Introduction To Lattice QCD

Bálint Joó, Jefferson Lab

QCDNA
June 20, 2014

QCD

- Quarks and Gluons are ‘fields’ in space-time (Minkowski Space)
- QCD Is defined by the Action (S) over the fields
- Action enumerates potential interactions
  - quark-gluon, gluon-gluon etc.
- Observables can be computed through Path Integrals over the fields.

“Functional Integral” over all the possible states of the fields

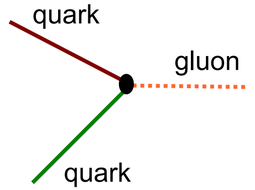
The “action” defining the theory

$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \int \overbrace{\mathcal{D}A \; \mathcal{D}\bar{\psi} \; \mathcal{D}\psi} \; \mathcal{O} \; e^{-S(A,\bar{\psi},\psi)}$$

Expectation value of an observable (eg: particle mass)

Value of observable on a concrete set of fields

$$S = \int dx \; dy \; \bar{\psi}(y) M(A; y, x) \psi(x)$$

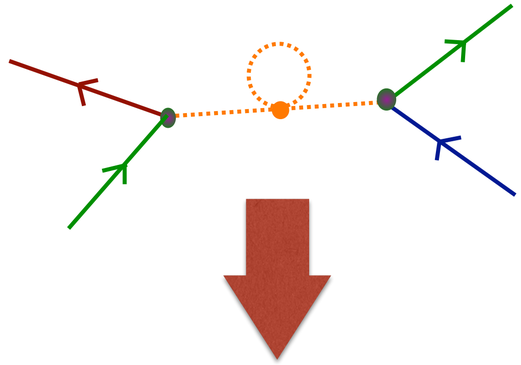


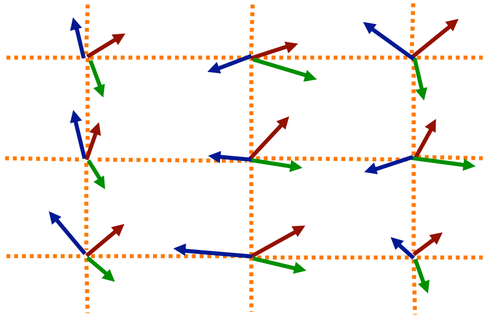
$$-\int dx \; \frac{1}{4} G_a^{\mu\nu}(x) G^a_{\mu\nu}(x)$$

Moving to the Lattice

- Replace “continuum” space time by 4D Lattice
- Discretize quark fields onto lattice sites
- Discretize gluon fields onto lattice links as SU(3) matrices
  - QCD local gauge symmetry: different color bases on each site
  - 3x3 matrices on links act as “parallel transporters” along links
  - rotate color basis at one site into that on another site.
- In the action:
  - use finite differences for derivatives
  - ‘imaginary’ time (t ⇒ it)
- Functional integrals become ‘regular’ integrals
- A configuration is a state of ‘gluon’ fields

$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \int \mathcal{D}A \; \mathcal{D}\bar{\psi} \; \mathcal{D}\psi \; \mathcal{O} \; e^{-S(A,\bar{\psi},\psi)}$$





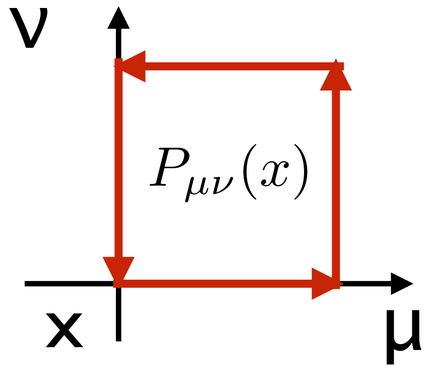
$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \int \prod_{\text{all links}} dU \prod_{\text{all sites}} d[\bar{\psi}, \psi] \; \mathcal{O} \; e^{-S(U,\bar{\psi},\psi)}$$

Gauge Actions

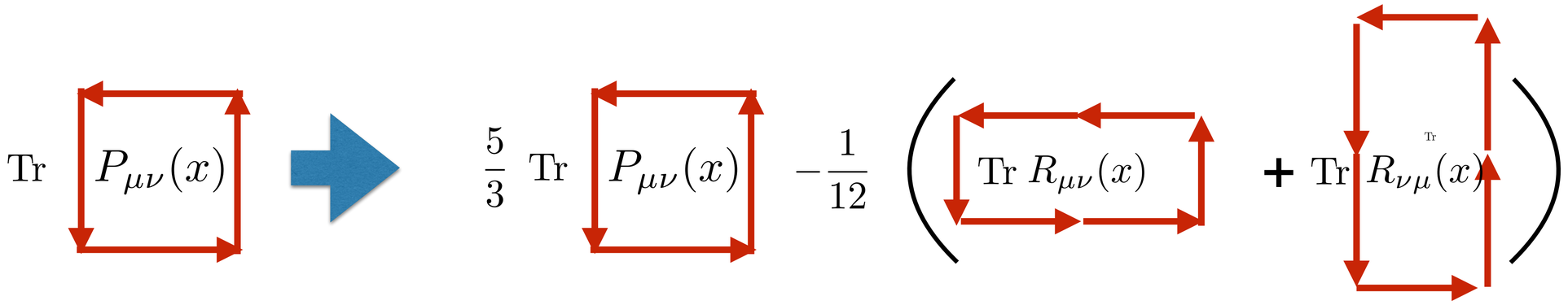
- Basic Wilson Plaquette Action
- as a -> 0 we have:

$$S_g = -\frac{\beta}{N_c}\sum_x\sum_{\mu<\nu}\mathrm{Re\ Tr}\ P_{\mu\nu}(x)$$

- $$S_g \to \int d^4x \frac{1}{8} F^a_{\mu\nu}(x) F^a_{\mu\nu}(x)$$



- Can further improve S_g by adding e.g. rectangle term:



Fermions & Pseudo-Fermions

- Fermions are Grassmann Numbers
  - but can do the Gaussian Integral
- Gives a determinant weight to the partition function
  - but determinants are nasty to evaluate so
- Bosonize the determinant
  - write as an integral over ‘pseudo fermion’ (boson) fields
  - NB: Now fermion matrix appears as an inverse: $\left(M^\dagger M\right)^{-1}$

$$\mathcal{Z} = \int \mathcal{D}U \; \mathcal{D}\bar{\psi} \; \mathcal{D}\psi \; e^{-\bar{\psi} M^\dagger(U) M(U) \psi - S_g(U)}$$

$$\mathcal{Z} = \int \mathcal{D}U \; \det\left(M^\dagger M\right) \; e^{-S_g(U)}$$

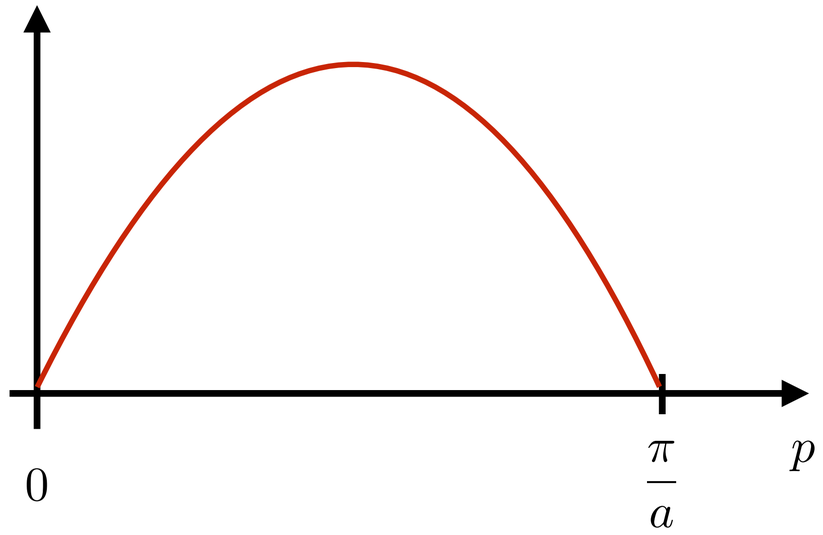
$$\mathcal{Z} = \int \mathcal{D}U \; \mathcal{D}\phi^\dagger \; \mathcal{D}\phi \; e^{-\phi^\dagger \left(M^\dagger M\right)^{-1} \phi - S_g(U)}$$

