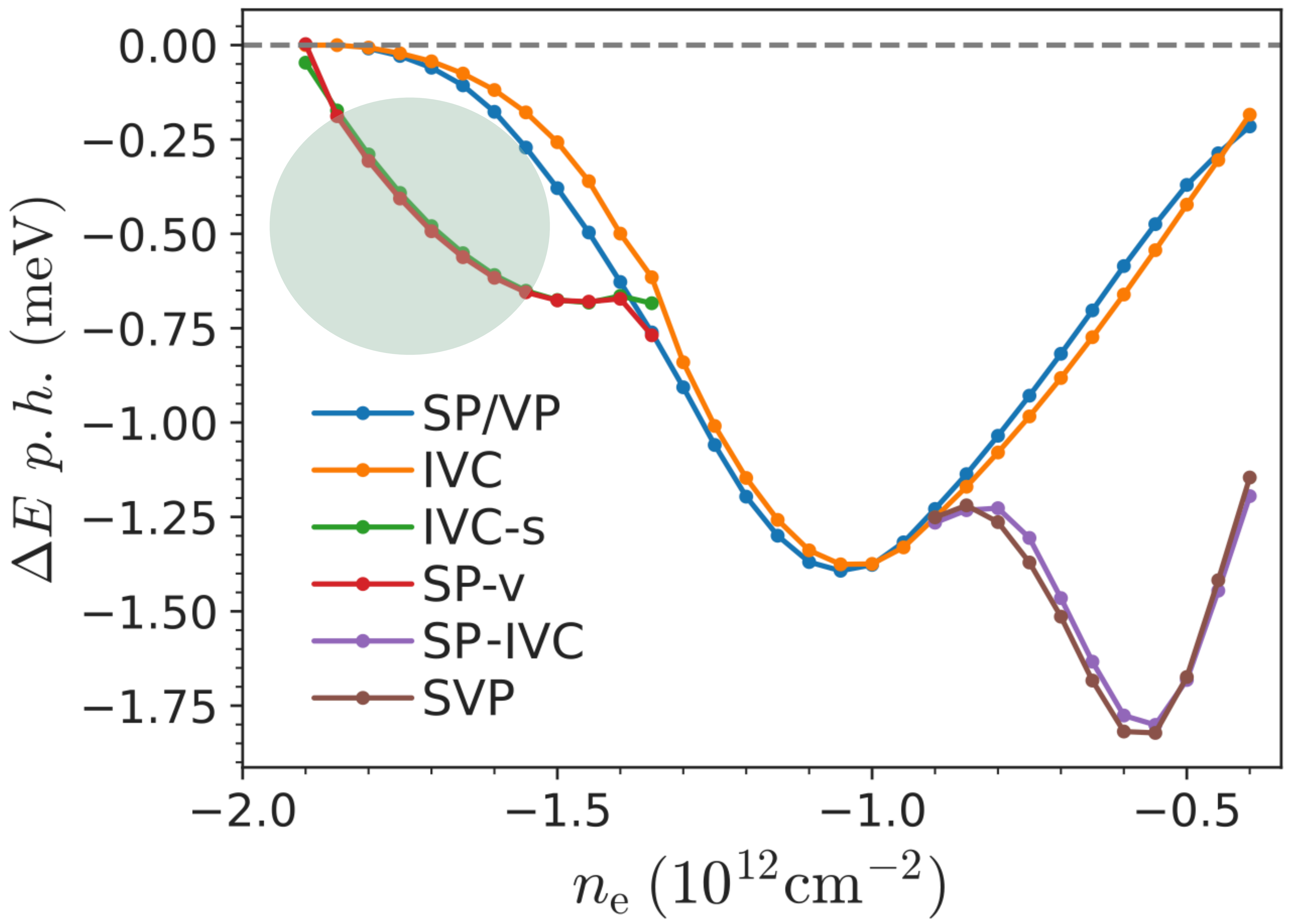
Physics-informed transformers for electronic quantum states

João Sobral
Michael Perle
Mathias S. Scheurer

Universität Stuttgart

Motivation: Hartree-Fock and Mean Field Approximation

- Characterization of ground states in Moiré systems and other materials.



Chatterjee, S., Wang, T., Berg, E. *et al.* *Nat Commun* 13, 6013 (**2022**).

$$\hat{H} = \sum_{\substack{\zeta,\eta \\ \vec{k}_1,\vec{k}_2}} h^{\vec{k}_1,\vec{k}_2}_{\zeta,\eta} d^\dagger_{\vec{k}_1} d_{\vec{k}_2} + $$

$$+ \sum_{\substack{\alpha,\beta,\gamma,\delta \\ \vec{k}_1,\vec{k}_2,\vec{k}_3,\vec{k}_4}} V^{\vec{k}_1,\vec{k}_2,\vec{k}_3,\vec{k}_4}_{\alpha,\beta,\gamma,\delta} d^\dagger_{\alpha,\vec{k}_1} d^\dagger_{\beta,\vec{k}_2} d_{\gamma,\vec{k}_3} d_{\delta,\vec{k}_4}$$

Interacting term

$$\hat{H} = \sum_{\vec{k},\alpha} \epsilon_{\vec{k},\alpha} \bar{d}^\dagger_{\vec{k},\alpha} \bar{d}_{\vec{k},\alpha} + ✗$$

MFT

New basis

$$\bar{d}_{\vec{k},\alpha} = \sum_p \left(U_{\vec{k}}\right)_{\alpha,p} d_{\vec{k},p}$$

Can generative models help?

Variational Principle and NQS

NN parameters

$$\arg\min_{\vec{\theta}} E(\vec{\theta}) = \arg\min_{\vec{\theta}} \frac{\left\langle \Psi_{\vec{\theta}} \middle| \hat{H} \middle| \Psi_{\vec{\theta}} \right\rangle}{\left\langle \Psi_{\vec{\theta}} \mid \Psi_{\vec{\theta}} \right\rangle},$$

Bounded by NN representational power

$$\min E\left(\vec{\theta}\right) \geq E_{\mathrm{GS}}$$

e.g., $|10101\rangle$

$$q_{\vec{\theta}}(\mathbf{s}) = \left|\psi_{\vec{\theta}}\left(\vec{s}\right)\right|^2 / \sum_{\vec{s}'} \left|\psi_{\vec{\theta}}\left(\vec{s}'\right)\right|^2$$

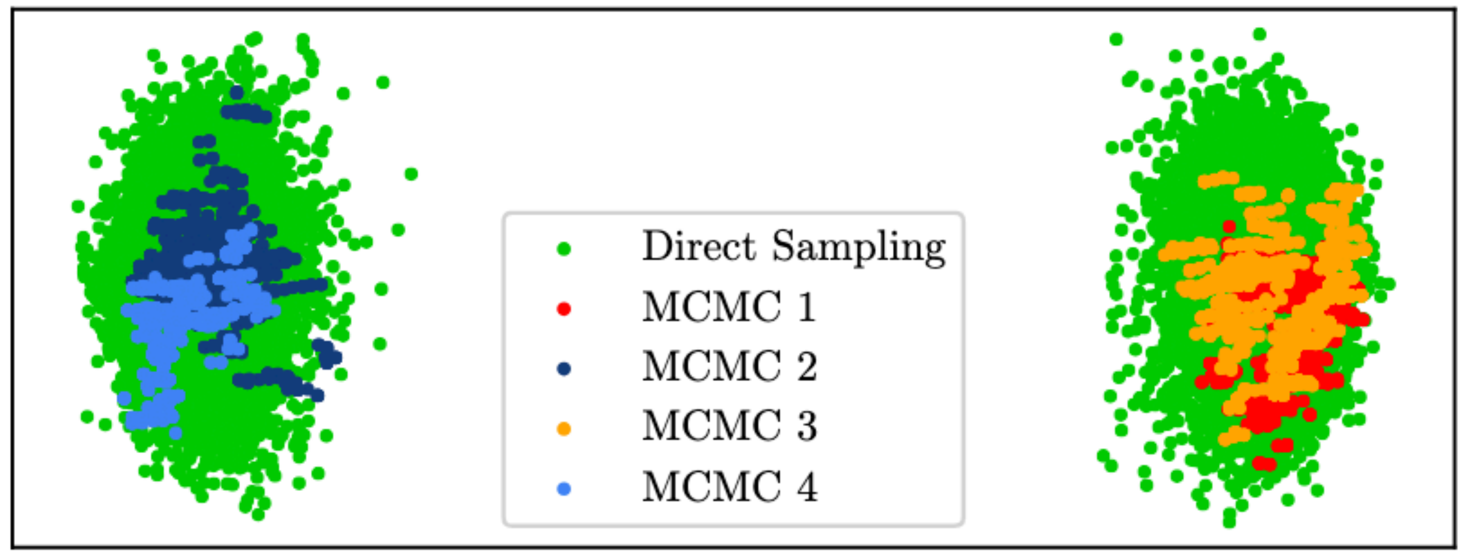
$$\psi_{\vec{\theta}}\left(\vec{s}\right) = \sqrt{q_{\vec{\theta}}\left(\vec{s}\right)}\, e^{i\phi_{\vec{\theta}}(\vec{s})}$$

$$q_{\vec{\theta}}\left(\vec{s}\right) = \prod_{i=1}^{N} q\left(x_i \mid x_{i-1}, \ldots, x_1\right).$$

Autoregressive NQS

Efficient (and parallel) sampling

Sharir, Levine et al. Phys. Rev. Lett. **124**, 020503 (2020)



Transformer quantum states

Viteritti, Rende & Becca. Phys. Rev. Lett. **130**, 236401 (2023)

Carleo G. & Troyer. *Science* **355**, 602-606 (2017) | Medvidović, M. & Moreno, J.R. *Eur. Phys. J. Plus* **139**, 631 (2024). | Hannah Lange et al. arXiv:2402.09402 (2024)

Modification to Variational Framework

$$\left|\Psi_{\{\vec{\theta},\alpha\}}\right\rangle = \alpha\left|\text{RS}\right\rangle + \sqrt{1-\alpha^2}\sum_{\vec{s}\neq\text{RS}} \psi_{\vec{\theta}}(\vec{s})\,|\,\vec{s}\rangle,$$

