Domain Wall Fermion QCD with the Exact One Flavor Algorithm

David Murphy
Columbia University

The 35th International Symposium on Lattice Field Theory (Granada, Spain)

June 20th, 2017

Talk based on: C. Jung, C. Kelly, R.D. Mawhinney, and D.J. Murphy, *Domain Wall Fermion QCD with the Exact One Flavor Algorithm* [arXiv:1706.05843]

Motivation

- Standard algorithm for lattice ensemble generation is hybrid Monte Carlo (HMC)
- Dynamical quark effects are described by determinant of fermion matrix

$$\langle \mathscr{O} \rangle = \frac{1}{\mathcal{Z}} \int \mathscr{D}U \, \mathcal{O}[U] \det\left(\mathcal{M}[U]\right) e^{-S[U]}$$

- Typically choose Hermitian $\mathcal{M} = D^{\dagger} D$, where D is the lattice Dirac operator, but $\det(D^{\dagger} D)$ describes two degenerate quark flavors with same mass
- Take square root for one flavor simulation $\Longrightarrow$ rational HMC (RHMC)
- Multishift CG makes RHMC practical, but...
  - Generally more expensive than ordinary CG, due to extra linear algebra at each iteration
  - Can't use nonzero initial guesses, so techniques such as restarted solvers and forecasted initial guesses are not applicable
  - Empirically find Hasenbusch mass preconditioning to be prohibitively expensive
- For RBC/UKQCD $\Delta I = 1/2$ $K \to \pi\pi$ calculation with G-parity boundary conditions $\det(D^{\dagger} D)$ describes four flavors, and RHMC is needed for light quarks as well
- **Goal:** Use EOFA to accelerate $K \to \pi\pi$ ensemble generation
  - New preconditioning scheme for EOFA Dirac equation
  - Timing benchmarks on production ensembles
- More detail in our paper!

The Exact One Flavor Algorithm (EOFA)

- RHMC: Form rational approximation $x^{1/2} \simeq \alpha_0 + \sum_k \alpha_k / (x + \beta_k)$ and compute

$$\det \left[\frac{D(m_1)}{D(m_2)} \right] = \left\{ \det \left[\frac{D^\dagger D(m_1)}{D^\dagger D(m_2)} \right] \right\}^{1/2}$$

  Multishift CG allows $D^\dagger D + \beta_k$ to be inverted for all k simultaneously

- EOFA: Use Schur determinant identity applied to spin structure of D to factorize

$$\det \left[\frac{D(m_1)}{D(m_2)} \right] = \frac{1}{\det(H_1)} \cdot \frac{1}{\det(H_2)}$$

  with H_1 and H_2 Hermitian and positive-definite [TWQCD, arXiv:1403.1683]

- No rational approximation, so no need for multishift CG
- Expect [TWQCD, arXiv:1412.0819]:
  - Reduced memory footprint
  - Faster algorithm since we avoid extra linear algebra overhead

Lattice Ensembles

We use two RBC/UKQCD ensembles with physical mass Möbius DWF (MDWF) quarks:

1. **24ID:** Coarse simulation with periodic boundary conditions (RHMC or EOFA strange quark) [see Bob Mawhinney's talk, Tue. @ 17:30]
2. **32ID-G:** Production ensemble for $\Delta I = 1/2$ $K \to \pi\pi$ calculation with G-parity boundary conditions (RHMC or EOFA light quarks) [Bai et al., arXiv:1505.07863]

Ensemble	N_f	Action	β	$L^3 \times T \times L_s$	Möbius Scale	GPBC	am_l	am_h
24ID	2+1	MDWF + ID	1.633	$24^3 \times 64 \times 24$	4.0	—	0.00789	0.085
32ID-G	2+1	MDWF + ID	1.75	$32^3 \times 64 \times 12$	2.67	x,y,z	0.0001	0.045

Table: Summary of ensembles and input parameters. Both ensembles use the Iwasaki + dislocation suppressing determinant ratio (DSDR) gauge action (ID). Here β is the gauge coupling, and "GPBC" denotes spatial directions with G-parity boundary conditions.

Ensemble	a^{-1} (GeV)	L (fm)	m_π (MeV)	m_K (MeV)
24ID	0.981(39)	4.82(19)	137.1(1)	494.6(1)
32ID-G	1.378(7)	4.57(2)	143.1(2.0)	490.6(2.4)

Table: Measured properties of both ensembles.

