Generating QCD Gauge Configurations

The Fourth International Workshop on Numerical Analysis and Lattice QCD

Yale University

Michael Clark
Boston University

Talk Outline

- Introduction
- Hybrid Monte Carlo
- Algorithm Improvement
- “Non-local” Algorithms
- Conclusion

Introduction

- Lattice QCD path integral

$$\langle\Omega\rangle = \frac{1}{Z}\int [dU] e^{-S_{\mathrm{g}}(U)} [\det \mathcal{M}(U)]^{\alpha}\Omega(U)$$

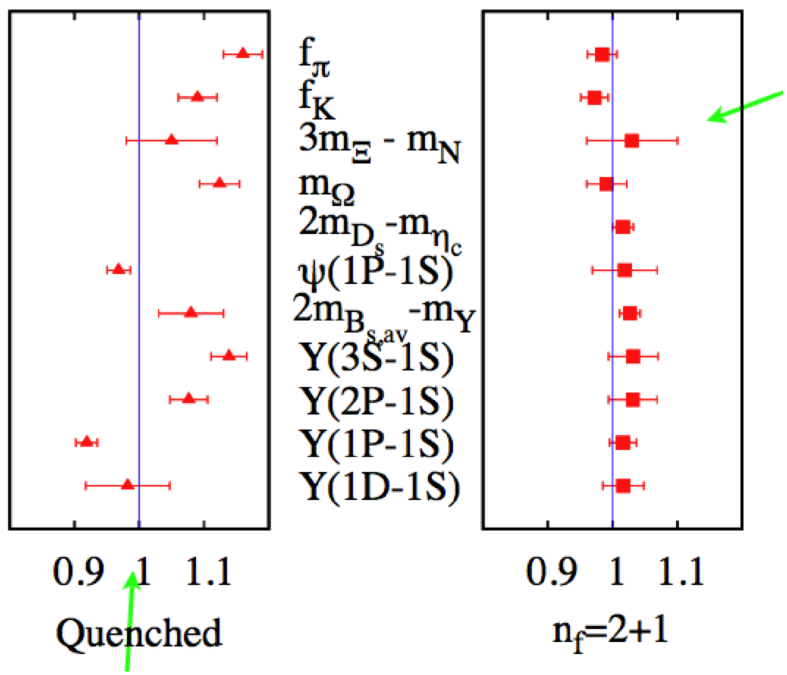
$\alpha = \frac{N_f}{2}$ $(\frac{N_f}{4})$ for Wilson (staggered) fermions, $\mathcal{M} = M^{\dagger}M$

- $10^8 - 10^9$ degrees of freedom $\Rightarrow$ Monte Carlo integration
- Interpret $e^{-S_{\mathrm{g}}} \det \mathcal{M}^{\alpha}$ as a Boltzmann weight, and use importance sampling

$$\langle\Omega\rangle \approx \frac{1}{N}\sum_{i=1}^{N}\Omega(U_i)$$

Quenched Approximation

- Fermion determinant extremely non-local object
- Quenched approximation: set det $\mathcal{M} = 1$
- Gauge action local: over relaxed heatbath algorithms very efficient
- Just plain wrong!



$\Rightarrow$ Must include dynamical fermions to obtain QCD

The HMC Algorithm (Duane *et al*)

- *De facto* algorithm for including dynamical fermions
- Rewrite determinant in terms of pseudo-fermions

$$\det \mathcal{M} = \int D\phi^\dagger D\phi \, e^{-\phi^\dagger \mathcal{M}^{-1} \phi} = \int D\phi^\dagger D\phi \, e^{-S_\mathrm{f}}$$

- Need global updates since psuedofermion action is non-local
- Introduce fictitious momentum field π and define a Hamiltonian

$$H = \frac{1}{2} \mathrm{tr}\, \pi^2 + S_\mathrm{g} + S_\mathrm{f} = T + S$$

- Integrate Hamilton's equations to propose a new configuration
- Global Accept / Reject to obtain desired probability distribution

$$P(U, \phi) = \frac{1}{Z} e^{-S_\mathrm{g} - S_\mathrm{f}}$$

The HMC Algorithm (Duane *et al*)

- Each update consists of
  - Hybrid Molecular Dynamics Trajectory
    - Momentum refreshment heatbath ($P(\pi) \propto e^{-\pi^*\pi/2}$).
    - Pseudo-fermion heatbath ($\phi \propto M^\dagger \xi$, where $P(\xi) \propto e^{-\xi^*\xi}$).
    - MD trajectory with $\tau/\delta\tau$ steps.
  - Metropolis Acceptance Test $P_{\text{acc}} = \min(1, e^{-\delta H})$

Molecular Dynamics

- Hamilton's equations $\frac{dU}{d\tau} = \frac{dT}{d\pi} = \pi$ and $\frac{d\pi}{d\tau} = -\frac{dS}{dU} = F$
- Must discretize the "fictious time" τ and integrate numerically
- Define integrators in terms of evolution operators Q and P

$$Q \equiv \frac{dT}{d\pi}\frac{\partial}{\partial U} \quad \text{with} \quad e^{\delta\tau Q} : f(U,\pi) \to f(U + \delta\tau T'(\pi), \pi)$$

$$P \equiv -\frac{dS}{dU}\frac{\partial}{\partial \pi} \quad \text{with} \quad e^{\delta\tau P} : f(U,\pi) \to f(U, \pi - \delta\tau S'(U))$$

- Metropolis requires Detailed Balance
  - Integration must be reversible and area preserving
  - Use Symmetric Symplectic Integrators, e.g., leapfrog

$$U(\delta\tau)^{\tau/dt} = \left(e^{\delta\tau\, P/2} e^{\delta\tau\, Q} e^{\delta\tau\, P/2}\right)^{\tau/\delta\tau} + O(\delta\tau^2)$$

Molecular Dynamics Forces

- Pure Gauge Force ($\frac{dS_{\mathsf{g}}}{dU}$) local analytic quantity
  - CHEAP
- Fermion force

