Ensemble Quasi-Newton HMC

Xiao-Yong Jin and James Osborn
Argonne National Laboratory

July 23, 2018
The 36th International Symposium on Lattice Field Theory
East Lansing, MI

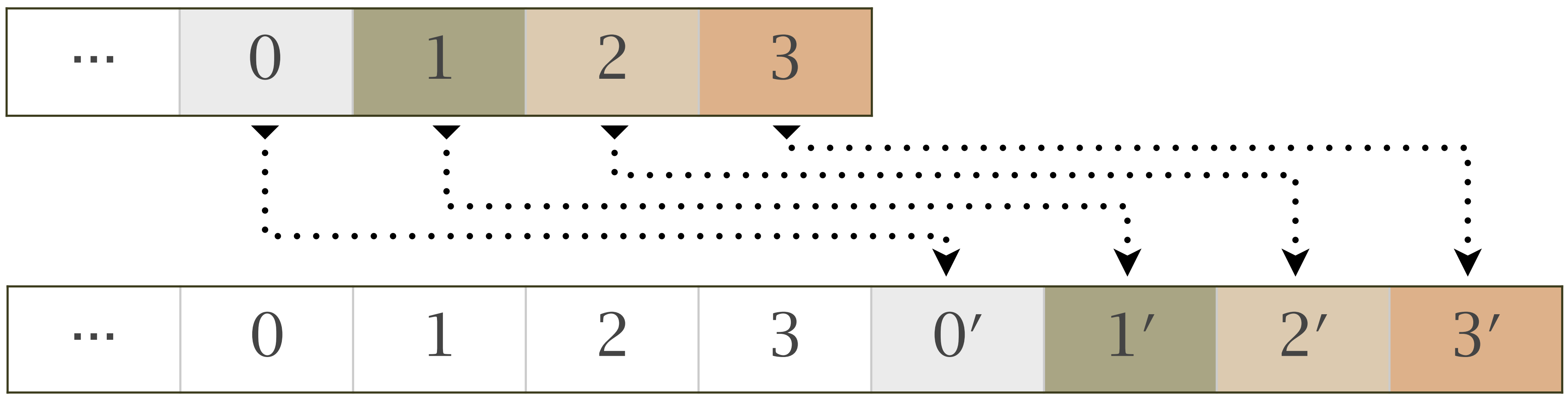
Reduce critical slowing down

- Part of US DOE-funded
  Exascale Computing Project (ECP)
- Support research in lattice QCD
  to prepare for exascale
- Reducing critical slowing down,
  lead by Norman Christ,
  is part of the USQCD's effort in ECP
- See Norman's slides
  for a list of people actively involved