Reference state

$$\arg\min_{\{\vec{\theta},\alpha\}} E(\vec{\theta},\alpha) = \arg\min_{\{\vec{\theta},\alpha\}} \frac{\left\langle\Psi_{\{\vec{\theta},\alpha\}}\left|\hat{H}'\right|\Psi_{\{\vec{\theta},\alpha\}}\right\rangle}{\left\langle\Psi_{\{\vec{\theta},\alpha\}}\,|\,\Psi_{\{\vec{\theta},\alpha\}}\right\rangle}. \qquad \hat{H}' = \mathscr{U}_{\vec{k}}^{\dagger}\hat{H}\mathscr{U}_{\vec{k}}$$

HF-basis

$$\mathscr{U}_{\vec{k}} = \otimes_j^{N_e} U_{\vec{k}_j}$$

1. $E(\vec{\theta},\alpha) = \alpha^2 E_{RS} + (1-\alpha^2)\,\mathbb{E}_{\vec{s}\sim q_{\vec{\theta}}}\left[\hat{H}'_{\text{loc}}(\mathbf{s})\right] + \alpha\sqrt{1-\alpha^2}\,2\text{Re}\left(\mathbb{E}_{\vec{s}\sim q_{\vec{\theta}}}\left[\hat{H}'^{RS}_{loc}(\vec{s})\right]\right).$ Calculate energy functional

2. $\nabla_{\vec{\theta}}\left\langle\hat{H}'\right\rangle = 2\text{Re}\left(\mathbb{E}_{\mathbf{s}\sim q_{\vec{\theta}}}\left[\hat{H}_{loc}(\vec{s},\alpha)\cdot\nabla_{\vec{\theta}}\log\psi^*_{\vec{\theta}}(\vec{s})\right]\right),$ Optimize NN parameters

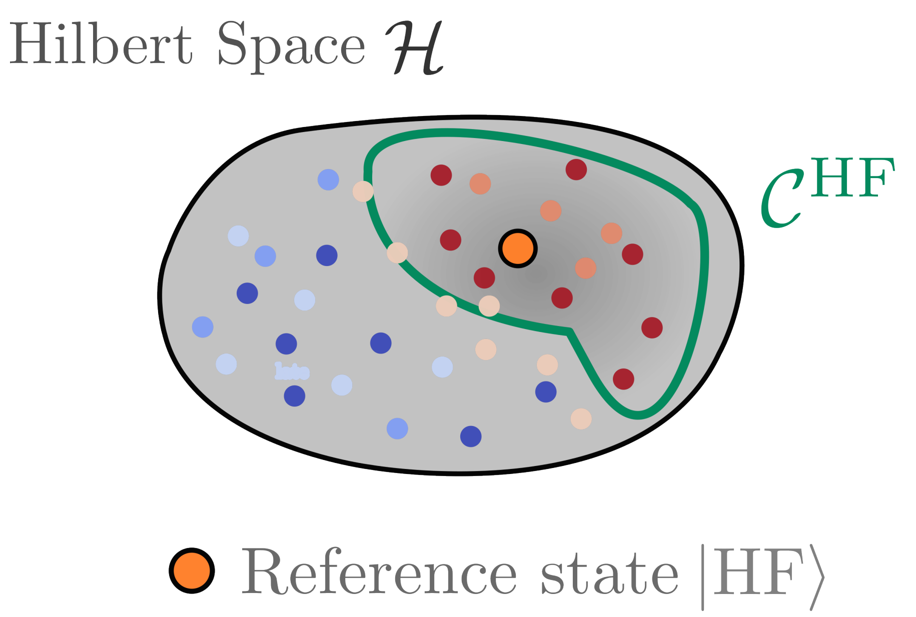
3. $\nabla_{\alpha_0}E(\vec{\theta},\alpha) = 2\alpha\nabla_{\alpha_0}\alpha\left[E_{RS} - E_{\vec{s}\vec{s}'} + \frac{E_{\vec{s}RS}(1-2\alpha^2)}{2\alpha\sqrt{1-\alpha^2}}\right],$ Optimize $\alpha = (1+\tanh\alpha_0)/2$

Overview

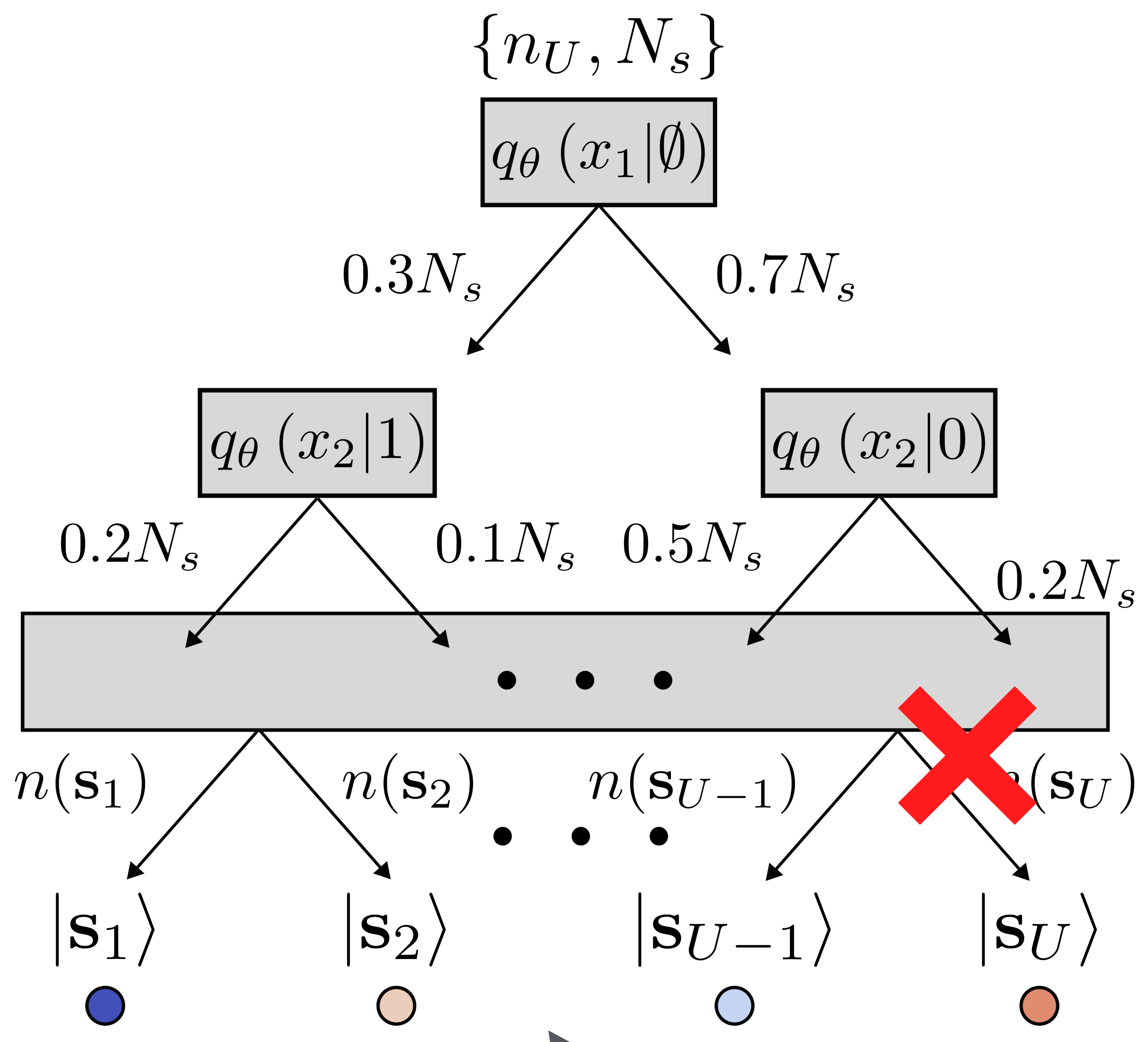
Barrett, Malyshev and Leovsky et al.
Nat. Mach. Intell. vol. 4, p 351–358 (2022)