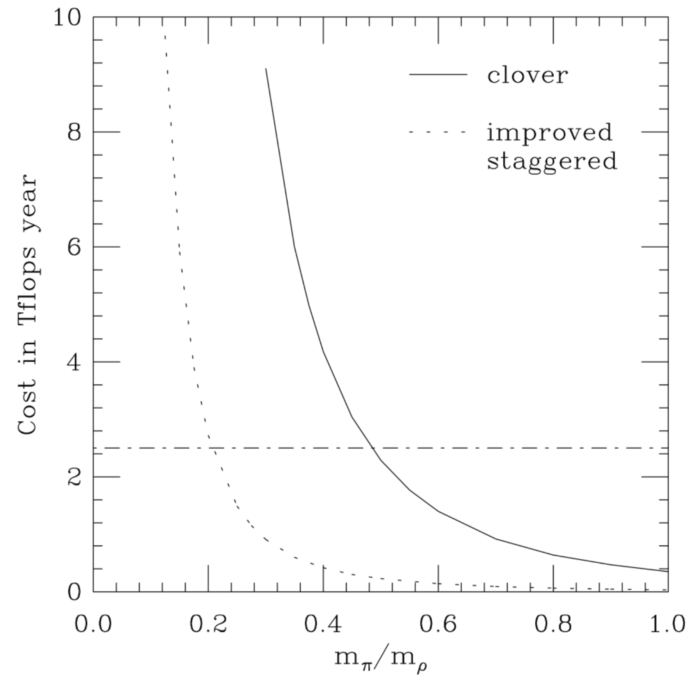
$$\frac{dS_{\mathsf{f}}}{dU} = \frac{d}{dU}\phi^{\dagger}\mathcal{M}^{-1}\phi = -\phi^{\dagger}\mathcal{M}^{-1}\frac{d\mathcal{M}}{dU}\mathcal{M}^{-1}\phi$$

  - Each update to the momentum requires solution to $\mathcal{M}\chi = \phi$
  - Generally calculated using a Krylov solver, e.g., CG
  - EXPENSIVE

Cost of HMC

- Condition number blows up as $m \to 0$
- Force $\propto 1/m$, requires $\delta\tau \to 0$ to maintain acceptance rate
- Also, as $m \to 0$, correlation lengths diverge
- $C \propto \left(\frac{m_\pi}{m_\rho}\right)^{-6} L^5 a^{-7}$

  CP-PACS and JLQCD, 2002



$\Rightarrow$ Require huge computers OR better algorithms

Chronological Inverter

- Solution $\chi(\tau) = \mathcal{M}(\tau)^{-1}\phi$ is a smooth function
- Idea: Use previous solutions to act as an initial guess
- Minimize over the space of previous solutions (Broweret *al*) :

  for

$$x_0 = \sum_i c_i \mathcal{M} \chi_i$$

  solve

$$\chi_j^\dagger \phi = \sum_i c_i \chi_j^\dagger \mathcal{M} \chi_i$$

- Requires high precision solutions to maintain reversibility
- Gain around a factor of 2

Higher Order Integrator

- Potential for gain from using $O(\delta\tau^4)$ integrator, e.g., Campostrini

$$U(\delta\tau)^{\tau/dt} = \left(e^{\delta\tau\,\epsilon P/2} e^{\delta\tau\,\epsilon Q} e^{\delta\tau\,(1-\sigma)P/2} e^{-\delta\tau\,\epsilon\sigma Q}\right. \\ \left. e^{\delta\tau\,(1-\sigma)P/2} e^{\delta\tau\,\epsilon Q} e^{\delta\tau\,\epsilon P/2}\right)^{\tau/\delta\tau} + O(\delta\tau^4)$$

- Better volume scaling $V^{9/8}$ vs. $V^{5/4}$
- Constructed from sub-leapfrog steps with $\delta\tau^{sub} > \delta\tau$
- Sub-leapfrog integrator can go unstable much sooner than $\delta\tau$ suggests (Joó *et al*)
- Higher order integrators are very unstable
- $O(\delta\tau^2)$ usually found to be optimal

Multiple Timescale Integration (Sexton/Weingarten, 1992)

- For Hamiltonians of the form $H = T + S_1 + S_2$
- Integrate S_1 and S_2 force contributions on different timescales

$$U(\delta\tau)^{\tau/\delta\tau} = \left(\left(e^{\delta\tau\, P_1/4m} e^{\delta\tau\, Q/2m} e^{\delta\tau\, P_1/4m}\right)^m e^{\delta\tau P_2} \left(e^{\delta\tau\, P_1/4m} e^{\delta\tau\, Q/2m} e^{\delta\tau\, P_1/4m}\right)^m\right)^{\tau/\delta\tau}$$

- Two separate timescales $\delta\tau^{S_1} = \delta\tau/m$, $\delta\tau^{S_2} = \delta\tau$
- Large and cheap force $= P_1$, Small and expensive force $= P_2$,
- Naïve partitioning: $S_1 = S_{\mathrm{g}}, S_2 = S_{\mathrm{f}}$ fails as $m \to 0$
- Can extend this recursively for N timescales
  $H = T + S_1 + \ldots + S_N$
- Great idea, but dormant for 10 years...

Multiple Pseudofermions with Mass Preconditioning

- Mass-precondition the fermion determinant (Hasenbusch)

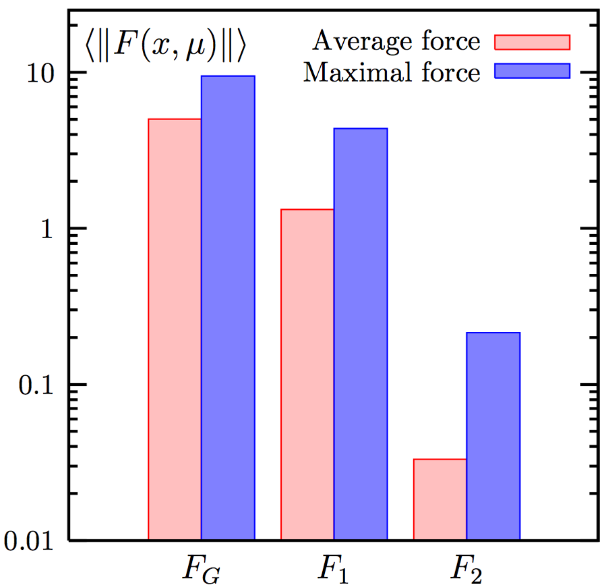
$$\det(M^\dagger M) = \det(\widehat{M}^\dagger \widehat{M}) \det\left(\widehat{M}(M^\dagger M)^{-1}\widehat{M}^\dagger\right)$$