EOFA Mathematical Formalism

EOFA Action [TWQCD, arXiv:1403.1683]

$$S_{\mathrm{EOFA}} = \phi^{\dagger} \Bigg[\underbrace{\mathbb{1} - k P_{-} \Omega_{-}^{\dagger} \left[H(m_1) \right]^{-1} \Omega_{-} P_{-} + k P_{+} \Omega_{+}^{\dagger} \left[H(m_2) - \Delta_{+} P_{+} \right]^{-1} \Omega_{+} P_{+}}_{\equiv \mathcal{M}_{\mathrm{EOFA}}} \Bigg] \phi$$

- $\boldsymbol{D_{\mathrm{DWF}}}$ (standard DWF Dirac operator) and $\boldsymbol{D_{\mathrm{EOFA}}}$ (TWQCD's EOFA Dirac operator) are related by

$$\boxed{\boldsymbol{D_{\mathrm{DWF}} = D_{\mathrm{EOFA}} \cdot \widetilde{D}}}$$

  where $\boldsymbol{\widetilde{D} = \delta_{xx'} \delta_{\alpha\alpha'} \delta_{aa'} (\widetilde{D})_{ss'}}$ is block diagonal in 4D spacetime, spin, and color indices, and has no gauge field dependence
- $\boldsymbol{H \equiv \gamma_5 R_5 D_{\mathrm{EOFA}}}$ is Hermitian EOFA Dirac operator
- Introducing D_{EOFA} trades explicit γ_5-Hermiticity for any choice of parameters for dense $L_s \times L_s$ block structure in fifth dimension: D_{DWF} has tridiagonal ss' stencil but $(\gamma_5 R_5 D_{\mathrm{DWF}})^{\dagger} \neq \gamma_5 R_5 D_{\mathrm{DWF}}$ in general
- Evaluating Hamiltonian or pseudofermion force requires two (ordinary) CG inversions
- Heatbath requires evaluating $\phi = \mathcal{M}_{\mathrm{EOFA}}^{-1/2} \eta$ (still need rational approximation!)

Cayley Preconditioning: Formalism

Generic EOFA Linear System

$$\big(\gamma_5 R_5 D_{\rm EOFA} + \beta \Delta_{\pm} P_{\pm}\big) \psi = \phi$$

- Möbius $D_{\rm EOFA}$ is dense in ss' and thus expensive compared to (tridiagonal) $D_{\rm DWF}$
- Can use $\widetilde{D}^{-1}$ as a preconditioner to write EOFA system in terms of $D_{\rm DWF}$ for $\beta = 0$:

$$\boxed{\gamma_5 R_5 D_{\rm EOFA} \psi = \phi} \iff \underbrace{D_{\rm EOFA} \cdot \widetilde{D}}_{= D_{\rm DWF}} \cdot \underbrace{\widetilde{D}^{-1} \psi}_{\equiv \psi'} = \gamma_5 R_5 \phi \iff \boxed{D_{\rm DWF} \psi' = \phi'}$$

- $\beta \neq 0$ system can be treated as slight generalization of $D_{\rm DWF}$ to a four-point stencil in ss' by noticing that $\Delta_{\pm} \widetilde{D} = \vec{u} \otimes \vec{v}$ is rank-one [Jung et al., arXiv:1706.05843]
- **"Cayley preconditioning":** solve substantially cheaper preconditioned system for ψ' and recover ψ for one additional matrix multiplication by $\widetilde{D}$
- Allows simple EOFA implementation reusing existing high-performance $D_{\rm DWF}$ code

Cayley Preconditioning: 24ID Benchmark

Benchmark: Compare analogous solves required to compute the RHMC or EOFA Hamiltonian for the 24ID strange quark on a 256-node BG/Q partition

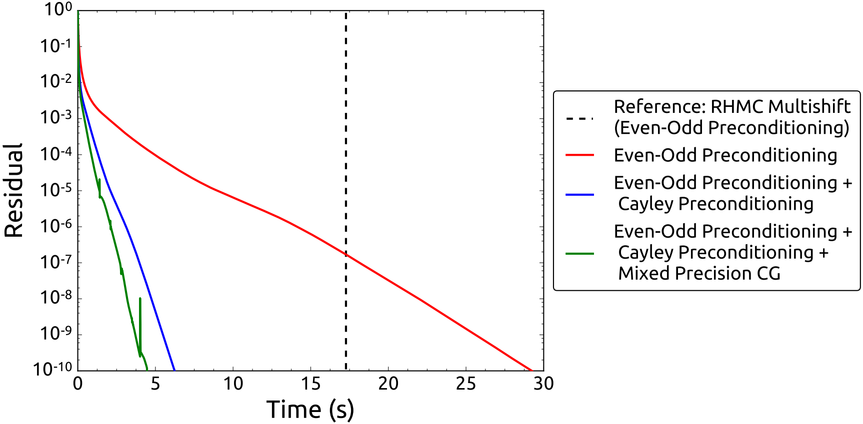


Figure: Wall clock time required to invert $\gamma_5 R_5 D_{\rm EOFA} \psi = \phi$ at the physical strange quark mass on the 24ID ensemble. The dashed vertical line corresponds to the total time required to invert $(D_{\rm DWF}^{\dagger} D_{\rm DWF})^{1/2} \psi = \phi$ at the same mass using multishift CG.

