Algorithms for gauge evolution and solvers

Peter Boyle (BNL)

- Gauge evolution
  - Critical slowing down
    - GFHMC & RMHMC
    - FTHMC
  - Communication avoiding
    - Domain decomposition
  - Multilevel integration & Master field simulation
  - Parallel tempering
- Solvers
- Outlook

Brookhaven National Laboratory

Hardware trends: how does this interact with algorithms

- Incredible growth in accelerated node performance continues
- Huge floating point throughput; Reduced precision arithmetic
- Enormous (but less flexible) dense matrix throughput: Machine learning / DNN
- High speed intranode networks; Lower speed internode networks
- Cost of computation not just flop/s: memory, GPU-GPU & node-node bandwidth important
    - Growing penalty for communication
    - Multi-scale physics and large volumes

  $\Rightarrow$ Changes to algorithms

Location	System	Interconnect (GB/s) per node (X+R)	Floating point performance (GF/s) per node	Memory Bandwidth (GB/s) per node	Year	System peak (PF/s)	FP / Interconnect	FP / Memory	Memory / Interconnect
LLNL	BlueGene/L	2.1	5.6	5.5	2004	0.58	2.7	1.0	2.6
ANL	BlueGene/P	5.1	13.6	13.6	2008	0.56	2.7	1.0	2.7
ANL	BlueGene/Q	40	205	42.6	2012	20	5.1	4.8	1.1
ORNL	Titan	9.6	1445	250	2012	27	150.5	5.8	26.0
NERSC	Edison	32	460	100	2013	2	14.4	4.6	3.1
NERSC	Cori/KNL	32	3050	450	2016	28	95.3	6.8	14.1
ORNL	Summit	50	42000	5400	2018	194	840.0	7.8	108.0
RIKEN	Fugaku	70	3072	1024	2021	488	43.9	3.0	14.6
NERSC	Perlmutter/GPU	200	38800	6220	2022	58	194.0	6.2	31.1
ORNL	Frontier	200	181200	12800	2022	>1630	906.0	14.2	64.0

Hybrid Monte Carlo

- Auxiliary Gaussian integral over conjugate momentum field $\int d\pi e^{\frac{-\pi^2}{2}}$

$$\int d\pi \int d\phi \int dU \quad e^{-\frac{\pi^2}{2}} e^{-S_G[U]} e^{-\phi^*(M^\dagger M)^{-1}\phi}$$

- Outer Metropolis Monte Carlo algorithm
  - Draw gaussian momenta and pseudofermion as gaussian $\eta = M^{-1}\phi$
  - Metropolis acceptance step
  - Proposal includes inner molecular dynamics at constant Hamiltonian:

$$H = \frac{\pi^2}{2} + S_G[U] + \phi^*(M^\dagger M)^{-1}\phi$$

- $\dot{U} = i\pi U$, derive HMC EOM from:

$$\dot{H} = 0 = \pi[\dot{\pi} + iU \cdot \nabla_U S_{TA}]$$

- Finite timestep performed in Lie algebra, keeps U on group manifold:

$$U' = e^{i\pi dt} U$$

- Force terms $\nabla_U S$; Invert $M^\dagger M$ at each timestep of evolution in MD force
- $\Delta H \sim 0.1$, $H \sim 10^9$
  HMC designed to conserve probability
  $\Rightarrow$ Can expect wicked volume scaling for approximate algorithms:

Avenues actively investigated by community (incomplete list)

- Trivialising flows: covered in Gurtej Kanwar's plenary, Mon 10am
- Critical slowing down / Fourier acceleration
  - GFFA-HMC
  - FTHMC
  - RMHMC
- Domain decomposition
- Parallel tempering & topological sampling
- Multi-level integration & Master field simulations
- Solvers & multigrid:
  - DWF
  - Staggered multigrid
  - Staggered communication avoiding preconditioning
  - Equivariant (Covariant) neural networks