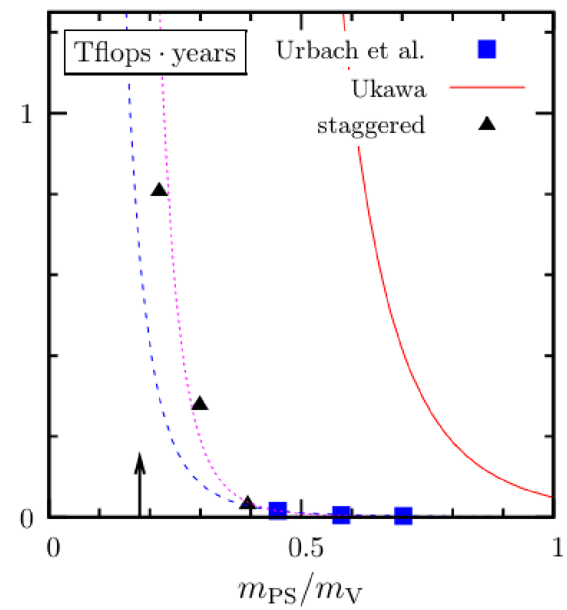
with $m(\widehat{M}) > m(M)$

- Tune $\kappa(\widehat{M}^\dagger\widehat{M}) \approx \kappa(\widehat{M}(M^\dagger M)^{-1}M^\dagger) \approx \sqrt{\kappa(M^\dagger M)}$ (Hasenbusch-Jansen)
- Why does this work?
  - Better sampling of Gaussion integral using multiple pseudo-fermions
  - This reduces fluctuations in the fermion force
  - Fermion force $F \propto \kappa^\nu$
- Factor 2 improvement through $\delta\tau$ increase
- Use more than one dummy operator
- Gain increases as $m_l \to 0$

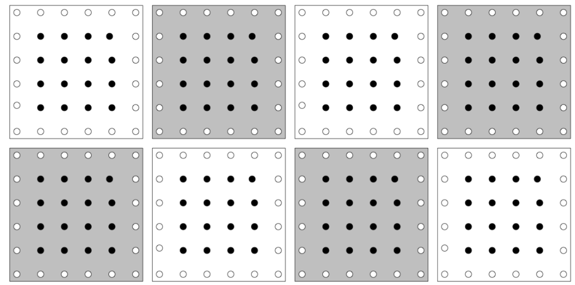
Multi-timescale Mass Preconditioning (QCDSF, Urbach *et al*)

- Tune dummy operators so that most expensive (preconditioned) force contributes the least
- Use multi-timescale integrator ($S_1 = S_\mathrm{g}$, $S_2 = \hat{S}$, $S_3 = S_\mathrm{f}$)
- Factor 10 improvement at light quark mass





Domain Decomposition (Lüscher)



- Can rewrite the Dirac operator $M = \begin{pmatrix} D_{\Omega} & D_{\delta\Omega} \\ D_{\delta\Omega'} & D_{\Omega'} \end{pmatrix}$
- D_{Ω} ($D_{\Omega'}$) is Dirac operator on black (white) blocks with Direchlet boundaries
- $D_{\delta\Omega}$ and $D_{\delta\Omega'}$ is the Dirac operator connecting these blocks
- Rewrite determinant

$$\det M = \det \begin{pmatrix} D_{\Omega} & 0 \\ 0 & D_{\Omega'} \end{pmatrix} \det \begin{pmatrix} D_{\Omega} & D_{\delta\Omega} \\ D_{\delta\Omega'} & D_{\Omega'} \end{pmatrix} \begin{pmatrix} D_{\Omega}^{-1} & 0 \\ 0 & D_{\Omega'}^{-1} \end{pmatrix}$$

- Separation into large and cheap, and small but expensive $\Rightarrow$ multi timescale integrator
- Large speed up over naïve algorithm

Determinant Preconditioning is Key

- Other methods
  - U.V. Filtering (de Forcrand)
  - Polynomial filtering (Peardon and Sexton)
  - Multistep stochastic correction (see talk by Enno)
- All improvements rely on determinant preconditioning

Non-local Actions

- Strange quark inclusion requires $\det \mathcal{M}^{\frac{1}{2}}$
- Finite temperature calculations typically use staggered quarks
  - Remnant chiral symmetry important here
  - Non-local action: $\det \mathcal{M}^{\alpha}$, $\alpha = \frac{1}{2}, \frac{1}{4}$
- HMC cannot be applied for these cases
- Inexact algorithms traditionally used
- Is using exact algorithms more expensive?
- Are the improvements in "Local actions" applicable here?

The R Algorithm (Gottlieb *et al*)

- Rewrite fermionic determinant:

$$\det \mathcal{M}^{\alpha} = \exp\left(\alpha \,\mathrm{tr} \ln \mathcal{M}\right) = \exp\left(-S_{\mathrm{eff}}\right)$$

- Integrate Hamilton's equations as before
- Use noisy estimator ξ for trace $\equiv$ pseudo-fermion force
- Leading order error term $\mathrm{O}(\delta\tau)$!
- Recover $\mathrm{O}(\delta\tau^2)$ with N_{f} dependent ξ updating
  - Non-reversible
  - Jacobian $\neq 1$
- Cannot include Metropolis acceptance test

  $\Rightarrow$ Algorithm is inexact
- Naïve cost = HMC, but requires extrapolation to zero stepsize
- Stepsize rule of thumb $\delta\tau \sim \frac{2}{3} m_{\mathrm{l}}$

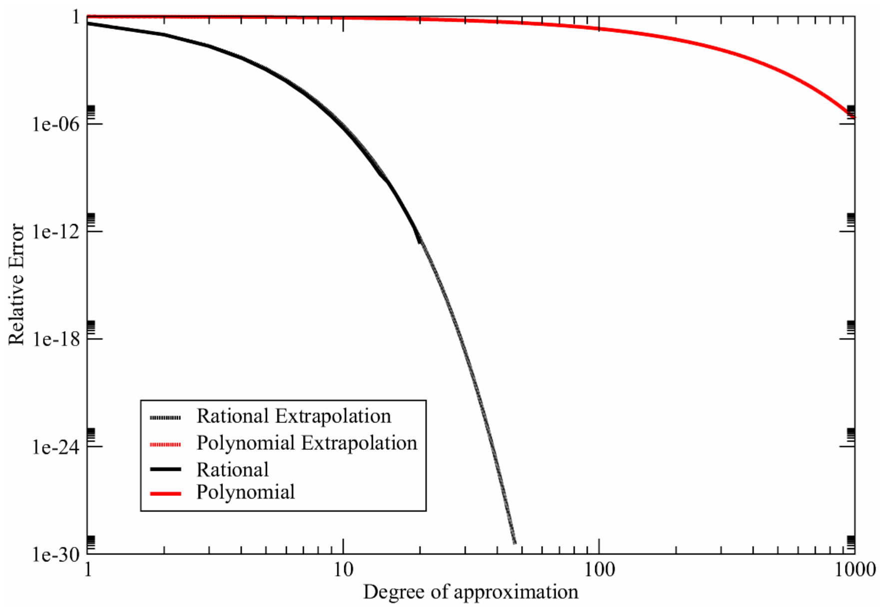
Polynomial Hybrid Monte Carlo (de Forcrand-Takaishi, Frezzotti-Jansen)

