FIGHTING CRITICAL SLOWING DOWN WITH NEURAL ENHANCED OUT-OF-EQUILIBRIUM SAMPLING

ELIA CELLINI

18/05/2026
Nordic Lattice Meeting 2026

Based on:

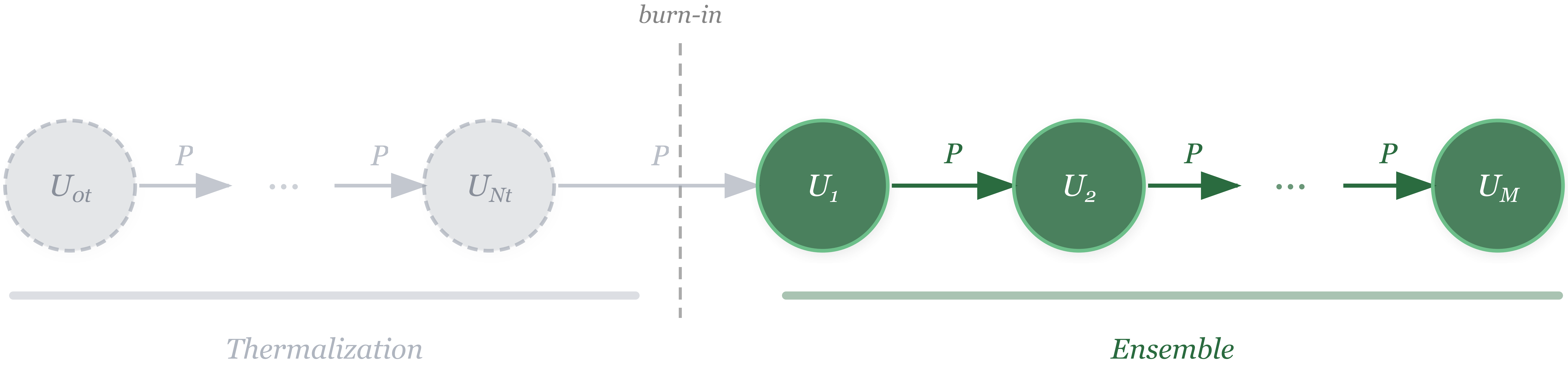
A. Bulgarelli, A. Nada	*Phys. Rev. D* 111 (2025) 7, 7, 2412.00200
+ C. Bonanno, D. Panfalone, D. Vadacchino, L. Verzichelli	JHEP 04 (2026) 051, 2510.25704

THE UNIVERSITY
of EDINBURGH

INTRODUCTION AND MOTIVATIONS

MARKOV CHAIN MONTE CARLO (MCMC)

In MCMC methods, a stochastic **Markov kernel** $P \propto \exp(-S)$ is applied recursively to obtain new configurations

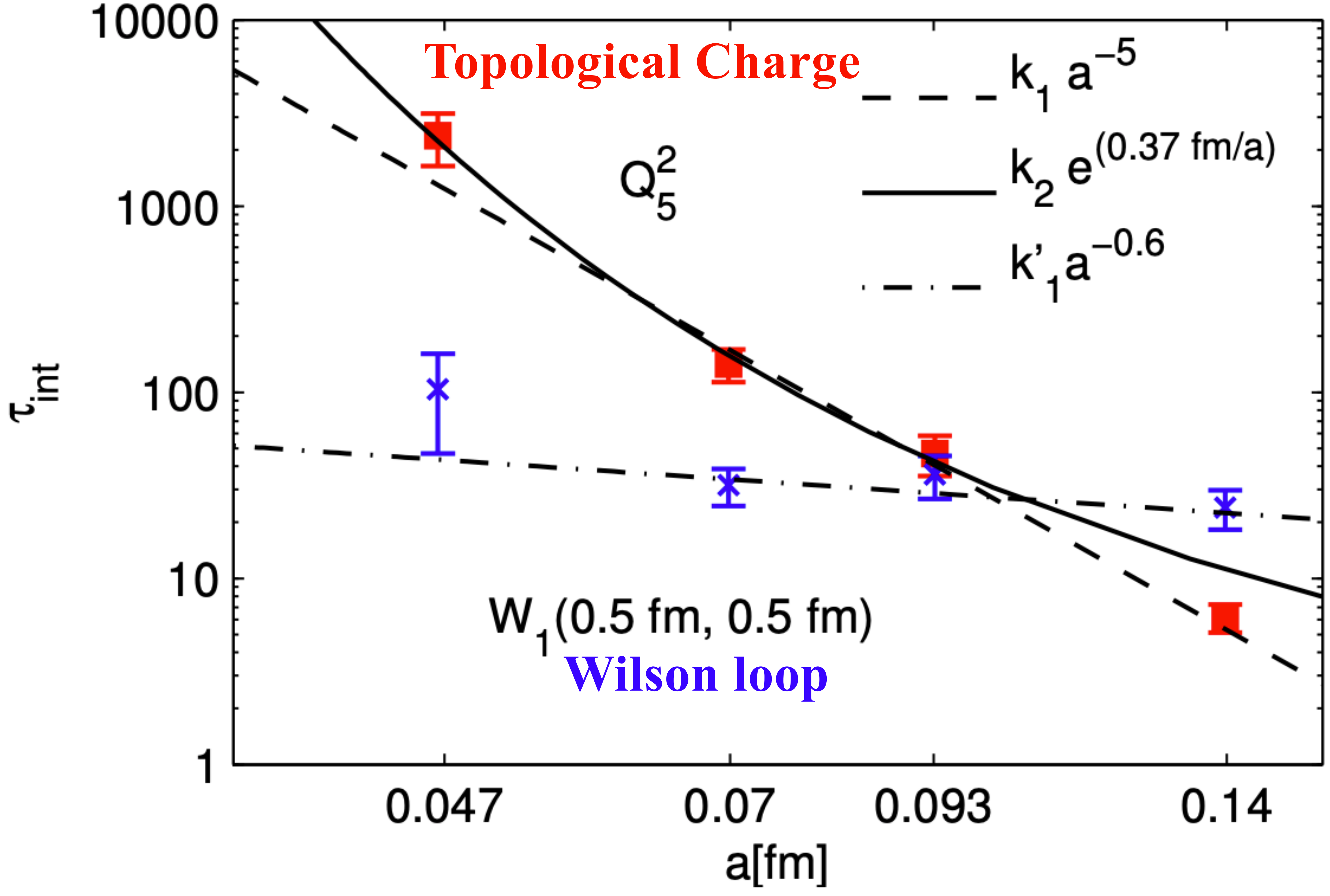


TOPOLOGICAL FREEZING

The configurations generated with MCMC are correlated, and the integrated correlation time τ_{int} is related to the lattice step a of the target theory

$$\tau_{int} \sim a^{-z}$$

[Schaefer et al.; 1009.5228]



