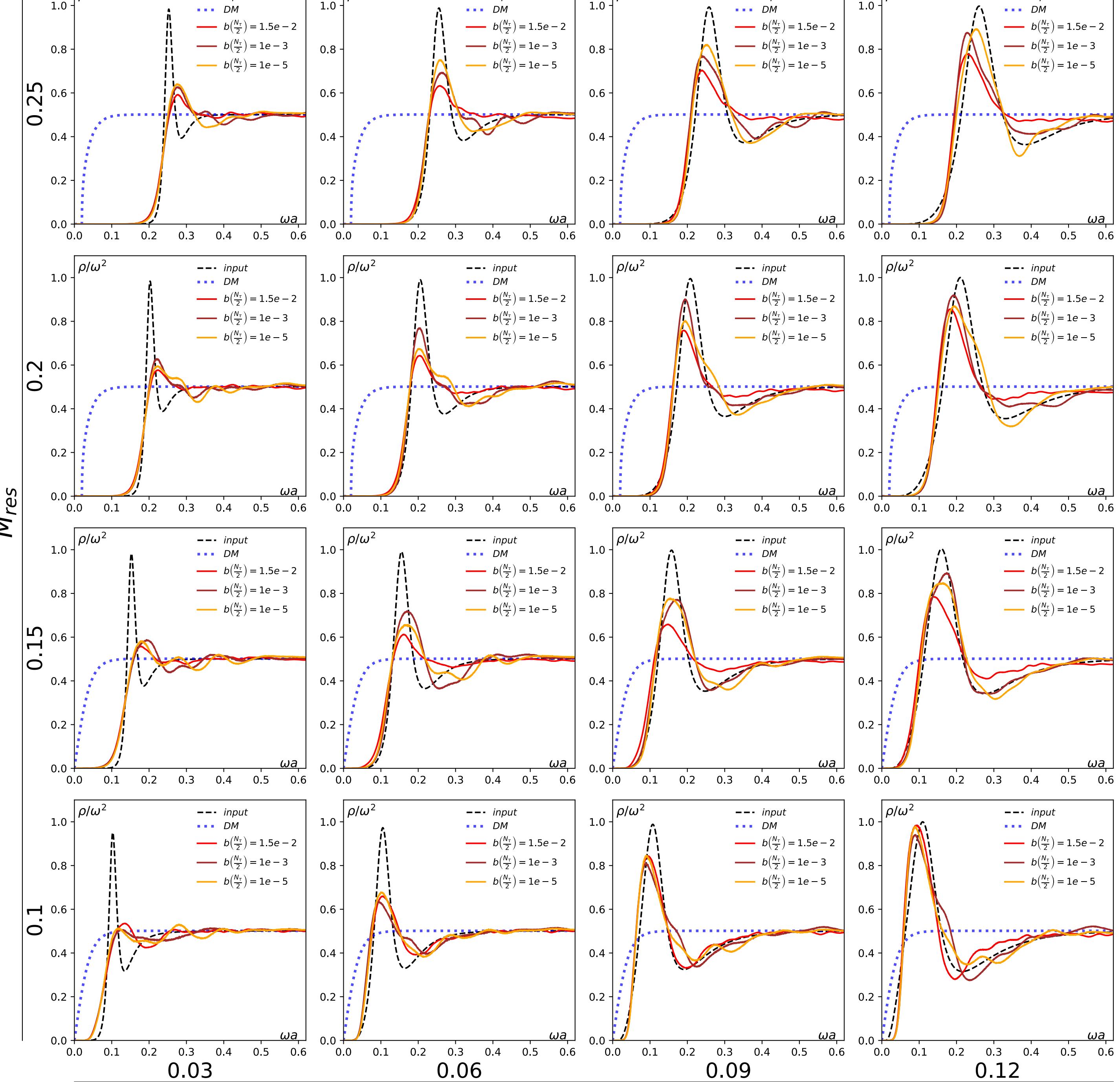
# Summary & Outlook

- ➊ Mock tests shows that SVAE is comparable to MEM and even outperforms MEM in certain cases
- ➋ Application to lattice QCD data of charmonium correlator in the pseudo-scalar channel