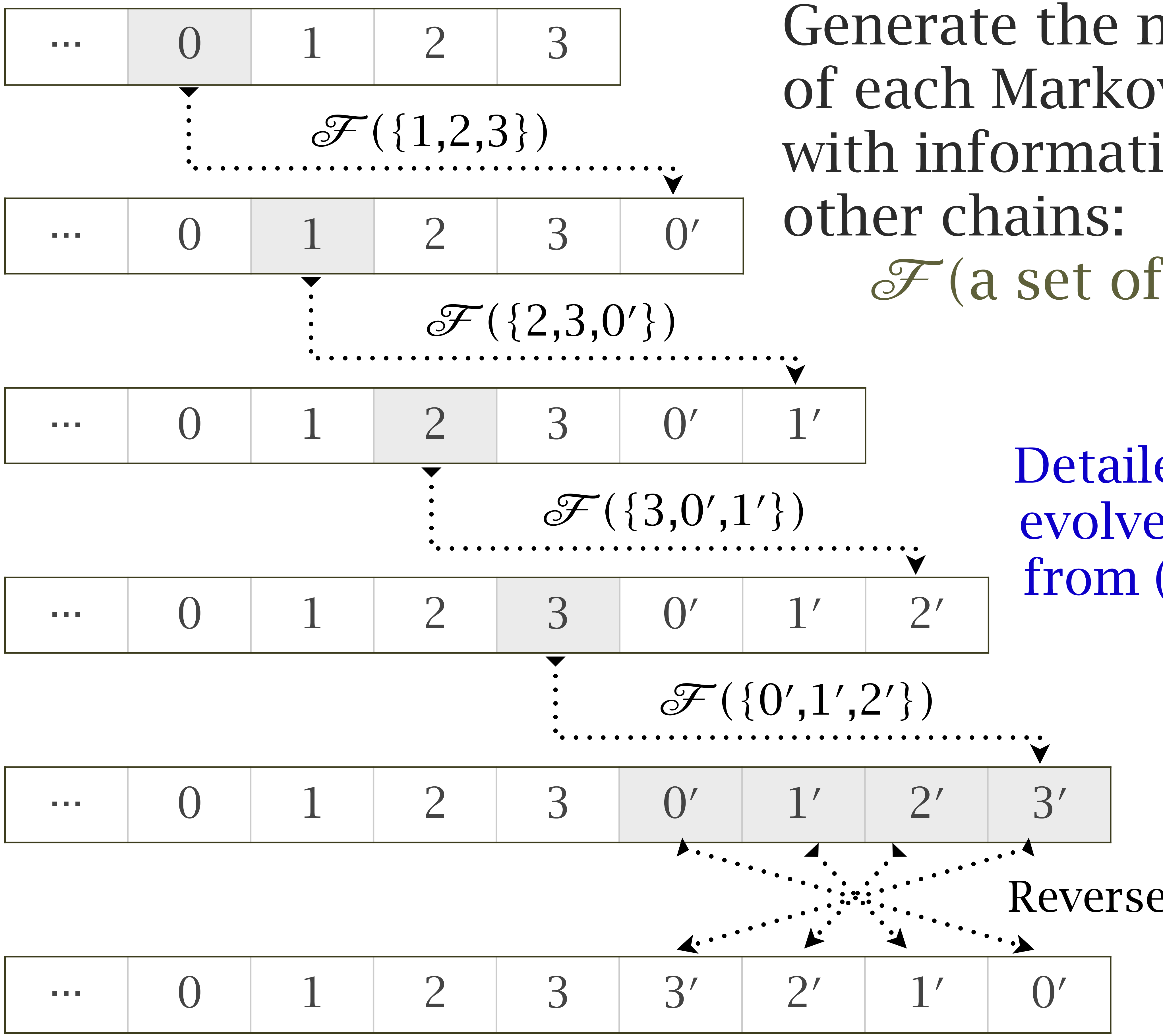
Outline

- Generate ensemble assisted Markov chains
- Apply Quasi-Newton HMC
- Test on 2D U(1) pure gauge theory
  (work in progress)

Generate multiple Markov chains



- Can we exchange information between chains?
- Use info from other chains
- Extra info from itself (not explored in this talk)
- Any advantage?