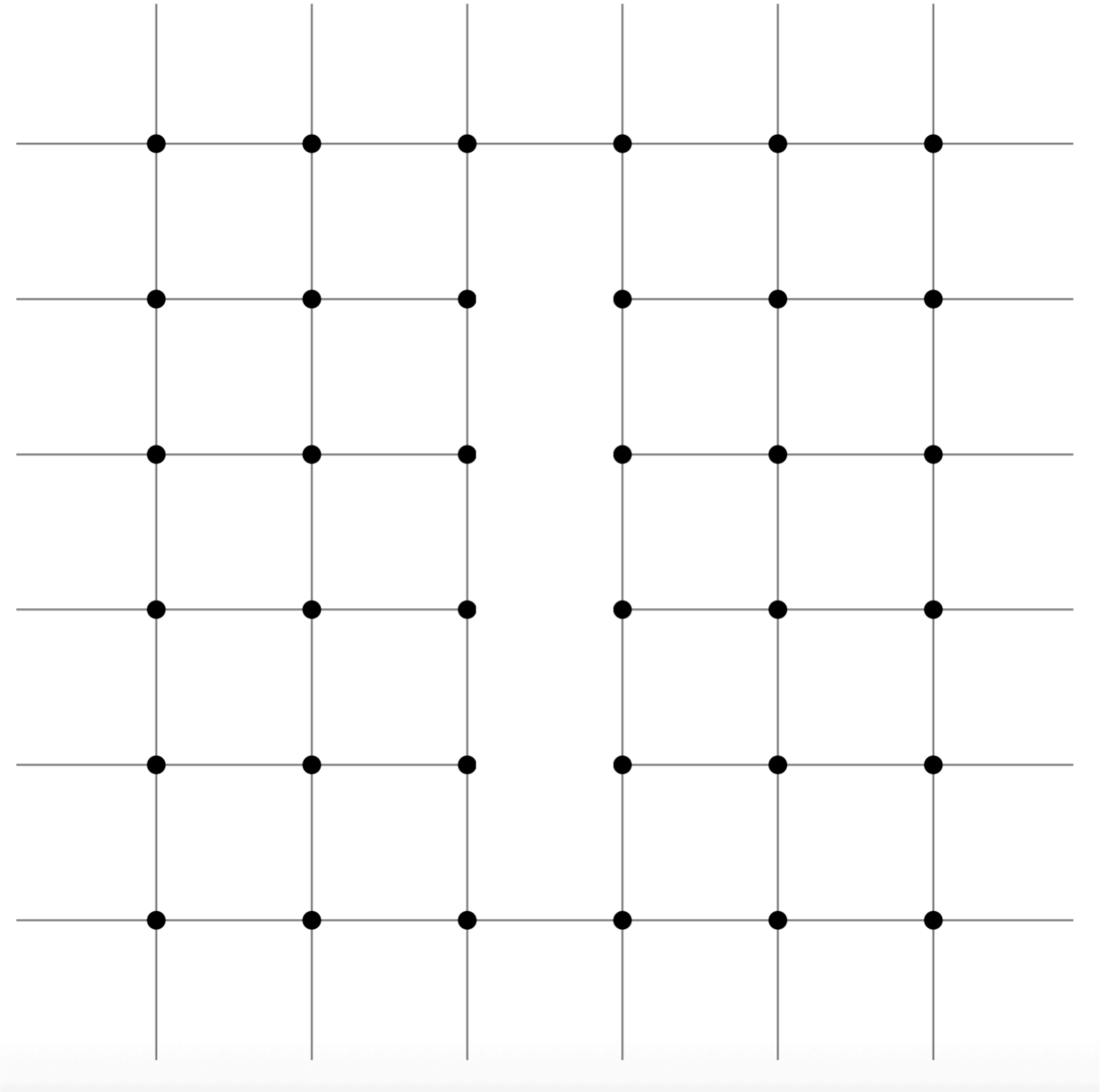
⟵ **continuum limit**

Approaching the continuum, τ_{int} diverge. This problem is known as the **Critical Slowing Down (CSD).** In particular, a severe CSD is the one on the topological charge called **Topological Freezing**

OPEN BOUNDARY CONDITIONS

One way to deal with the topological freezing is by adding a defect on the theory: the **Open Boundary Conditions (OBC)**. OBC are usually mapped to physical, Periodic Boundary Conditions (PBC) theories, using **Parallel Tempering (PT)**.

[Lüscher and Schaefer; 1105.4749],[Hasenbusch; 1706.04443]



Transport problem! OBC → PBC

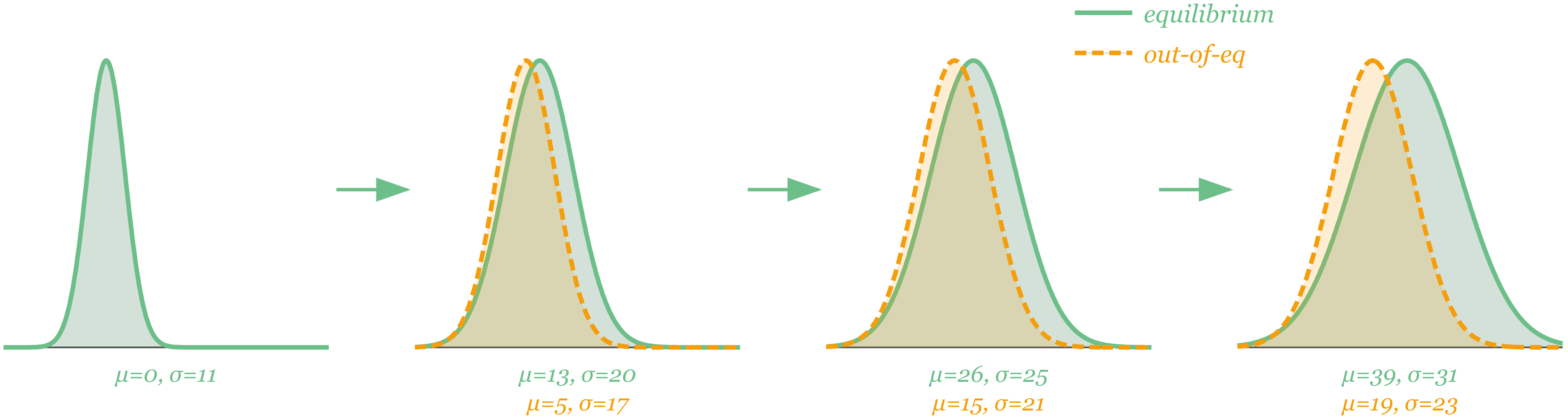
In recent years, two promising new sampling approaches, based on transporting distributions, have been introduced in lattice field theory:

- **Non-Equilibrium MCMC (NE-MCMC)** → Dual to PT, similar performances on OBC!
  [Bonanno et al.; 2402.06561, 2411.00620],[Syed et al., 2408.12057]
- **Normalizing Flows (NFs)**

OUT-OF-EQUILIBRIUM STATISTICAL MECHANICS: CROOKS' THEOREM AND JARZYNSKI'S EQUALITY

OUT-OF-EQUILIBRIUM TRANSFORMATIONS

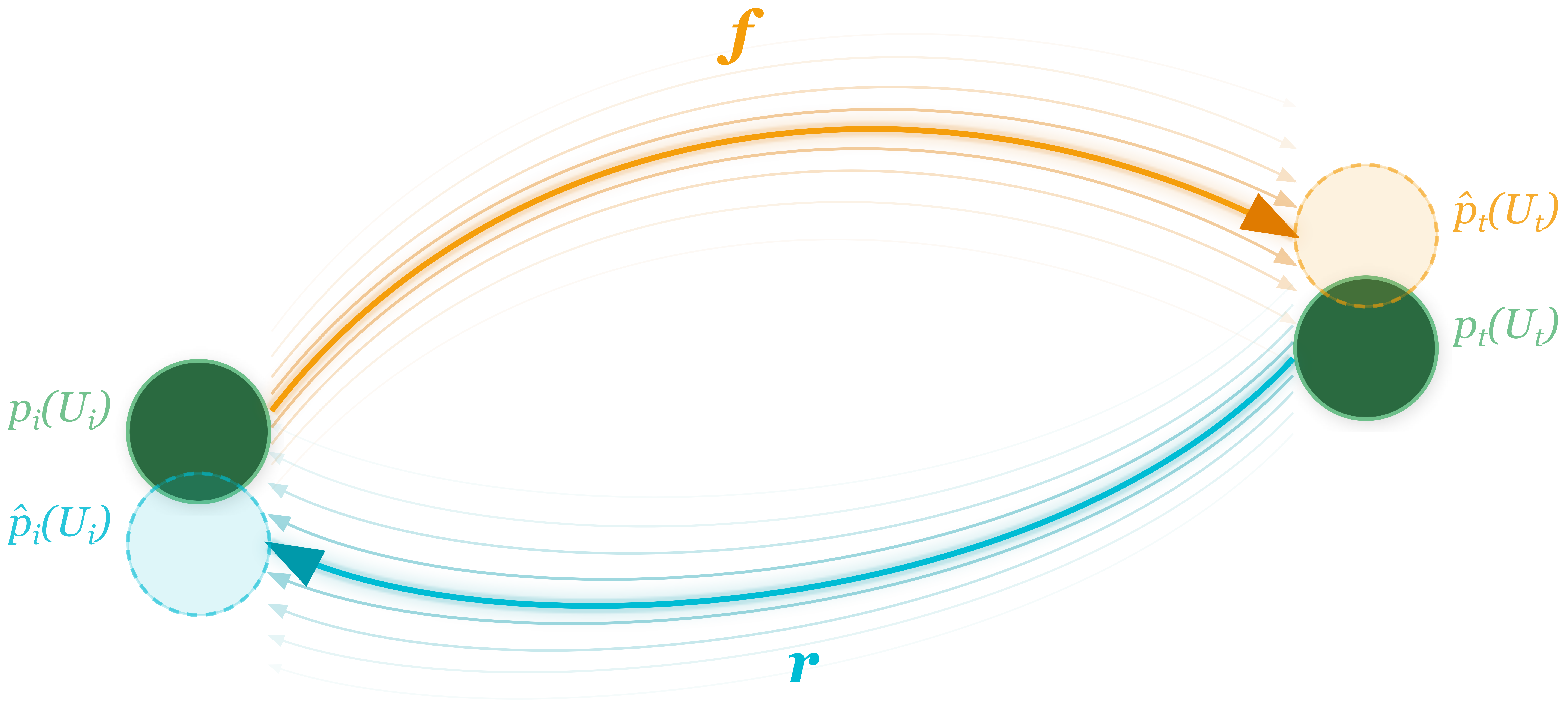
A system is driven out of equilibrium when it is **perturbed faster than its relaxation timescale**, so it never settles into the stationary distribution of the current conditions.