SciDAC-5 project

- "Multiscale acceleration: Powering future discoveries in High Energy Physics"
- 5 year project 2022/10 to 2027/10
- Algorithmic research collab. between USQCD HET and SciDAC institutes
- USQCD:
  - 3 Labs: ANL, BNL (lead), FNAL
  - 8 Universites: Columbia, BU, MSU, Illinois, UIUC, Utah
- SciDAC Fastmath :
  - LBNL, MIT, SUNY Buffalo
- 3 work packages:
  - WP1: Multigrid for Domain Wall and Staggered Fermions
  - WP2: Transformational sampling algorithms
  - WP3: Large domain decomposed HMC: minimise communication

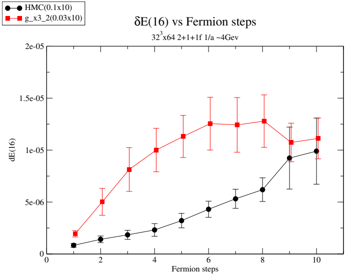
Critical slowing down

Free field: decoupled harmonic oscillators, period in MD time dependent on wavelength:

$$H(\tilde{\pi}_p, \tilde{\phi}_p) = \frac{1}{2}\sum_p \pi_p^\dagger \pi_p + \omega_p^2 \phi_p^\dagger \phi_p \qquad ; \quad \omega_p^2 = m^2 + p^2$$

- (soft covariant) Gauge fixed fourier accelerated HMC Y. Huo Tue 14:30. arXiv:2108.05486
- Riemanian Manifold HMC ECP and SciDAC-5, C. Jung Tue 15:10
  - arXiv:2112.04556, arXiv:1710.07036
  - momentum distribution depends on covariant Laplacian wavelength
  - Fourier accelerate with arbitrary profile function $H[D^2]$ of the adjoint Laplacian
  - Implicit integrator (Girolami & Calderhead)
  - c.f. *Gauge Invariant Fourier Acceleration* Duane & Pendleton, PLB 206 (1988)

$$\int d\pi_U \int d\pi_\phi \int d\phi \int dU \quad \underbrace{e^{-\frac{\pi_U H \pi_U}{2}} e^{-\frac{\pi_\phi H^{-1}\pi_\phi}{2}}}_{\text{cancelling} \det H[U]/\det H[U]} e^{-S_{QCD}[U]} e^{-\phi^*\phi}$$



- Long distance quantities decorrelate faster with RMHMC, measured in Fermion force evaluations
- Gauge implicit integrator needs further optimisation for "wall clock" gain

Flowed HMC

- Much activity based on Luscher's Wilson flowed HMC (arXiv:1009.5877)

Two possible directions to address critical slowing down:

- Complex IR flow: learned generative/trivialising maps; (Gurtej Kanwar plenary, Mon 10am)
- Simple UV flow: retain momentum based local update with field transformation FT-HMC
  - Might be substantially easier to map QCD to QCD than trivialising to strong coupling limit

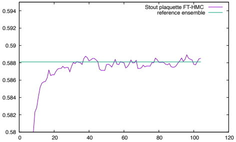
UV smearing function $U(V)$ brings tunable Fourier acceleration with *incomplete trivialisation*

$$\int dU e^{-S[U]} = \int dV \left| \frac{dU}{dV} \right| e^{-S[U(V)]}$$

- FTHMC :
  Gaussian momentum distribution
  Covariant smearing $\Rightarrow$ wavelength dependent transformation to physical gauge field
  Computable exact log det Jacobian
  - Quenched FT-HMC; general Wilson loops (Matsumoto, Jin, Izubuchi, Tomiya et al)

Simple flow FT-HMC with domain wall fermions

- (masked) stout plaquette smeared FT-HMC in Grid: enables Fermion simulations (PB, Jin)
  - Reimplements Luchang Jin's Qlat FT-HMC; adds many options for fermions
- $L_s = 16$ Domain wall fermions + Iwasaki gauge action
- $16^3 \times 48$, $\beta = 2.13$, $m_{ud} = 0.01$, $m_s = 0.04$ 2+1 flavour
  - TWQCD's exact one flavor algorithm for strange (arXiv:1403.1683)
- $2 \times Nd$ subsets for plaquette stout smearing, $\rho = 0.1$
  - Developed under SciDAC-5 WP2 ; single node of 4xAMD GPUs (Lumi-G)
    - *Field transformation overhead significant but sub-dominant*
  - Reproduces reference plaquette in smeared links
  - Plan to investigate critical slowing down on 32^3 at 3 GeV
- Exascale consideration: the Jacobian force parallelises; Fermion solvers do not.
  - FT-HMC overhead is scalable.