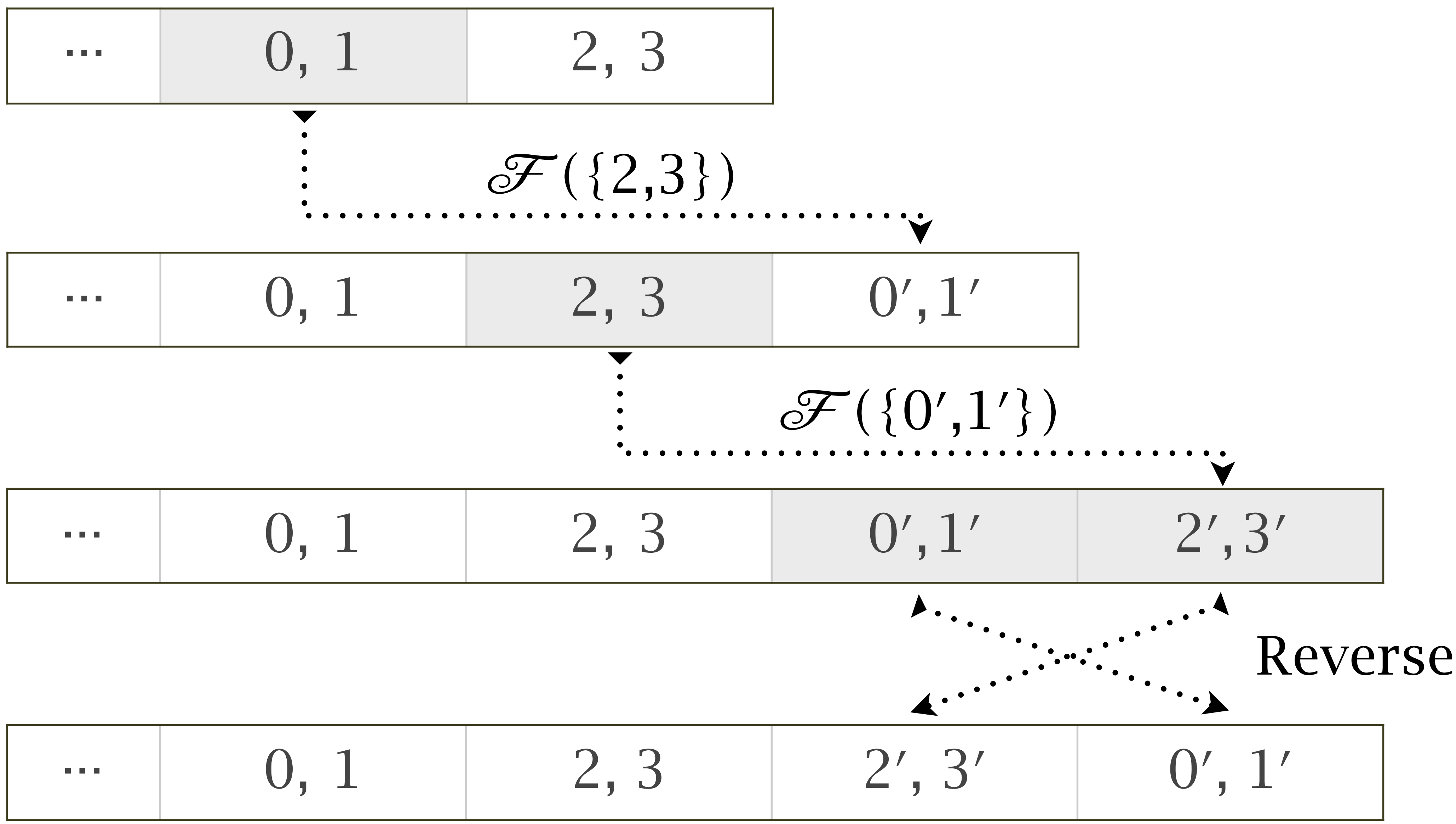


Generate the next state of each Markov chain with information from other chains:

$\mathscr{F}$(a set of configs)

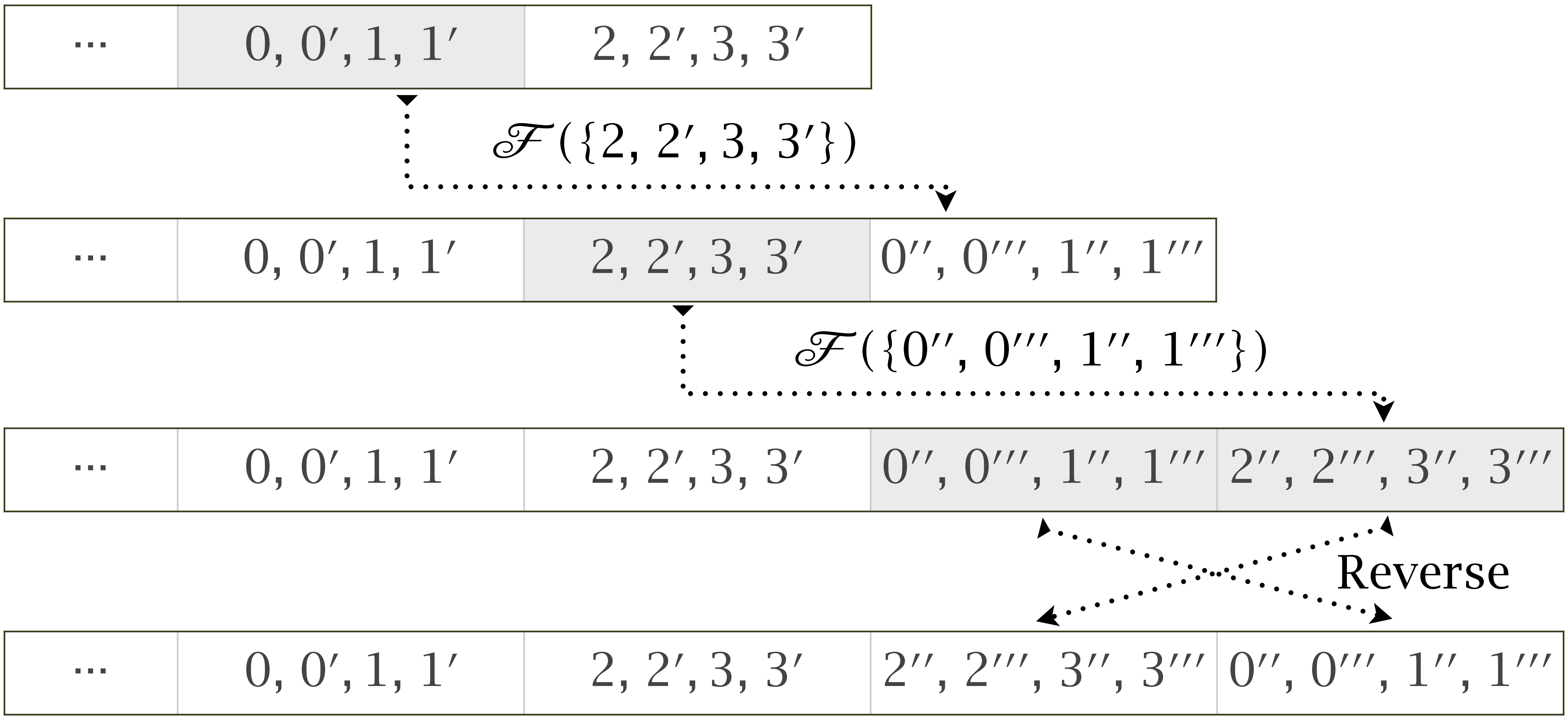
Detailed balance: evolve backward from $(3',2',1',0')$