THE UNIVERSITY *of* EDINBURGH

TRANSIENT ENSEMBLE

Given an initial and final distributions p_i and p_t, driving the system out of equilibrium generates a **transient ensemble of stochastic trajectories**, a biased transport between the two distributions.



For each trajectory, we can compute the **thermodynamic work** done on the system: $W(U) = \int_0^T dt \, \frac{\partial S(U(t), \lambda(t))}{\partial \lambda} \dot{\lambda}(t)$

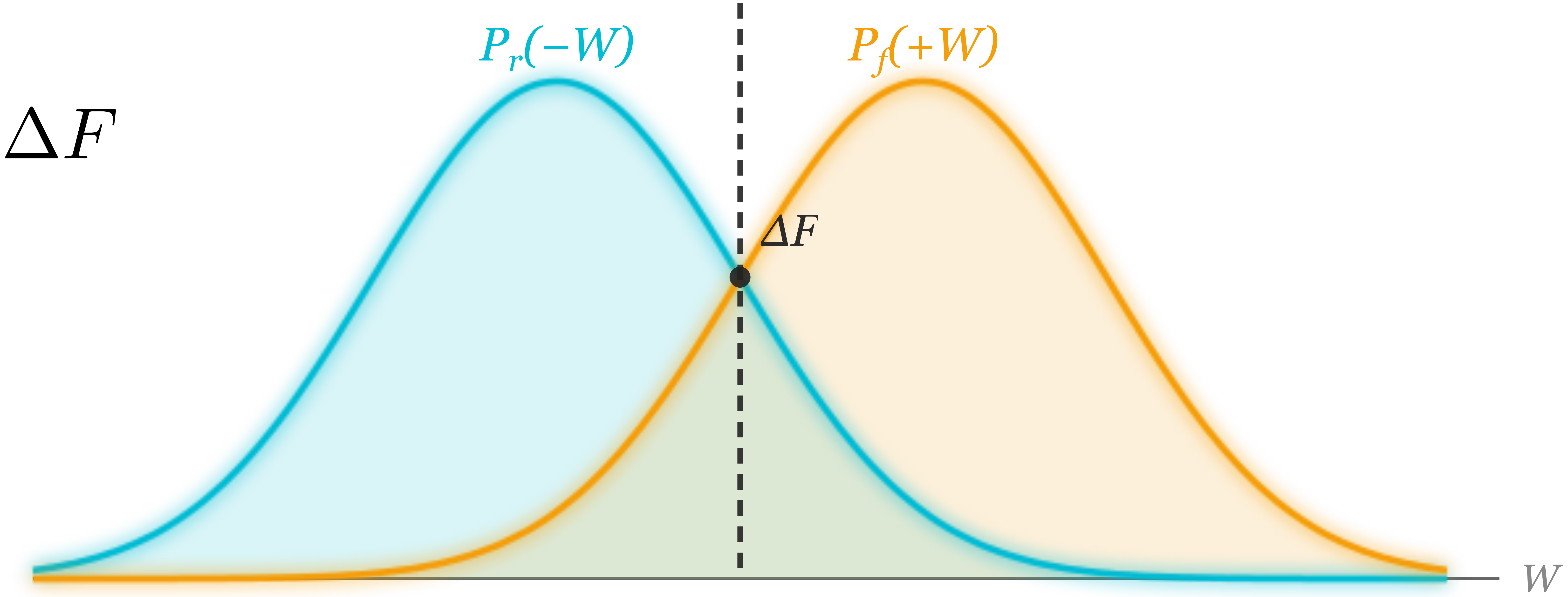
CROOKS' FLUCTUATION THEOREM

The ratio of **forward and reverse work distributions** gives direct access to the **dissipated work** W_d along each trajectory.

[Crooks; cond-mat/9901352]

$$W_d = W - \Delta F$$

$$\frac{P_f(+W)}{P_r(-W)} = e^{W - \Delta F}$$



JARZYNKSI'S EQUALITY

Jarzynski's equality provides a powerful reweighing to **compute the free energy** and **correct the bias** of the non-equilibrium transport

[Jarzynski; cond-mat/9610209]

$$1 = \int dW\, P_f(W) \left(\frac{P_r(-W)}{P_f(W)} \right) = \int dW\, P_f(W)\, e^{-(W-\Delta F)} = e^{\Delta F} \langle e^{-W} \rangle_f$$

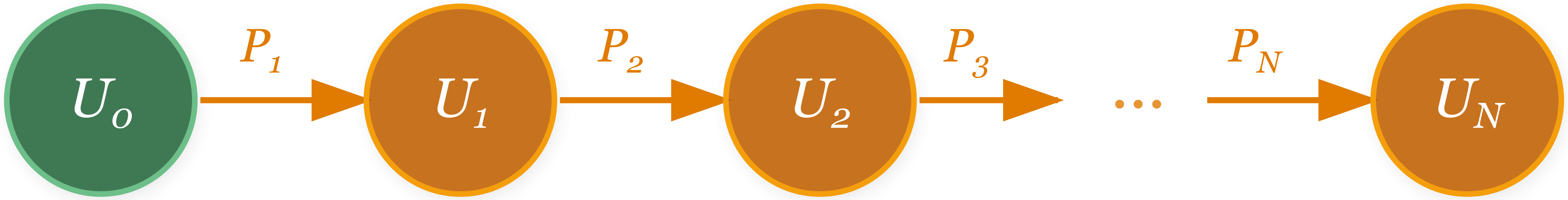
$$\langle e^{-W} \rangle_f = e^{-\Delta F} \qquad \langle \mathcal{O}\, e^{-W_d} \rangle_f = \langle \mathcal{O} \rangle_{p_t}$$

THE UNIVERSITY
of EDINBURGH

NON-EQUILIBRIUM MCMC

NON-EQUILIBRIUM MCMC (NE-MCMC)

$$q_0 \simeq e^{-S_0} \xrightarrow{P_1} e^{-S_1} \xrightarrow{P_2} \dots \xrightarrow{P_N} e^{-S_N} \simeq p$$



e.g. OBC

e.g. PBC

1. Thermalized q_0 "prior"
2. $P_i \propto \exp(-S_i)$ change along the processes following a protocol
3. $p = \exp(-S_N)/Z_N \rightarrow$ "target" distribution

Remark: no thermalization during the processes.

$$W(U_0, \dots, U_N) = \sum_{n=0}^{N-1} \left\{ S_{(n+1)}\left(U_n\right) - S_{(n)}\left(U_n\right) \right\}$$

Equivalent to: **Annealed Importance Sampling**

[Neal; physics/9803008]

FORWARD AND REVERSE DENSITIES

Let us define a sequence of samples as:

$$\mathcal{U} \equiv [U_0, U_1, \cdots, U_N]$$

Forward probability density:

$$q_0(U_0) \prod_{n=0}^{N-1} P[U_i \rightarrow U_{i+1}] = P_f(\mathcal{U})$$

Reverse probability density:

$$p(U_N) \prod_{n=N-1}^{0} P[U_{i+1} \rightarrow U_i] = P_r(\mathcal{U})$$

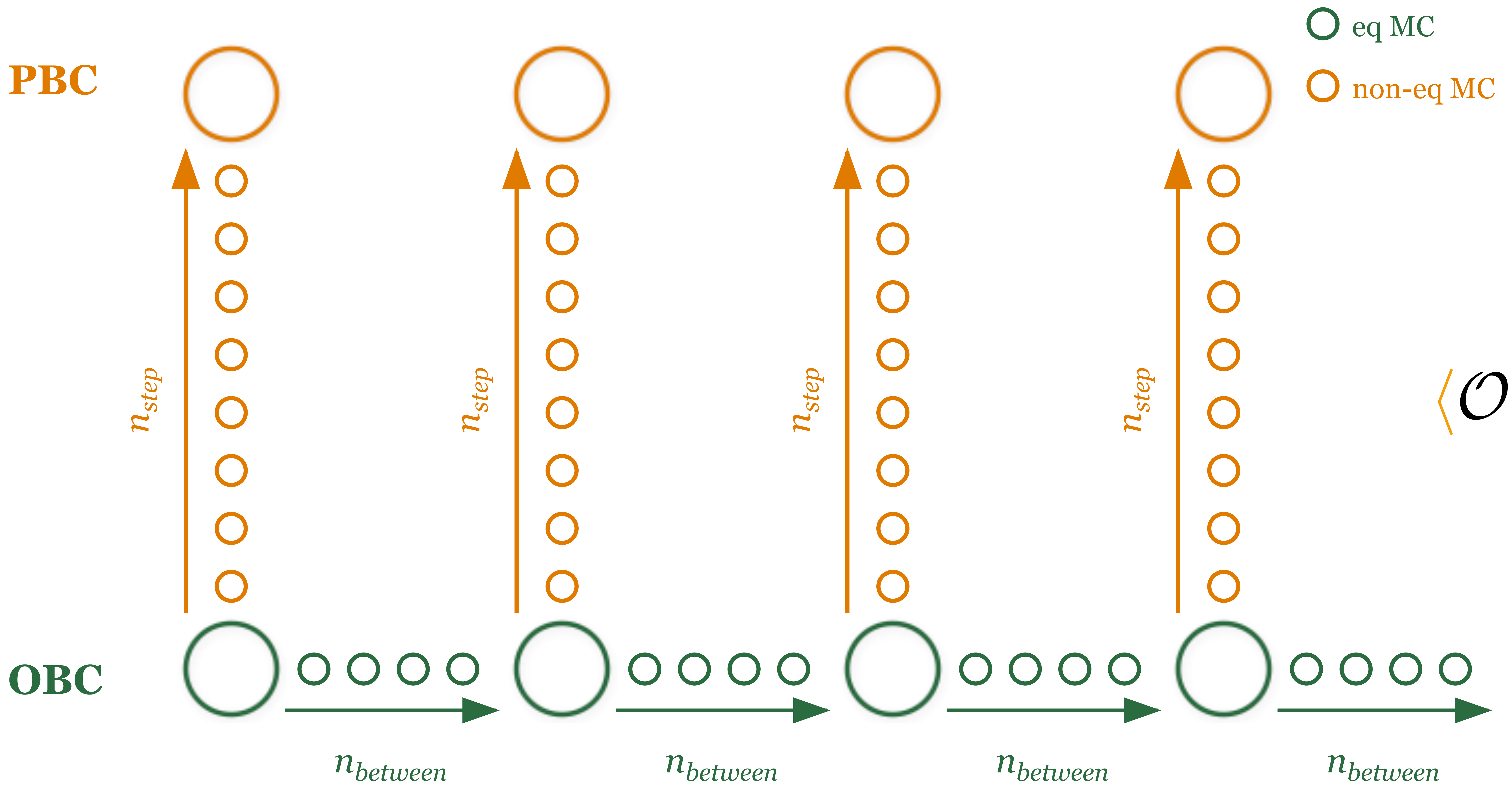
PROOF OF CROOKS' THEOREM

$$\ln \frac{P_f(\mathscr{U})}{P_r(\mathscr{U})} = \underbrace{S_N(U_N) - S_0(U_0) - Q}_{\textbf{Work } W} - \Delta F = W(\mathscr{U}) - \Delta F = W_d(\mathscr{U})$$

Where:

$$\underbrace{Q = -\sum_{n=0}^{N-1} \ln \frac{P[U_n \to U_{n+1}]}{P[U_{n+1} \to U_n]} = -\sum_{n=0}^{N-1} \ln \frac{q_{n+1}(U_{n+1})}{q_{n+1}(U_n)}}_{\textbf{Detailed Balance}} = \sum_{n=0}^{N-1} \left(S_{n+1}(U_{n+1}) - S_{n+1}(U_n) \right)$$

NON-EQUILIBRIUM MCMC (NE-MCMC)



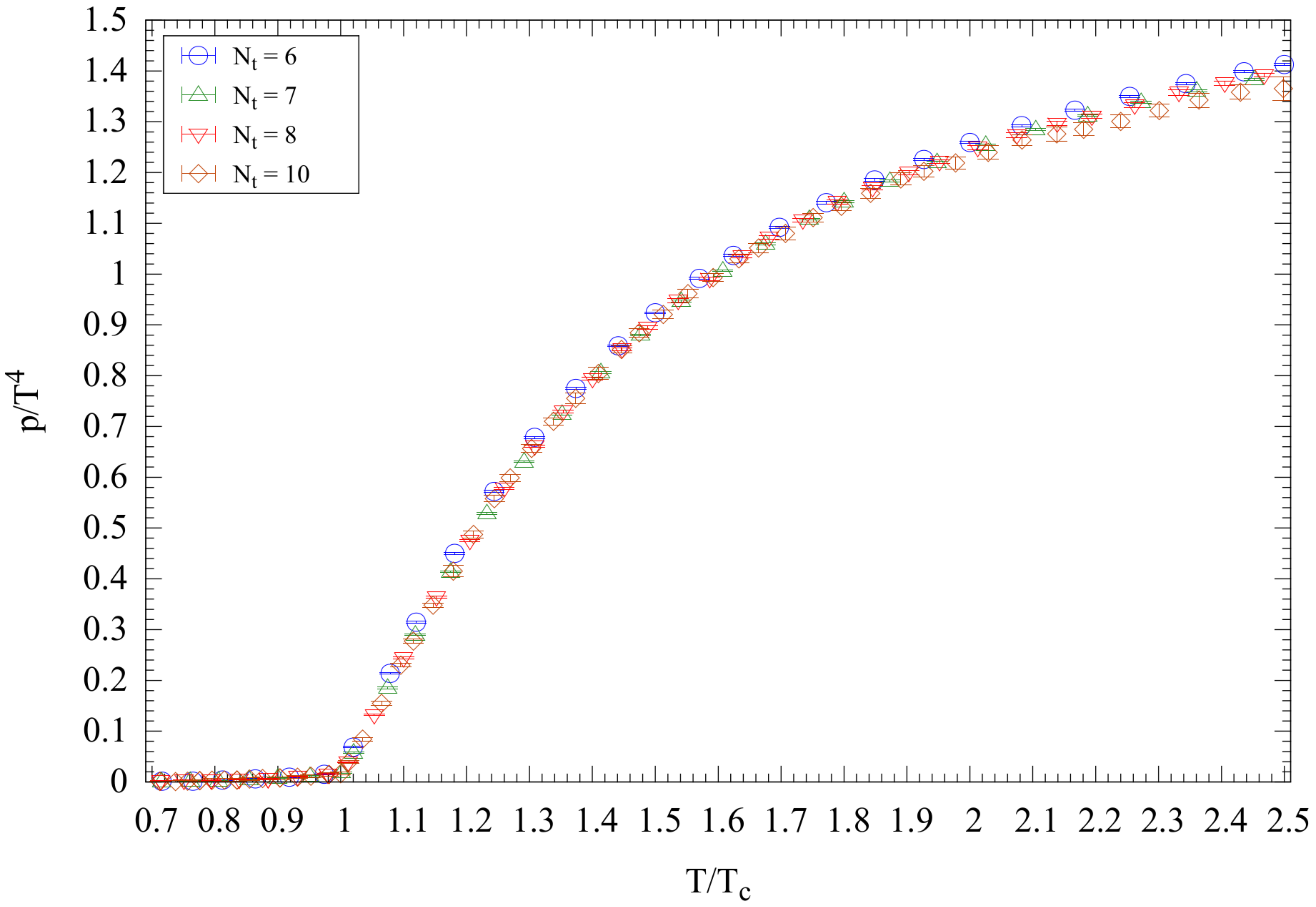
$$\langle \mathcal{O}\, e^{-W_d} \rangle_f = \langle \mathcal{O} \rangle_{p_t}$$

Autocorrelation on the generated samples cannot be larger than the one of the prior configurations!

NE-MCMC FOR LFT

NE-MCMC have been exploited to obtain **state-of-the-arts results** in LFT:

- Interface free energy.
  [Caselle et al.; 1604.05544]
- *SU*(3) e.o.s.
  [Caselle et al.; 1801.03110]
- Running coupling
  [Francesconi et al.; 2003.13734]
- Entanglement entropy
  [Bulgarelli and Panero; 2304.03311, 2404.01987]
- Topological freezing
  [Bonanno et al.; 2402.06561, 2411.00620]
- Casimir energy
  [Bulgarelli et al.; 2505.20403]



[Caselle et al.; 1801.03110]

THE UNIVERSITY *of* EDINBURGH

DRAWBACKS

The identities derived before are exact, however, the **exponential average:**

$$\langle e^{-W_d} \rangle_f$$

can be highly inefficient when W_d **is large** and the **statistic is finite**.