  - ➌ SVAE is consistent with MEM results using a featureless DM
  - ➌ Fate of  $\eta_c$  at 1.5 Tc is difficult to resolve with  $N_\tau = 48$
- ▶ Trial with a different kernel  $\omega^2/(\omega^2 + p^2)$  or  $e^{-\tau\omega}$
- ▶ Application to other physical problems

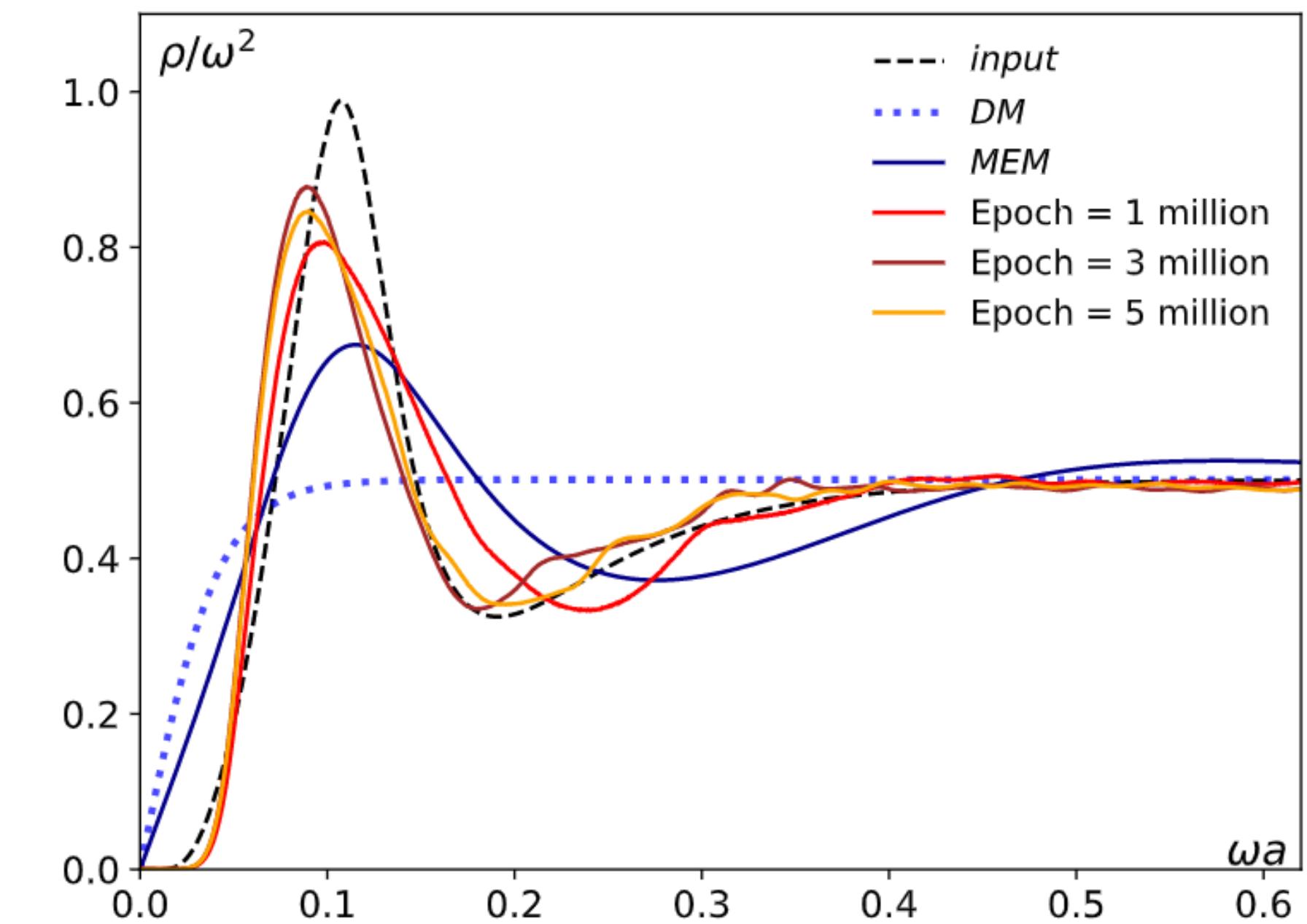