Ensemble assisted Markov chains: in parallel



- Embedding Markov chains in Markov chains

Ensemble assisted Markov chains: multi-state



- Embedding Markov chains in Markov chains

What kind of information from other chains?

- How do we generate the next state?
- Modify MD evolution
  “Quasi-Newton MCMC” — Zhang & Sutton (2011)
  “Ensemble precondition” — Matthews et al (2016)
  “Quasi-Newton Langevin” — Simsekli et al (2016)
  “Magnetic HMC” — Tripuraneni et al (2016)
  “Wormhole” — Lan et al (2013)
- Modify Metropolis-Hastings
  “Multi-try” — Liu, Liang, and Wong (2000)
- Other techniques? Machine learning!!!

Quasi-Newton method for HMC Hamiltonian

- BFGS approximation of the Hessian: $G's = y$
  Update an old approximation to a new one

$$G' = G + \frac{yy^\dagger}{y^\dagger s} - \frac{Gss^\dagger G}{s^\dagger Gs} \qquad \begin{array}{ll} s = \ln U'U^\dagger & \text{step} \\ y = \nabla S(U') - \nabla S(U) & \text{yield} \end{array}$$

- Approximate Hessian from configs of other MC
  Repeatedly apply the update according to N_{stream}

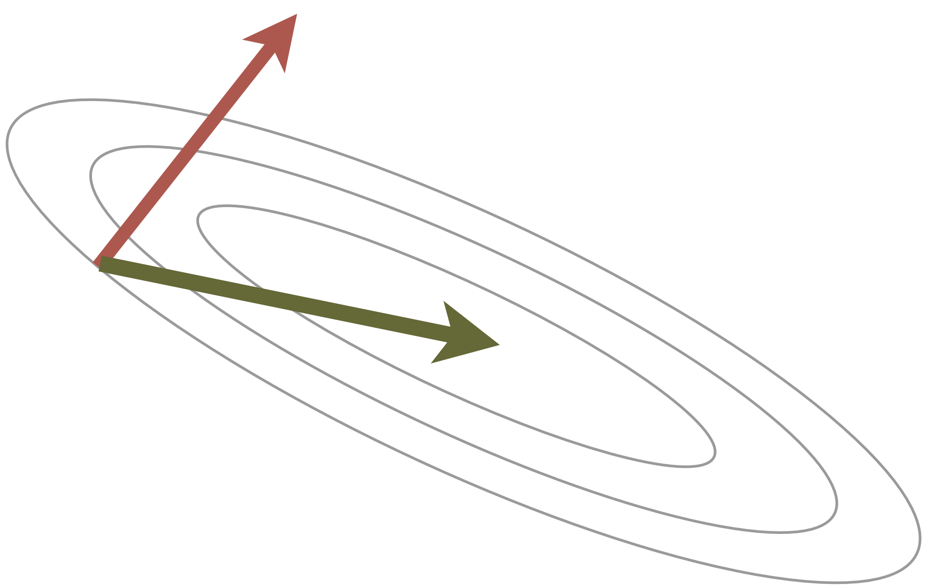
- Use the approximate Hessian for the mass matrix

