Algorithms for the Petaflops Era

R. C. Brower, JLab Oct 6, 2005

I don’t known --- have a nice trip home!

Ok let me try to make a few suggestions/guesses for amusement

<u>“BU Algorithm Group”:</u> Brower, Clark, Fleming, Orginos, Osborn, Rebbi et al in collaboration with mathematicians Brannick at LLNL’s and associates of the Institute for Scientific Computing Research (ISCR) Interim director: Prof. David Keyes, Test Platform BlueGene/L

K. Wilson (1989 Capri):

"One lesson is that lattice gauge theory could also require a 10^8 increase in computer power AND spectacular algorithmic advances before useful interactions with experiment ..."

- ab initio Chemistry

1. 1930+50 = 1980
2. 0.1 flops → 10 Mflops
3. Gaussian Basis functions

vs

- ab initio QCD

1. 1980 + 50 = 2030?*
2. 10 Mflops → 1000 Tflops
3. Clever Collective Variable?

**Much sooner but need less than $10/Gflops!*

Year 2015

- End of H. Clinton’s 2nd term
- Cost is $1 per Gigaflops
- Lattices sizes are up to 128x 64^3 & at “physical quark mass” [scaling:Time $\sim$ $(1/m^6_\pi)(1/a^7)$]
- Algorithms: All are Multi-scale!
- The non-SUSY limit is still a challenge for die hard QCD/String theorists.

Outline

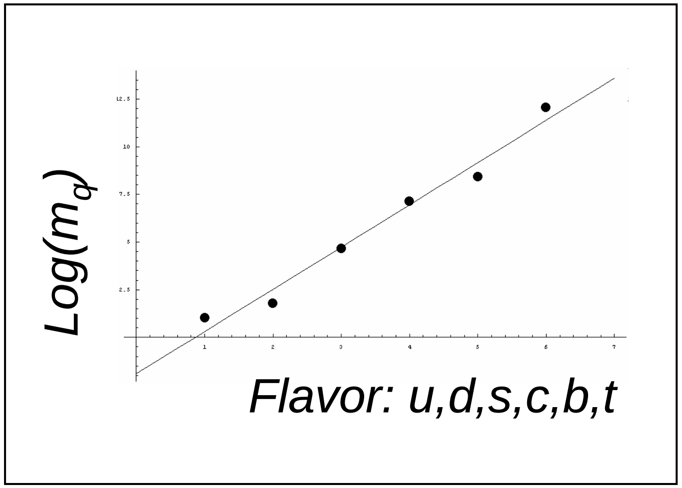
1. Multiple Scales in QCD:
   Space-Time and HMC Time