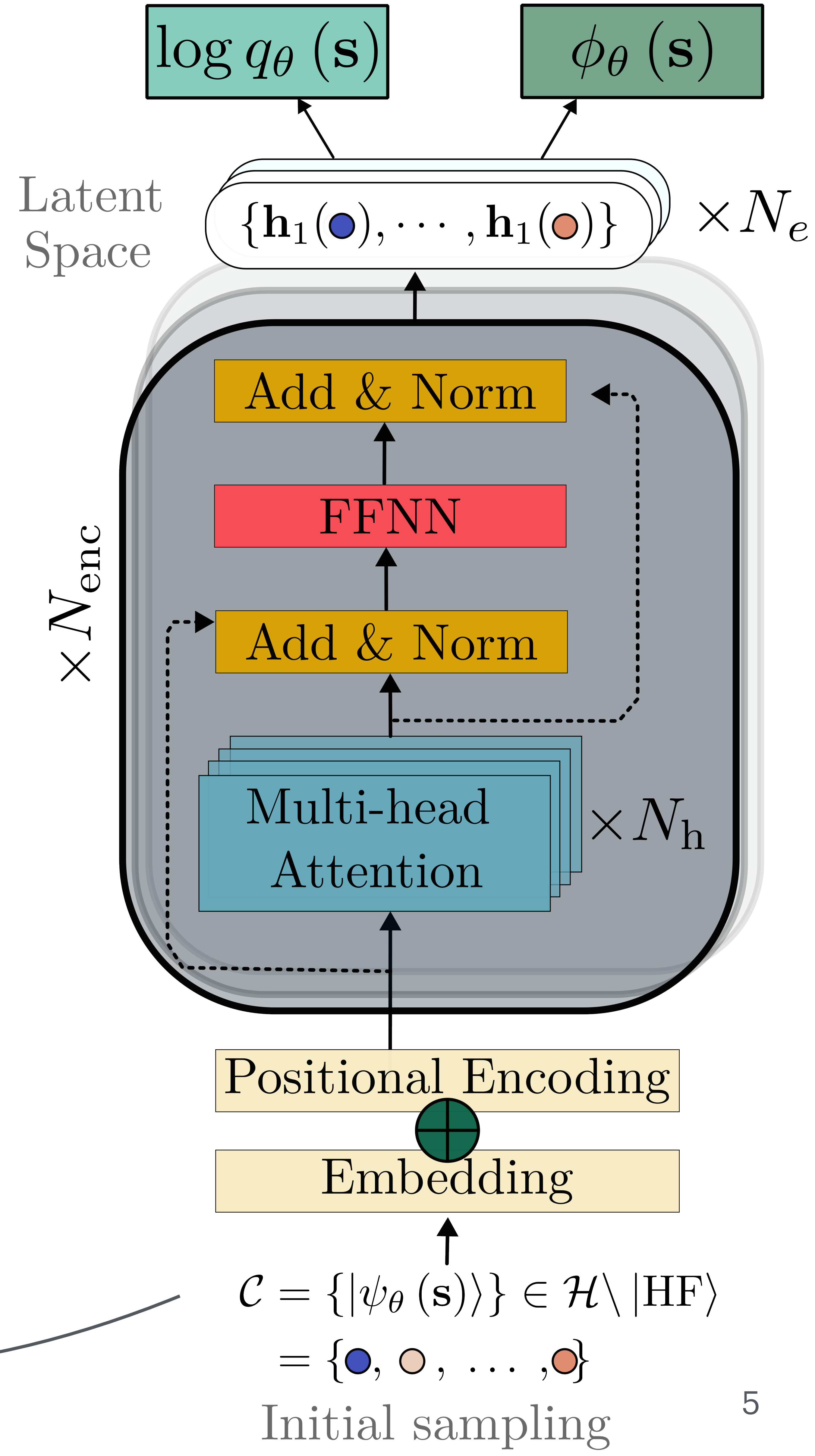
Hartree-Fock

$$\hat{H}_{\mathrm{HF}} = \mathcal{U}_{\mathrm{HF}}^{\dagger} \hat{H} \mathcal{U}_{\mathrm{HF}}$$



Get E_{HF} and $\mathcal{U}_{\mathbf{k}}$





Relative frequency

$$\mathbb{E}_{\vec{s}\sim q_{\vec{\theta}}}\left[\hat{H}'_{\mathrm{loc}}(\mathbf{s})\right] \approx \sum_{\mathbf{s}\in\mathcal{U}\neq \mathrm{RS}} \hat{H}'_{\mathrm{loc}}(\mathbf{s}) \frac{n(s)}{N_s}.$$

Repeat for N epochs

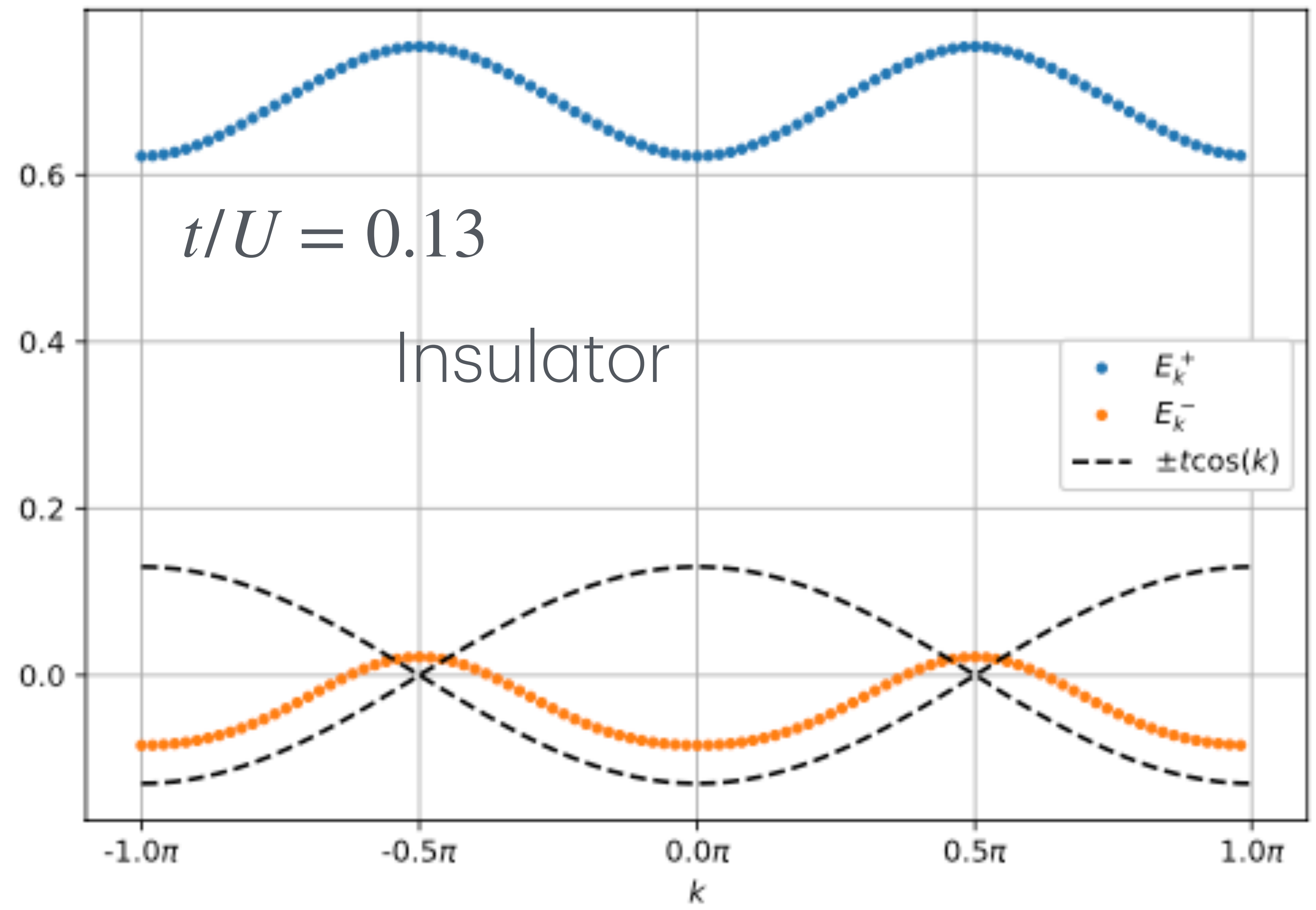
Example: Toy model with metal-insulator transition

Band | Chiral

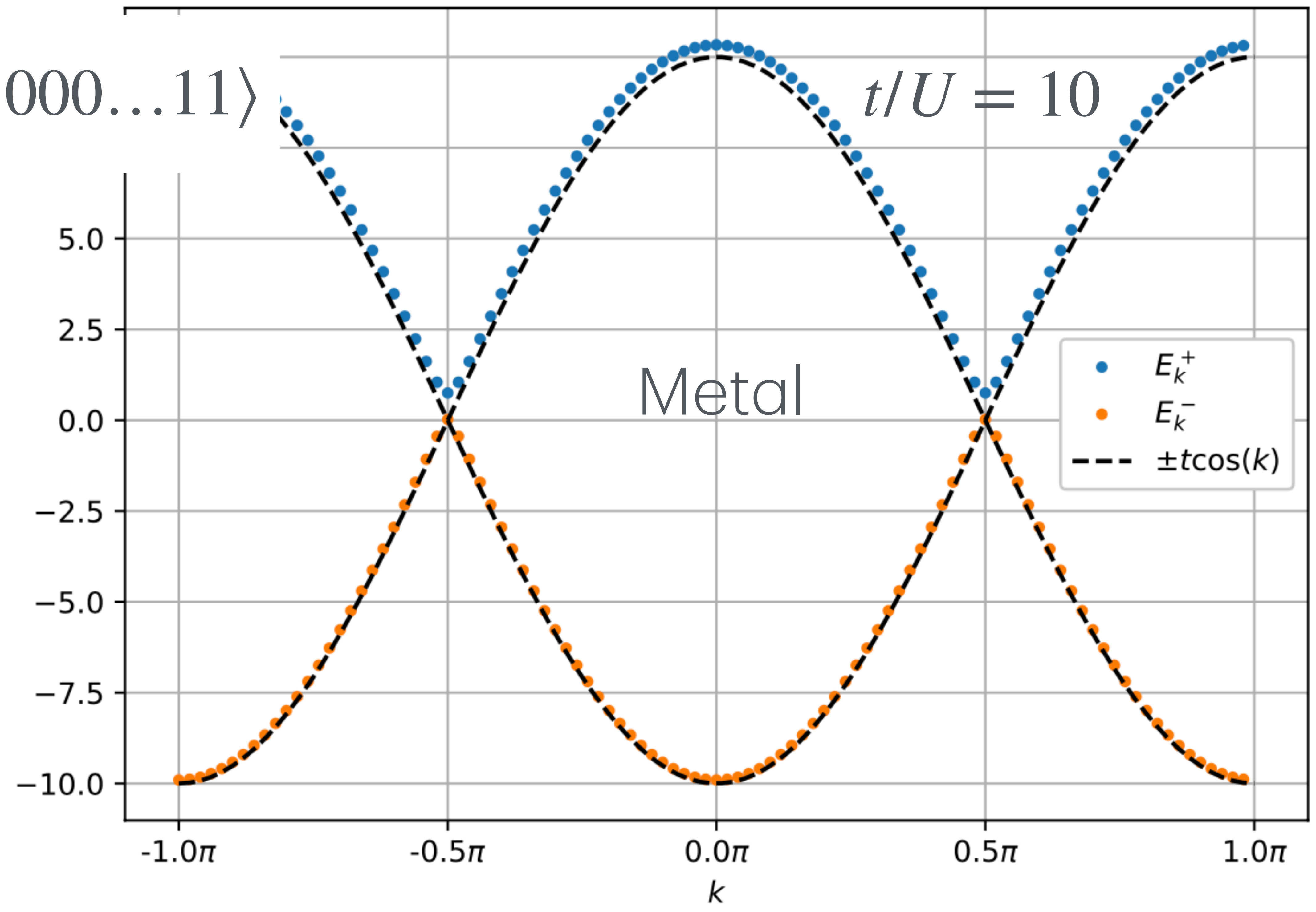
$$\hat{H} = \frac{t}{U}\sum_{k\in \mathrm{BZ}} \cos(k)\, c_k^\dagger \sigma_z c_k + \sum_{q\in\mathbb{R}} V(q)\rho_q\rho_{-q},$$