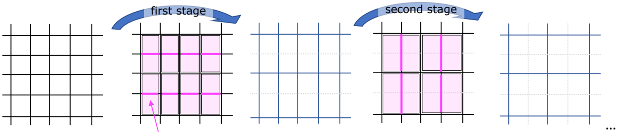


Trajectory	Jacobian	Force smearing	Fermion action + force
1530s	295s	166s	900s

Decimation

- arXiv:2212.11387, (Nobu Matusmoto Lattice 2022) + Akio Tomiya, Luchang Jin, Taku Izubuchi, PB, Christoph Lehner, Chulwoo Jung
- 2D U(1): trivialise subsets of gauge links sequentially

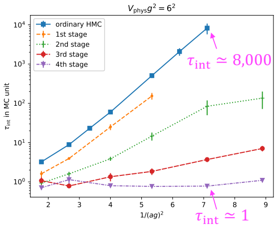
Decimation map in 2D for accelerating HMC   Nobuyuki Matsumoto, Mon 4:20



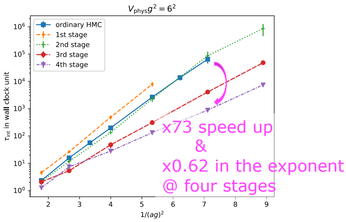
Lüscher's trivializing map applied to one link at a time
Lüscher 09

Combined with guided HMC: **Duane,Kennedy,Pendleton,Roweth 87**
Horowitz 91

$\tau_{\rm int}$ in Monte Carlo time



$\tau_{\rm int}$ in wall clock time



Fermion determinant factorisation

- Can split 2f determinants into multiple factors: closer to identity, independently estimated
  - Reduces forces, whose stochastic sum is estimated more precisely
  - Ideally reduce the computational cost of each factor

- Examples:

hep-lat/0107019 Hasenbusch mass preconditioning:

$$\det M^\dagger(m_l)M(m_l) = \det \frac{M^\dagger(m_l)M(m_l)}{M^\dagger(m_h)M(m_h)} \det M^\dagger(m_h)M(m_h)$$

hep-lat/0409134 Clark and Kennedy n-roots: $\det M^\dagger(m_l)M(m_l) = \left(\det\left[M(m_l)^\dagger M(m_l)\right]^{\frac{1}{n}}\right)^n$

hep-lat/0409106 Luscher's domain decomposition HMC

- Hasenbusch preconditioning has been most widely used
- DDHMC provides clean localisation of determinant and computational benefit
- n-roots did not initially reduce cost of each factor

arXiv:1808.01829 De Forcrand and Keegan use multi-RHS multi-shift solver for computational benefit

- Exascale considerations:
  - Domain decomposition is of particular interest for exascale as it reduces global commmunication
  - n-roots may be compined with machine subdivision for trivial parallelism

Domain decomposition: hep-lat/0409106, CLS, PACS-CS

- Divide space time into "black" and "white" blocks and decompose Dirac operator

$$\begin{pmatrix} D_\Omega & D_\partial \\ D_{\bar\partial} & D_{\bar\Omega} \end{pmatrix} = \begin{pmatrix} 1 & D_\partial D_{\bar\Omega}^{-1} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} D_\Omega - D_\partial D_{\bar\Omega}^{-1} D_{\bar\partial} & 0 \\ 0 & D_{\bar\Omega} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ D_{\bar\Omega}^{-1} D_{\bar\partial} & 1 \end{pmatrix}.$$

$$\det D = \det D_\Omega \det D_{\bar\Omega} \det \left\{ 1 - D_\Omega^{-1} D_\partial D_{\bar\Omega}^{-1} D_{\bar\partial} \right\},$$

- "boundary" determinant is projected to exterior boundary of Ω

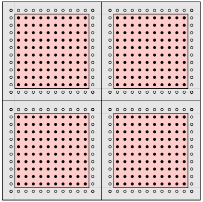
$$R = \mathbb{P}_{\bar\partial} - \mathbb{P}_{\bar\partial} D_\Omega^{-1} D_\partial D_{\bar\Omega}^{-1} D_{\bar\partial} \qquad R^{-1} = \hat{\mathbb{P}}_{\bar\partial} - \hat{\mathbb{P}}_{\bar\partial} D^{-1} \hat{D}_{\bar\partial}$$