In order to fight this problem, we want W_d to be “small”

Solution 1) Infinite MCMC steps → **quasi-static transformations** → “small” W_d

Solution 2) use Machine Learning to minimize W_d → **Neural-Enhanced Out-of Equilibrium (NEO)** sampling

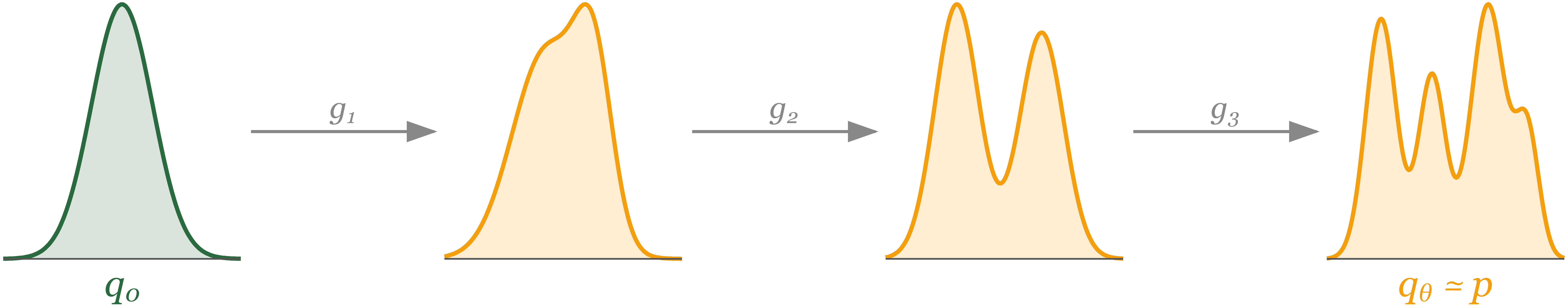
STOCHASTIC NORMALIZING FLOWS

NORMALIZING FLOWS

A **Normalizing Flow (NF)** g_θ is a **parametric**, **invertible** and **differentiable** function:

[Tabak and Vanden-Eijnden; 2010], [Tabak and Turner; 2013],[Rezende et al.; 1505.05770]

$$g_\theta : \; q_0 \to q_\theta \simeq p \qquad U = g_\theta(z) \qquad q_\theta(U) = q_0(g^{-1}(U))\,|\det J_g|^{-1}$$



TRAINING AND SAMPLING

NGSs can be trained to $q_\theta \simeq p(\phi)$ with $p(\phi) = \exp(-S[\phi])/Z$ by minimizing the reverse **Kullback-Leibler divergence**:

[Wu et al.; 1809.10606],[Noé et al.; 1812.01729],[Albergo et al.; 1904.12072],[Nicoli et al.; 1910.13496, 2007.07115]

$$D_{KL}(q_\theta || p) = \int dU q_\theta(U) \log \frac{q_\theta(U)}{p(U)} = \int dU q_\theta(U) \big(S(U) + \log q_\theta(U) \big) - F \geq 0.$$

Partition functions and **observables** can be computed using a re-weighting procedure also called **Importance Sampling**:

[Nicoli et al.; 1910.13496, 2007.07115]

$$\langle \mathcal{O} \rangle_{U \sim p} = \frac{1}{Z} \langle \mathcal{O} \tilde{w} \rangle_{U \sim q_\theta} \qquad Z = \langle \tilde{w} \rangle_{U \sim q_\theta} \qquad \tilde{w} = \frac{e^{-S(U)}}{q_\theta(U)}$$

Autocorrelation on the generated samples cannot be larger than the one of the prior configurations!

NFS FOR LFT AND MCMC

NFs have been widely **investigated in LFT**, however, as general sampler, they **struggle to scale to large volume.** A promising application is their use for **variance reduction in lattice QCD**.

See P. Butti Talk

[Abbot et al.; 2401.10874],[Abbott et al.; 2603.02984]

Combination with MCMC → **Scaling up powerful algorithms**

Standard (Equilibrium) MCMC

↓

- **Deep learning HMC**
  [Foreman et al.; 2105.03418]
- **Adaptive MCMC**
  [Gabrié et al.; 2105.12603]
- **Parallel Tempering**
  [Invernizzi et al.; 2210.14104]
  [Abbott et al.; 2404.11674]
  [Zhang et al.; 2502.10328]

Non-Equilibrium MCMC

↓

- **Stochastic NFs**
  [Wu et al; 2002.06707]
  [Caselle, EC, Nada, Panero; 2201.08862]
- **Continual Repeated Annealed Flow Transport**
  [Matthews et al.; 2201.13117]
- **Non-Equilibrium Transport Sampler**
  [Albergo et al.; 2410.02711] See S. Widyanto Talk

STOCHASTIC NORMALIZING FLOWS

Stochastic Normalizing Flows (SNFs) combine **NE-MCMC updates** and **NF layers**:

$$U_0 \longrightarrow g_\theta^1(U_0) \xrightarrow{P_1} U_1 \longrightarrow g_\theta^2(U_1) \xrightarrow{P_2} \dots \xrightarrow{P_N} U_N \equiv U$$

Where g_θ^i are NF layers and P_i are MCMC update

[Wu et al; 2002.06707],[Caselle, EC, Nada, Panero; 2201.08862],[Bulgarelli, EC, Nada.; 2412.00200]

VARIATIONAL DISSIPATED WORK

We have now:

$$W_d^\theta = W_\theta(U_0, \ldots, U_N) - \Delta F = S_N(U_N) - S_0(U_0) - Q_\theta - \Delta F$$

Where:

$$Q_\theta = \sum_{n=0}^{N-1} \left(S_{n+1}(U_{n+1}) - S_{n+1}(U_n) + \ln|\det J_{g_\theta^n}| \right)$$

[Vaikuntanathan and Jazynski; 1101.2612], [Wu et al; 2002.06707],[Caselle, EC, Nada, Panero; 2201.08862],[Bulgarelli, EC, Nada.; 2412.00200]

THE UNIVERSITY
of EDINBURGH

TRAINING OF SNF

We can now train a SNF by minimizing:

$$\mathscr{L}(\theta) = \langle W_d^\theta \rangle_f = D_{KL}(P_f || P_r) \geq 0$$

$$\langle W \rangle_f \geq \Delta F$$

Second Law!

More reversible trajectories → smaller work → better error!

Successful application to Effective String Theory (scalar theory)

[Caselle, EC, Nada; 2307.01107, 2409.15937]

$d = 3 + 1$ SU(3): OPEN BOUNDARY CONDITIONS

Bonanno, Bulgarelli, EC, Nada, Panfalone, Vadacchino, Verzichelli;
2510.25704
JHEP 04 (2026) 051

GAUGE EQUIVARIANT LAYER

[Morningstar and Peardon; hep-lat/0311018],[Nagai and Tomya.;2103.11965],[Abbott et al.;2305.02402]

(Masked) Stout smearing:

$$U'_\mu(x) = g_n(U_\mu(x)) = \exp\left(iQ^{(n)}_\mu(x)\right)\, U_\mu(x),$$

Traceless Hermitian

$$Q^{(n)}_\mu(x) = \frac{i}{2}\left((\Omega^{(n)}_\mu(x))^\dagger - \Omega^{(n)}_\mu(x)\right) - \frac{i}{2N}\mathrm{Tr}\left((\Omega^{(n)}_\mu(x))^\dagger - \Omega^{(n)}_\mu(x)\right)$$