- Write in pseudo-fermion notation

$$\det \mathcal{M}^\alpha = \int \mathcal{D}\bar\psi \mathcal{D}\psi e^{-\bar\psi \mathcal{M}^{-\alpha}\psi}$$
$$\approx \int \mathcal{D}\bar\psi \mathcal{D}\psi e^{-\bar\psi P(\mathcal{M})\psi},$$

  where $P(\mathcal{M})$ is valid over spectrum
- Pseudo-fermion heatbath easily realised $P(\mathcal{M}) = p^\dagger(\mathcal{M})p(\mathcal{M})$
- Use standard MD leapfrog $\Rightarrow$ exact
- Generally polynomial degree $m > N_{\text{iter}}$ CG iterations
- Use low degree polynomial
  - Reweight acceptance test or observable
- Use high degree polynomial
  - $\alpha \neq 1$ derivative uses Leibniz rule $\Rightarrow$ Memory, rounding

Optimal rational approximations



- Generated using Remez algorithm
- Real non-degenerate roots (poles are always +ve)
- Partial fractions - $r(x) = \sum_{k=1}^{m} \frac{\alpha_k}{x+\beta_k}$
- Evaluate using multi-shift solver
- Numerically stable (α_k have same sign)

Rational Hybrid Monte Carlo (Clark-Kennedy)

- Rewrite fermionic determinant

$$\det \mathcal{M}^{\alpha} = \int \mathcal{D}\bar{\phi}\mathcal{D}\phi e^{-\bar{\phi}\mathcal{M}^{-\alpha}\phi} \approx \int \mathcal{D}\bar{\phi}\mathcal{D}\phi e^{-\bar{\phi}r^2(\mathcal{M})\phi},$$

  with $r(x) = x^{-\alpha/2}$
- Precision is cheap: Conventional Metropolis
- RHMC:
  - Hybrid Molecular Dynamics Trajectory
    * Momentum refreshment heatbath ($P(\pi) \propto e^{-\pi^*\pi/2}$).
    * Pseudo-fermion heatbath ($\phi \propto r(\mathcal{M})^{-1}\xi$, where $P(\xi) \propto e^{-\xi^*\xi}$).
    * MD trajectory with $\tau/\delta\tau$ steps.
  - Metropolis Acceptance Test $P_{\text{acc}} = \min(1, e^{-\delta H})$

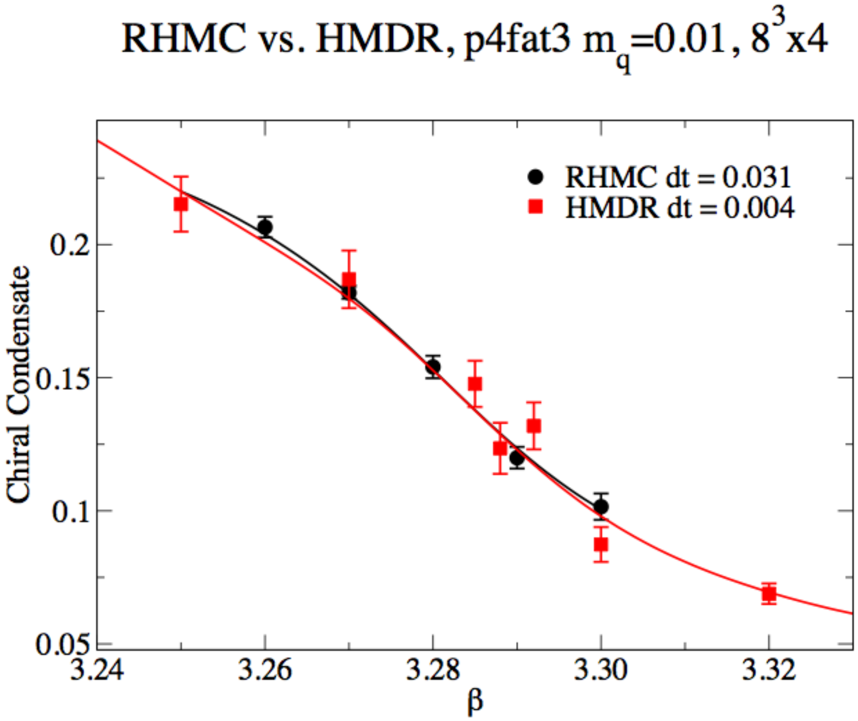
Rational Hybrid Monte Carlo (Clark-Kennedy)

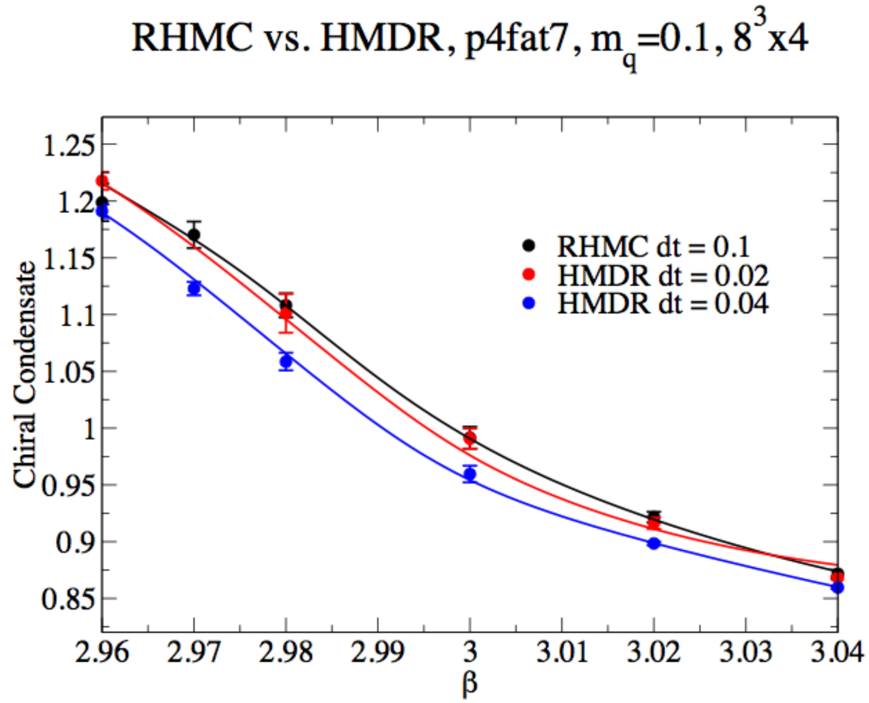
- MD trajectory
  - Double inversion from $r^2(\mathcal{M})$
  - Use low degree approx $\bar{r} \approx \mathcal{M}^{-\alpha} \approx r^2$
  - Pseudo-fermion force