$$H = S(U) + \frac{1}{2} p^\dagger G^{-1} p$$

- Note: Fourier acceleration≃Local free field Hessian

Quasi-Newton method

- Avoids the slow down of the steepest decent in narrow valleys



- Caveat in the current study:
  - The approximated Hessian is global
  - We do not use the current location

Benefits of rank-2 update (BFGS style)

- Factorizable matrix (Brodlie et al 1973)
  - Initializing random momenta

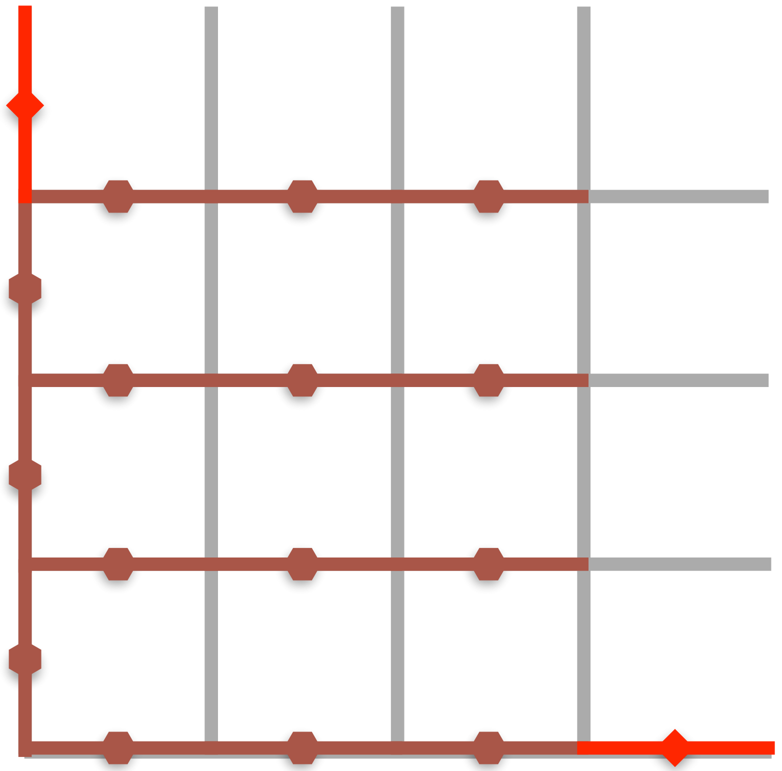
$$G' = G + ww^\dagger - zz^\dagger \quad \rightarrow \quad G' = (1 - uv^\dagger)G(1 - vu^\dagger)$$

- Exactly invertible
  - MD evolution
  - Computing the kinetic energy

$$G'^{-1} = \left(1 - \frac{vu^\dagger}{v^\dagger u - 1}\right)G^{-1}\left(1 - \frac{uv^\dagger}{v^\dagger u - 1}\right)$$

Gauge fixing of 2D U(1) lattice

- Removes exact zero modes from the real Hessian
- Frozen degrees of freedom take the same values
- We choose maximal tree gauge fixing
- Fix two more non-gauge degree of freedom



Regulate the approximated Hessian matrix

- Remove low modes in the approximate global Hessian
- Add one more term to keep the rank-2 update