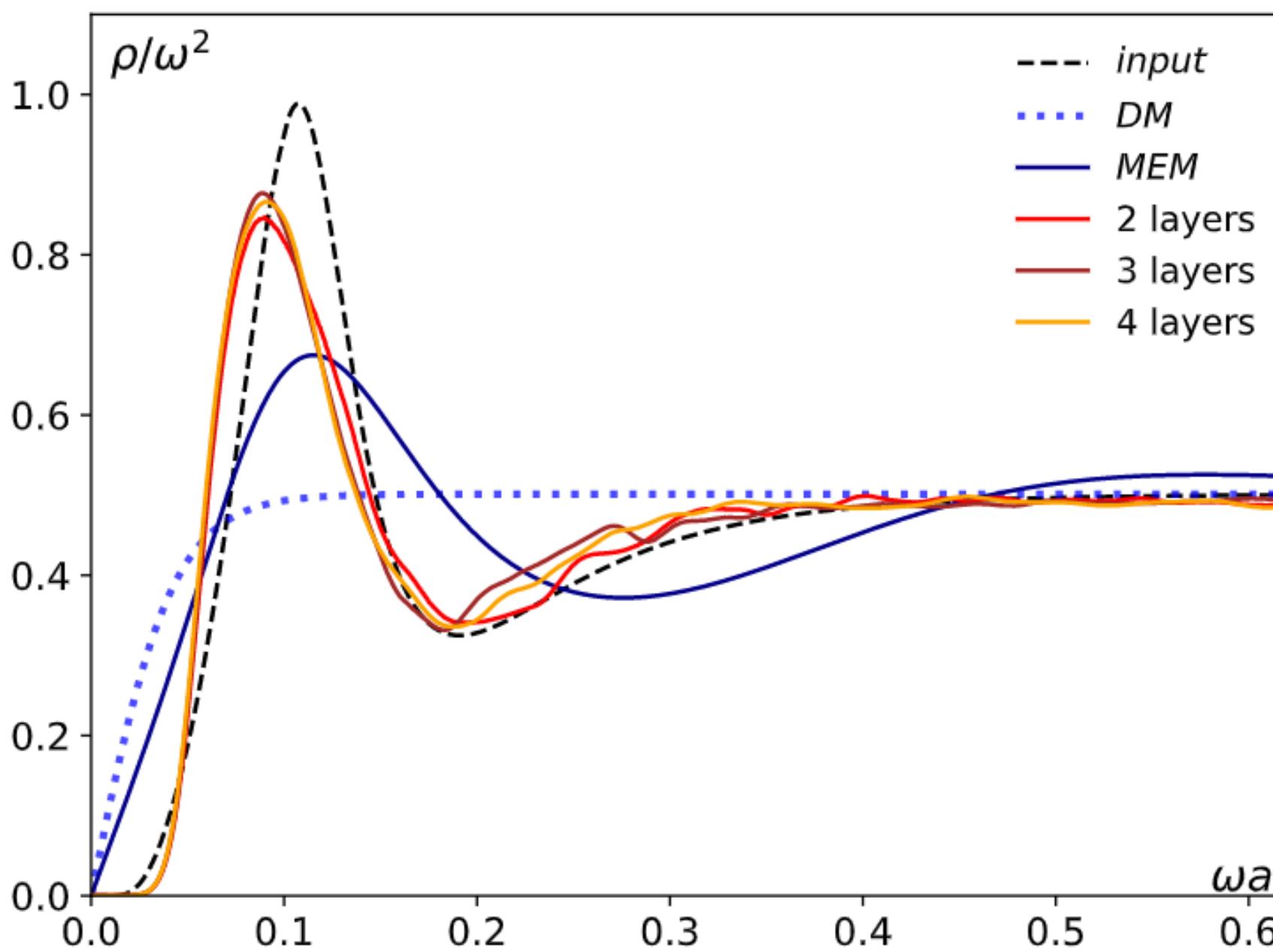
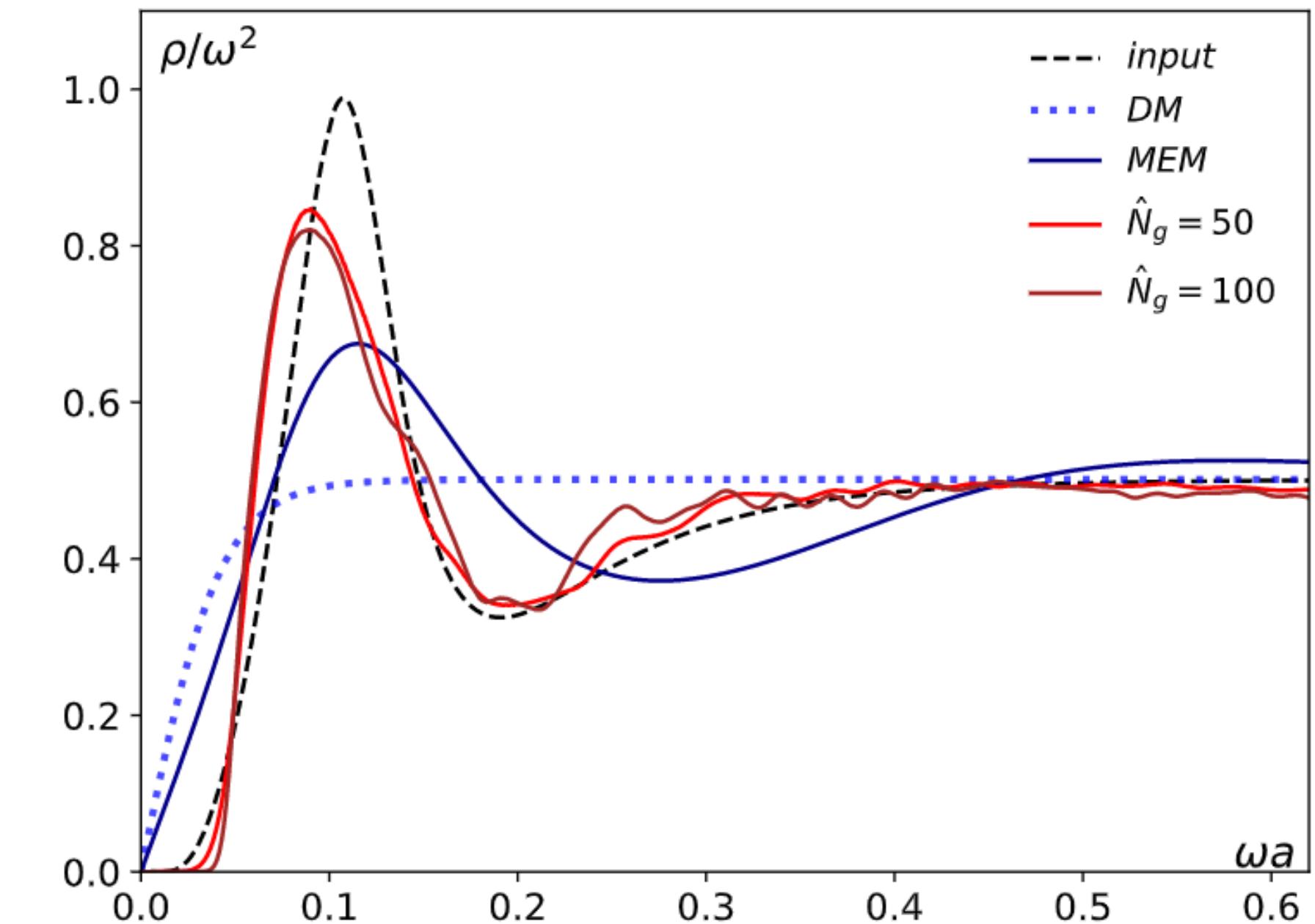
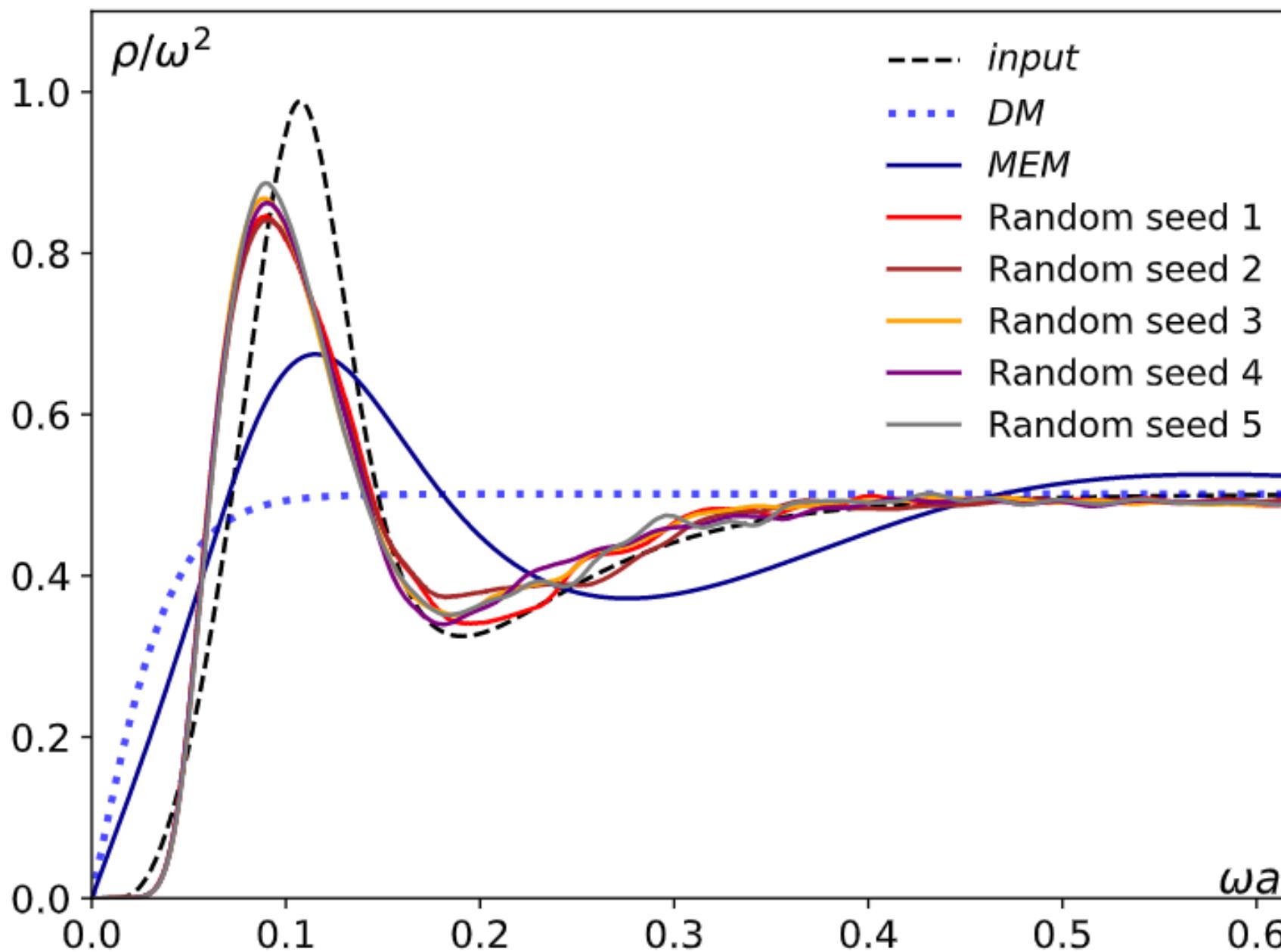
# Backup

# Noise dependences

$N_\tau = 48$



## Mild dependence on hyperparameters



# Hyper-parameters

Subnet \ Layer	Input layer	Hidden Layer 1	Hidden Layer 2	Output layer
Encoder 1	$\rho_{gt} \& G_{gt} : N_\omega + N_\tau/2 - \tau_{min} + 1$	500, Relu	250, Hard sigmoid	$\mu : N_g \times N_z$ , none
				$\sigma : N_g \times N_z$ , softplus
				$\pi : N_g$ , softmax
Encoder 2	$G : N_\tau/2 - \tau_{min} + 1$	100, Relu	250, Hard sigmoid	$\hat{\mu} : N_g \times N_z$ , none
				$\hat{\sigma} : N_g \times N_z$ , softplus
				$\hat{\pi} : N_g$ , softmax
Decoder	$z : N_z$	250, Relu	500, Hard sigmoid	$\rho(z) : N_\omega$ , softplus
				$\alpha(z) : N_\tau/2 - \tau_{min} + 1$ , softplus