Active link transformed
with frozen staples

$$\Omega^{(n)}_\mu(x) = C^{(n)}_\mu(x)U^\dagger_\mu(x)$$

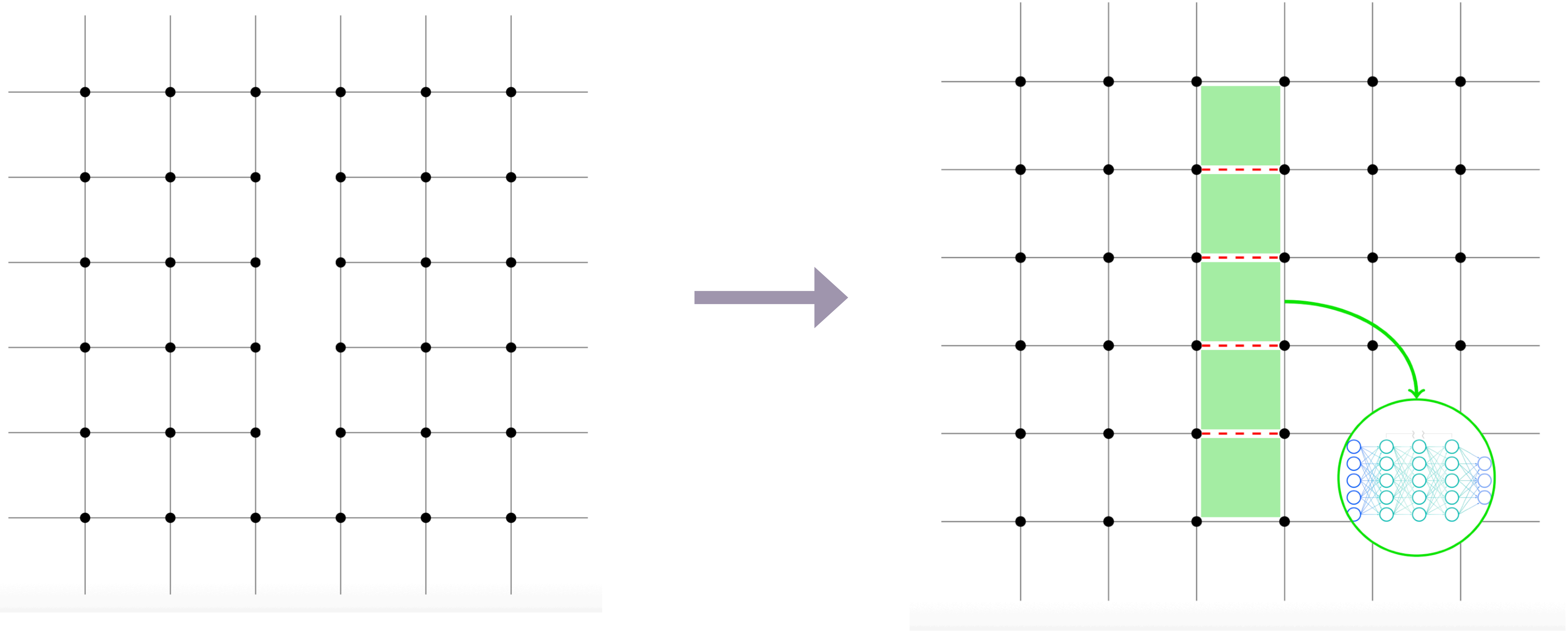
$$C^{(n)}_\mu(x) = \sum_{\nu\neq\mu} \rho^{(n)}_{\mu\nu}(x)\left(U_\nu(x)U_\mu(x+\hat{\nu})U^\dagger_\nu(x+\hat{\mu}) + U^\dagger_\nu(x-\hat{\nu})U_\mu(x-\hat{\nu})U_\nu(x-\hat{\nu}+\hat{\mu})\right)$$

Few parameters!

MCMC: Heatbath + Over relaxation

THE UNIVERSITY
of EDINBURGH

DEFECT COUPLING LAYER

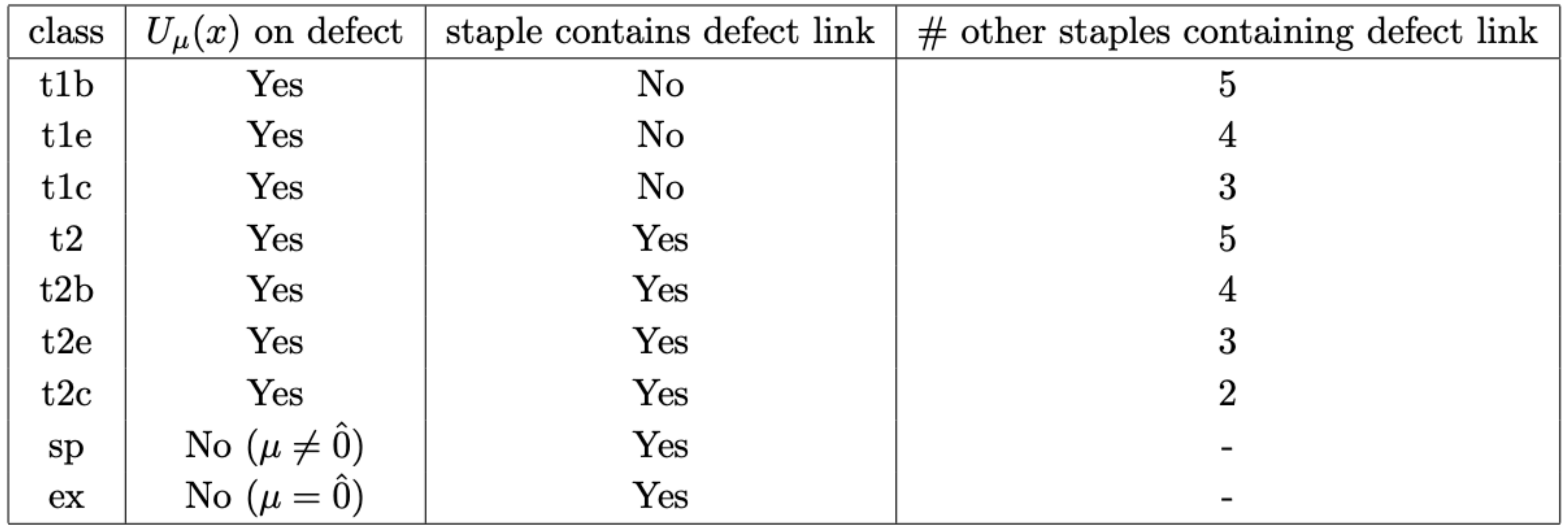


Protocol: $\beta_{defect} : 0 \to \beta_{bulk}$

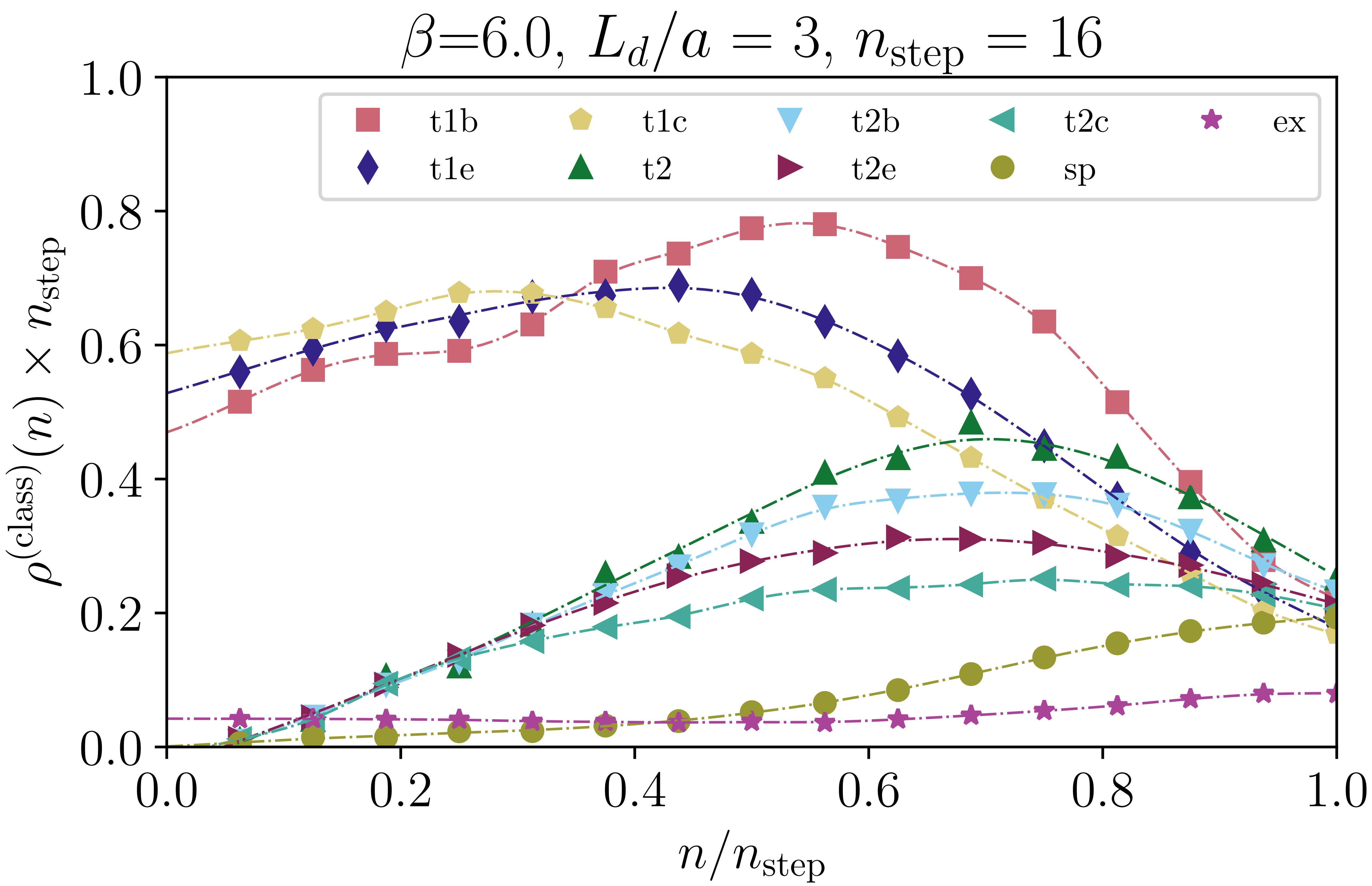
[Bulgarelli, EC et al.; 2410.14466]