24ID Evolution Benchmark

Benchmark: compare one molecular dynamics trajectory of the 24ID ensemble using either RHMC or EOFA for the strange quark

- Both simulations use identical force gradient (FG) integration schemes
- Can use FG forecasting [Yin and Mawhinney, arXiv:1111.5059] for EOFA evolution (ordinary CG) but not for RHMC evolution (multishift CG)
- Also compare EOFA with and without Cayley preconditioning

	RHMC		**EOFA** (Dense)		**EOFA** (Preconditioned)	
Step	Time (s)	%	Time (s)	%	Time (s)	%
Heatbath	42.6	2.7	340.6	15.1	68.9	15.5
Force gradient integration (total)	1485.6	94.8	1840.6	81.8	355.9	80.1
Final Hamiltonian evaluation	39.4	2.5	68.8	3.1	19.8	4.4
Total	1567.6	—	2250.0	—	444.6	—
(Total RHMC) / Total	**1.0**	—	**0.7**	—	**3.5**	—

Table: Strange quark timings for a single molecular dynamics (MD) trajectory of the 24ID ensemble on a 256-node BG/Q partition.

$3.5\times$ speed-up in strange quark determinant $\Longleftrightarrow \sim 20\%$ speed-up in job time

32ID-G Evolution Tuning: Strategy

First large-scale RBC/UKQCD EOFA calculation: switch to EOFA for light (G-parity) quarks in ongoing $\Delta I = 1/2 \; K \to \pi\pi$ calculation and re-tune integrator

- Previous RHMC run used four-level Omelyan integrator, with quark determinants split across two time steps, and a single Hasenbusch preconditioning mass for the light quark determinant [Bai et al., arXiv:1505.07863]

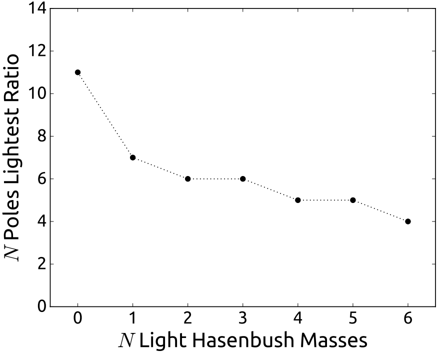
Tuning Strategy

1. Switch to EOFA for the light quarks and three-level force gradient integrator with all quark determinants on same time step
2. Re-tune Hasenbusch mass preconditioning for (EOFA) light quark determinant
3. Tune heatbath rational approximations and CG tolerances
4. Re-tune step size and molecular dynamics CG tolerances

32ID-G Evolution Tuning: Hasenbusch Mass Preconditioning

In contrast to RHMC, Hasenbusch mass preconditioning works extremely well for EOFA!

- Smaller pseudofermion forces $\Longrightarrow$ larger integration step size
- Also observe that spectral range of $\mathcal{M}_{\rm EOFA}$ contracts as $m_2 \to m_1$, allowing cheaper rational approximations to be used in the preconditioned heatbath



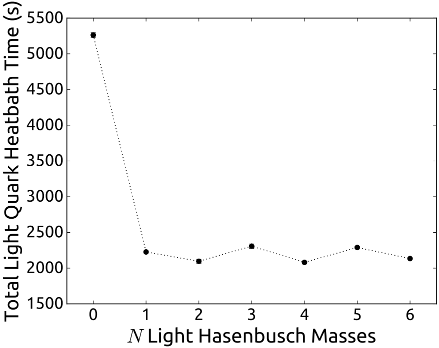
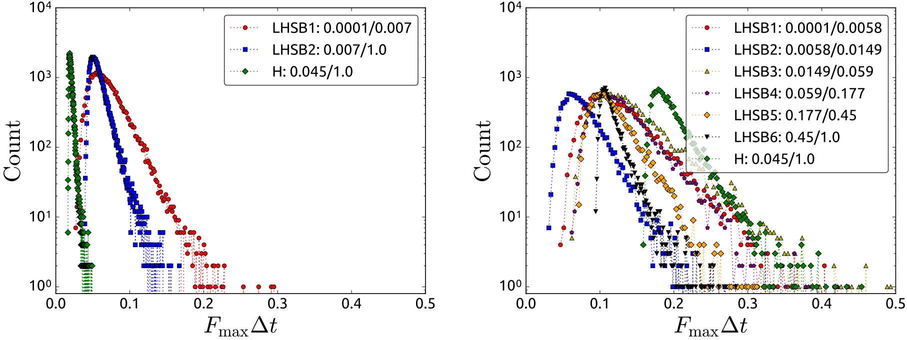