- Boundary gauge links are frozen (cross domain & surface plane)
  - 3D Pseudofermion action $\phi_{\bar\partial}^\dagger (RR^\dagger)^{-1} \phi_{\bar\partial}$ involves two solves due to boundary projector
  - MD evolve gauge action & local det without communication
  - Small cell IR regulates Dirichlet solve, fits in cache
- Drawbacks:
  - Doesn't work for odd flavours (nested solve needed due to projector)
  - Too many inactive links loses efficiency
  - Too few inactive links has spikes in δH
  - Small domains: good for CPUs but not for GPUs

Large domain DDHMC:

arXiv:2203.17119 PB, D. Bollweg, C. Kelly, A. Yamaguchi, SciDAC-5

- Gain of communication avoidance, do not use *small* cell
- Make domain as big as required for good performance on a *multi-GPU exascale node*
- Make inactive region as broad required to suppress communication
  - Expect to lose IR bound on subdomain solves
  - Expect to gain fidelity and *large force suppression* with bigger inactive zones/more efficiency
  - Perturbative massless zero momentum two point function $\propto t^{-3}$: suppress with wider inactive link zones
- Must be more efficient on a large enough volume: asymptotically guaranteed to win if communications is a limit
- Suited to master-field ideas



Modified pseudofermion approach

Write $\tilde{D}$ as Dirichlet operator. Use

$$\det\left\{1 - D_{\Omega}^{-1} D_{\partial} D_{\bar{\Omega}}^{-1} D_{\bar{\partial}}\right\} = \frac{\det D}{\det \tilde{D}},$$

- 2 flavour (4d) boundary pseudofermion action:

$$S_{2f} = \phi^{\dagger} D_{\text{dirichlet}} (D^{\dagger} D)^{-1} D_{\text{dirichlet}}^{\dagger} \phi$$

- Introduce domain wall pseudofermions in style of (dirichlet) Hasenbuch intermediate operators:

$$\det P(D_l^{\dagger} D_l)^{-1} P^{\dagger} \quad = \quad \det \tilde{D}_l (D_l^{\dagger} D_l)^{-1} \tilde{D}^{\dagger} \cdot \det \tilde{P} (\tilde{D}_l^{\dagger} \tilde{D}_l)^{-1} \tilde{P}^{\dagger} \cdot \det P (\tilde{P}^{\dagger} \tilde{P})^{-1} P^{\dagger}$$

Can now take fractional powers of boundary determinant!

- RHMC boundary pseudofermion:

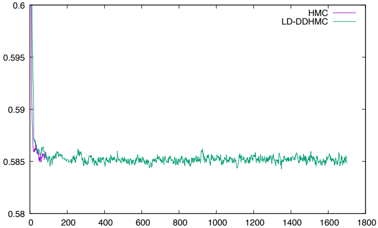
$$S_{1f}^{B} \quad = \quad \phi_1^{\dagger} (D_{\text{dirichlet}}^{\dagger} D_{\text{dirichlet}})^{\frac{1}{4}} (D^{\dagger} D)^{-\frac{1}{2}} (D_{\text{dirichlet}}^{\dagger} D_{\text{dirichlet}})^{\frac{1}{4}} \phi_1$$

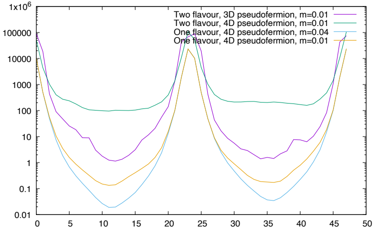
- RHMC local 1 flavor determinant ratio

$$S_{1f}^{L} \quad = \quad \phi_2^{\dagger} (P_{\text{dirichlet}} P_{\text{dirichlet}})^{\frac{1}{4}} (D_{\text{dirichlet}}^{\dagger} D_{\text{dirichlet}})^{-\frac{1}{2}} (P_{\text{dirichlet}} P_{\text{dirichlet}})^{\frac{1}{4}} \phi_2$$