More Fermion Nastyness

- Naive fermion discretization leads to massless free field propagator

$$G(p) = \frac{i}{a} \sum_{\mu} \gamma_\mu \sin\left(p_\mu a\right)$$

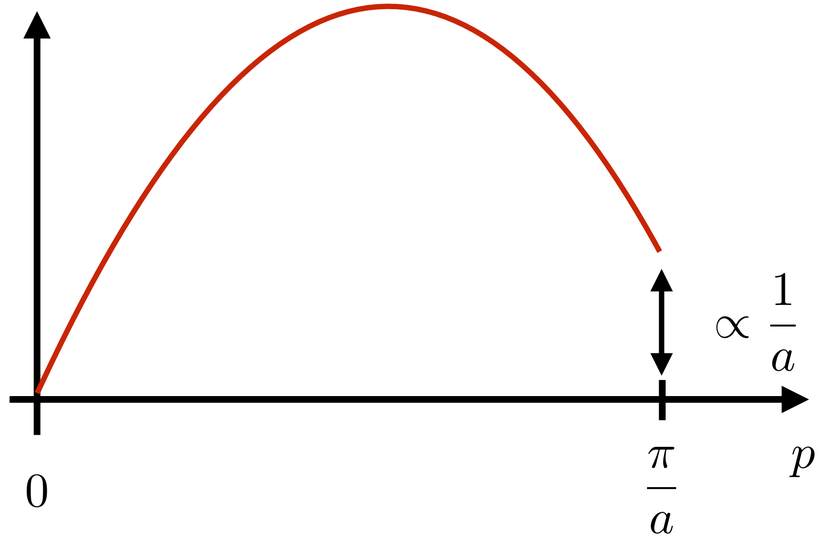
- Poles at p=0, p=π/a in each dimension
- Correspond to 2 species of fermion per dimension
- Fermion doubling problem
- Nielsen-Ninomiya No Go Theorem:
  - one cannot simultaneously have all the following:



Ultra-locality
Chiral Symmetry
No Doublers
Still look like a fermion prop

Solutions…

- Wilson-like Fermions:
  - give doubler modes mass proportional to 1/a
  - doublers decouple in continuum limit
  - explicitly break Chiral symmetry
  - discretization errors of O(a) for naive Wilson Fermions
  - Clover term can remove O(a) errors with appropriate c_{sw}
- Staggered Fermions:
  - distribute spin components to corners of a hypercube
  - reduce 16 flavors (in 4D) to 4 flavors (tastes?)
  - take the square root to get 2 flavors
    - this was the source of much controversy
  - taste symmetry breaking
    - reduce this through “improvement” with fat links (AsqTAD, HiSQ)
  - remnant U(1) “Chiral Symmetry”, $O(a^2)$ discretization errors

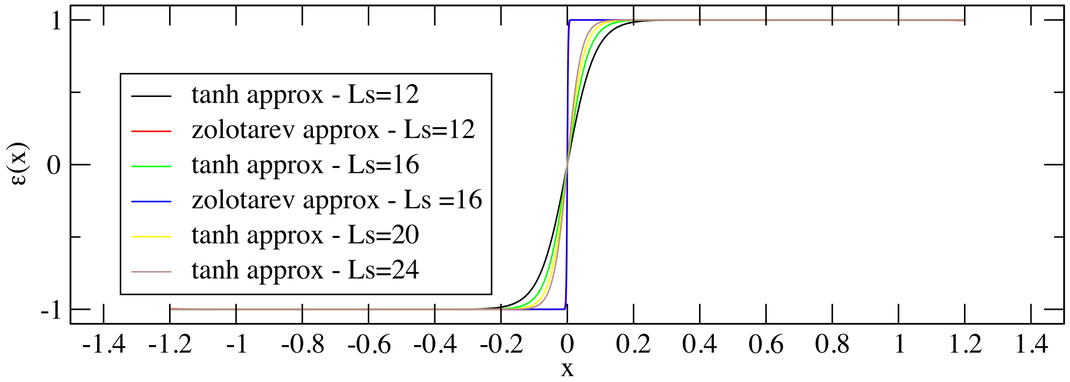


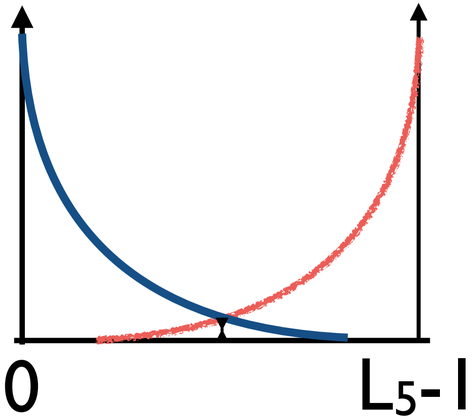
Details about staggered fermion implementation by Ruizi Li on Friday

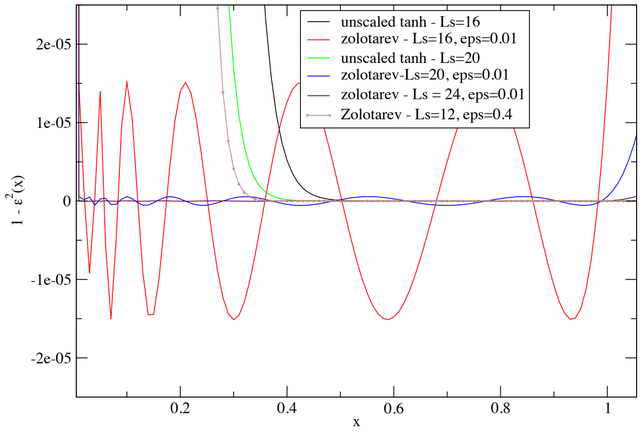
Solutions

- Overlap Fermions:
  - define a lattice version of Chiral symmetry through Ginsparg Wilson relation
  - Overap fermion is a solution of GW relation
  - involves a matrix sgn function (Talk by Frommer)
- Domain Wall and 5D fermions
  - ‘physical interpretation’ - 4D chiral modes bound to walls in 5D
  - or… a cunning 5D way of inverting a 4D matrix Sign Function.
- Both Overlap and DWF have O(a²) discretization errors

$$\gamma_5 D + D\gamma_5 = D\gamma_5 D \qquad D_{ov} = 1 + \gamma_5 \mathrm{sgn}(\gamma_5 D_W(-M_0))$$







Observables

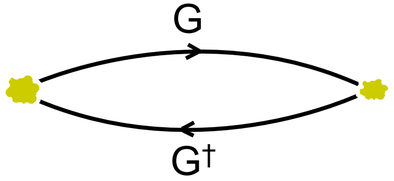
- Lattice QCD Observables are “correlation functions”
- E.g. For mesons (quark-antiquark pairings):

$$C(\vec{p},t) = \sum e^{i\vec{p}.\vec{x}}\ \mathrm{Tr}\ \Gamma\ G^{\dagger}(\vec{x},t;0,0)\ \Gamma\ G(\vec{x},t;0,0)$$