$$S'_{\mathsf{pf}} = -\sum_{i=1}^{\bar{m}} \bar{\alpha}_i \phi^\dagger (\mathcal{M} + \bar{\beta}_i)^{-1} \mathcal{M}' (\mathcal{M} + \bar{\beta}_i)^{-1} \phi.$$

- CG cost per trajectory $\approx$ HMC
  - One extra inversion required for heatbath

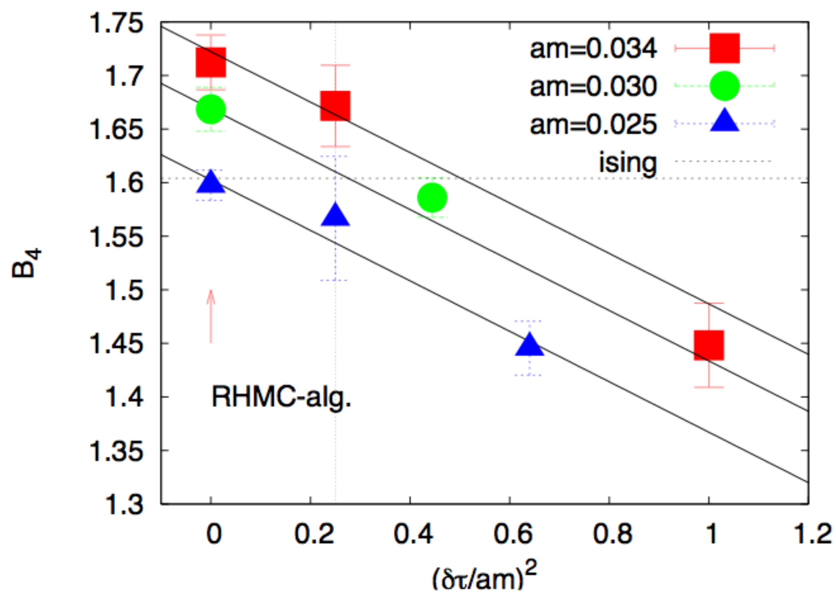
Exact vs. Inexact (RBC-Bielefeld)

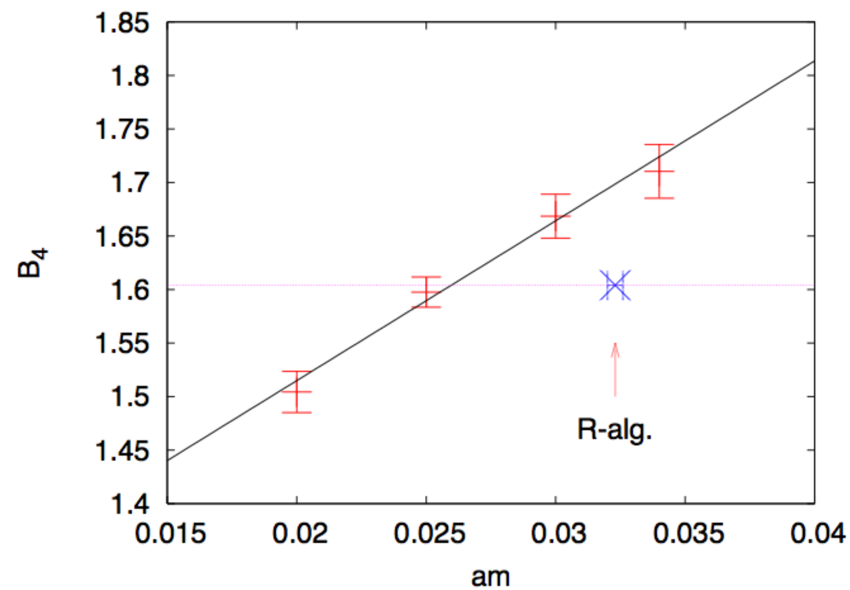




- P4 staggered fermions
- RHMC allows an $O(10)$ increase in stepsize
- Speedup greater as $m_l \to 0$

Exact vs. Inexact (de Forcrand-Philipsen)





(Naïve Staggered Fermions, $N_{\mathrm{f}} = 3$, $V = 8^3 4$)

- Results:
  - Binder cumulant increases
  - Stepsize extrapolation is vital for R algorithm
  - 25% reduction in critical quark mass at $\delta\tau^R = \frac{1}{2} m_{\mathrm{l}}$
  - 20% change in renormalised quark mass
- Conclusion: "an exact algorithm is mandatory"

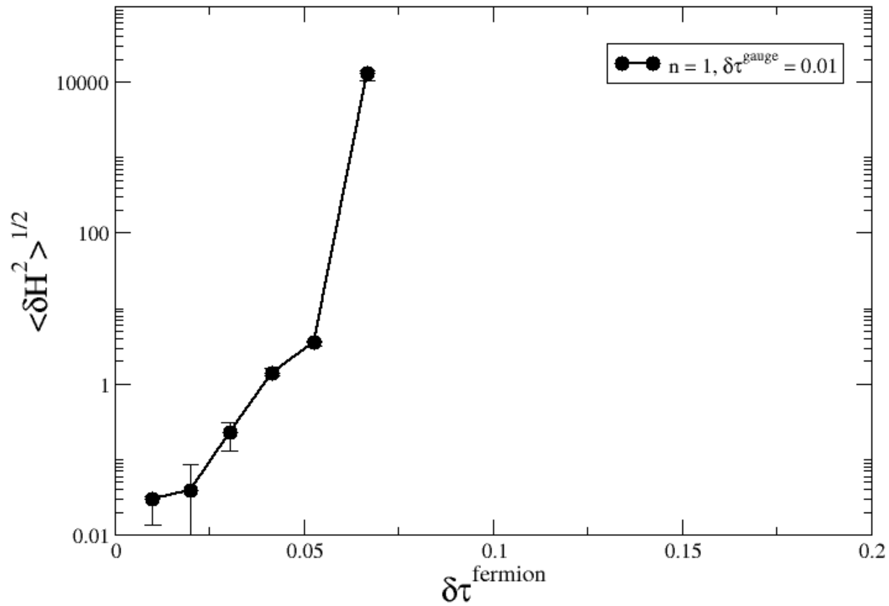
Multiple Pseudofermions with RHMC (Clark-Kennedy)