1. Multi-scale Algorithms for QCD:
   1990 vs 2005 vs 2015

2. Disconnected diagrams
   How to compute Tr[O D^{-1}] ?

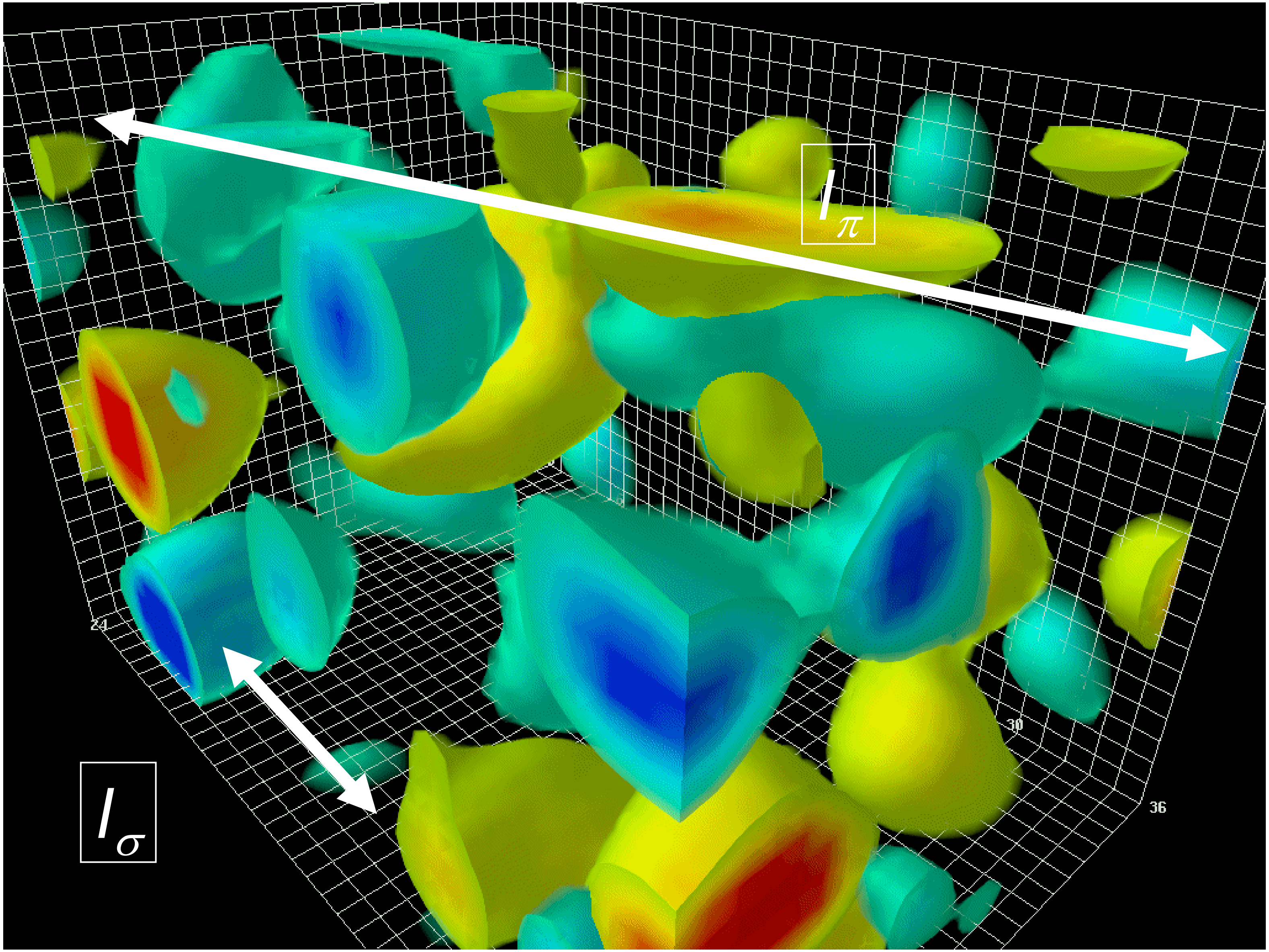
3. Taking the 5th Dimension Seriously
   Gluons and Hadron in 5-d

1. Length/Mass Scales in QCD



- Quarks Masses: (197 fm Mev)
  - 2, 8, 100, 1200, 4200, 175,000 Mev
- String Length:
  - 1000 Mev ($\sim$ .2 fm)
- Chiral limit: m_π =140 Mev ($\sim$1.4 fm)
- Nuclear: scattering length/effective range
  - $a_{singlet}$ = - 23.714 fm ($\sim$ 8 Mev) & r = 2.73
  - $a_{triplet}$ = 5.425 fm ($\sim$ 36 Mev) & r = 1.749 fm
- Deuteron Binding = 50 Mev. ($\sim$ 4 fm)
- Finite T, finite μ etc

Confinement length vs Pion Compton length



Back to the Future ~ QCD: Present Paradigm

Three different Fermions "Inversion" problems:

1. **Propagator**
2. **Trace**
3. **Determinant**

$$\int dU d\bar{\Psi} d\Psi \quad [\Psi(x_1)\bar{\Psi}(x_2)\cdots\bar{\Psi}(x_n)] \quad e^{-\frac{1}{g^2}S[U,\bar{\Psi},\Psi]} =$$

Nucleon

$$\int dU \; [\underbrace{D^{-1}D^{-1}D^{-1}}_{\text{Nucleon}}](x,0) \times Tr[\mathcal{O}D^{-1}] \times Det[D]\; e^{-\frac{1}{g^2}S[U]}$$