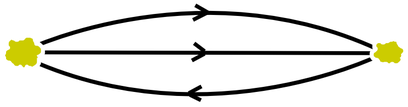
- G is the quark propagator defined as:

$$G(x,y) = M^{-1}_{x,y} S(x)$$

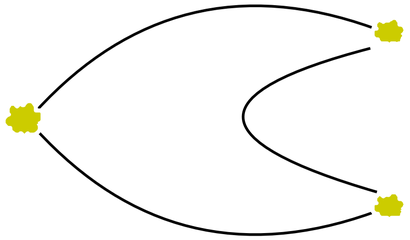
- M is the Fermion matrix
- Computing G involves solving a system of linear equations (Solvers)



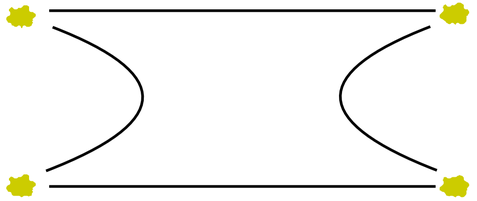
Meson: e.g. the π meson (a.k.a pion)



Baryon: e.g. proton or neutron



Meson to 2 meson decay: e.g. ρ → 2π

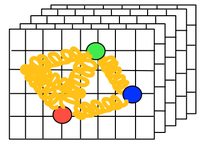


2 meson in to 2 meson out: e.g. 2π → 2π

LQCD Calculation Workflow

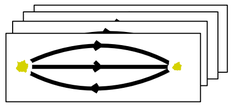
Gauge Generation

Gauge Configurations

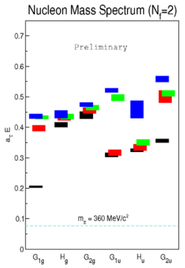


Analysis Phase 1

Propagators, Correlation Functions



Analysis Phase 2



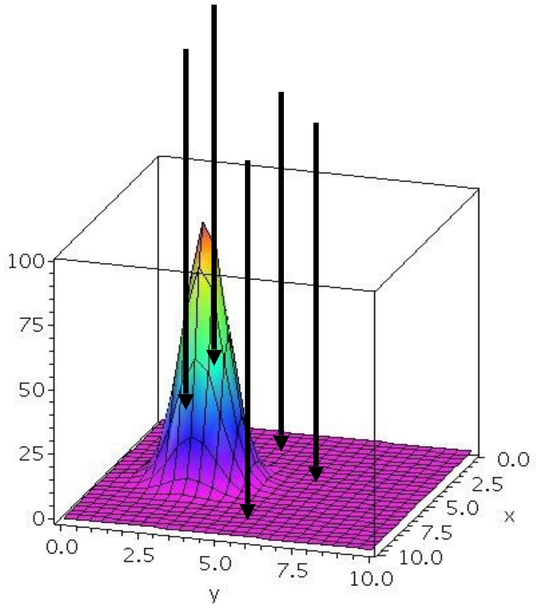
Physics Result

- Gauge Generation: Capability Computing on Leadership Facilities
  - configurations generated in sequence using Markov Chain Monte Carlo technique
  - focus the power of leadership computing onto single task exploiting data parallelism
- Analysis: Capacity computing, cost effective on Clusters
  - task parallelize over gauge configurations in addition to data parallelism
  - can use clusters, but also LCFs in throughput (ensemble) mode.

Evaluating Path Integrals: Monte Carlo

- On a lattice we have 4xVolume links.
  - e.g. 32^3x256 Lattice: ~33.6M Links
- Carrying out a 4V dimensional integral directly is unfeasible
- Turn to Monte-Carlo methods

$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \int \prod_{\text{all links}} dU_i \; \mathcal{O} \;\; e^{-S(U)} \longrightarrow \bar{O} = \frac{1}{Z} \sum_{\text{configuration}} \mathcal{O}(U) \; P(U)$$



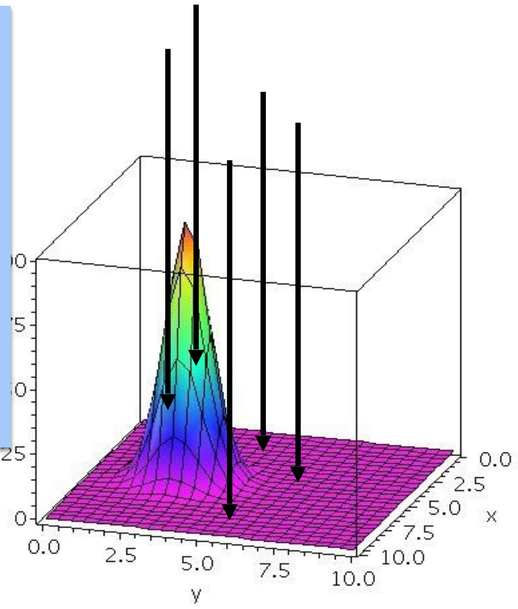
- Recipe:
  - Generate Configurations: U
  - Evaluate the Observable on each configuration
  - Form the “ensemble average” - which is the approximation to the Path Integral

Evaluating Path Integrals: Monte Carlo

- On a lattice we have 4xVolume links.
  - e.g. 32^3x256 Lattice: ~33.6M Links
- Carrying out a 4V dimensional integral directly is unfeasible
- Turn to Monte

$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \int_{\text{all}}$$

Problem: Since equilibrium probability is sharply peaked, random sampling can pick samples that are not very important and contribute little to the average



- Recipe:
  - Generate Configurations: U
  - Evaluate the Observable on each configuration
  - Form the “ensemble average” - which is the approximation to the Path Integral

Importance Sampling

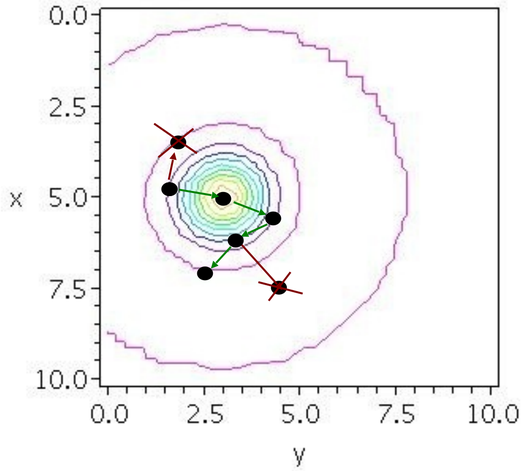
- Pick Configuration ‘U’ with probability P(U)
- Ensemble average then becomes a ‘regular average’

$$\langle \mathcal{O} \rangle = \frac{1}{\mathcal{Z}} \int \prod_{\text{all links}} dU_i \; \mathcal{O} \; e^{-S(U)} \longrightarrow \bar{O} = \frac{1}{N} \sum_{\text{N}} \mathcal{O}(U) \qquad \sigma(\bar{\mathcal{O}}) \propto \frac{1}{\sqrt{N}}$$

- E.g.: Metropolis Algorithm
  - Start from some initial configuration U
  - Pick trial config U’ from U reversibly: ie $P_c(U \rightarrow U') = P_c(U' \rightarrow U)$
  - Accept with Metropolis probability

$$P(U' \leftarrow U) = \min\left(1, \frac{e^{-S(U')}}{e^{-S(U)}}\right)$$

  - If we reject, next config is U again



Global Updating

- Metropolis Algorithm would proceed link by link
- For each link one would need to evaluate the quark part of the action

$$S_f = \phi^\dagger \left(M^\dagger M\right)^{-1} \phi = \langle \phi | X \rangle$$

- where

$$\left(M^\dagger M\right)\, X = \phi$$

- and again, M is the fermion matrix
- With 4V links this is prohibitive and so one needs a global update method

Hybrid Monte Carlo

- Big Trick: Update all links at once using Molecular Dynamics
  - Treat each link as ‘canonical coordinate’
  - Assign to each link a ‘canonical momntum’ in the lie algebra su(3)
- Construct a fictitious Hamiltonian

$$H = \frac{1}{2}\sum_{\text{links}} p^2 + S(U)$$

- Simulate Hamiltonian System with partition function:

$$\mathcal{Z} = \int \mathcal{D}U\ \mathcal{D}p\ e^{-H} = \int \mathcal{D}U\ e^{-S} \int \mathcal{D}p\ e^{-\frac{1}{2}\sum_{\text{links}} p^2} = C \int \mathcal{D}U e^{-S(U)}$$