Interacting term

$$\rho_q = \sum_{k\in \mathrm{BZ}} \left(c^\dagger_{\mathrm{BZ}(k+q)}[f_1(k,q) + i\sigma_y f_2(k,q)]c_k - \sum_{G\in \mathrm{RL}} \delta_{q,G} f_1(k,G) \right),$$

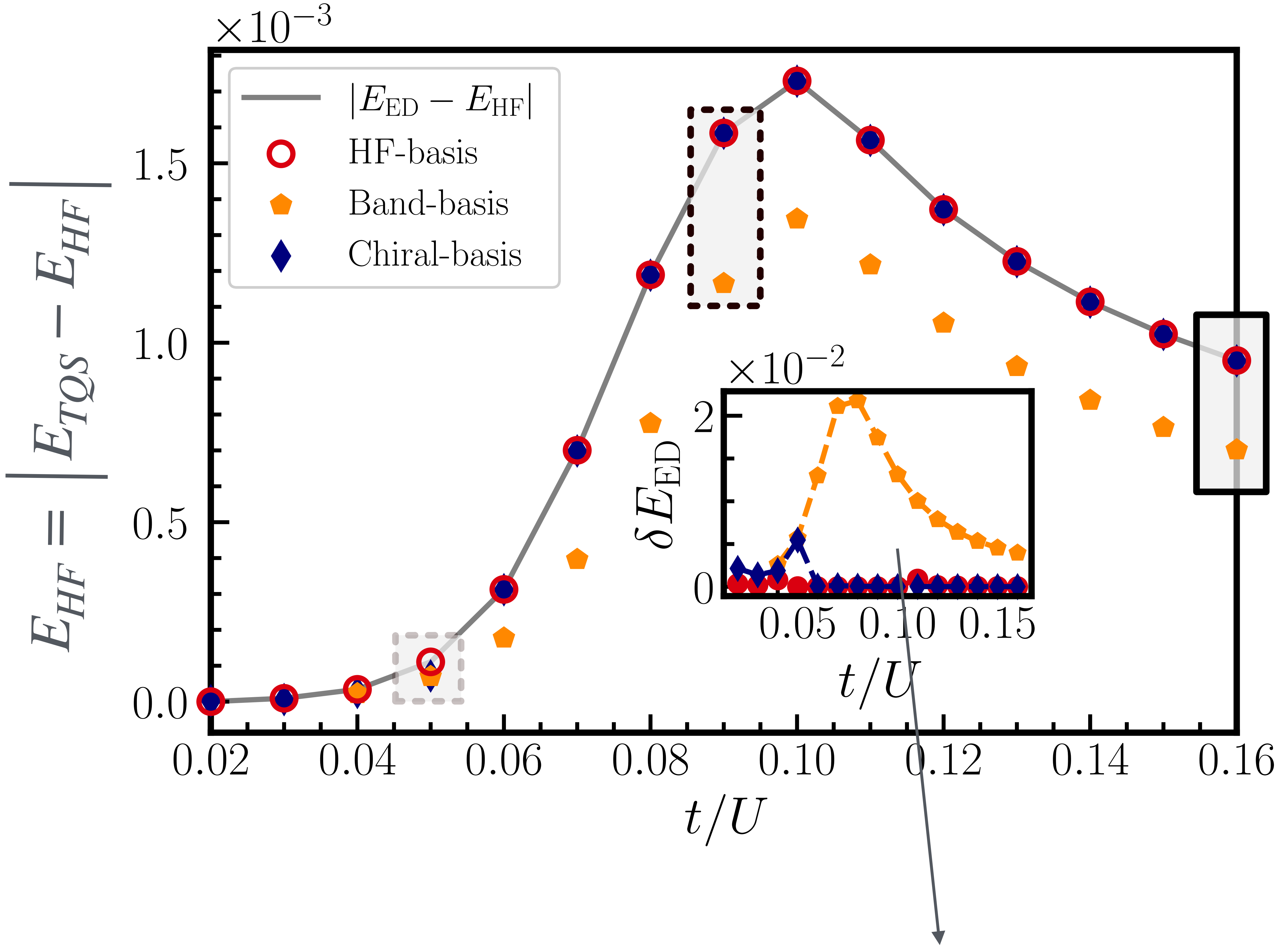


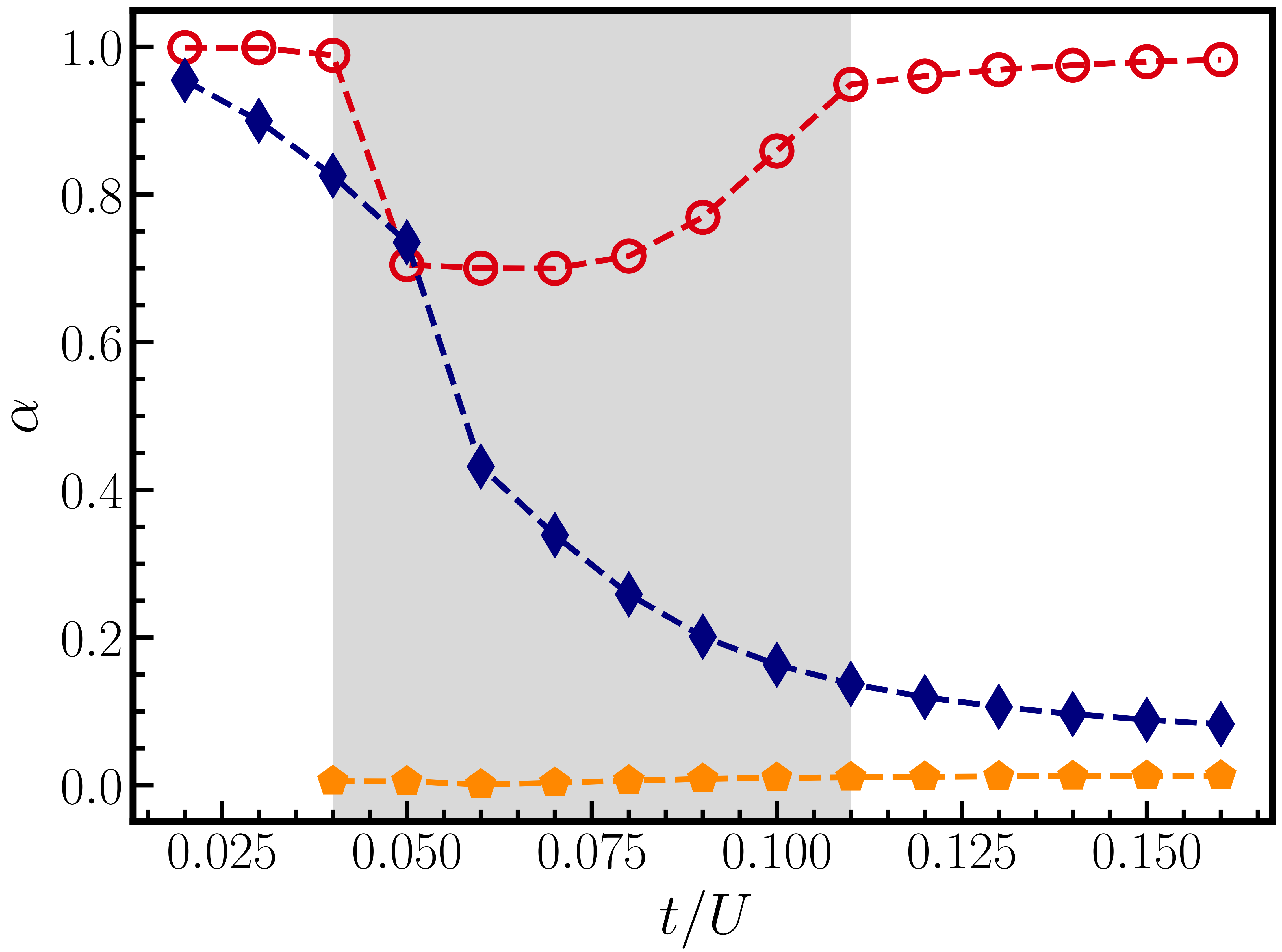
$|11\ldots000\ldots11\rangle$



Comparison in different bases

$N_e = 10 \quad \left|\mathrm{RS}\right\rangle = \left|11\ldots1111\right\rangle$