Figure: Left: minimum number of poles in the rational approximation to $\mathcal{M}_{\rm EOFA}^{-1/2}$ needed to achieve a relative error $\varepsilon < 10^{-10}$ for the lightest Hasenbusch ratio. Right: total light quark heatbath time including all Hasenbusch ratios.

32ID-G Evolution Tuning: Force Distributions

Strategy: Empirically observe that tails of force distributions control acceptance, so add intermediate Hasenbusch masses until light and heavy quark force distributions have comparable tails.



(a) RHMC Ensemble (b) EOFA Ensemble

Figure: Histograms of the maximum force, $F_{\max} \equiv \max_{x,\mu} \|F_\mu(x)\|$, measured on 850 trajectories of the 32ID-G RHMC and EOFA ensembles. "LHSB" denotes the various mass ratios entering into our mass preconditioning scheme for the light quark determinant, and "H" denotes the strange quark determinant.

32ID-G Evolution Benchmark

Light Quark Action	Integrator	Light Hasenbusch Masses	Δt	r_{MD}	N_{traj}	Acceptance	Efficiency
RHMC	Omelyan	0.007	0.0625	10^{-8}	850	88%	—
EOFA	FG	0.0058, 0.0149, 0.059, 0.177, 0.45	0.1667	10^{-7}	850	93%	4.2

Table: Fully tuned RHMC and EOFA schemes. Δt is the outermost time step, r_{MD} is the MD CG tolerance, and **efficiency is the speed-up in the total job time relative to the RHMC scheme**.

RHMC

- Omelyan integrator ($\lambda = 0.22$)
- One light quark Hasenbusch mass
- Multishift CG with single precision $\not{D}$ but accumulating solution and search vectors in double precision, coupled with reliable update to correct residual
- Even-odd preconditioning

EOFA

- Force gradient integrator
- Five light quark Hasenbusch masses
- Mixed precision defect correction CG
- Even-odd preconditioning
- Cayley preconditioning
- Force gradient forecasting
- Heatbath forecasting
- Heatbath CG tolerance tuning

Figure: Comparison of RHMC and EOFA integration schemes.

Conclusions

- We have introduced a new preconditioner for EOFA which relates inversions of (dense) D_{EOFA} to significantly cheaper inversions of (tridiagonal) D_{DWF}
- We have also explored the performance of standard lattice preconditioners — even-odd, Hasenbusch — and additional tuning and forecasting techniques
  - EOFA appears much more amenable to Hasenbusch preconditioning than RHMC
  - No multishift CG in EOFA calculations means a larger set of tricks can be used (implicitly restarted solvers, forecasted CG initial guesses, ...)
- We find that physical quark mass Möbius DWF simulations can be substantially cheaper with EOFA compared to RHMC
  - 24ID ensemble: $\sim 20\%$ speed-up in total job time per trajectory
  - 32ID-G ensemble: $4.2\times$ speed-up in total job time per trajectory
- Ongoing $\Delta I = 1/2$ $K \rightarrow \pi\pi$ ensemble generation is now being performed with EOFA evolution code [see Chris Kelly's talk, Wed. @ 13:10, for the physics!]
- Where else might EOFA be useful?
  - $N_f = 2+1+1$ simulations with dynamical (EOFA) strange and charm quarks
  - Simulations with physical, isospin-broken light quarks (hadron spectra, $K_{\ell 3}$, ...)
  - Simulations with $SU(3)$-symmetric light quarks (χPT studies, thermodynamics, ...)

Thank you!