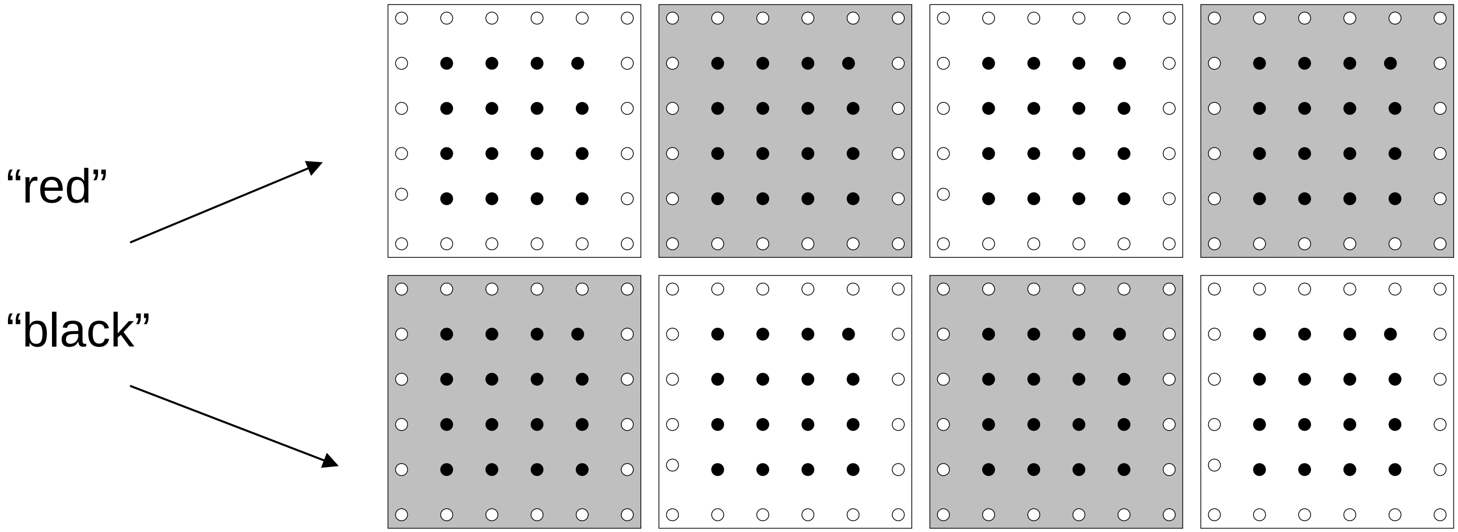
Propagators: 1 to All

Trace; All to All

Determinant

2. Multi-scale Methods Schwarz/Hasenbush/RHMC

Schwarz Precondition CG Inverter



- Schwarz Alternating Procedure (SAP)
- Red/Black partition with non-overlapping 6^4 blocks
- Use even/odd precondition inside blocks and Schwarz as preconditioner for outer CG iterations

† M. Luscher, "Schwarz-preconditioned HMC algorithm for two-flavour lattice QCD", hep-lat/0409106.

Shur Factorization† of Determinant

$$Det[D] = Det[D_{bb}]\ Det[D_{rr}]\ Det[1 - D_{rr}^{-1} D_{rb} D_{bb}^{-1} D_{br}]$$

Follows from Shur decomposition

$$D \;=\; \begin{bmatrix} D_{bb} & D_{br} \\ D_{rb} & D_{rr} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ D_{rb} D_{bb}^{-1} & 1 \end{bmatrix} \begin{bmatrix} D_{bb} & 0 \\ 0 & D_{rr} - D_{rb} D_{bb}^{-1} D_{br} \end{bmatrix} \begin{bmatrix} 1 & D_{bb}^{-1} D_{br} \\ 0 & 1 \end{bmatrix}$$

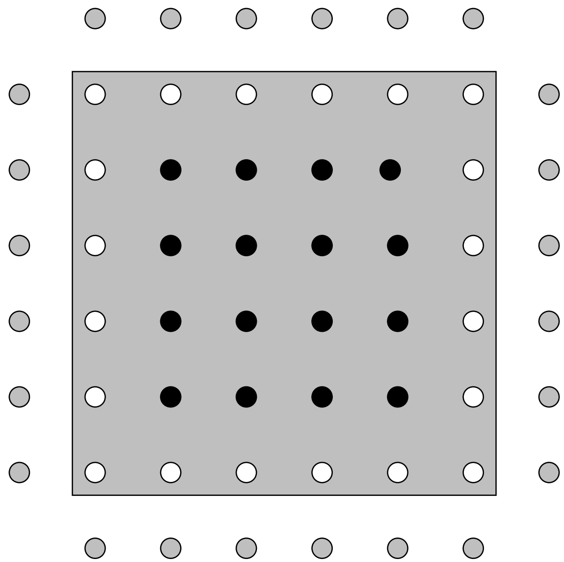
$$\begin{bmatrix} D_{bb} & 0 \\ 0 & D_{rr} - D_{rb} D_{bb} D_{br} \end{bmatrix} = \begin{bmatrix} D_{bb} & 0 \\ 0 & D_{rr} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 - D_{rr}^{-1} D_{rb} D_{bb}^{-1} D_{br} \end{bmatrix}$$