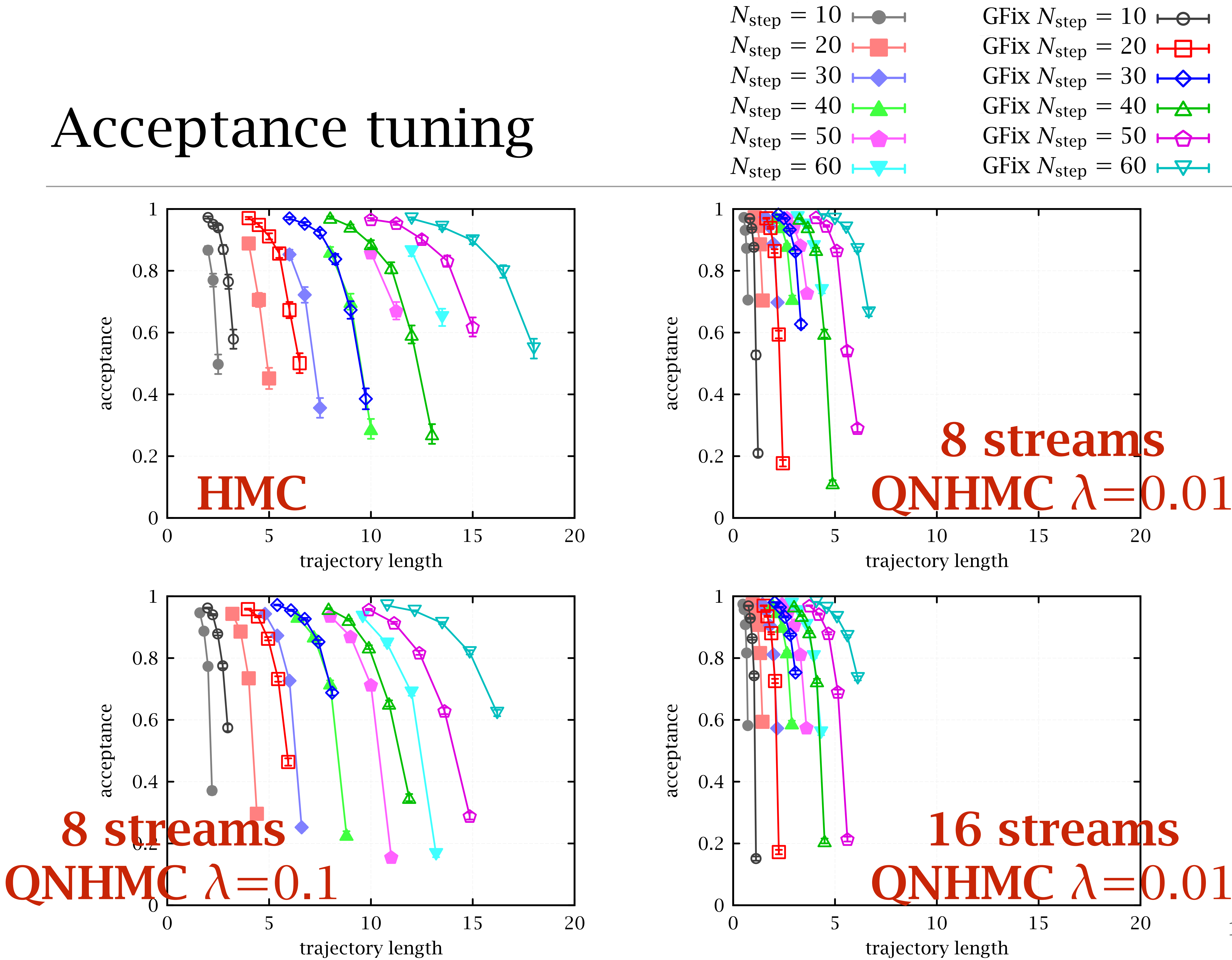
$$G' = G + \frac{yy^\dagger}{y^\dagger s} - \left(1 - \lambda \frac{s^\dagger s}{s^\dagger G s}\right) \frac{G s s^\dagger G}{s^\dagger G s}$$

- Works in practice, but not a strict bound
- Caveat:
  - Mildly violates $G's = y$
  - Still no upper bound