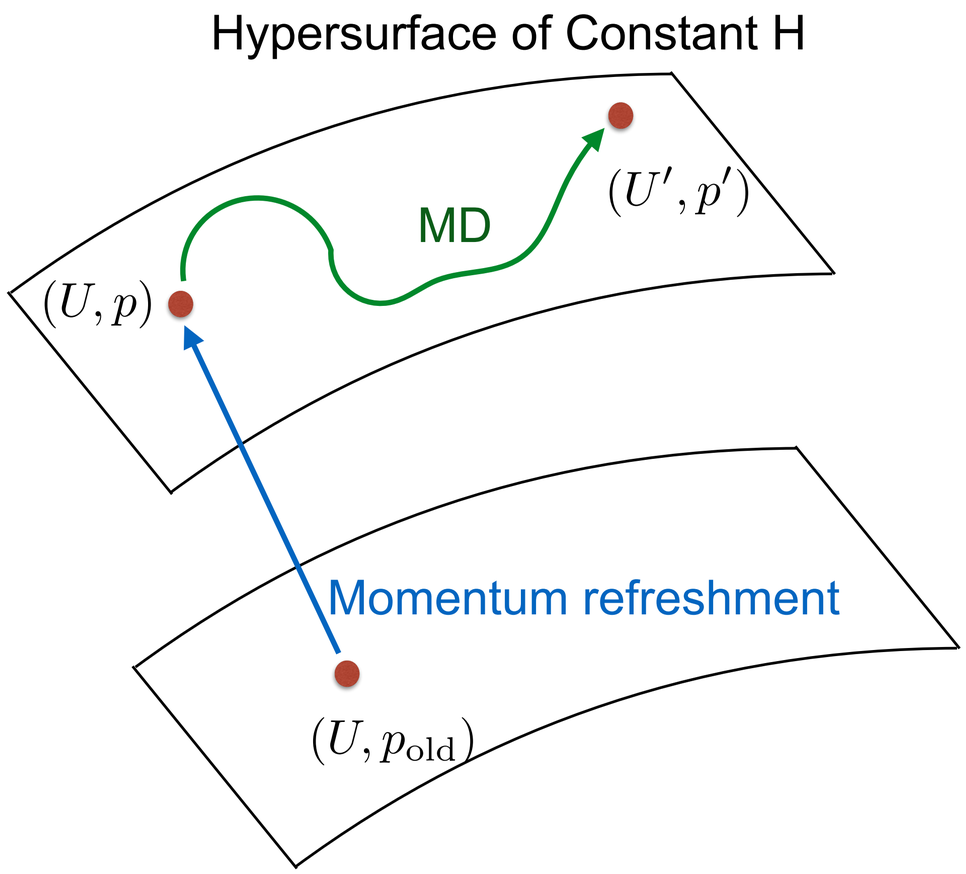
- Momenta have gaussian distribution: easy to generate from heatbath
- Integral over momenta produce a constant, which will cancel in the path integral

Hybrid Monte Carlo (HMC)

1. Refresh momenta from Gaussian Heatbath
   - generate (U,p) from (U,p_{old})
2. Compute H = H(U,p)
3. Perform Molecular Dynamics trajectory
   - generate (U',p')
   - MD must be reversible and 'area preserving'
4. Compute H' = H(U',p')
5. Accept with Metropolis probability

$$P = \min\left(1, e^{-H(U',p')+H(U,p)}\right)$$

6. If rejected new state is (U,p)



HMC Accept Rate

- Acceptance Rate goes as:

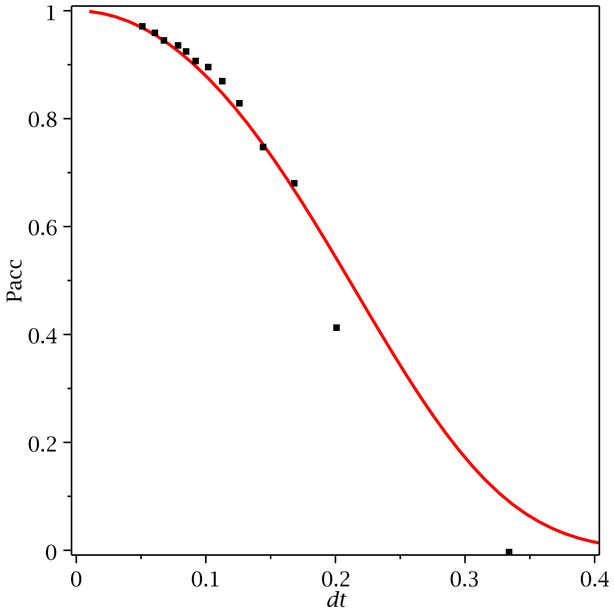
$$\langle P_{acc} \rangle = \mathrm{erfc}\left(\sqrt{\frac{\langle \Delta H^2 \rangle}{8}}\right)$$

- and $< \Delta H^2 >$ depends on the integration scheme. For n-th order integrator over a unit length trajectory:

$$\langle \Delta H \rangle \approx \langle \Delta H^2 \rangle \propto V \delta\tau^{2n}$$

- In general:
  - cost ~ number of steps ~ 1 / δτ
  - tuning: allow increased step size, without lowering acceptance

Figure from: “Improving dynamical lattice QCD simulations through integrator tuning, using Poisson Brackets and a force-gradient Integrator”, M. A. Clark, B. Joo, A.D. Kennedy, P.J. Silva Phys Rev.D84,071502

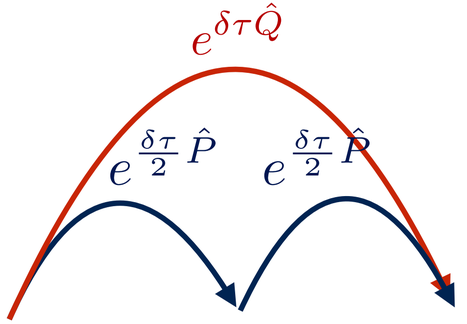


(b)Acceptance rate as a function of $\delta\tau$, with $\lambda = 0.18$.

MD Integrators

- MD Integrators must be
  - reversible
  - area (phase space measure) preserving
- Can't use some sophisticated ODE integrators such as Runge-Kutta etc.
- Can use symplectic integrators
  - composed of symplectic update pieces
  - gauge update: $e^{\delta\tau\hat{Q}}$
  - momentum update: $e^{\frac{\delta\tau}{2}\hat{P}}$
- Examples: 2nd order Leap-Frog, etc.
- Some have tunable parameters e.g. λ

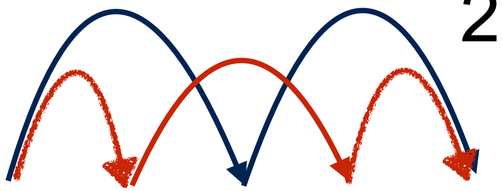
Leapfrog:



3 step
2nd order

$$e^{\frac{\delta\tau}{2}\hat{P}}e^{\delta\tau\hat{Q}}e^{\frac{\delta\tau}{2}\hat{P}}$$

Omelyan:



5 step
2nd order

$$e^{\lambda\delta\tau\hat{Q}}e^{\frac{\delta\tau}{2}\hat{P}}e^{(1-2\lambda)\delta\tau\hat{Q}}e^{\frac{\delta\tau}{2}\hat{P}}e^{\lambda\delta\tau\hat{Q}}$$

HMC With Fermions

- Fermions have pseudofermion action: $S_f = \phi^\dagger \left(M^\dagger M\right)^{-1} \phi$
- Can draw pseudofermions from Heat Bath
  - write: $S_f = \eta^\dagger \eta$
  - then: $\phi = M^\dagger \eta$
- We draw new pseudofermions at the start of each trajectory to sample the integral: $\int \mathcal{D}\phi^\dagger \mathcal{D}\phi$
- We typically keep pseudofermions fixed along a trajectory.