D_1 D_2

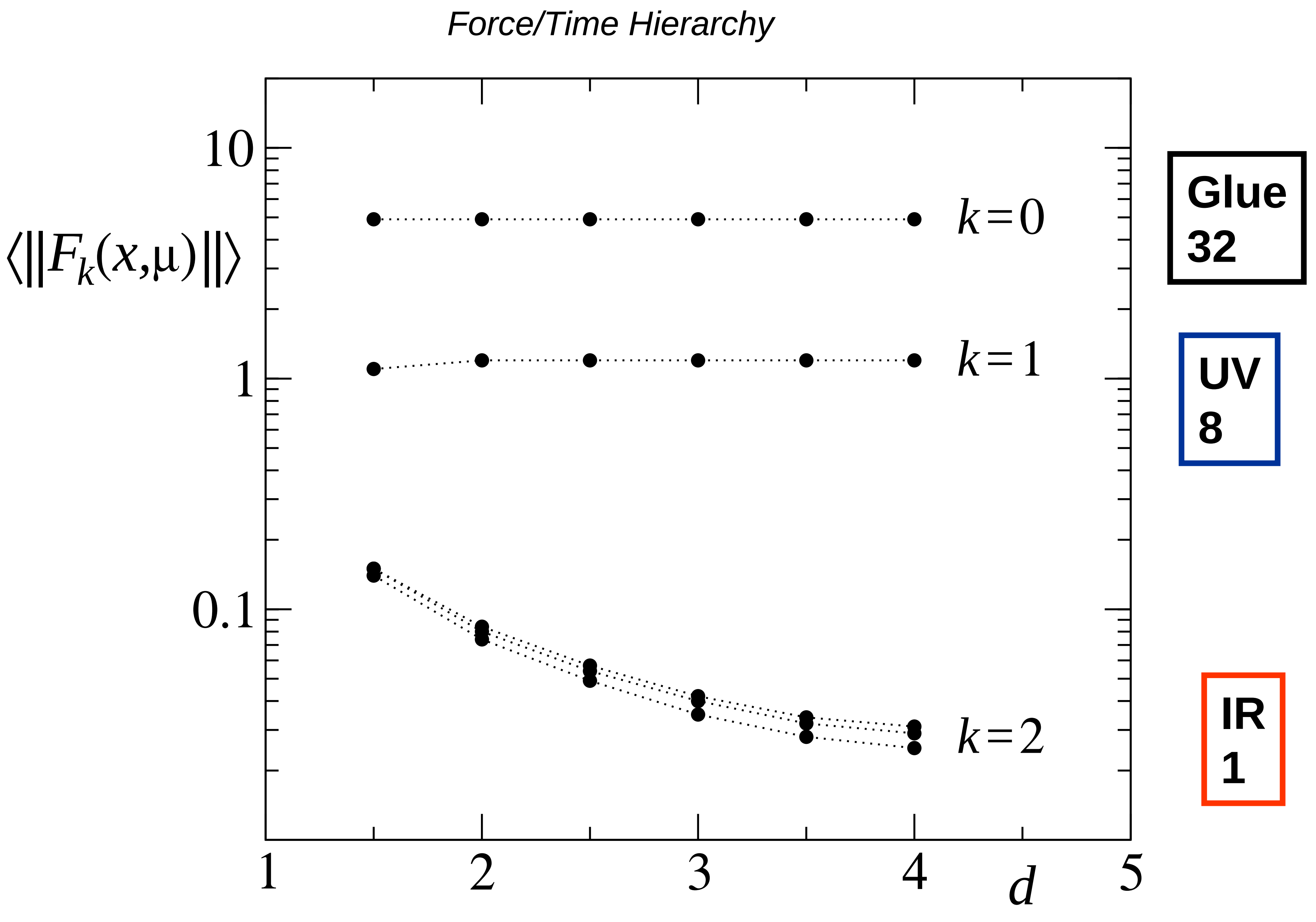
Short Distance (UV) Inverter

Even 6^4 Block with Dirichlet Boundary Condition

$$D_{bb}\Psi_b = b_b - D_{br}\Psi_r \rightarrow \Psi = D_{bb}^{-1}(b - D\Psi)$$