16^3 test

- $16^3 \times 48$, $\beta = 2.13$, $m_{ud} = 0.01$ 2+1 flavour

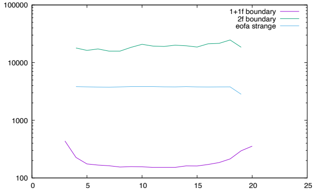




- 2 flavour 4D pseudofermion has larger force than 2f 3D pseudofermion
- 1+1 flavour 4D pseudofermion has smaller force than 2f 3D pseudofermion
- Odd flavour domain decomposition is now possible

Large volume run $\beta = 2.25$, $48^3 \times 96 \times 12$, $m_{ud} = 0.00078$

- Close to physical, but $L_s = 12$ not $L_s = 24$
- Frontier/Lumi-G have *good* communications: Time gain will be strictly limited
  - Certain existing (and future) machines are less balanced: e.g. Polaris, Perlmutter



Bdy	Cell	Iters	Conds
pppa	48x48x48x96	12351	antiperiodic
pppd	48x48x48x96	11864	open time
dddd	48x48x48x96	7349	open 4 dirs
pppd	48x48x48x48	10015	dirichlet
pppd	48x48x48x24	7416	dirichlet
ppdd	48x48x24x24	5150	dirichlet
pddd	48x24x24x24	4324	dirichlet
dddd	24x24x24x24	3692	dirichlet

- Again, reduced force with rational 1+1 f boundary
- Introduced **partial** dirichlet BCs with further force reduction
  - Surface physical fields remain connected, 5d bulk is disconnected

- Nice surprise: Subdomain solves are MUCH faster than full solves
- Lower iteration count MULTIPLIES communications gain
- Applies to all intermediate Hasenbusch factors in our DDHMC scheme

HMC and DDHMC comparison on Lumi-G

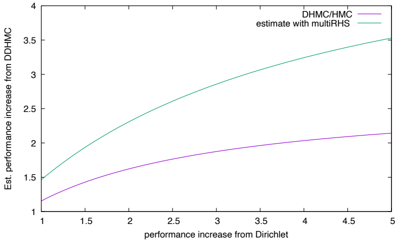
- Compare double precision algorithms for simplicity
  - Loosely tuned to similar δH; preliminary integrator hierarchy (work in progress!)
  - DDHMC block size : $24^3 \times 48$, active link block: $16^3 \times 40$

algorithm	HMC	DDHMC
integrator	Force Gradient	Force Gradient
rms dH	0.04	0.09
L1/2/3 steps	6/2/8	3/3/12
$D_{\mathrm{dirichlet}}$	0	1176892
D_{full}	2432132	884884

- Take cost of full communicating operator to dirichlet as R
- Relative speed up $S(R)$ is

$$S = \frac{N_{full}^{HMC} R}{N_{full}^{DDHMC} R + N_{dir}^{DDHMC}} \sim \frac{2.4R}{0.9R + 1.2} \geq 2.2$$

- Additional gain with multishift/multiRHS is possible (de Forcrand and Keegan, arXiv:1808.01829)

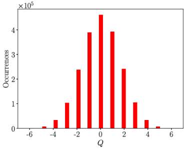


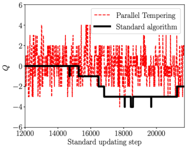
Topology sampling

- Topological sampling is a sign of non-ergodicity and algorithm breakage.
  - Symptomatic treatment may not address all sources of non-ergodicity.
  - Opening boundary conditions cures the symptom by amputation
  - Pay with larger volume and loss of translation averaging
    We had to destroy topology in order to save topology

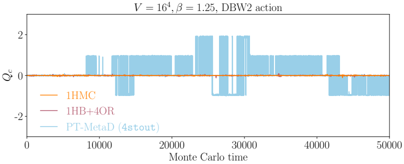
Parallel tempering and topology sampling

- Early UKQCD work: arXiv:hep-lat/9810032
  - Jointly sample multiple ensembles with nearby parameters
  - Replica switching requires suff. similar actions and scales poorly with volume
- Recent interest in parallel tempering as algorithmic solution
- Hasenbusch: arXiv:1706.04443
- Bonnano and D'Elia: arXiv:1911.03384, arXiv:2212.02330, arXiv:2012.14000, arXiv:2205.06190
  - Localised action difference at boundary: temper between open and periodic
  - Does not change periodic distribution, but gives access to tunneling
  - Thursday 14:50, talk by Dasilva Golan, Twisted BC + PT
- Eichorn et al: arXiv:2307.04742 Metadynamics + PT
- Also: out of equilibrium simulation: A. Nada, Monday, 14:50.