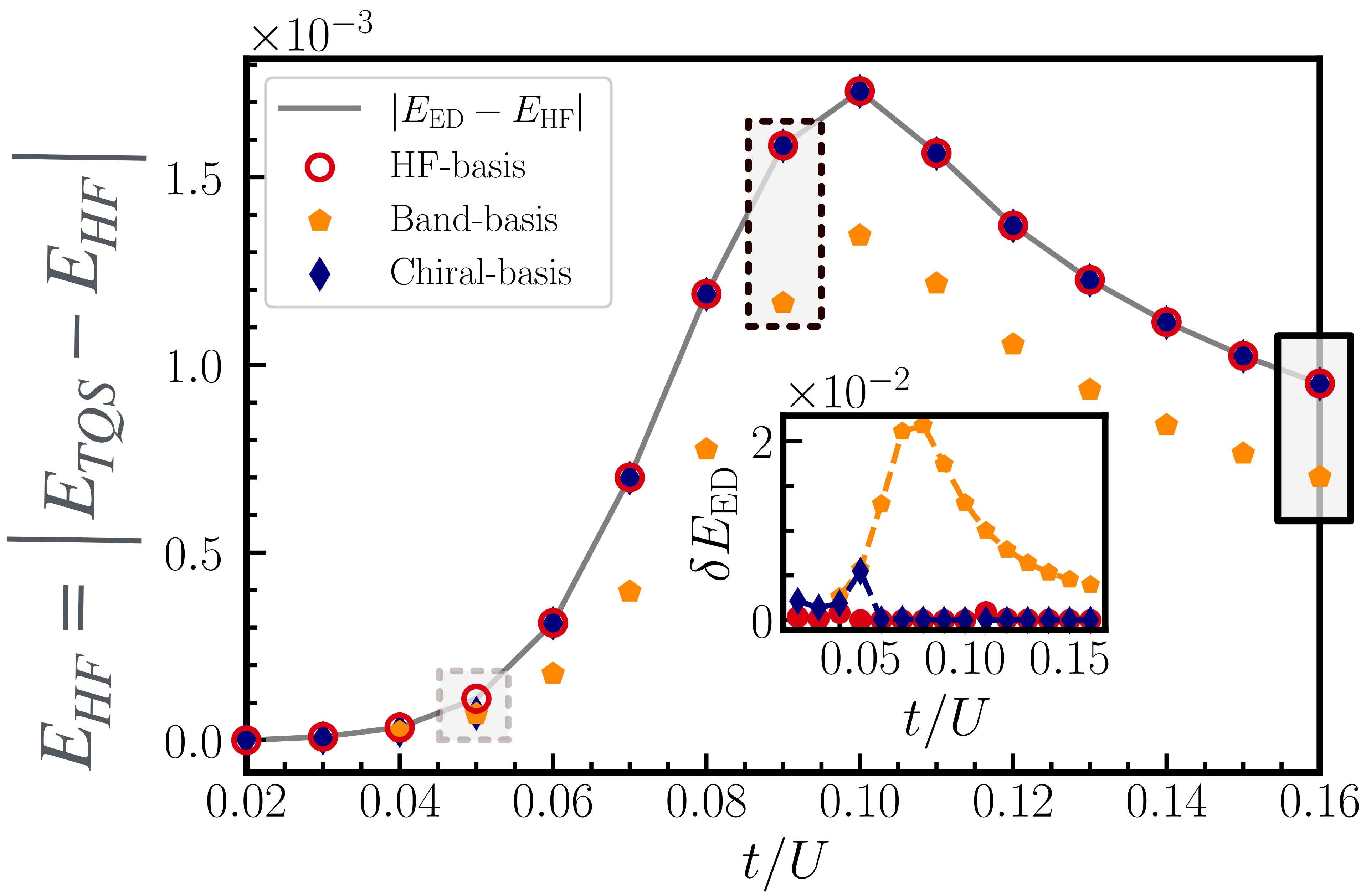


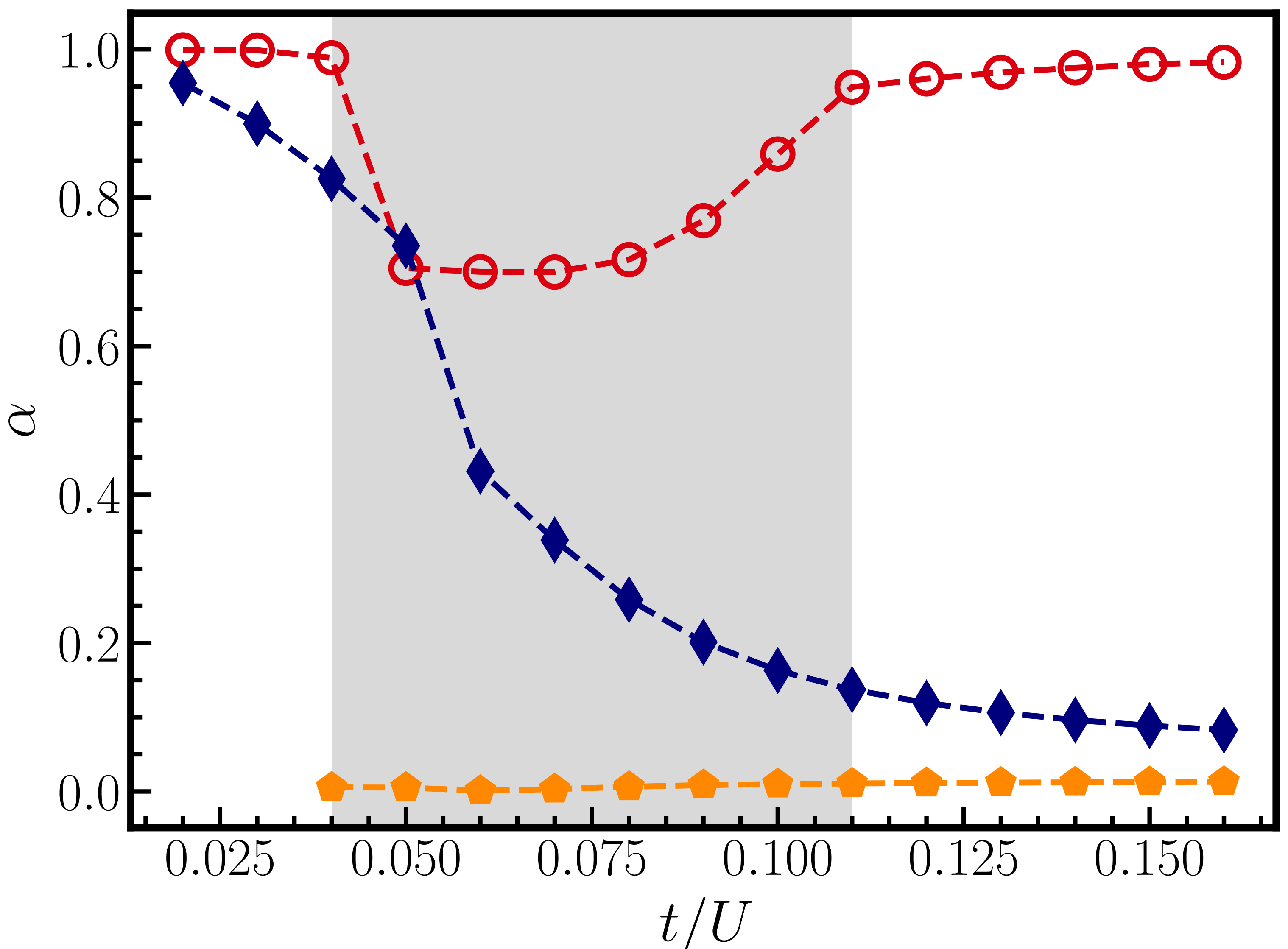


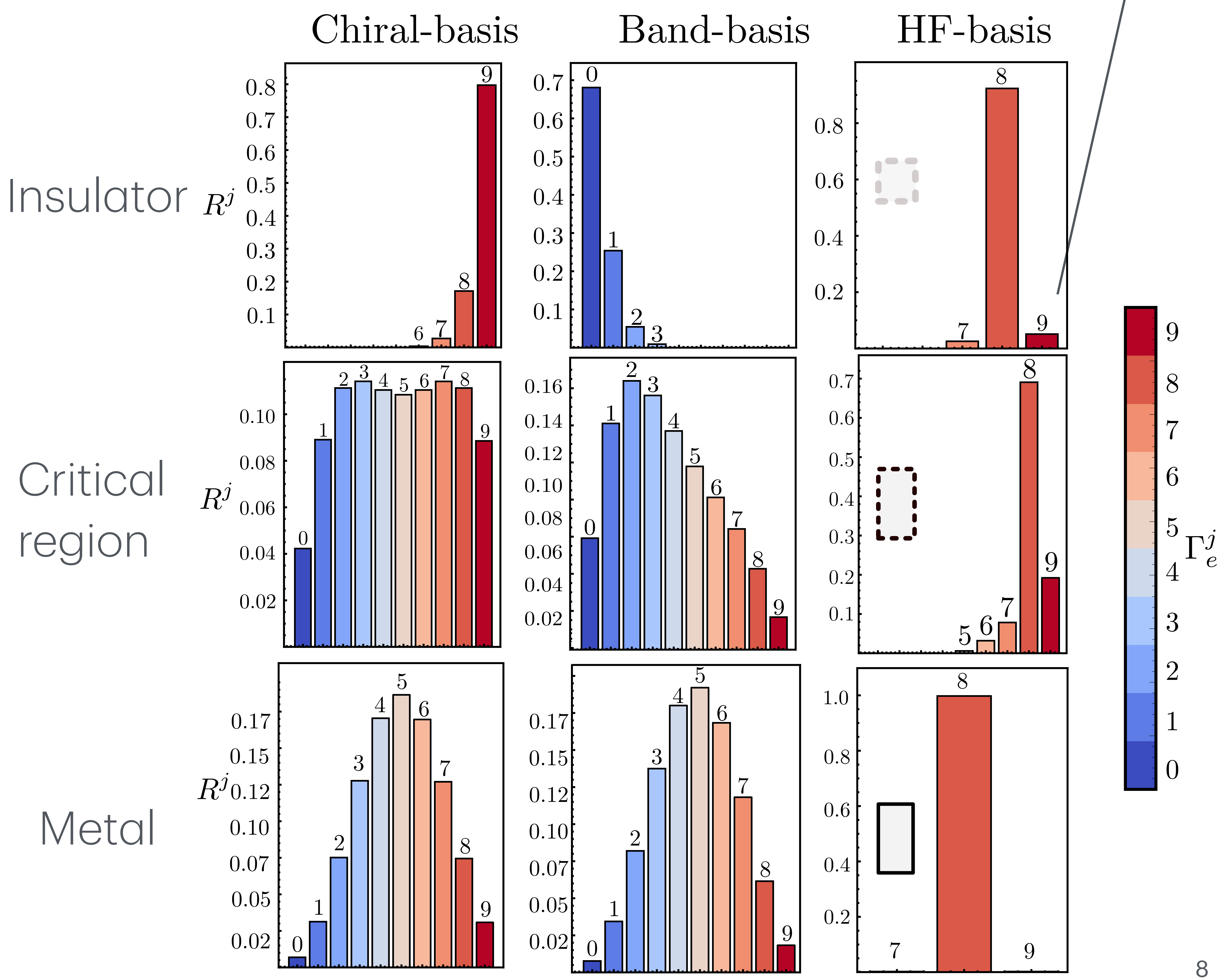
$\delta E_{ED} = \left|(E_{TQS} - E_{ED})/E_{ED}\right|$