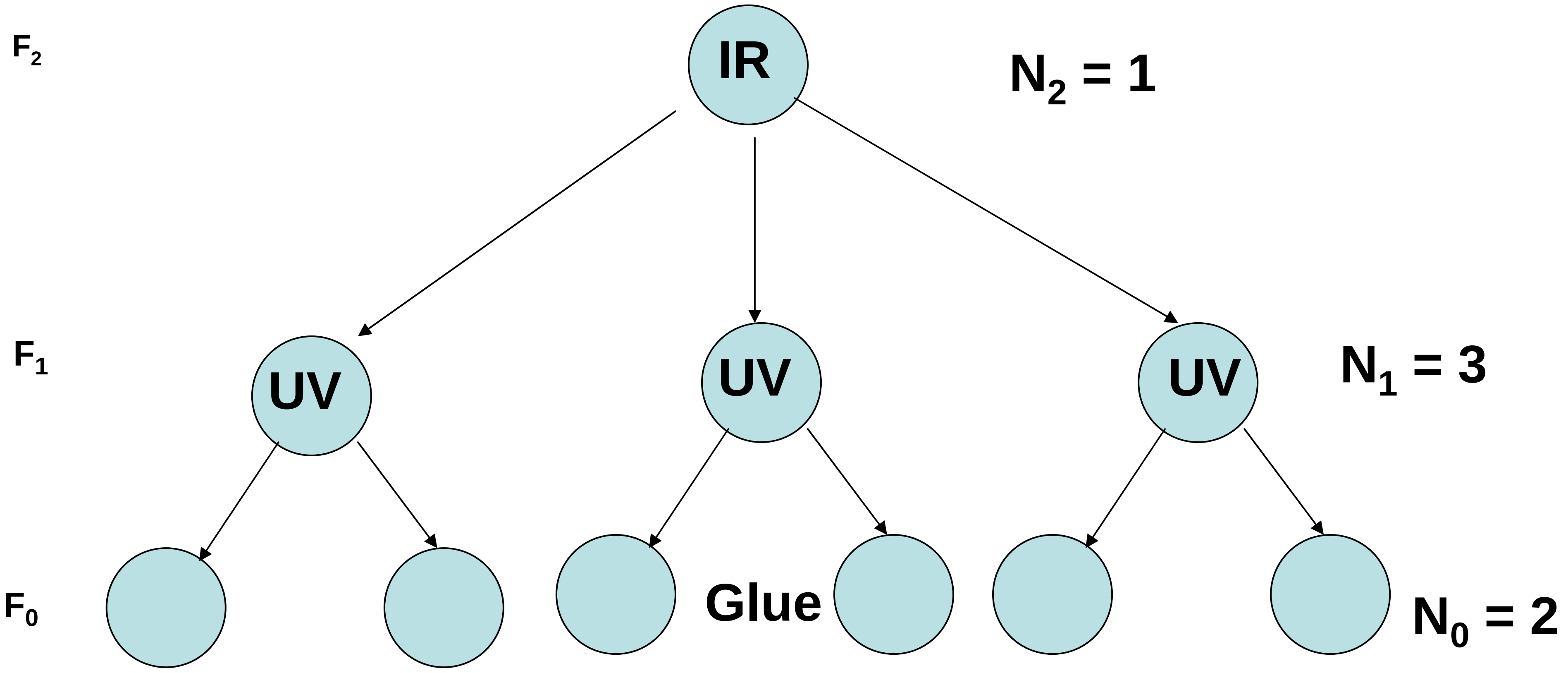


$$D_{rr}\Psi_r = b_r - D_{rb}\Psi_b \rightarrow \Psi = D_{rr}^{-1}(b - D\Psi)$$



Force/Time Hierarchy
10
1
0.1
⟨‖F_k(x,μ)‖⟩
k=0
k=1
k=2
Glue
32
UV
8
IR
1
1
2
3
4
5
d