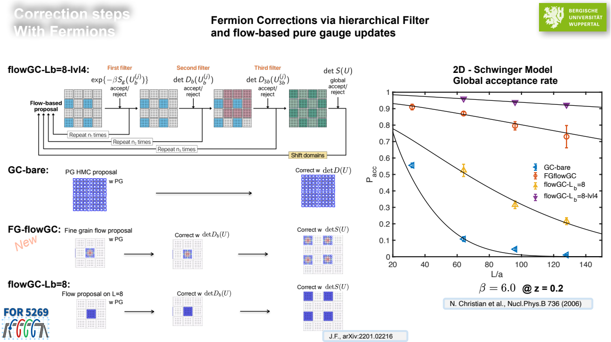
arXiv:2210.07622, SU(6) glueball spectrum with no topology systematic



arXiv:2307.04742 : Quenched DBW2 gauge evolution topological sampling gain

Parallel tempering and topology sampling

- Parallel tempering illustrates problem of *approximate* algorithms
  - Suffers as volume increases *or* change in action grows
  - c.f. accurate MD integration in HMC conserves probability to high accuracy
- Field transformation/normalizing flow may bridge action changes: will suffer if approximate
- Difficult to avoid poor acceptance (exact alg.) / low effective sample size (reweighted)
- Volume scaling in 4D means simple 2D models could mislead
- Subvolume switching for topology is good solid plan:
  in future likely involve domain decomposition for non-local fermion action ?
  - D. Hackett Thu 15:10
  - J. Finkenrath Mon 16:40, local updates, $N_f = 2$ 2D Schwinger model



Master field simulation

- arXiv:1707.09758, 1812.02062, 1911.04533 2111.11544
- Lüscher; Lüscher, Fritzsch, Bulava, Cé, Francis, Rago
- Premise:
  very large lattice gauge field, the master field, can replace a conventional Markov chain Monte-Carlo ensemble of a discretised field theory with a mass gap
- Larger volumes than traditional have certain advantages and disadvantages
- Advantages:
  - *Global* topology: effects fall as $\frac{1}{V}$ and can be made negligible
    - do we care if there is an instanton behind the moon?
  - Suited to large GPU systems
  - Variance: reduces with volume so equivalent to gaining statistics
  - Estimate statistical error, in part, from *spatial decorrelation*
- Disadvantages:
  - Thermalisation overhead: low configuration count argument somewhat oversold
    - To be sure of thermalisation, must run for many autocorrelation times. Less waste to accumulate statistics
    - Thermalisation bias falls as $\frac{1}{N_{\rm conf}}$; variance falls as $\frac{1}{\sqrt{N_{\rm conf}}}$
      Evolution removes bias faster than error falls
  - Intractable for eigenvector methods & other V^2 algorithms: idea best with a good multigrid

Multilevel integration

- arXiv:1601.04587, arXiv:1612.06424, arXiv:1812.01875, arXiv:2112.02647

$$\det Q = \frac{\det(1-w)\det Q_0 \det Q_2}{\det Q_{\Lambda_1}}$$

- N_1 samples in interstitial Λ_1 domain
- For each sample N_2 independently from each of Λ_0 and Λ_2
- Reweight factor $(1-w)$ regularised by width of intermediate zone
- $N_1 \times N_2^2$ samples
- Lots avenues to explore in valence measurements:
  - e.g. multigrid subspace reuse

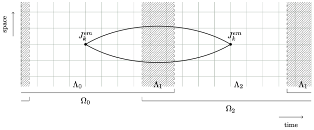


Figure 1: Decomposition of the lattice in domains as described in the text; periodic and anti-periodic boundary conditions in the time direction are enforced for gluons and fermions, respectively.