MD Forces

- Momentum Update: $e^{\delta\tau \hat{P}} : \; p(\tau + \delta\tau) \rightarrow p(\tau) + \delta\tau F$
- For 2 Flavor Quark Action:

$$F = -\phi^\dagger \left(M^\dagger M\right)^{-1} \left[\dot{M}^\dagger M + M^\dagger \dot{M}\right] \left(M^\dagger M\right)^{-1} \phi$$

- Need to evaluate:

$$\left(M^\dagger M\right) X = \phi$$

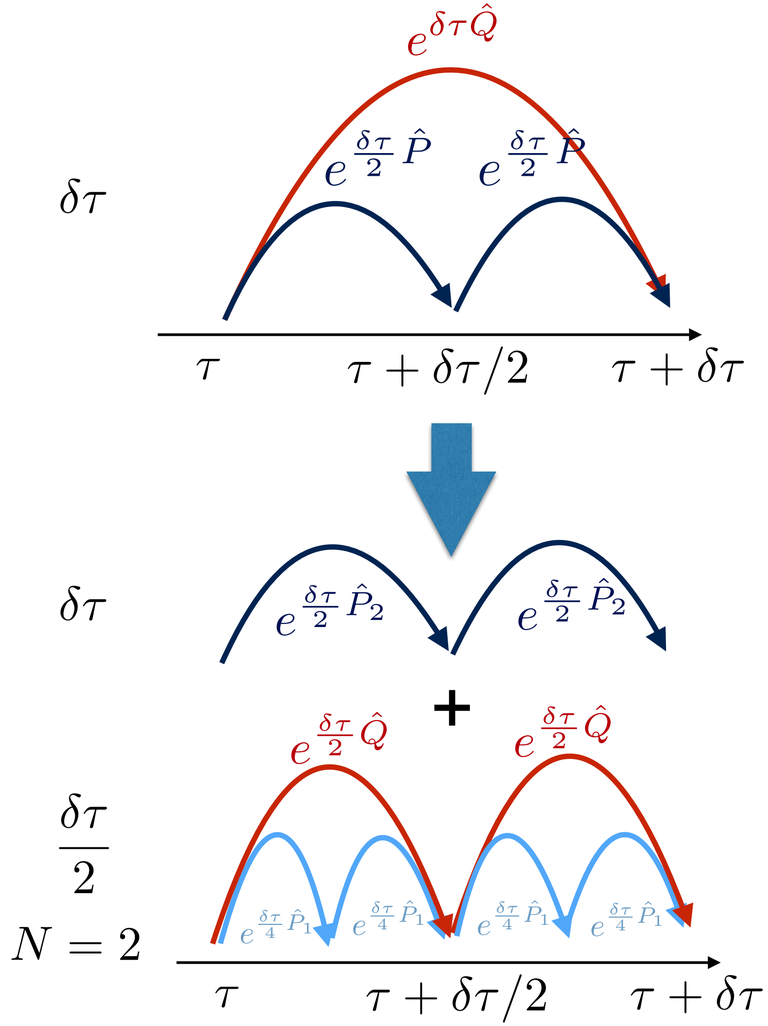
- Here again we need a solver but:
  - System is manifestly Hermitian and Positive definite
  - Common to use two step solve: $M^\dagger Y = \phi$ followed by $M X = Y$ (reduced condition number)
  - M will change as we perform the MD gauge field update, long set-up times for solver may not be as easily amortized as for propagators.

Multiple Time Scales

- Sexton & Weingarten introduced a way to have multiple time scales in the MD
- Split action as S = S_1 + S_2

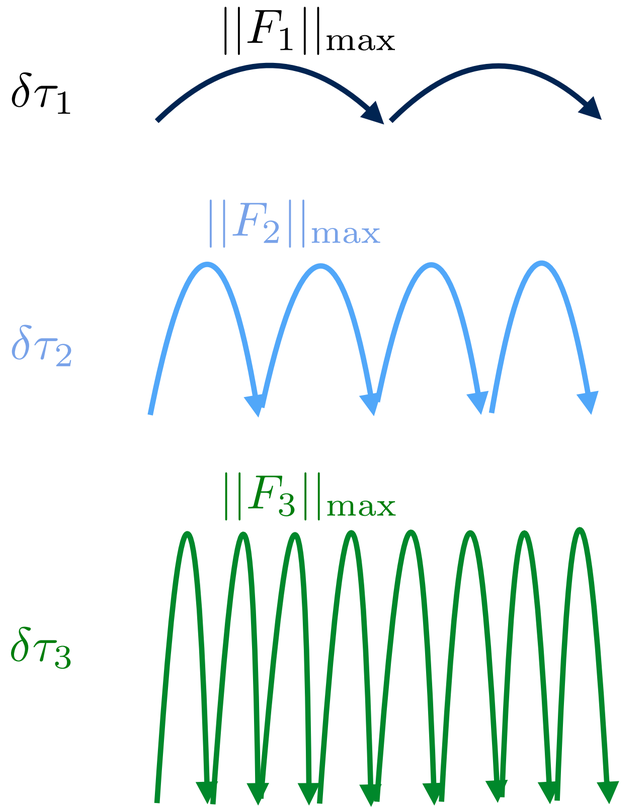
$$U^{(2)} = e^{\frac{\delta\tau}{2}\hat{P}_2}\left[U\left(\hat{P}_1, \frac{\delta\tau}{N}\right)\right]^N e^{\frac{\delta\tau}{2}\hat{P}_2}$$

- Here, P_2 updates momenta with the Force from S_2 with steps of length dt
- And the U is a full update with S_1 taking N steps of length dt/N



Using Multiple Time Scales

- Action contains several pieces:
  - Gauge Action for gluons
  - Light quark action
  - Strange quark action
- These all have different sized Forces
- Heuristic Tuning
  - want to run at largest dt, for which integrator is stable
  - for smaller forces this will be a larger dt
  - for larger forces it will be a smaller dt
  - group together pieces with similar sized forces and



$$\delta\tau_1 ||F_1||_{\max} \sim \delta\tau_2 ||F_2||_{\max} \sim \delta\tau_3 ||F_3||_{\max}$$

Fermion Determinant Splitting

- Hasenbusch Trick (Hasenbusch)
  - introduce auxiliary M_1 similar to M
  - e.g. add small twist or slightly different mass
  - get two determinents simulate with two p.f.-s
  - ratio term “close to” identity: small forces, long steps
  - run cancellation terms on different time scale.
    - mass preconditioning (Jansen, Urbach, Shindler, Wenger)
  - cancellation term heavier than original 2 flavor term
- Multi-pseudofermion Trick (Clark, Kennedy)
  - a pseudofermion field for each N-th root of the det.
  - N-th root typically implemented with Rational approx.

$$\det(M^\dagger M) = \frac{\det(M^\dagger M)}{\det(M_1^\dagger M_1)} \det(M_1^\dagger M_1)$$

$$\phi^\dagger \left(M^\dagger M\right)^{-1} \phi \to \phi^\dagger M_1 \left(M^\dagger M\right)^{-1} M_1^\dagger \phi + \phi_1^\dagger \left(M_1^\dagger M_1\right) \phi_1$$

$$M_1 \left(M^\dagger M\right)^{-1} M_1^\dagger \approx 1 + \delta M$$

$$\det(M^\dagger M) = \prod_{i=1}^{N} \det(M^\dagger M)^{1/N}$$

$$\phi^\dagger \left(M^\dagger M\right)^{-1} \phi \to \sum_{i=1}^{N} \phi_i^\dagger \left(M^\dagger M\right)^{-1/N} \phi_i$$