Recursive Integrator: $\Delta t_{k+1} = N_k \Delta t_k$



$\Delta t_2 = N_1 N_0 \Delta t_0 \quad > \quad \Delta t_1 = N_0 \Delta t_0 \quad > \quad \Delta t_0$

(depth first scan of recursive tree)

Cost ν of Hasenbush vs Schwarz vs HMC

	κ	ν	ν from [8]	ν from [9]
A	0.15750	0.09(3)	0.69(29)	1.8(8)
B	0.15800	0.11(3)	0.50(17)	5.1(5)
C	0.15825	0.23(9)	0.28(9)	-

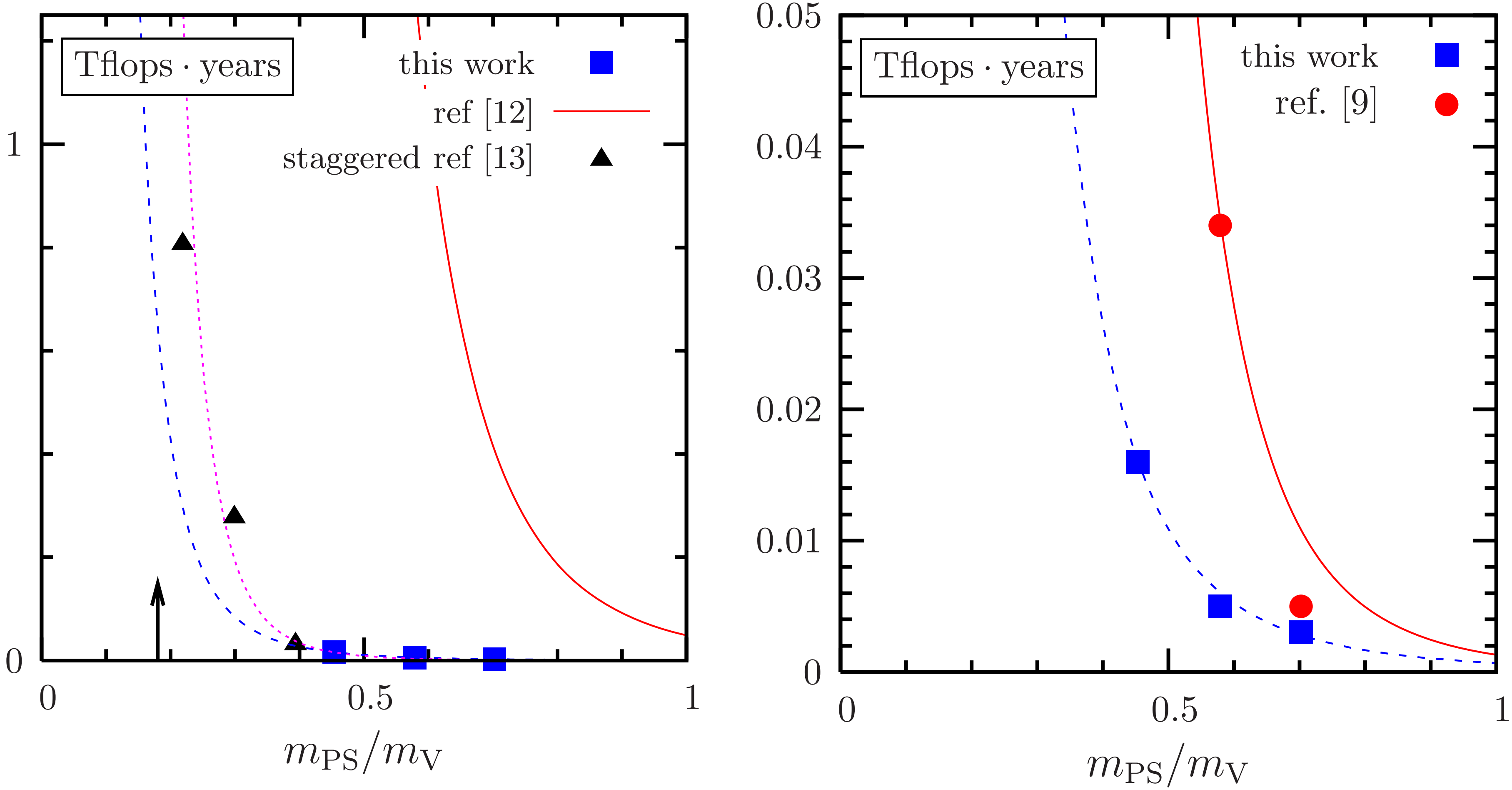
(Urbach, Jansen, Shindler, Wegner, hep-lat/0506011)

Luscher, hep-lat/0409106.