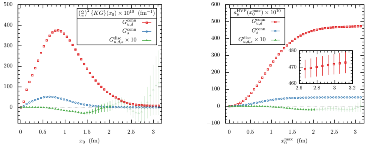


Figure 3: Best results for contribution of the light-connected (red squares), strange-connected (blue circles) and disconnected (green triangles) contractions to the integrand in Eq. (1) (left panel) and to $a_\mu^{\rm HVP}$ (right panel) as a function of the time coordinate $x_0^{\rm max}$.

Solvers

- Wilson multigrid is a solved problem for valence
- Faster setup for HMC needed
  - Encourage Wilson folks to try Chebyshev filter setup in arXiv:2103.05034
  - Both setup AND solved DWF twice faster than single RB-CGNE
- Domain wall and staggered multigrid research on-going
  - Currently algorithms exist for DWF and Staggered but less dramatic gains than for Wilson

Multigrid preconditioners

Low mode subspace vectors ϕ generated in some way

- Inverse iteration, Inverse iteration with self refinement, Chebyshev filters

$$\phi_k^b(x) = \begin{cases} \phi_k(x) & ; \quad x \in b \\ 0 & ; \quad x \notin b \end{cases} \tag{1}$$

$$\mathrm{span}\{\phi_k\} \subset \mathrm{span}\{\phi_k^b\}. \tag{2}$$

$$P_S = \sum_{k,b} |\phi_k^b\rangle\langle\phi_k^b| \qquad ; \qquad P_{\bar{S}} = 1 - P_S \tag{3}$$

$$M = \begin{pmatrix} M_{\bar{S}\bar{S}} & M_{S\bar{S}} \\ M_{\bar{S}S} & M_{SS} \end{pmatrix} = \begin{pmatrix} P_{\bar{S}} M P_{\bar{S}} & P_S M P_{\bar{S}} \\ P_{\bar{S}} M P_S & P_S M P_S \end{pmatrix} \tag{4}$$

We can represent the matrix M exactly on this subspace by computing its matrix elements, known as the *little Dirac operator* (coarse grid matrix in multi-grid)

$$A_{jk}^{ab} = \langle\phi_j^a|M|\phi_k^b\rangle \qquad ; \qquad (M_{SS}) = A_{ij}^{ab}|\phi_i^a\rangle\langle\phi_j^b|. \tag{5}$$

the subspace inverse can be solved by Krylov methods and is:

$$Q = \begin{pmatrix} 0 & 0 \\ 0 & M_{SS}^{-1} \end{pmatrix} \qquad ; \qquad M_{SS}^{-1} = (A^{-1})_{ij}^{ab}|\phi_i^a\rangle\langle\phi_j^b| \tag{6}$$

It is important to note that A inherits a sparse structure from M because well separated blocks do *not* connect through M.

DWF: multigrid enfant terrible

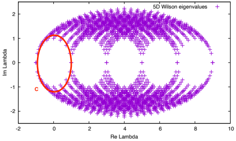
- Spectrum of DWF presents problems to non-Hermitian Krylov solvers
- Why? Krylov space is the span of polynomials of matrix M.
  Let $|i\rangle$ be the set of right eigenvectors, $\mathscr{P}(x) = c_n x^n$ a polynomial

$$\begin{aligned} M|i\rangle &= \lambda_i|i\rangle \\ \eta &= \eta_i|i\rangle \\ \psi^{\mathrm{Krylov}} &= \mathscr{P}(M)\eta = (c_n\lambda_i^n)\eta_i|i\rangle \\ \psi^{\mathrm{True}} &= \frac{1}{\lambda_i}\eta_i|i\rangle \end{aligned}$$

- There exists a contour C contained entirely within the (dense in large/infinite volume) spectrum such that

$$\oint_C \mathscr{P}^{\mathrm{Krylov}}(z)dz = 0$$

$$\oint_C F^{\mathrm{True}}(z)dz = \oint_C \frac{1}{z}dz = 2\pi i$$



- Thus the Krylov polynomial and true solution must differ *within the domain of the spectrum*
- Must differ from solution *between* discrete eigenvalues. Low order polynomial is inadequate
- Manifests as slow convergence, perhaps of order system size