- Rewrite determinant

$$\det \mathcal{M} = [\det \mathcal{M}^{1/n}]^n$$
$$\propto \prod_{j=1}^{n} d\phi_j \, d\phi_j^\dagger \exp\left(-\phi_j^\dagger \mathcal{M}^{-1/n} \phi_j\right),$$

- So called n^{th} root trick
- Speedup through $\delta\tau$ increase
- No dummy mass parameters to tune $\Rightarrow$ easy to increase n
- Single fermion timescale

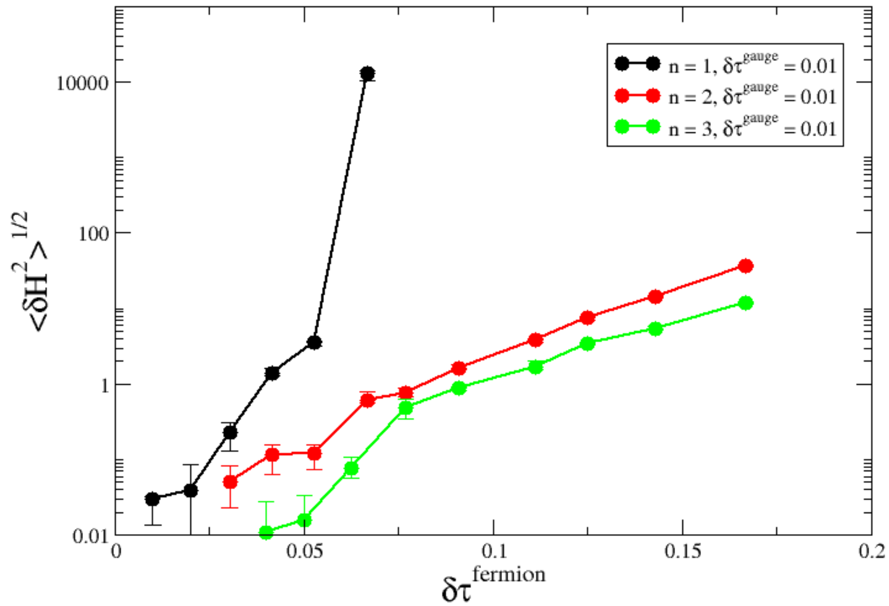
Integrator Instability (Clark-Kennedy)



(Staggered fermions, $V = 16^4$, $\beta = 5.6$, $N_f = 2$, $m = 0.005$)

- With $n = 1$ integrator breaks down as $\delta\tau$ is increased
- Instability "tickled" by low fermion modes $\sim O(\frac{1}{m_l})$ (Joó *et al*)
- Higher order integrators are more ticklish
- What happens with multiple pseudofermions?

Integrator Instability (Clark-Kennedy)



(Staggered fermions, $V = 16^4$, $\beta = 5.6$, $N_{\mathsf{f}} = 2$, $m = 0.005$)

- Removes instability in the integrator!
- Why does this work? Lowest modes now $O(\frac{1}{m_{\mathsf{l}}})^{1/n}$
- Force now bulk dominated
- Higher order integrators now beneficial

Who's the fastest of them all? ($N_f = 2$ Wilson)

- Compare multi-timescale mass preconditioning and high order RHMC
- Use popular testing parameters ($V = 24^3.32$, $\beta = 5.6$, Plaquette + Wilson fermions)
- Use measure $C = \tau_{\text{int}}^{\text{plaq}}.N_{\text{mv}}.10^{-4}$

	C		
κ	RHMC	Urbach *et al*	Orth *et al*
0.15750	9.6	9.0	19.1
0.15800	29.9*	17.4	128
0.15825	52.5*	56.5	-

*Using 4MN5 fourth order integrator (de Forcrand and Takaishi)

- RHMC similar in cost to mass preconditioning
- Look at integrator stability with mass preconditioning
  - Gain from higher order integrators also?

2+1 QCD

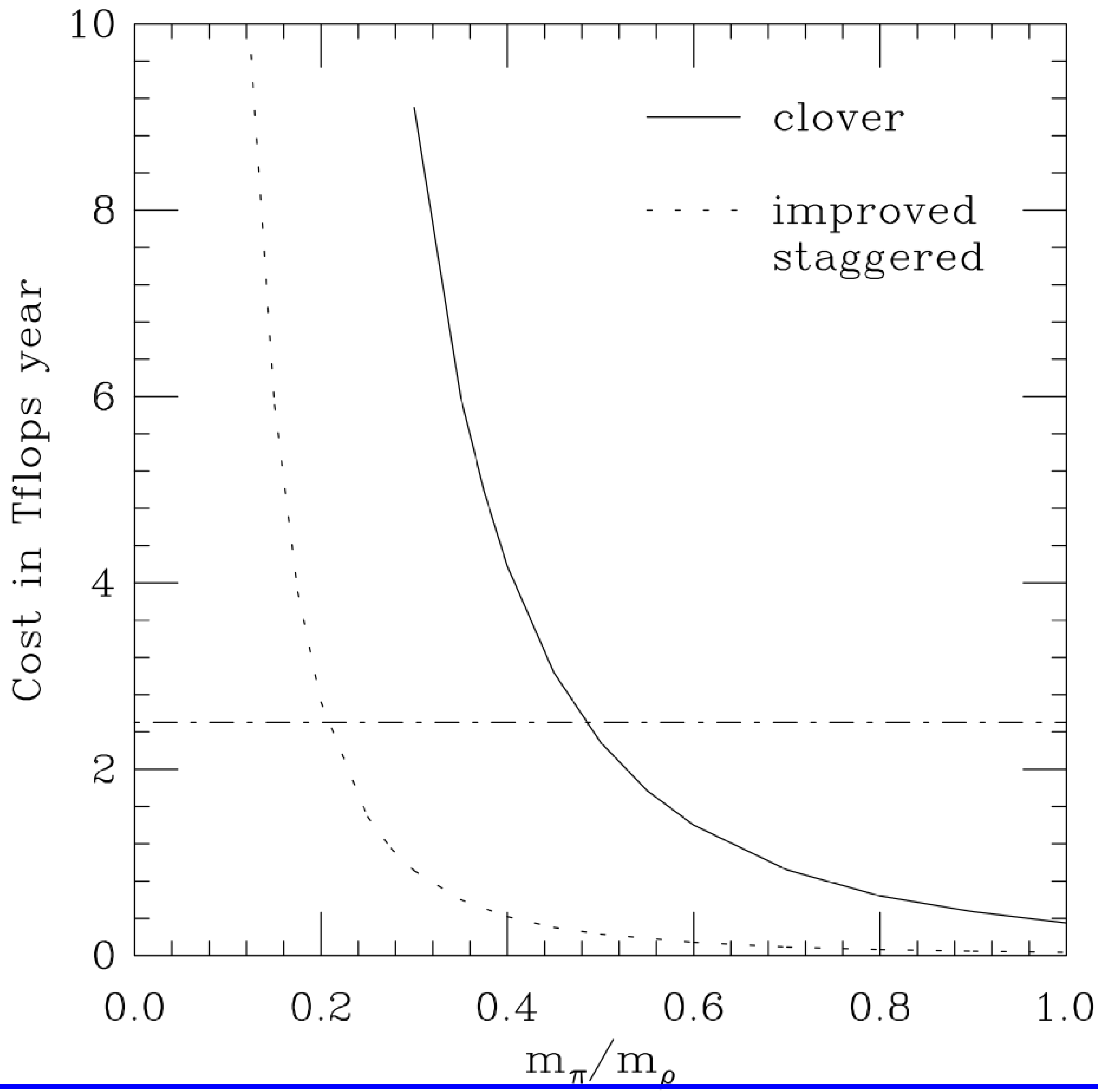
- e.g. Domain Wall 2+1 flavour determinant