Otho, Lippert, Schilling Hep-lat/0503016

Moving the Berlin Wall

(Urbach, Jansen, Shindler, Wegner, hep-lat/0506011)



Time for 1000 configurations: $\sim 1/m^4_\pi$ equals Schwarz algorithm

2. Multi-scale Methods
Multi-grid: Old vs New

References from early 1990's

[1] Richard C. Brower, Eric Myers, Claudio Rebbi, and K. J. M. Moriarty. The multigrid method for fermion calculations in quantum chromodynamics. Print-87-0335 (IAS,PRINCETON).

[2] R. Brower, K. Moriarty, C. Rebbi, and E. Vicari. Variational multigrid for nonabelian gauge theory. In *Tallahassee 1990, Proceedings, Lattice 90* 89-93. (see HIGH ENERGY PHYSICS INDEX 29 (1991) No. 11041).

[3] Richard C. Brower, Claudio Rebbi, and Ettore Vicari. Projective multigrid method for propagators in lattice gauge theory. *Phys. Rev.*, D43:1965–1973, 1991.

[4] R. C. Brower, K. J. M. Moriarty, C. Rebbi, and E. Vicari. Multigrid propagators in the presence of disordered u(1) gauge fields. *Phys. Rev.*, D43:1974–1977, 1991.

[5] Richard C. Brower, Robert G. Edwards, Claudio Rebbi, and Ettore Vicari. Projective multigrid for wilson fermions. *Nucl. Phys.*, B366:689–705, 1991.

[6] Richard C. Brower, Claudio Rebbi, and Ettore Vicari. Projective multigrid for propagators in lattice gauge theory. *Phys. Rev. Lett.*, 66:1263–1266, 1991.

[7] R. Ben-Av et al. Fermion simulations using parallel transported multigrid. *Phys. Lett.*, B253:185–192, 1991.

[8] R. Ben-Av, M. Harmatz, S. Solomon, and P. G. Lauwers. Parallel transported multigrid for inverting the dirac operator: Variants of the method and their efficiency. *Nucl. Phys.*, B405:623–666, 1993.

[9] A. Brandt. Multigrid methods in lattice field computations. *Nucl Phys. Proc. Suppl.*, 26:137–180, 1992.

[10] A. Brandt. Multigrid methods in lattice field computations. *Nuclear Phys. B Proc. Suppl.*, 26:137–180, 1992.

[11] A. Hulsebos, J. Smit, and J. C. Vink. Multigrid inversion of the staggered fermion matrix with u(1) and su(2) gauge fields. In *Juelich 1991, Proceedings, Fermion algorithms* (QCD161:W573:1991), 161-168.

Gauge Invariant Projective Multigrid†

† **R. C. Brower, R. Edwards, C.Rebbi,and E. Vicari,**
SPECTRAL AMG
"Projective multigrid forWilson fermions", Nucl. Phys.B366 (1991) 689