Domain wall and Staggered multigrid

- arXiv:1205.2933, Cohen, Brower: real positive ($M^\dagger$ precondition)
  coarsen $M^\dagger M$ (2hop)
- arXiv:1402.2585, PB: real indefinite (RB-NE precondition)
  coarsen $M_{pc}^\dagger M_{pc}$ (4hop)
- arXiv:1611.06944, arXiv:2203.17119, PB, Yamaguchi: real indefinite (Γ_5 precondition)
  coarsen $\Gamma_5 M$ (1hop)
- arXiv:2004.07732, Weinberg et al: complex positive half plane ($M_{PV}^\dagger$ precondition),
  coarsen M (1 hop) and $M_{PV}^\dagger$ (1 hop)
  - 2D U(1) arXiv:2004.07732
  - 4D SU(3) arXiv:2203.17119
- Staggered multigrid: arXiv:2212.12559, arXiv:1801.07823

SciDAC-5 / PETSc collaboration:

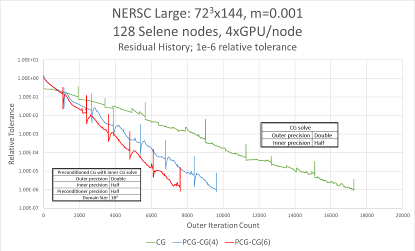
- Directly coarsen 2-hop matrix $M_{PV}^\dagger M$
- Work with PETSc team opens up many algebraic and other multigrid options
  - Apply good ideas arising from applied math community

Staggered Schwarz preconditioning

- Kate Clark, Thu 14:00 : bit reproducible and more precise reductions
- Evan Weinberg, Mon 15:20 : HISQy business, Schwarz Preconditioning HISQ

Schwarz Preconditioning for the HISQ operator

"HISQy Business"
Evan Weinberg, Monday 15:20

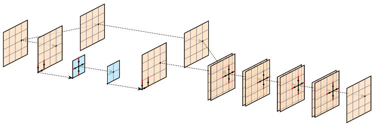


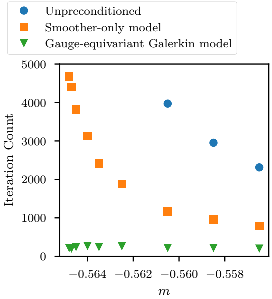
- Problem: not all HPC facilities prioritize network bandwidth
- Solution: use communication-avoiding preconditioning?
- Additive Schwarz preconditioning to improved staggered fermions is challenging
  - Operator is less local due to 3-hop Naik term
  - Typically solve the normal operator, doubling the radius of the operator
- Critical to account for "snake" terms
  - Contributions that start and end at the domain boundary but have non-zero support outside of the domain
  - Snake terms induce an effective "boundary clover" term
- Successfully deployed in **QUDA**
  - Stable convergence using half precision
  - Time-to-solution improvement on machines with poor network bandwidth
- Future work: application of Schwarz preconditioner as a smoother for staggered multigrid

NVIDIA 1

Covariant network approach

- Lehner & Wettig : arXiv:2302.05419 arXiv:2304.10438
- Key idea: NN directly represents *inverse* of matrix on coarse space
  - Similar coarsening to multigrid / inexact deflation
  - Replace *iterative* solve of coarse representation of operator with less local approxation of inverse (deep covariant network).
- Clearly reasonably for high mode smoother.
- Look forward to data scaling low mode coarse grid correction to larger volumes
  - currently $8^3 \times 16$





Summary

- Exciting range of research in gauge evolution
    - Trivialising flows
    - Critical slowing down: GFFAHMC, RMHMC, FTHMC
    - Parallel tempering to address topological sampling via subvolume "topology factories"
        - I am optimistic open BC's can soon be avoided
        - Conventional/subvolume PT could be combined with a learned flow mapping QCD to QCD
    - Domain decomposition & communication avoidance
        - DD determinant likely central to subvolume techniques with fermions
        - Possibly required to make flow methods exact and will impose a cost
- Solver research ongoing for MG in staggered and DWF continues
- Opportunities for ML ; pragmatism vs. ambition
- Communication avoidance continues to be important