Comparison in different bases

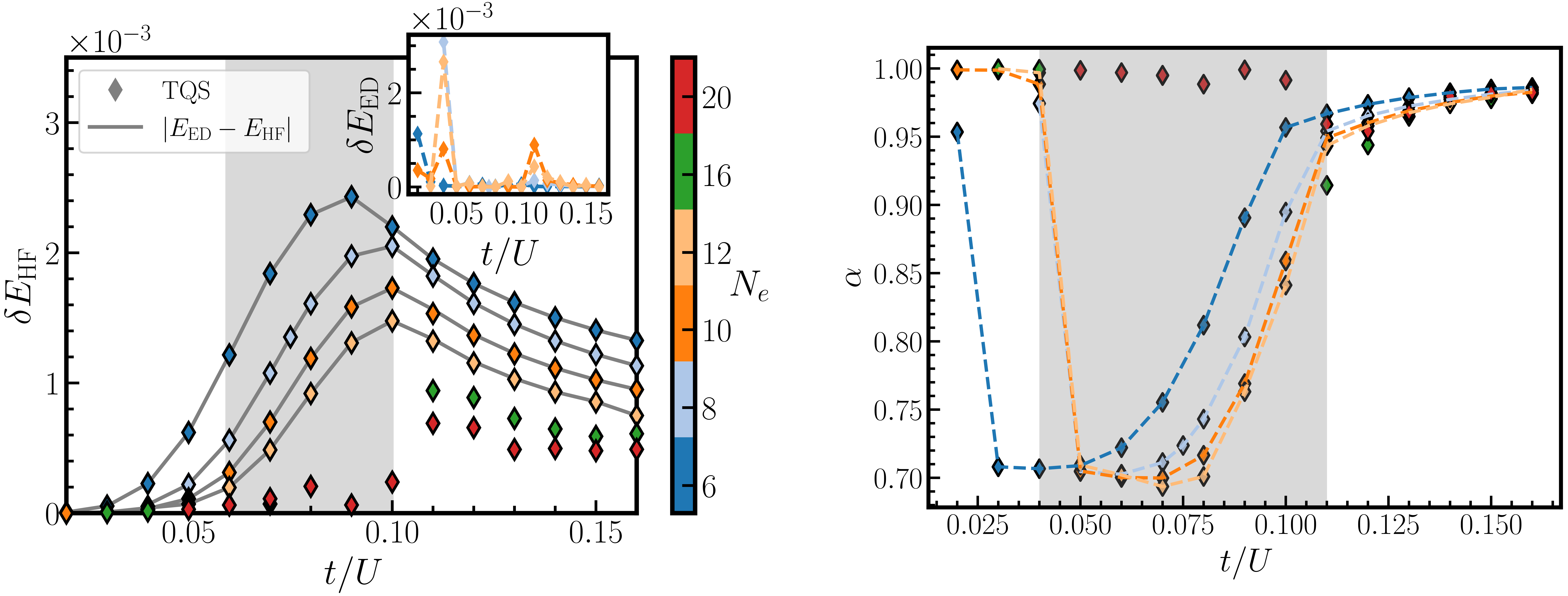
$|1111111110\rangle$



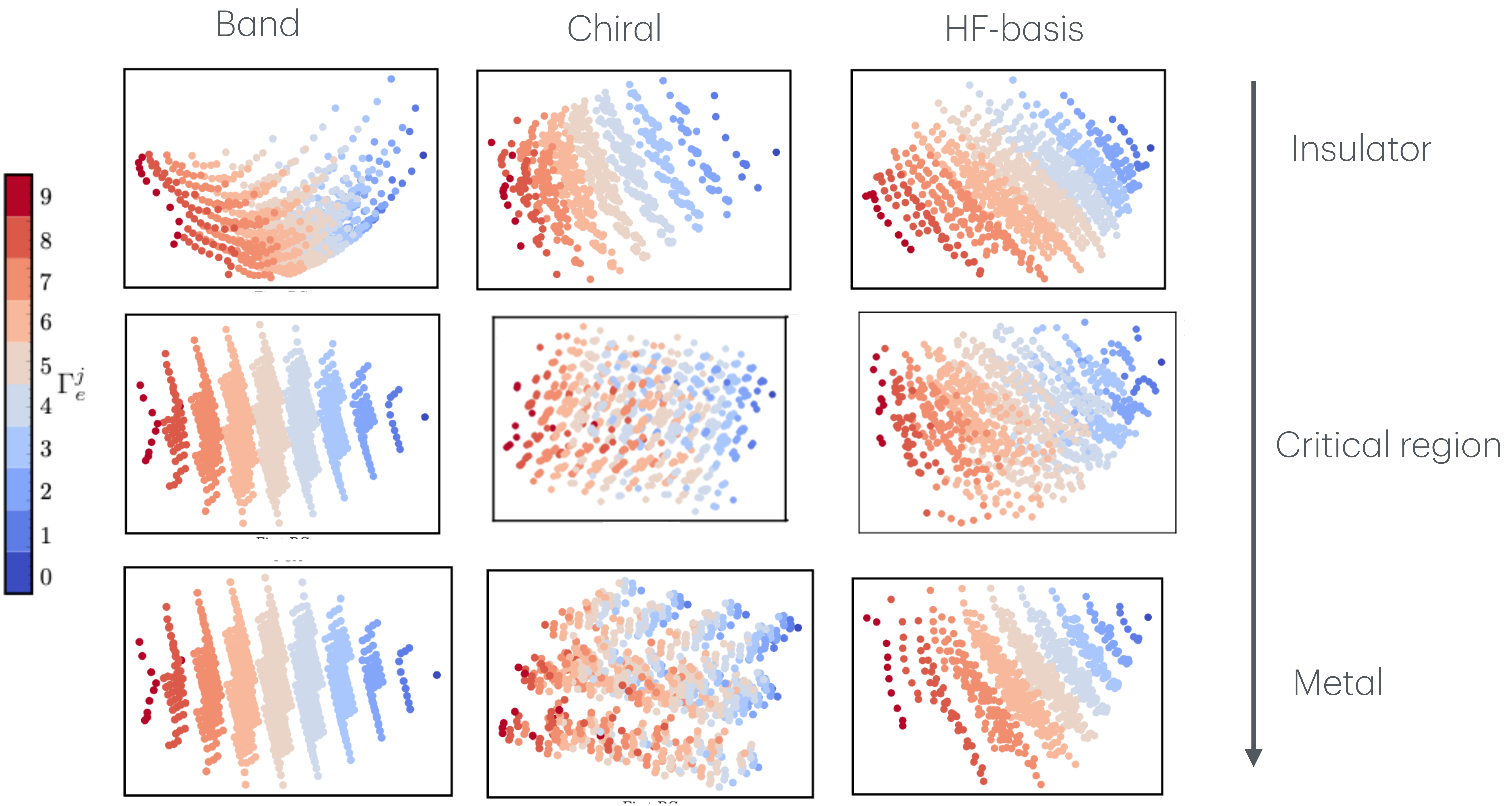




Scalability



Hidden representations for different bases



Conclusions & Outlook

HF-basis can induce TQS to find a more efficient (and naturally interpretable) GS representation;

→ What about other Hamiltonians in strongly correlated systems?

Future:

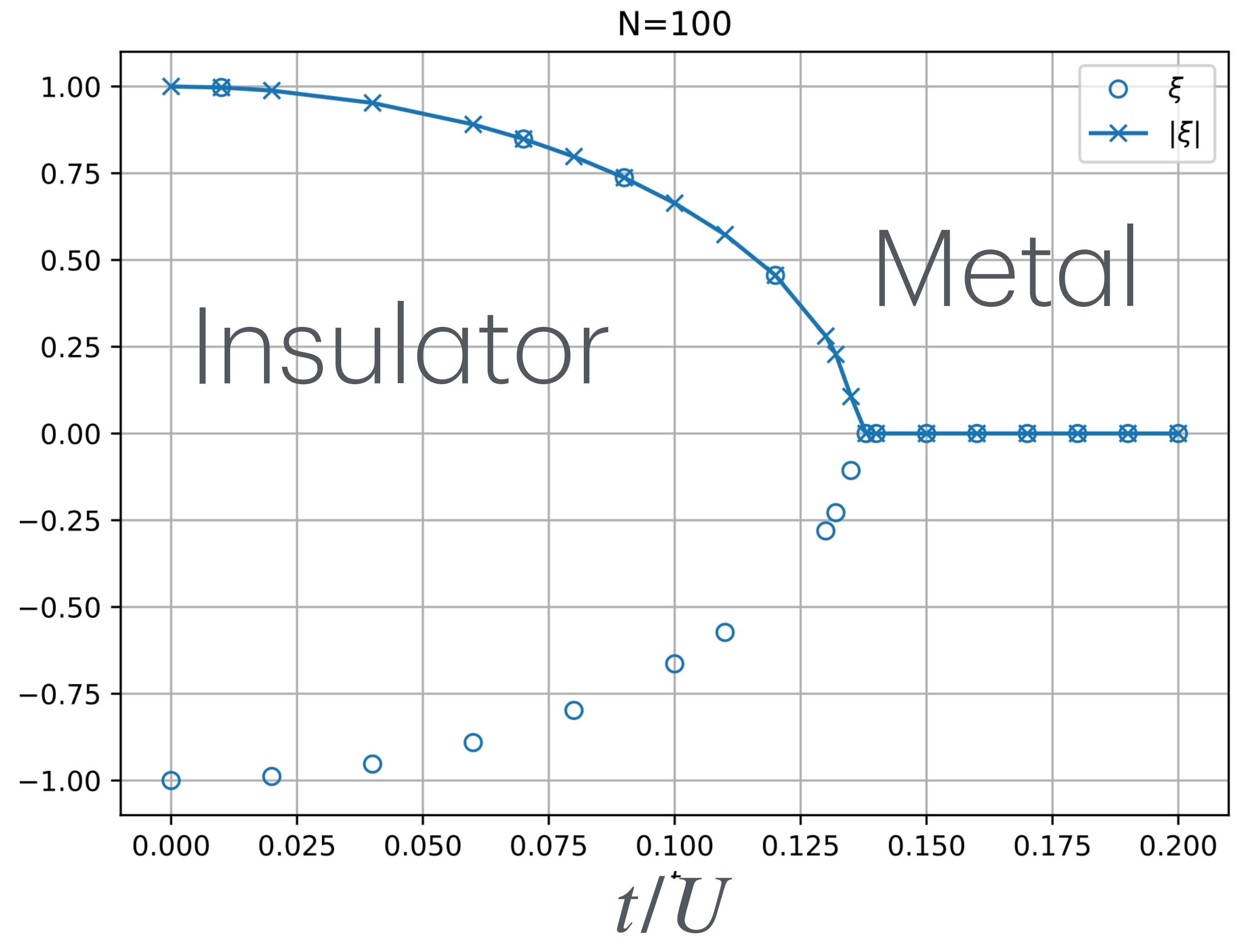
- Influence of stochastic reconfiguration on training;
  Chen & Heyl. *Nat. Physics 20, 1476-1481* (**2024**)
- Role of symmetries;
  Pescia et al. Phys. Rev. B **110**, 035108 (**2024**) | Shuai-Tin Bao et al arXiv:2407.20065 (**2024**)
- Scalability;
  Malyshev et al. arXiv:2408.07625 (**2024**)

Thanks for the attention :)

And to Michael and Mathias for the collaboration!