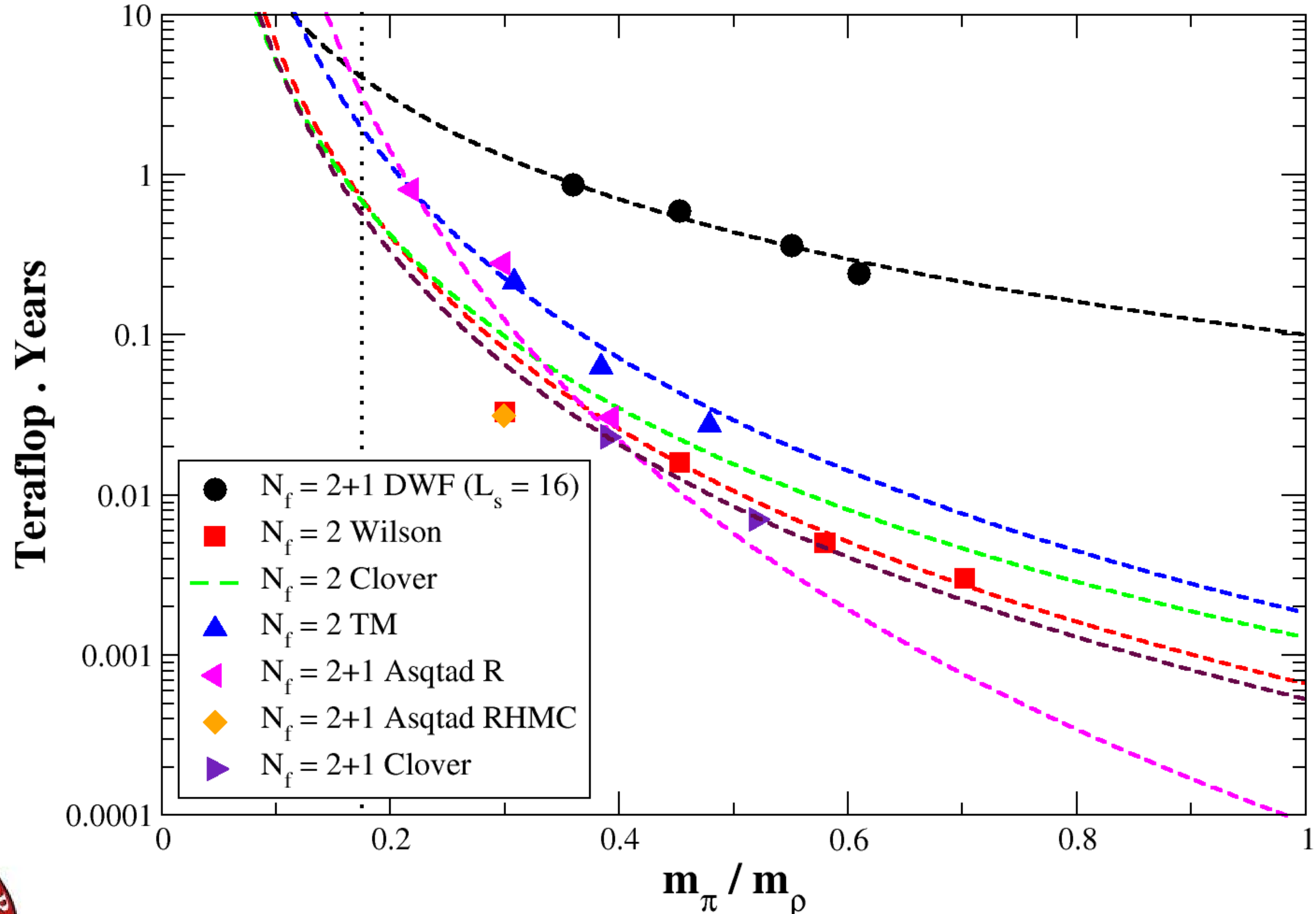
$$\left(\frac{\det M_l^\dagger M_l}{\det M_{pv}^\dagger M_{pv}}\right)\left(\frac{\det M_s^\dagger M_s}{\det M_{pv}^\dagger M_{pv}}\right)^{1/2} = \left(\frac{\det M_l^\dagger M_l}{\det M_s^\dagger M_s}\right)\left(\frac{\det M_s^\dagger M_s}{\det M_{pv}^\dagger M_{pv}}\right)^{3/2}$$

- Mass Precondition using the strange quark
- Use n^{th} root trick for triple strange
- NOT mutually exclusive improvements
- Use multi-timescale integrator (gauge, triple strange, light)
- Light quark mass constitutes around 10% CG cost
- Cost dependence on mass comes mostly from autocorrelation