EXTRAPOLATION OF THE PARAMETERS

We grouped the parameters based on their positions within the defect.



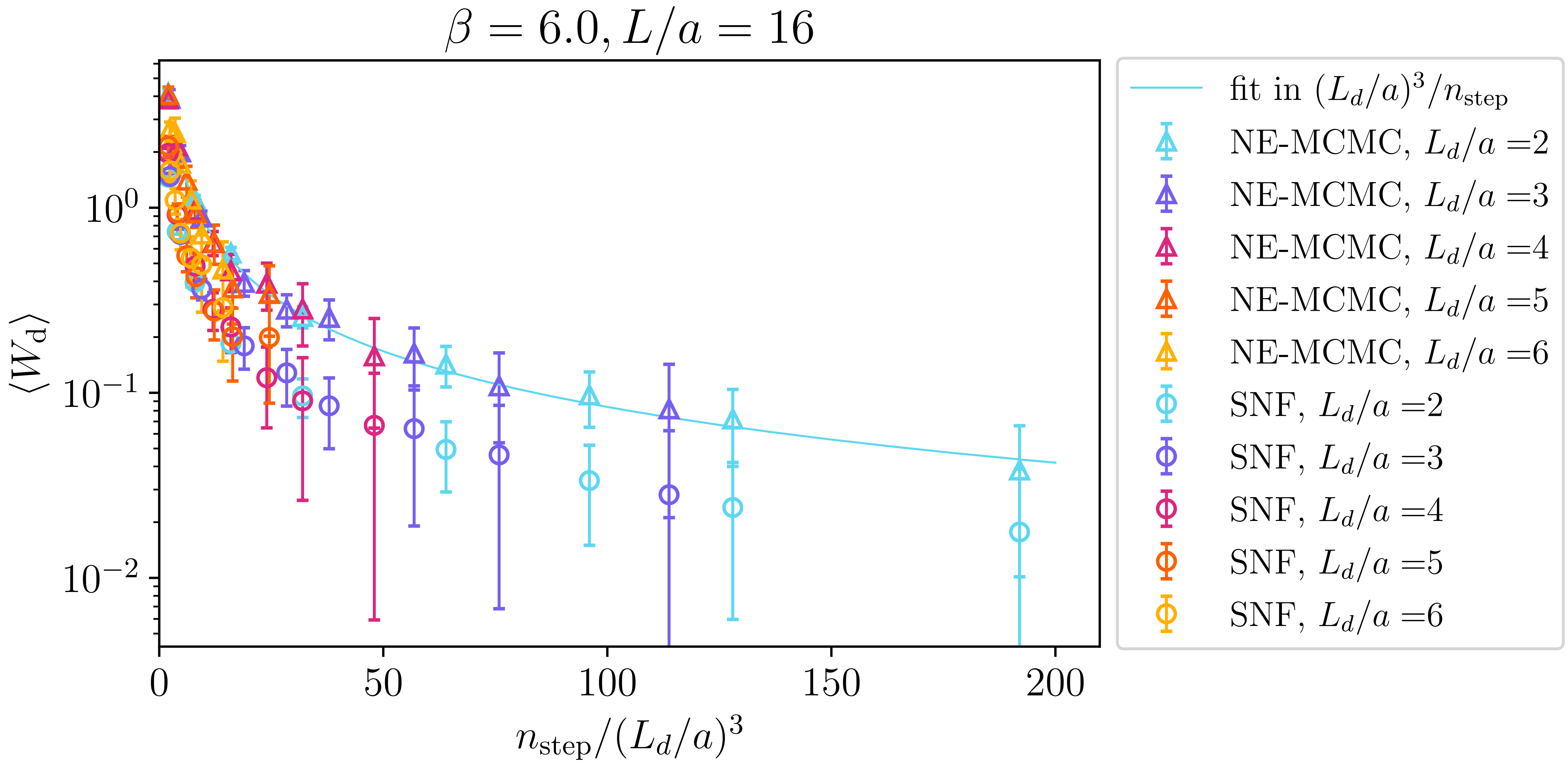
class	$U_\mu(x)$ on defect	staple contains defect link	# other staples containing defect link
t1b	Yes	No	5
t1e	Yes	No	4
t1c	Yes	No	3
t2	Yes	Yes	5
t2b	Yes	Yes	4
t2e	Yes	Yes	3
t2c	Yes	Yes	2
sp	No ($\mu \neq \hat{0}$)	Yes	-
ex	No ($\mu = \hat{0}$)	Yes	-



Same procedure used in:

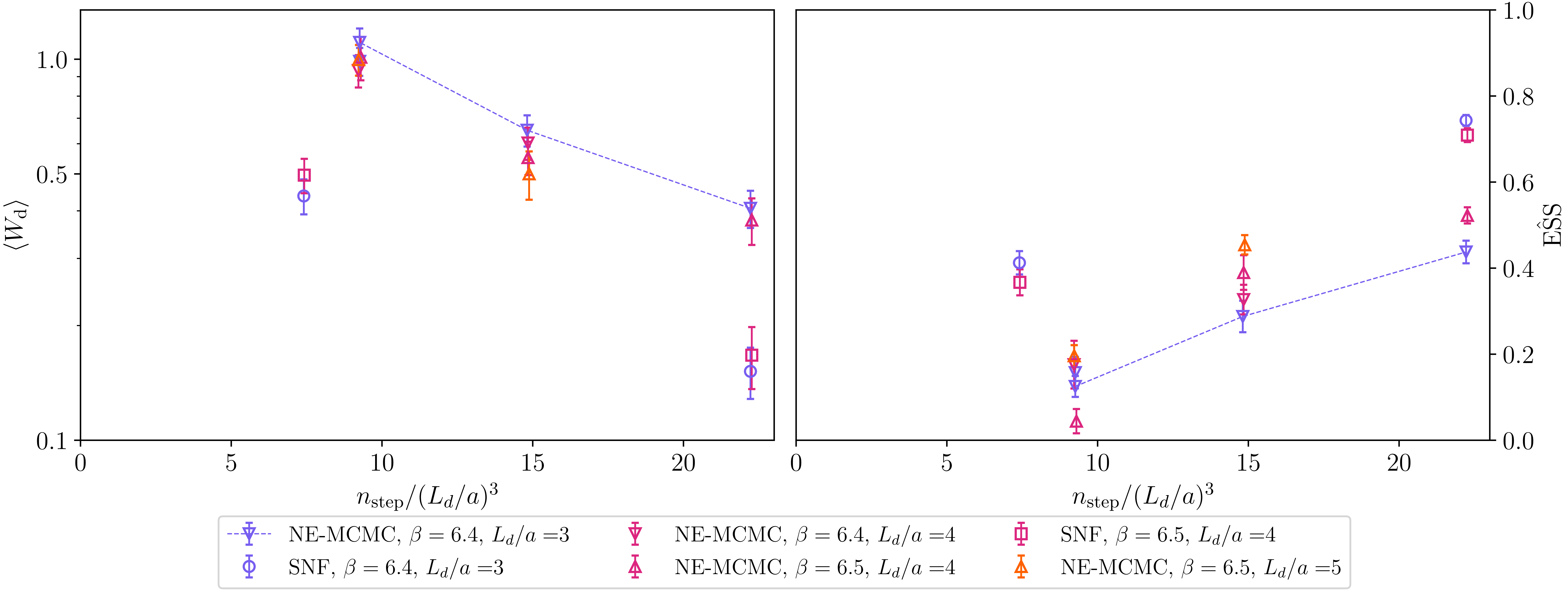
[Bulgarelli, EC, Nada.; 2412.00200]