Shadow Hamiltonians in LQCD

- Clark, Kennedy, Silva, Joo. Also, for DWF: H. Yin, R. D. Mawhinney
- Error in Symplectic Update can be computed using Baker Campbell Haussdorff formula:

$$e^{\hat{A}}e^{\hat{B}} = e^{\hat{A}+\hat{B}+\frac{1}{2}[\hat{A},\hat{B}]+\frac{1}{12}\left([\hat{A},[\hat{A},\hat{B}]]-[\hat{B},[\hat{A},\hat{B}]]\right)+\mathrm{h.o.t}}$$

- Using this we can show that the 3-step Omelyan integrator behaves as:

$$e^{\lambda\delta\tau\hat{Q}}e^{\frac{\delta\tau}{2}\hat{P}}e^{(1-2\lambda)\delta\tau\hat{Q}}e^{\frac{\delta\tau}{2}\hat{P}}e^{\lambda\delta\tau\hat{Q}} = e^{\hat{H}'\delta\tau}$$

- with

$$\hat{H}' = \hat{H} + \left(\frac{6\lambda^2-6\lambda+1}{12}[\hat{Q},[\hat{Q},\hat{P}]] + \frac{1-6\lambda}{24}[\hat{P},[\hat{Q},\hat{P}]\right)\delta\tau^2 + O(\delta\tau^4)$$

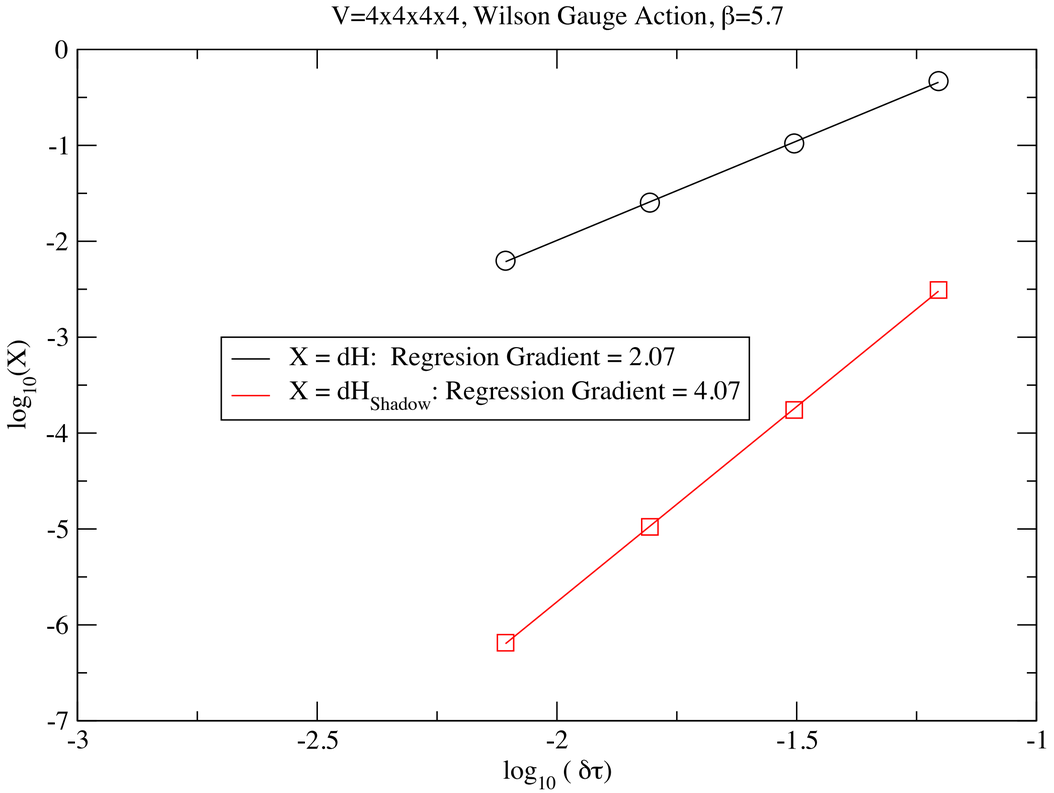
Shadow Hamiltonians

$$H' = H + \left(\frac{6\lambda^2 - 6\lambda + 1}{12} \{Q, \{Q, P\}\} + \frac{1 - 6\lambda}{24} \{P, \{Q, P\}\} \right) \delta\tau^2 + O(\delta\tau^4)$$

- H is the Hamiltonian function
- {Q,{Q,P}} and {P,{Q,P}} are Poisson Brackets (analogues to the commutators)
- H' is the "Shadow Hamiltonian"
- The Shadow Hamiltonian is preserved EXACTLY by the integrator
  - No step-size error!!
- If one could measure the Poisson brackets one could
  - optimize λ to minimize $\delta\tau^3$ term
  - construct a higher order $O(\delta\tau^5)$ "Force Gradient Integrator",

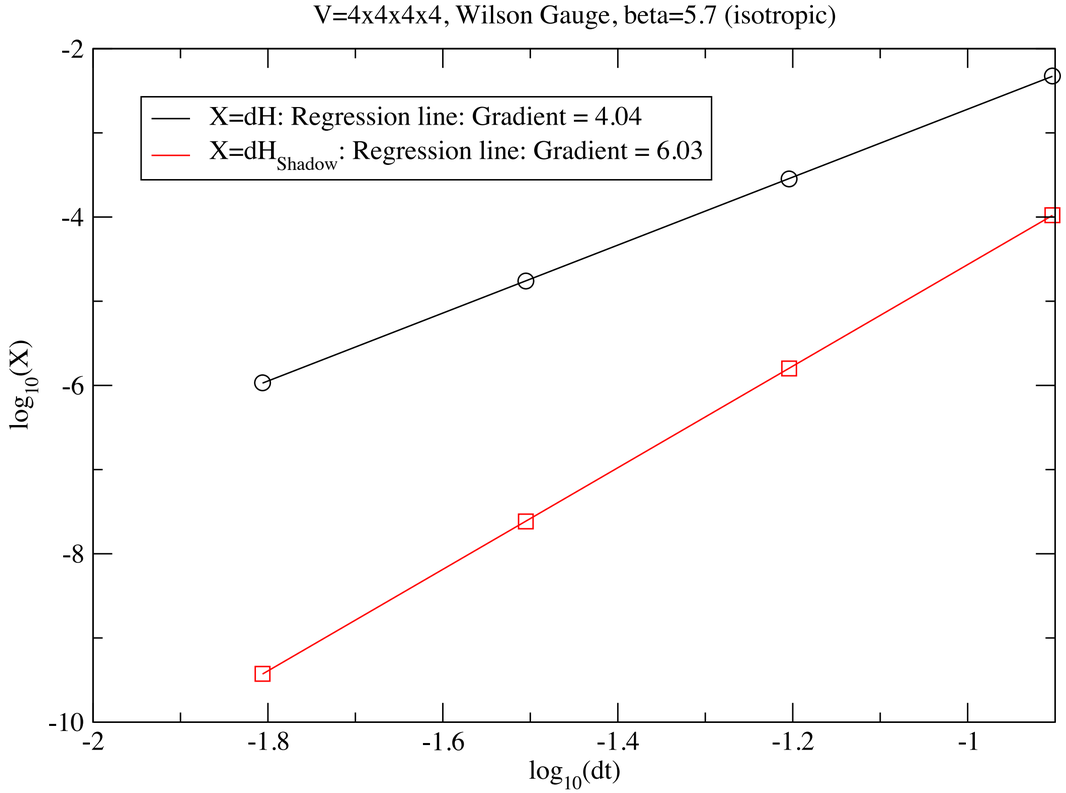
Shadow Hamiltonian Examples

Leapfrog Integrator



Leapfrog: $O(\delta\tau^3)$ per step, $O(\delta\tau^2)$ over unit trajectory,
Shadow: $O(\delta\tau^5)$ per step, $O(\delta\tau^4)$ over unit trajectory