Extra Slides

Example: Toy model with metal-insulator transition



$$\hat{H} = \frac{t}{U} \sum_{k \in \mathrm{BZ}} \cos(k)\, c_k^\dagger \sigma_z c_k + \sum_{q \in \mathbb{R}} V(q) \rho_q \rho_{-q},$$

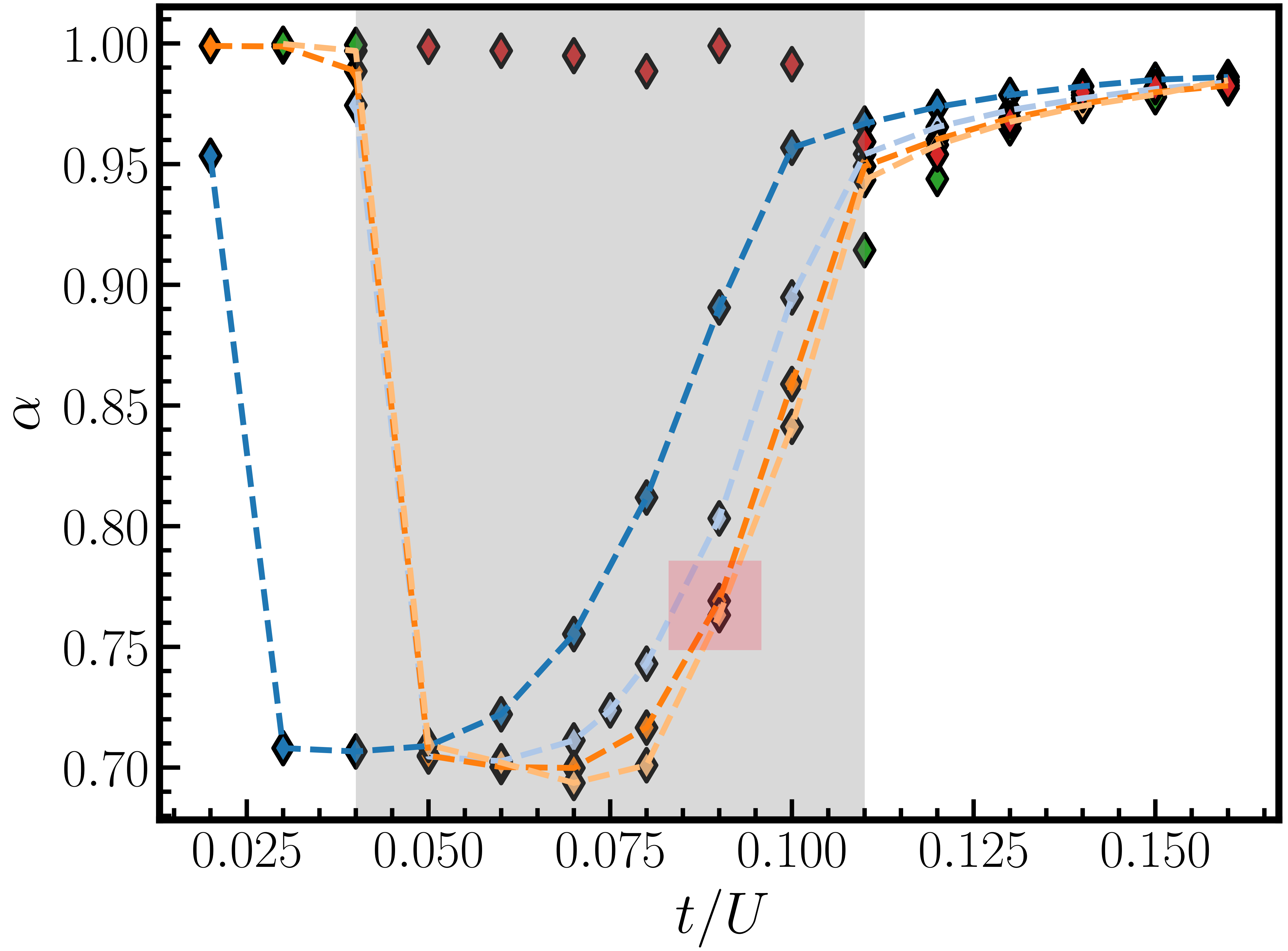
Interacting term

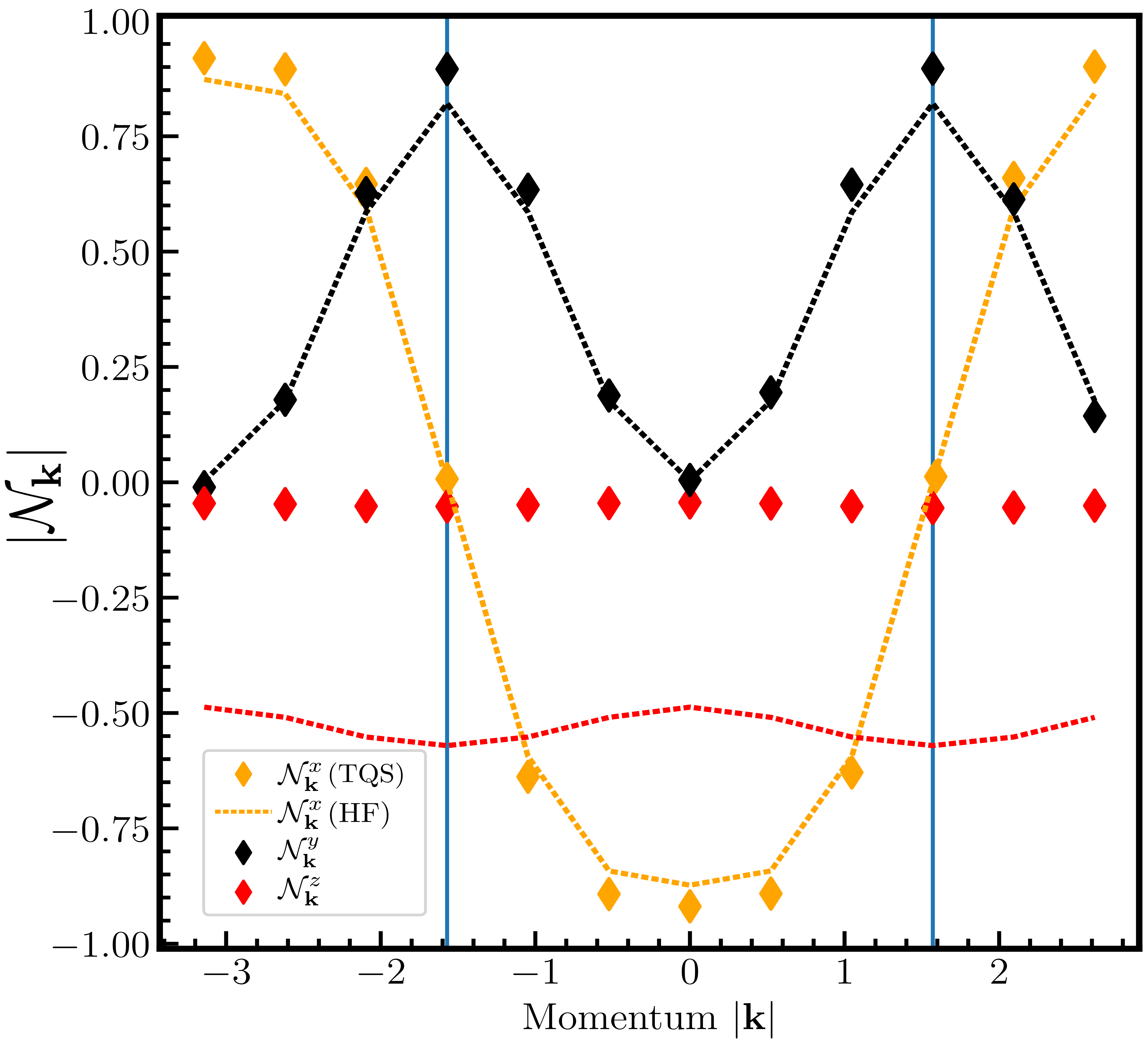
$$\rho_q = \sum_{k \in \mathrm{BZ}} \left(c^\dagger_{\mathrm{BZ}(k+q)} [f_1(k,q) + i\sigma_y f_2(k,q)] c_k - \sum_{G \in \mathrm{RL}} \delta_{q,G} f_1(k,G) \right),$$

"Ideal" basis for

$t/U \ll 1$	$t/U \neq 0$	$t/U \gg 1$
$\mathcal{U}_k = \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix}$	HF-basis?	$\mathcal{U}_k = \mathbb{I}$
Chiral		Band