Backup Slides

Refinement II: Heatbath Tuning

- EOFA heatbath: draw Gaussian field η and seed $\phi = \mathcal{M}_{\mathrm{EOFA}}^{-1/2} \eta$
- Can show

$$\mathcal{M}_{\mathrm{EOFA}}^{-1/2} \simeq \alpha_0 + \sum_{k=1}^{N_p} \alpha_l \gamma_l \Big\{ \mathbb{1} + k \gamma_l P_- \Omega_-^\dagger \left[H(m_1) - \gamma_l \Delta_-(m_1, m_2) P_- \right]^{-1} \Omega_- P_- \\ - k \gamma_l P_+ \Omega_+^\dagger \left[H(m_2) - \beta_l \gamma_l \Delta_+(m_1, m_2) P_+ \right]^{-1} \Omega_+ P_+ \Big\}$$

  for a rational approximation $x^{-1/2} \simeq \alpha_0 + \sum_l \alpha_l / (\beta_l + x)$, with $\gamma_l \equiv (1 + \beta_l)^{-1}$
- Requires $2N_p$ independent CG solves, since $\Delta_\pm$ operators are singular $\longrightarrow$ expensive!
- Cost can be partially ameliorated by forecasting initial guesses across poles [TWQCD, arXiv:1403.1683; Murphy, arXiv:1611.00298]
- Further reduction possible by observing that coefficients $\{\alpha_0, \alpha_l \gamma_l, k \alpha_l \gamma_l^2\}$ generally exhibit large separation of scales, suggesting loose CG tolerances can be used for inversions multiplied by small coefficients

Refinement II: Heatbath Tuning

We propose tuning the heatbath CG tolerances by considering the quantity

$$\varepsilon \equiv \frac{\left|\eta^{\dagger}\eta - \phi^{\dagger}\mathcal{M}_{\text{EOFA}}\phi\right|}{\eta^{\dagger}\eta} \qquad \left(\phi = \mathcal{M}_{\text{EOFA}}^{-1/2}\eta\right)$$

and using the following simple algorithm:

1. Choose a desired overall tolerance for the heatbath, ε_{tol}
2. Choose one of the inversions required to compute ϕ, and relax the stopping tolerance until the overall error in the heatbath (ε) reaches ε_{tol}
3. Iterate over each inversion until all stopping conditions have been tuned

Ensemble		ε	Total Heatbath Time	**Speed-Up**
24ID	Untuned	1.52×10^{-11}	129.5 s	**1.9**
	Tuned	7.79×10^{-11}	68.9 s	
32ID-G	Untuned	1.11×10^{-11}	2289.0 s	**2.7**
	Tuned	7.69×10^{-11}	840.1 s	

Table: Relative error (ε) and total running time for the EOFA heatbath before and after tuning.

24ID and 32ID-G Integrator Layouts

Ensemble	Level	Action	Update
24ID (RHMC)	1	$\mathrm{RatQuo}_{1/2}(0.085, 1.0)$	1:1
	2	$\mathrm{Quo}(0.00107, 0.00789) + \mathrm{Quo}(0.00789, 0.0291) + \mathrm{Quo}(0.0291, 0.095) + \mathrm{Quo}(0.095, 0.3) + \mathrm{Quo}(0.3, 0.548) + \mathrm{Quo}(0.548, 1.0)$	1:1
	3	Gauge + DSDR	1:1
24ID (EOFA)	1	$\mathrm{EOFA}(0.085, 1.0)$	1:1
	2	$\mathrm{Quo}(0.00107, 0.00789) + \mathrm{Quo}(0.00789, 0.0291) + \mathrm{Quo}(0.0291, 0.095) + \mathrm{Quo}(0.095, 0.3) + \mathrm{Quo}(0.3, 0.548) + \mathrm{Quo}(0.548, 1.0)$	1:1
	3	Gauge + DSDR	1:1
32ID-G (RHMC)	1	$\mathrm{RatQuo}_{1/2}(0.0001, 0.007)$	1:1
	2	$\mathrm{RatQuo}_{1/2}(0.007, 1.0) + \mathrm{RatQuo}_{1/4}(0.045, 1.0)$	1:2
	3	DSDR	1:2
	4	Gauge	1:1
32ID-G (EOFA)	1	$\mathrm{EOFA}(0.0001, 0.0058) + \mathrm{EOFA}(0.0058, 0.0149) + \mathrm{EOFA}(0.0149, 0.059) + \mathrm{EOFA}(0.059, 0.177) + \mathrm{EOFA}(0.177, 0.45) + \mathrm{EOFA}(0.45, 1.0) + \mathrm{RatQuo}_{1/4}(0.045, 1.0)$	5:1
	2	DSDR	1:2
	3	Gauge	1:1

Table: Integrator layouts for the 24ID and 32ID-G ensembles. The notation A:B for the update scheme denotes the number of steps of the next innermost integrator level (A) per step of the current level (B).