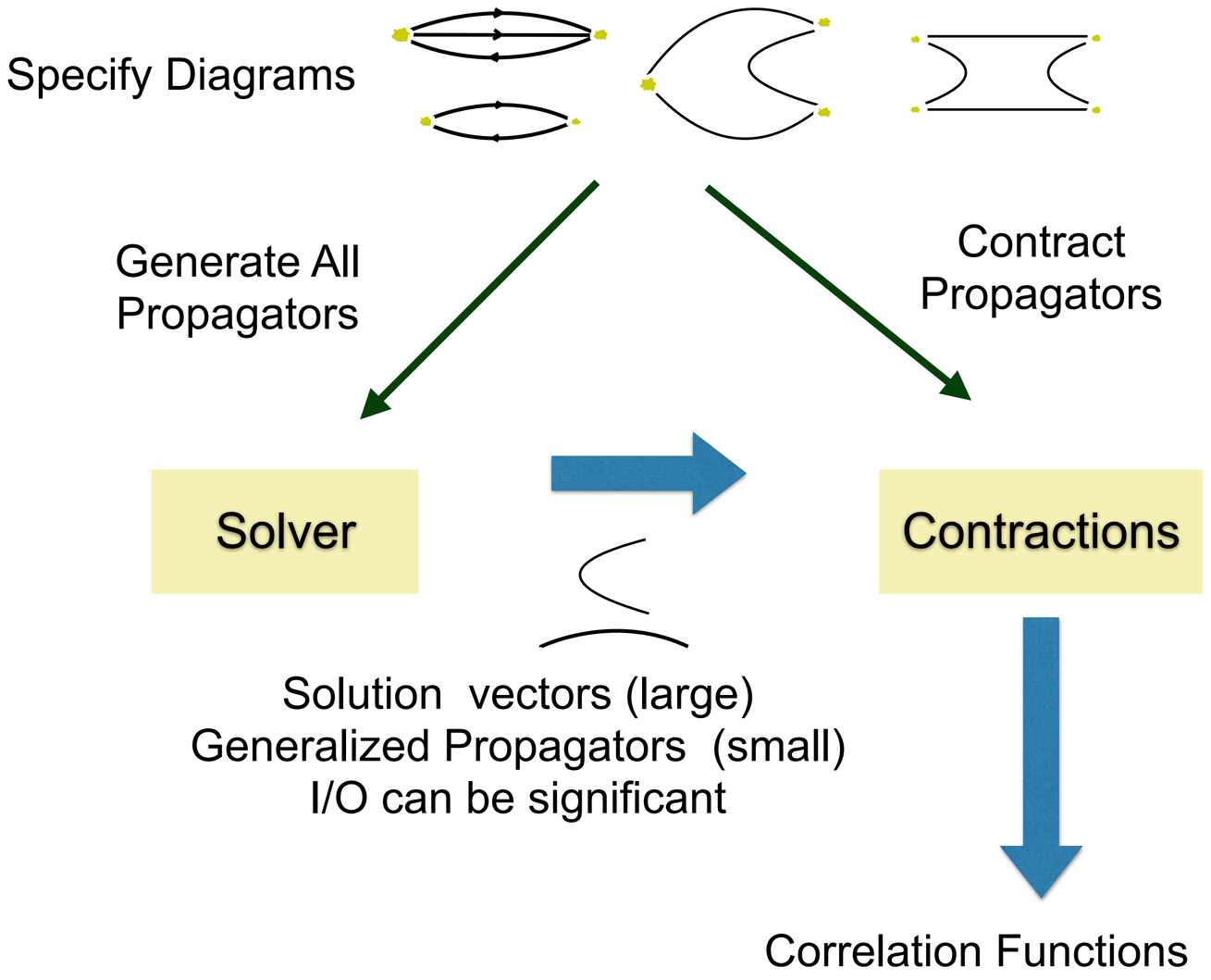
Force Gradient Integrator



Leapfrog: $O(\delta\tau^5)$ per step, $O(\delta\tau^4)$ over unit trajectory,
Shadow: $O(\delta\tau^7)$ per step, $O(\delta\tau^6)$ over unit trajectory

Analysis Pipeline

- Two main components
  - propagator calculations (solver)
  - contraction calculations
- Contractions use dense matrix mutliply
  - matrix dimension is O(100) (# sources)
- Many solves needed on single configuration:
  - #spin x #timeslice x #source x #quarks
- Typical Example
  - 4 spins, 256 timeslices, 386 source vectors and light + strange quarks
  - **790,528** individual solves per configuration
  - Even more for larger lattices, more complex diagrams



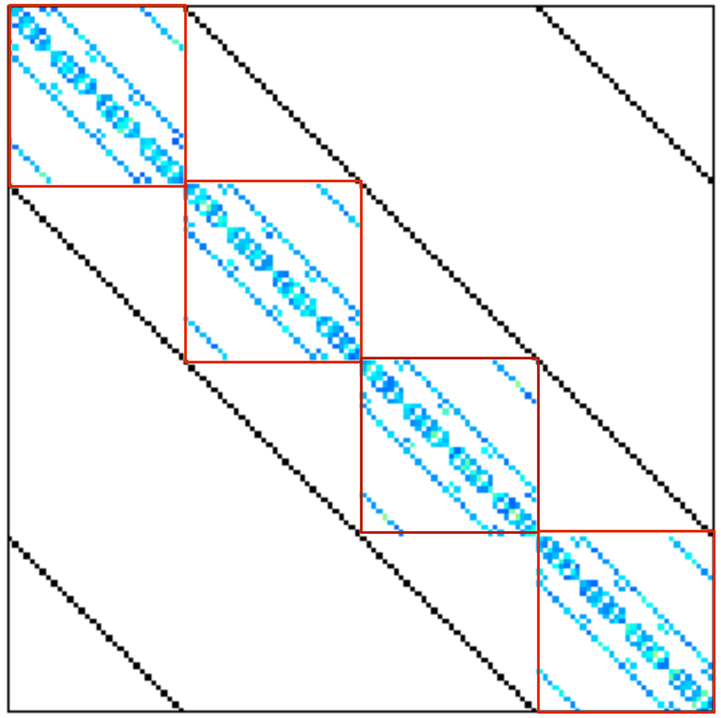
Solvers

- Traditionally we solve the Linear Systems with iterative Krylov Subspace solvers
- These can
  - work as black boxes
  - typically need only L1 BLAS and MV operations
  - typical candidates: Conjugate Gradients, BiCGStab
- Convergence depends on condition number of M
- As quark mass approaches the physical mass, M becomes more and more ill conditioned
- Critical Slowing Down in the Solver.