Other observables

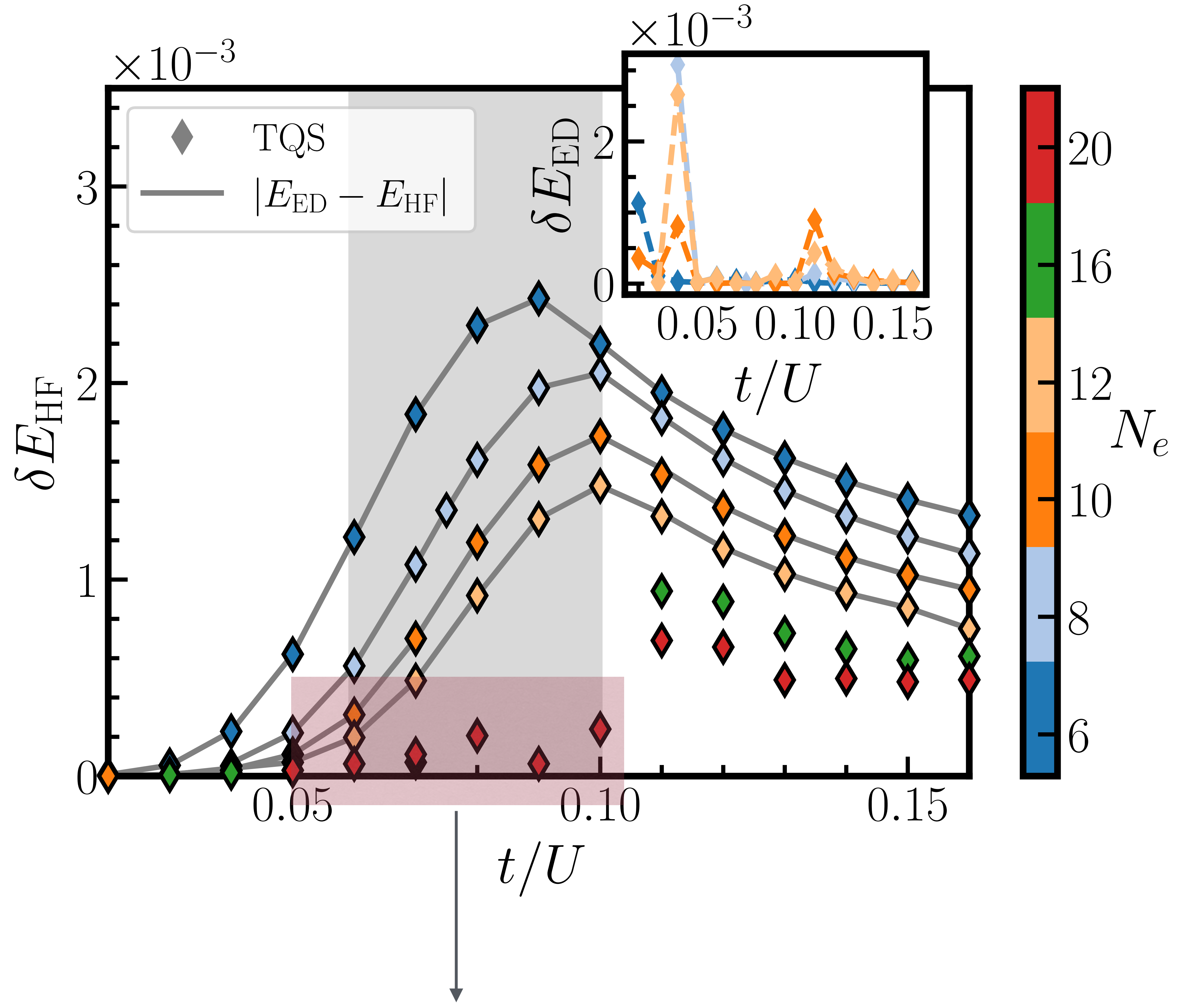




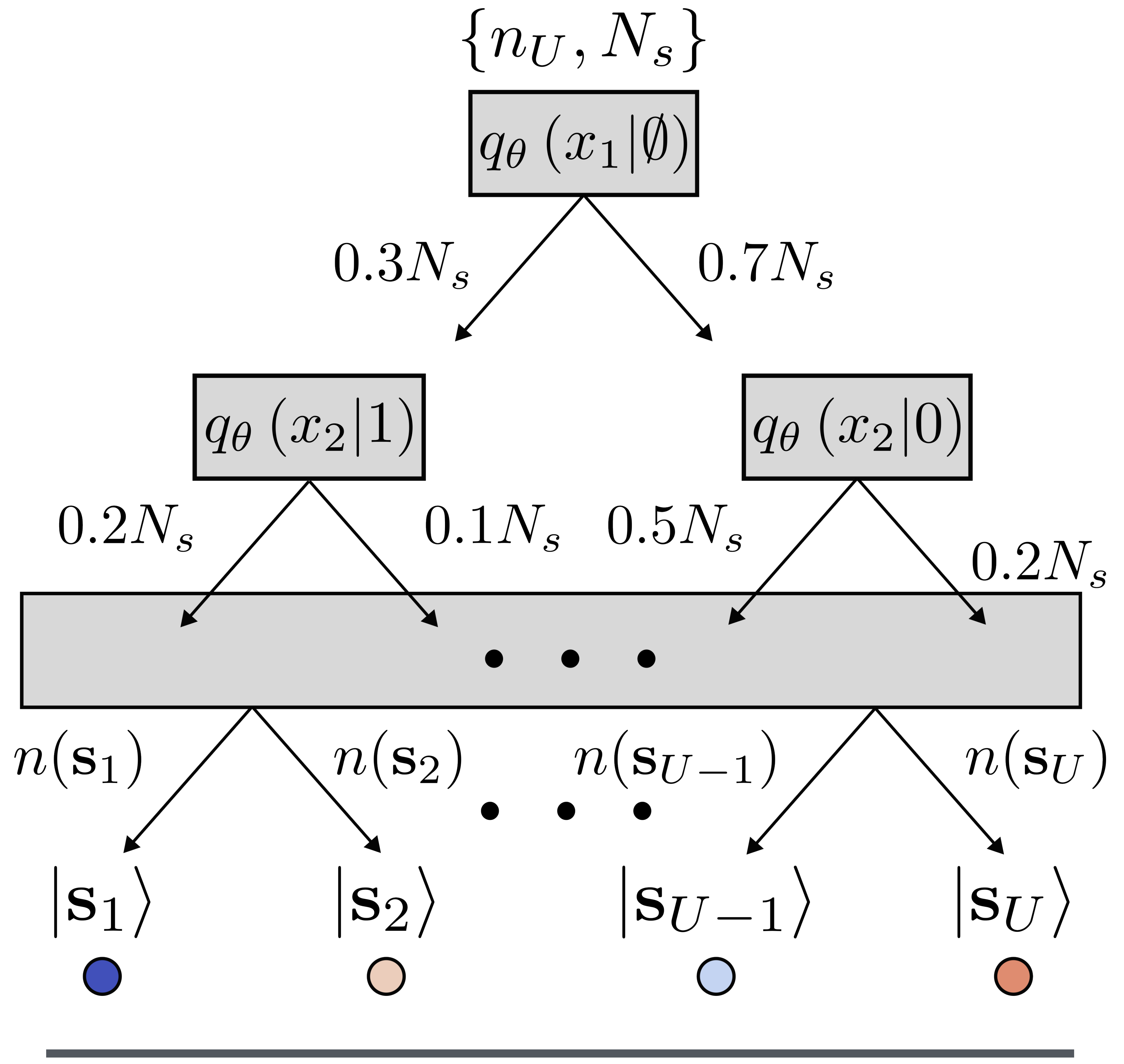
$$\overrightarrow{\mathcal{N}}_k = \mathcal{U}_k^\dagger \vec{\sigma} \mathcal{U}_k$$

$$\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$$

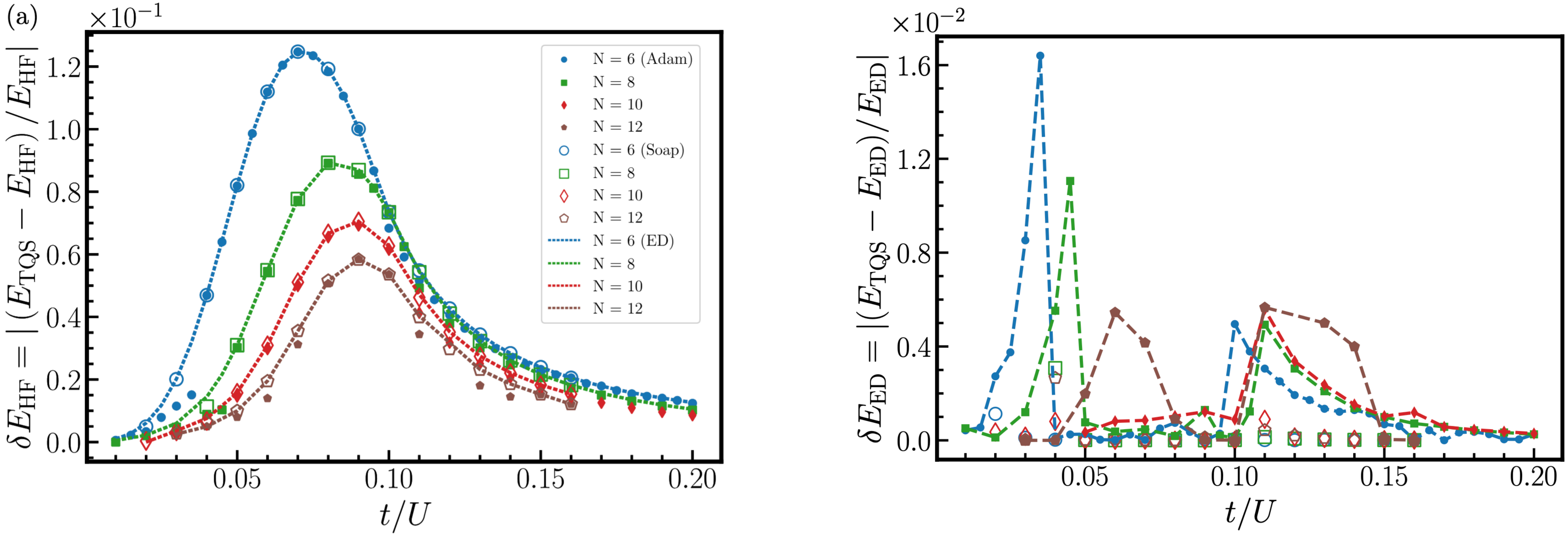
Scalability



$N_U = 4 \times 10^3 \ll 2^{N_e} = 2^{20}$



Soap vs ADAM


(a)
$\times 10^{-1}$
$\delta E_{\rm HF} = |(E_{\rm TQS} - E_{\rm HF})/E_{\rm HF}|$
t/U
N = 6 (Adam)
N = 8
N = 10
N = 12
N = 6 (Soap)
N = 8
N = 10
N = 12
N = 6 (ED)
N = 8
N = 10
N = 12
$\times 10^{-2}$
$\delta E_{\rm ED} = |(E_{\rm TQS} - E_{\rm ED})/E_{\rm ED}|$
t/U

Influence of d_{emb}, N_{enc}, and N_h