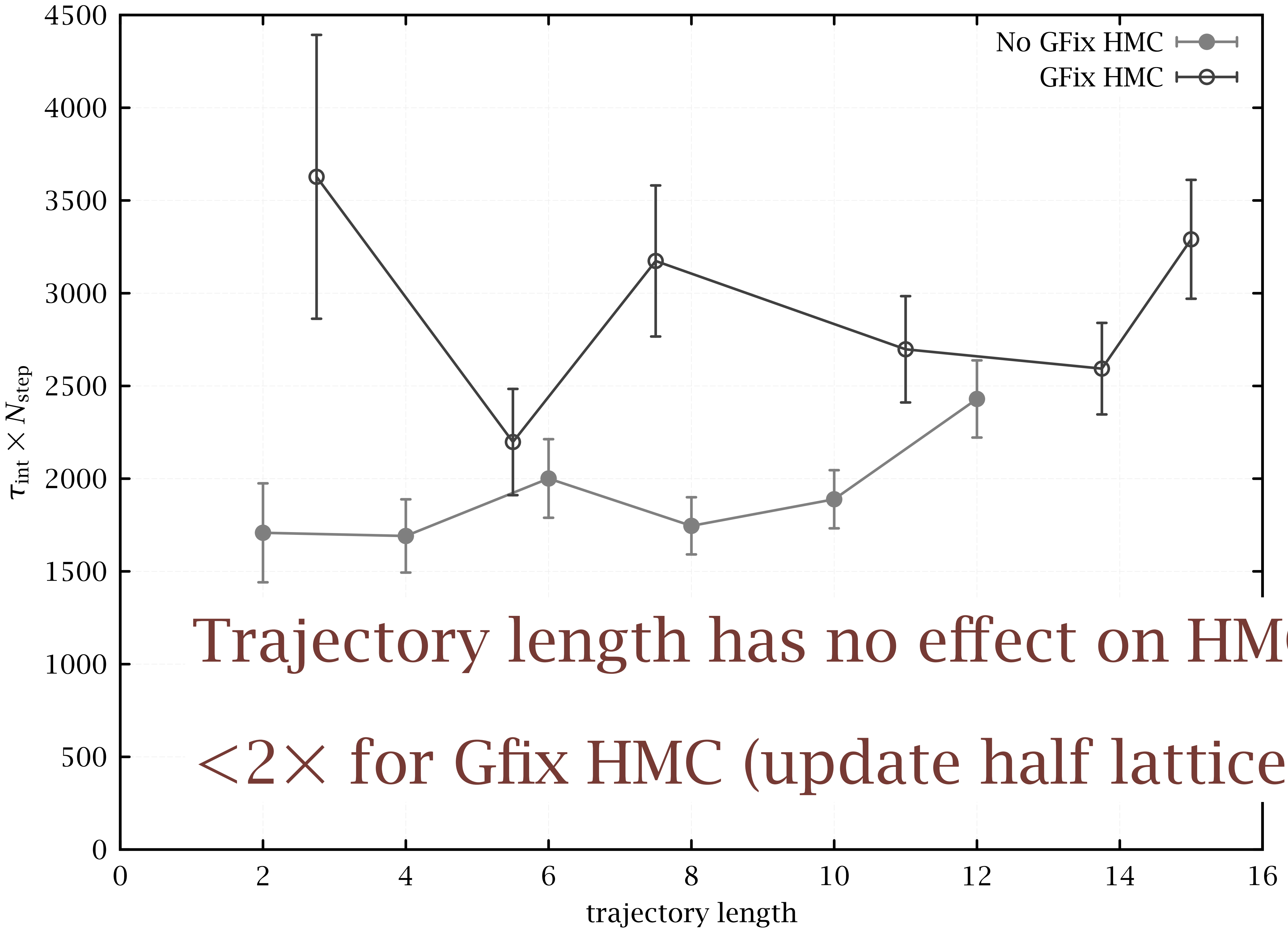
Test on 2D U(1) theory (work in progress)

- Fixed $\beta = 5.8$, lattice size 32×32
- Serial version of the ensemble Markov chain
- Second order Omelyan integrator (did not tune λ)
- Look at the autocorrelation of the topological susceptibility, $\langle Q^2/V \rangle$
- Topological charge, $Q = \dfrac{1}{2\pi} \sum_x \mathrm{Arg}\,\square_x$
  $\mathrm{Arg} : \mathbb{C} \mapsto (-\pi, \pi)$
- Topological charge is exact integer with periodic boundary conditions