($\phi_0 = \phi$ is an Initial Guess)

1. Compute $r_0 = \chi - M^\dagger M \phi_0,\ p_0 = r_0$
2. For $j = 0, 1, \ldots$ until convergence:
3. $\alpha_j = \dfrac{\langle r_j, r_j \rangle}{\langle M p_j, M p_j \rangle}$
4. $\phi_{j+1} = \phi_j + \alpha_j p_j$
5. $r_{j+1} = r_j - \alpha_j \left(M^\dagger M\right) p_j$
6. $\beta_j = \dfrac{\langle r_{j+1}, r_{j+1} \rangle}{\langle r_j, r_j \rangle}$
7. $p_{j+1} = r_{j+1} + \beta_j p_j$
8. End For

Domain Decomposition Preconditioner



Use a block-diagonal operator as a ‘preconditioner’ in the solver

- inner-outer scheme
- outer scheme needs to be ‘flexible’ (FGMRES,GCR)

Arrange to spend most time in the preconditioner.

block diagonal operator is a ‘wavelength filter’

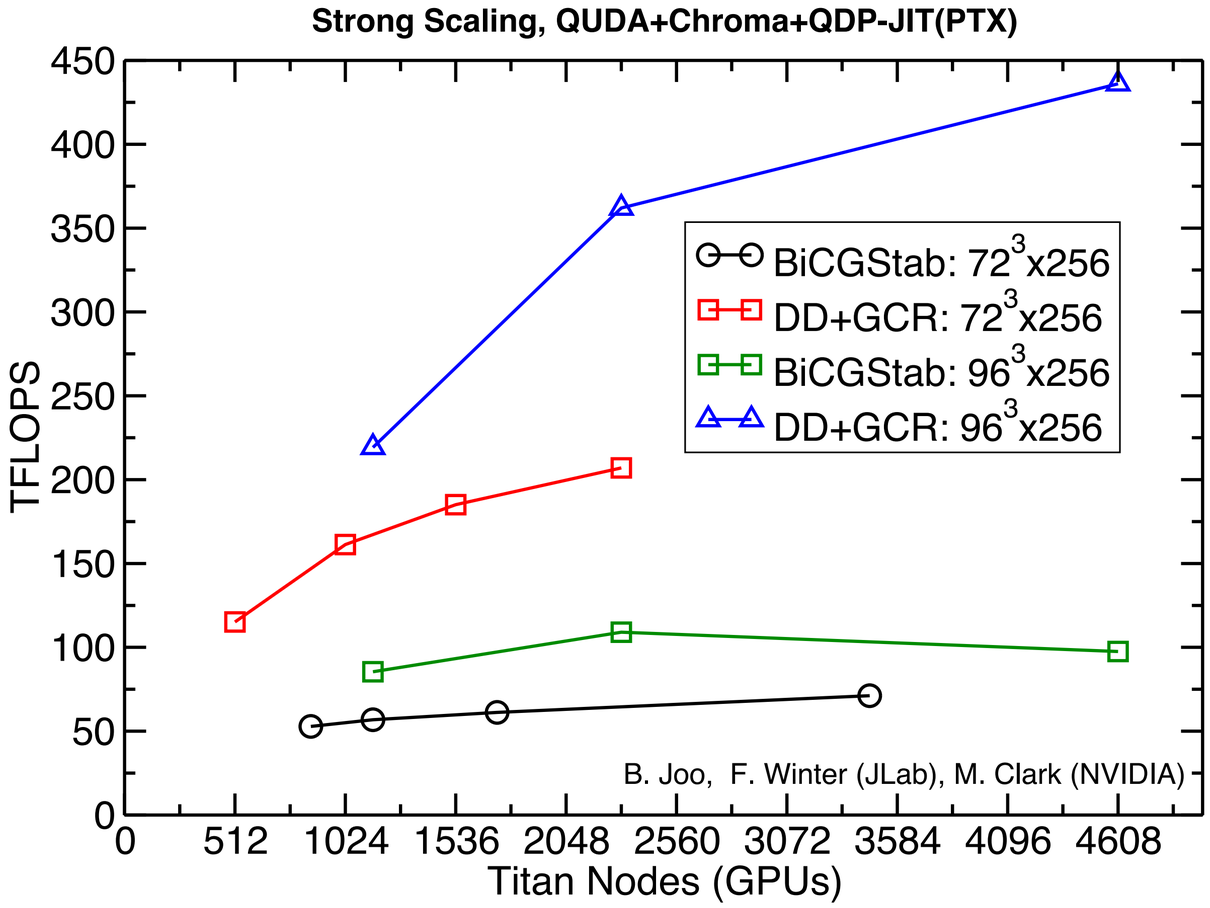
- reduces errors of “short” wavelength modes
- outer scheme still needs to deal with long wavelength modes (e.g. w. deflation — Frommer et. al)

Very suitable for architectures where communication is a bottleneck (e.g. GPU, Xeon Phi)

- Suitable as a smoother for Multi-Grid (talk later on by Rottmann)

Solver Performance

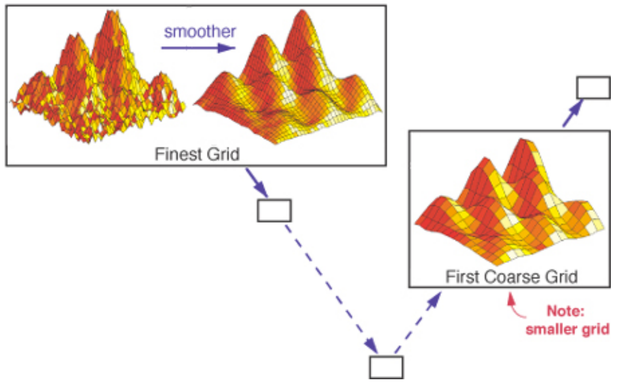
- QUDA Solver performance on Titan
  - Cray XK7 system
  - 1 NVIDIA K20X GPU per node
  - Gemini Interconnect
- The DD+GCR solver does considerably better than the standard BiCGStab
- But even DD+GCR is affected by strong scaling effects

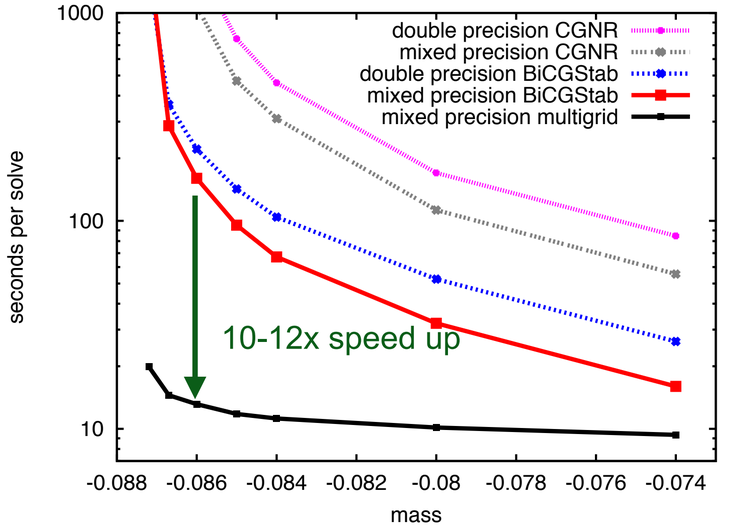


Algebraic Multi Grid

- Critical Slowing down is caused by ‘near zero’ modes of M
- Multi-Grid method
  - separate (project) low lying and high lying modes
  - solve for high lying modes with “smoother”
  - solve for low modes on coarse grid with reduced dimensional operator
  - Gauge field is ‘stochastic’, so no geometric smoothess on low modes => algebraic multigrid
  - Setting up restriction/prolongation operators is costly
  - Easily amortized in Analysis with O(100,000) solves
- Several more MG talks to follow (e.g. Brannick, Rottmann)

Image From: http://computation.llnl.gov/casc/sc2001_fliers/SLS/SLS01.html
Credit: LLNL, CASC





Multi-Grid. figure from J. C. Osborn et. al. PoS Lattice 2010:037,2010, R. Babich et. al. Phys. Rev. Lett, 105:201602,2010

Summary

- Fermions are troublesome
  - several approaches (Wilson, TM, Staggered, DWF, Overlap etc) scarifice different desiderata
- Dealing with fermions is a predominant cost of HMC simulations
  - split fermion determinants, multiple time-scales etc
- Shadow Hamiltonian Techniques promise
  - better tuning
  - And/Or cheaper 4th order “Force Gradient” MD integrators
- Always want better solvers:
  - AMG Has been highly successful for propagator calculations with Wilson Fermions (talks by Brannick, Rottmann)
  - See follow on talk by Kahl for details and applications to overlap fermions
  - Domain Decomposed preconditioners have shown themselves to be scalable for Wilson Fermions
  - Next frontier is to see if AMG can be applied to HMC (see talk by Lin)