DEFECT VOLUME SCALING DKL



$$\langle W_d \rangle_f \sim 3\beta k(\beta) \frac{n_{dof}}{n_{steps}}$$

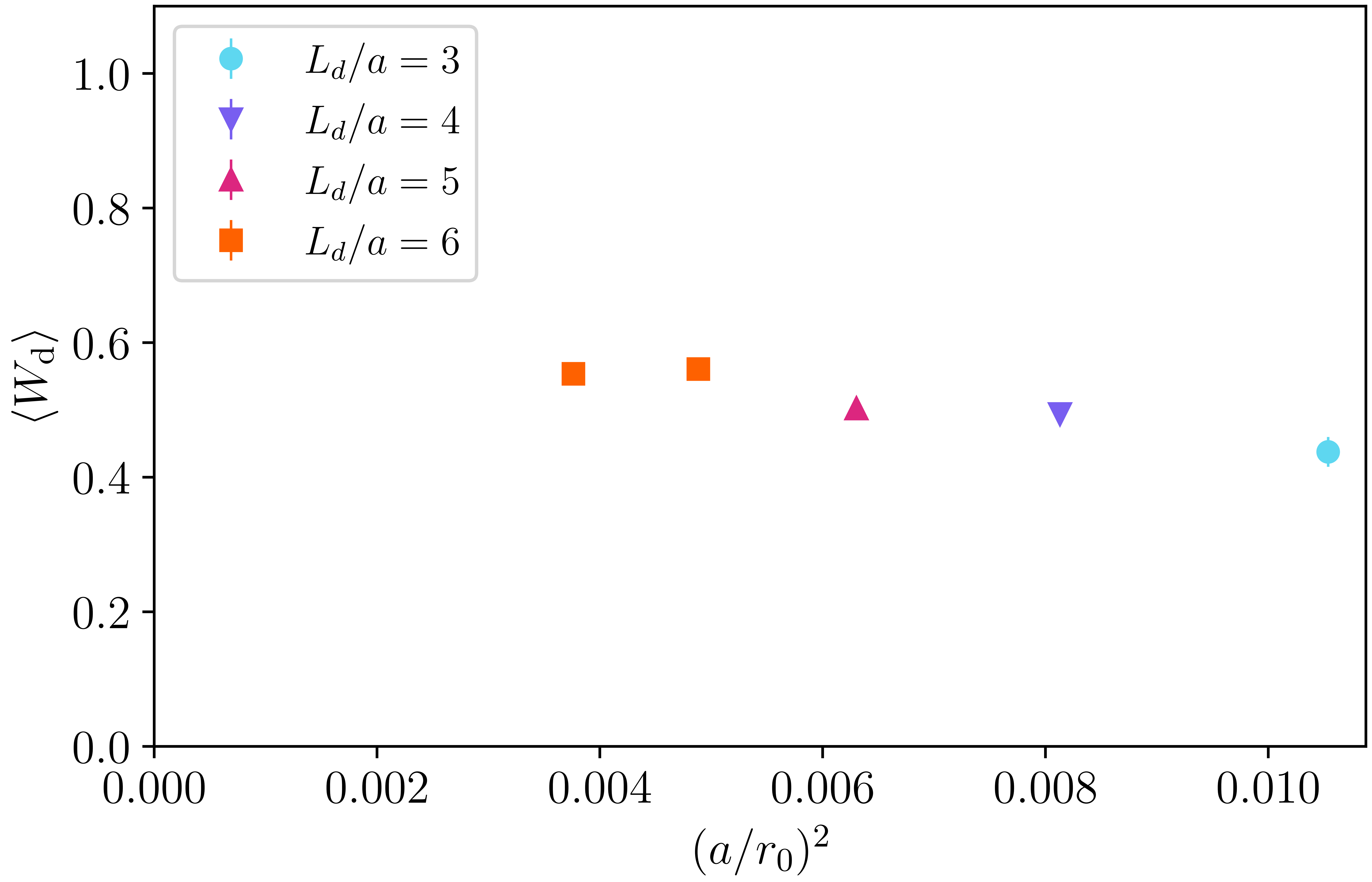
PERFORMANCES ON FINE LATTICE SPACINGS



$\beta = 6.4$, $(L/a)^4 = 30$, $a[\mathrm{fm}] = 0.051$

$\beta = 6.5$, $(L/a)^4 = 34$, $a[\mathrm{fm}] = 0.045$

TOWARD CONTINUUM EXTRAPOLATION OF χ



Preliminary results

PyTorch Notebook

arXiv:2510.25704

OUTLOOK

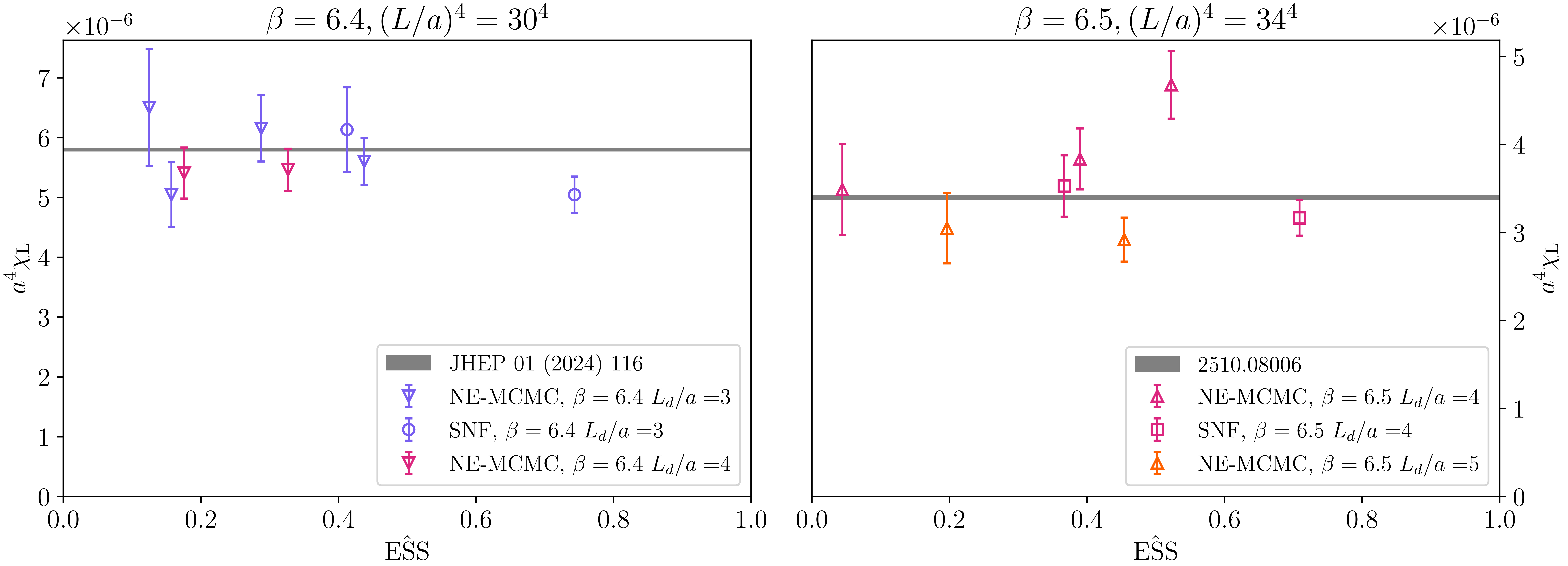
- Great **Scaling** of NE-MCMC and SNFs with d.o.f and n_{steps}!
- Defect SNFs allow for the **reduction the cost** of NE-MCMC by a **factor** $\simeq 3$
- **eXplainable AI** (XAI) →
  1. **Training costs of SNFs are negligible!**
  2. **NF set of layers are the same,** continuous in MC time, **vector field**

- Toward the **continuum extrapolation** of the topological susceptibility!
- New **NEO transports** (e.g. Stochastic Differential Equations) → See S. Widyanto Talk

THANK YOU FOR YOUR ATTENTION

BACKUP

TOPOLOGICAL SUSCEPTIBILITY



THE UNIVERSITY
of EDINBURGH