Acceptance tuning



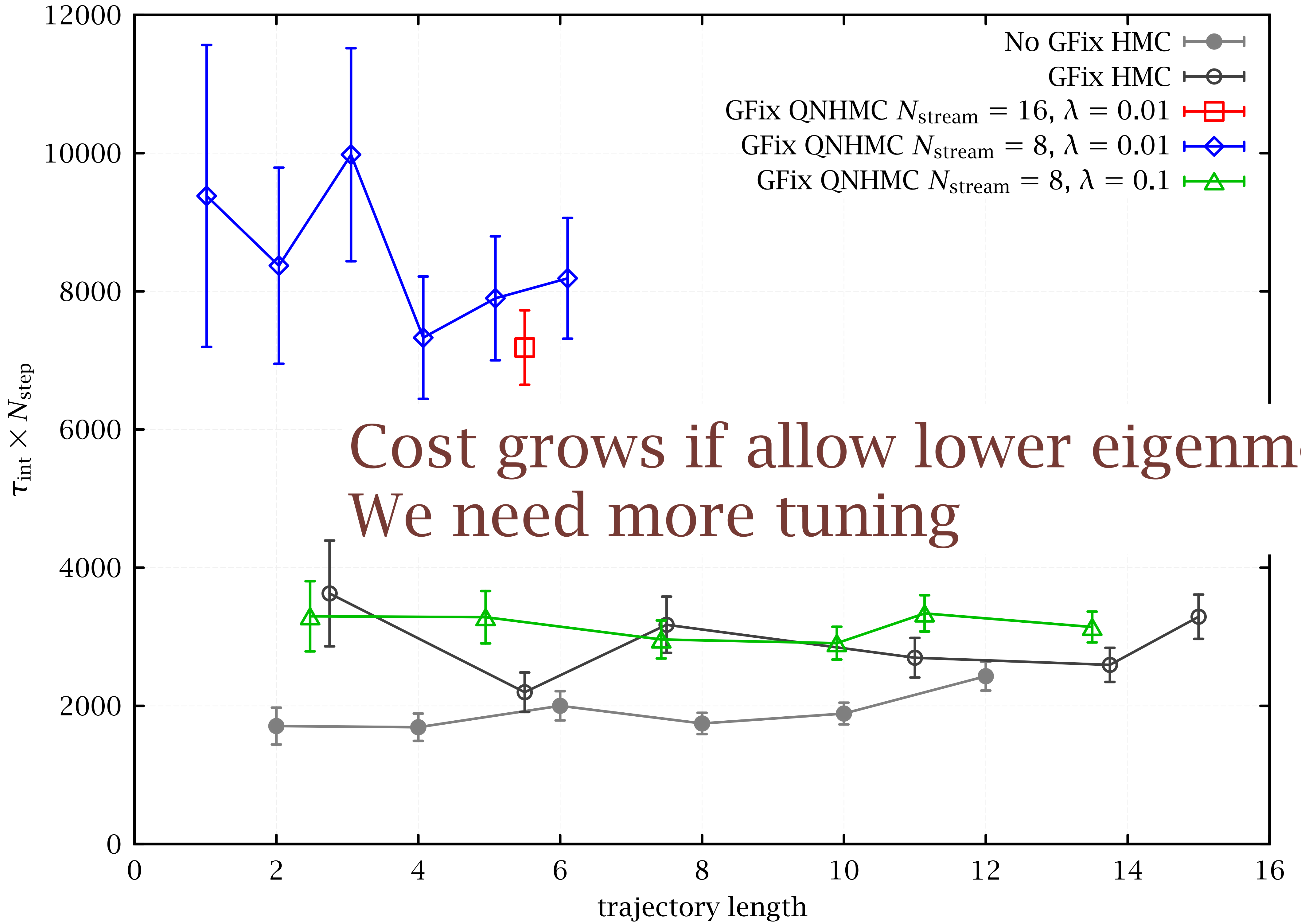
Autocorrelation of topological susceptibility



Trajectory length has no effect on HMC

<2× for Gfix HMC (update half lattice)

Autocorrelation of topological susceptibility



Cost grows if allow lower eigenmodes
We need more tuning

Summary & Outlook

- We devise an algorithm creating multiple Markov chains in parallel
  Allow exchange of information while generating the Markov chains
- We modify HMC to use information from neighboring Markov chains
  BFGS approximated Hessian as the mass matrix of the MD Hamiltonian
  Use a custom regulator for the approximated Hessian for stability
- We still need more tuning and testing (parameters / observables)
- Ways to improve the algorithm
  - Exploit the ensemble of Markov chains (multi-scale?)
  - Other method for constructing the mass matrix
  - Use other information / observables to augment MD / Metropolis
- Machine learning!