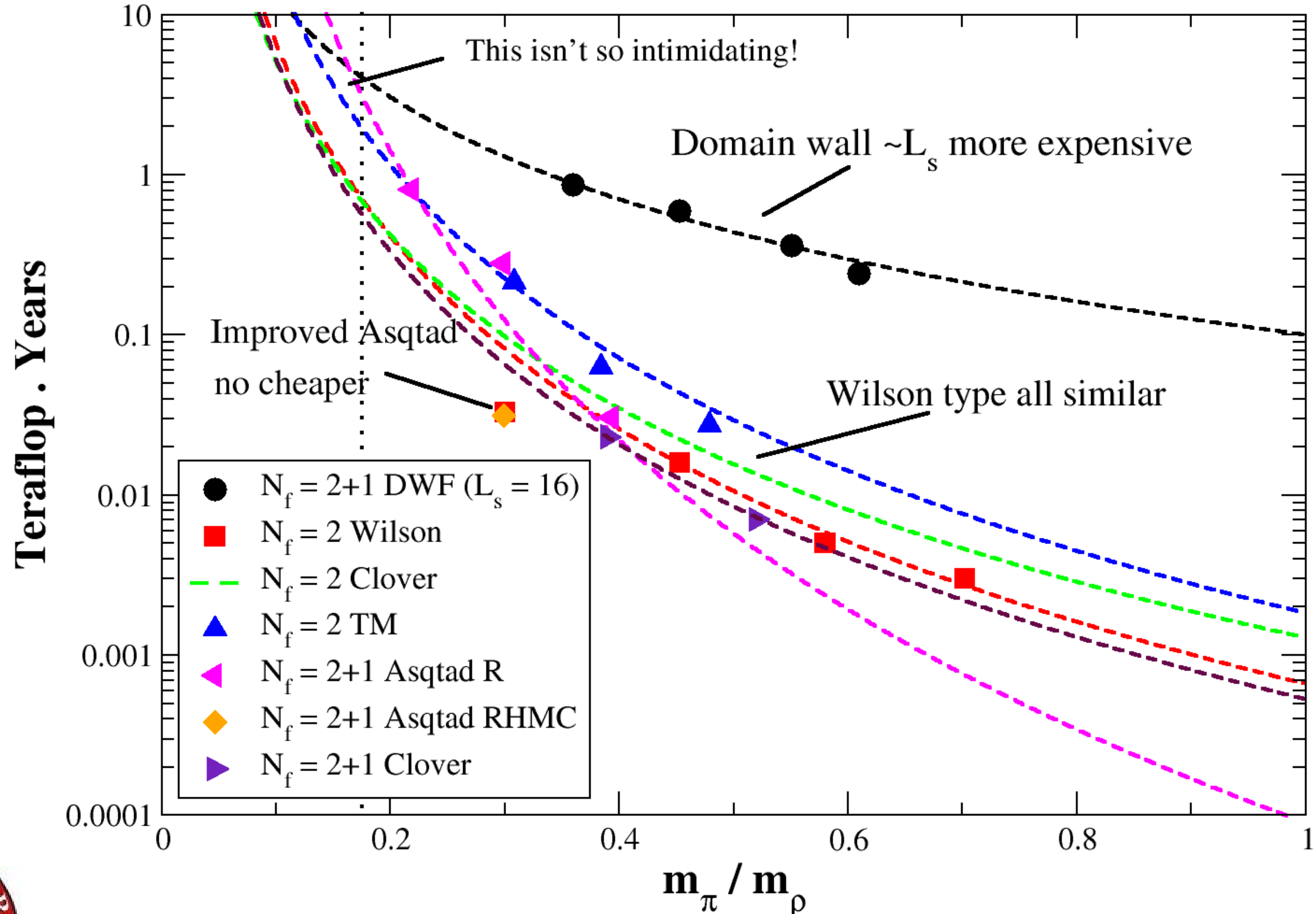
Berlin Wall Plot

- Compare the cost of fermion formulations and/or algorithms
  - $N_f = 2 + 1$ DWF RHMC (RBC-UKQCD)
  - $N_f = 2$ Mass preconditioned Wilson (Urbach *et al*)
  - $N_f = 2$ Mass preconditioned Clover (QCDSF)
  - $N_f = 2 + 1$ Mass preconditioned Clover + RHMC (Wuppertal-Jülich)
  - $N_f = 2$ Mass preconditioned Twisted Mass (ETM)
  - $N_f = 2 + 1$ AsqTad R (MILC)
  - $N_f = 2 + 1$ AsqTad RHMC (Clark-Kennedy)
- All data scaled to $V = 24^3 \times 40$, $a = 0.08$
- Box is too small and too coarse


10
8
6
4
2
0
0.0
0.2
0.4
0.6
0.8
1.0
Cost in Tflops year
m_π/m_ρ
clover
improved staggered



10
1
0.1
0.01
0.001
0.0001
0
0.2
0.4
0.6
0.8
1
Teraflop . Years
m_π / m_ρ
N_f = 2+1 DWF (L_s = 16)
N_f = 2 Wilson
N_f = 2 Clover
N_f = 2 TM
N_f = 2+1 Asqtad R
N_f = 2+1 Asqtad RHMC
N_f = 2+1 Clover


This isn't so intimidating!
Domain wall ~L_s more expensive
Improved Asqtad no cheaper
Wilson type all similar
Teraflop . Years
m_π / m_ρ
10
1
0.1
0.01
0.001
0.0001
0
0.2
0.4
0.6
0.8
1
N_f = 2+1 DWF (L_s = 16)
N_f = 2 Wilson
N_f = 2 Clover
N_f = 2 TM
N_f = 2+1 Asqtad R
N_f = 2+1 Asqtad RHMC
N_f = 2+1 Clover

Conclusions

- After 20 years HMC is still the best dynamical algorithm
- The last 5 years has seen an explosion in HMC improvement
- Determinant preconditioning is the key behind all improvement
- Non-local actions are no problem
- Pick and mix the most appropriate algorithm
  - Multiple time scale mass preconditioning
  - Domain decomposition
  - n^{th} rootary with RHMC
- HMC cost is now $C \propto \left(\frac{m_\pi}{m_\rho}\right)^{-2} L^5 a^{-6}$ (Lüscher)
- Physical point no longer a pipedream
- Further improvement must come from autocorrelation?