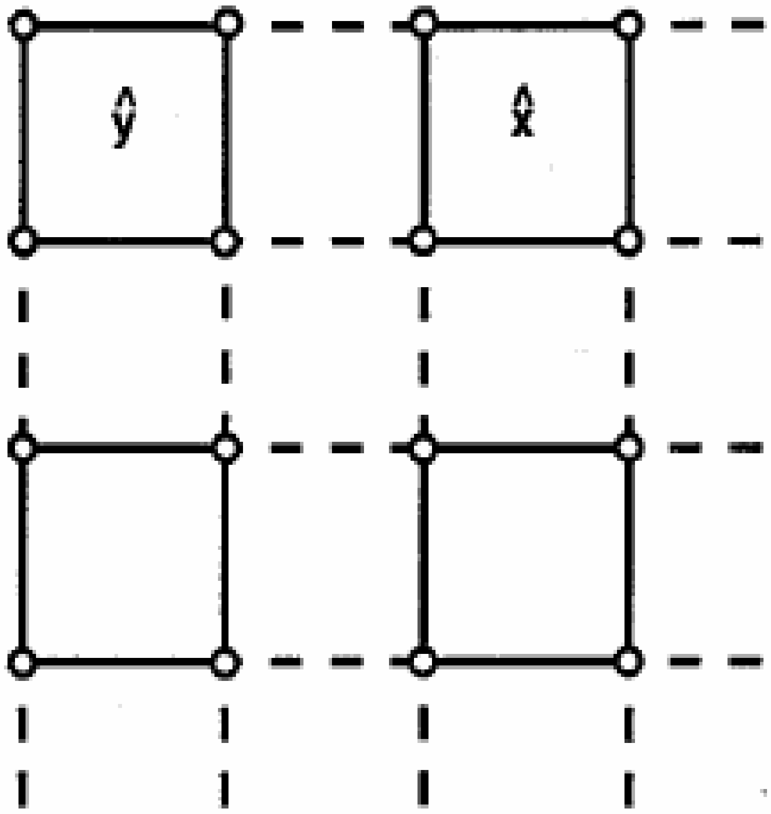
- Multigrid Scaling (a ➔ 2 a) ---- aka "renormalization group" in QCD
- Map should (must?) preserve long distance spectrum and symmetries.
- Operators $P(x_C,x_F)$ & $Q(x_F,x_C)$ should be "square" in spin / color space!

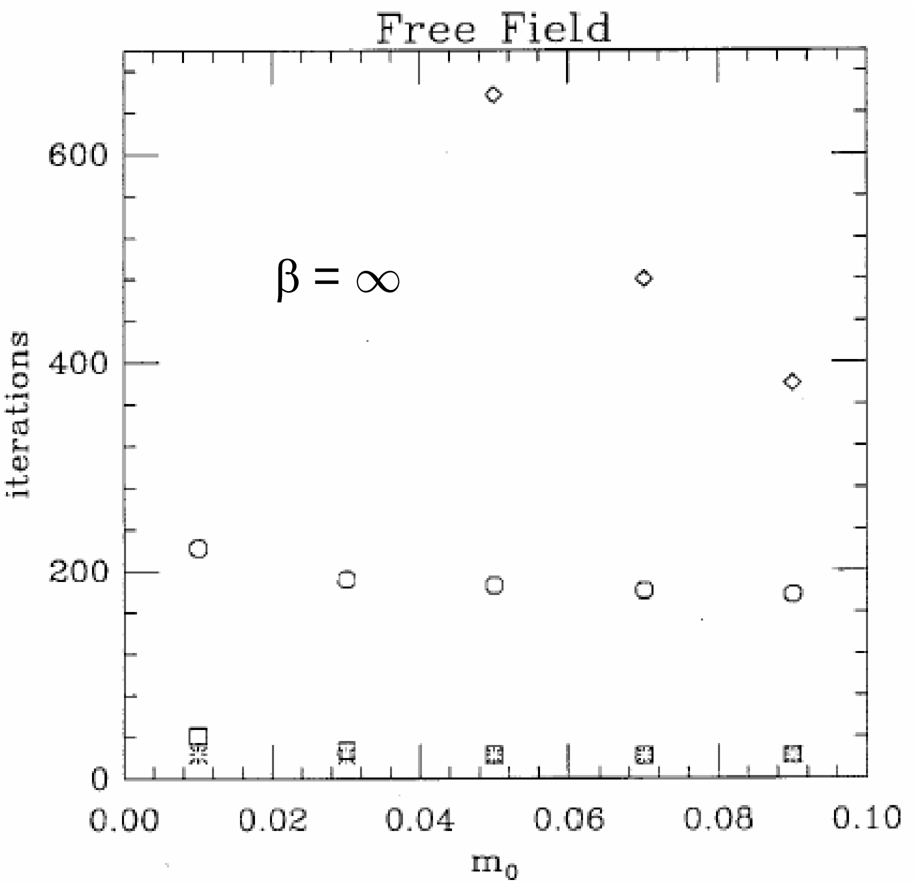
- Galerkin Example

$$A_{CC} = P\, A_{FF}\, Q \;\rightarrow\; A_{x_C,y_C} = P(x_C, x_F)\, A_{x_C,y_C}\, Q(y_C, y_F)$$

$\gamma 5$ Hermitcity constraint: $\gamma 5\, Q\, \gamma 5 = P^{\dagger}$

2x2 Blocks for U(1) Dirac

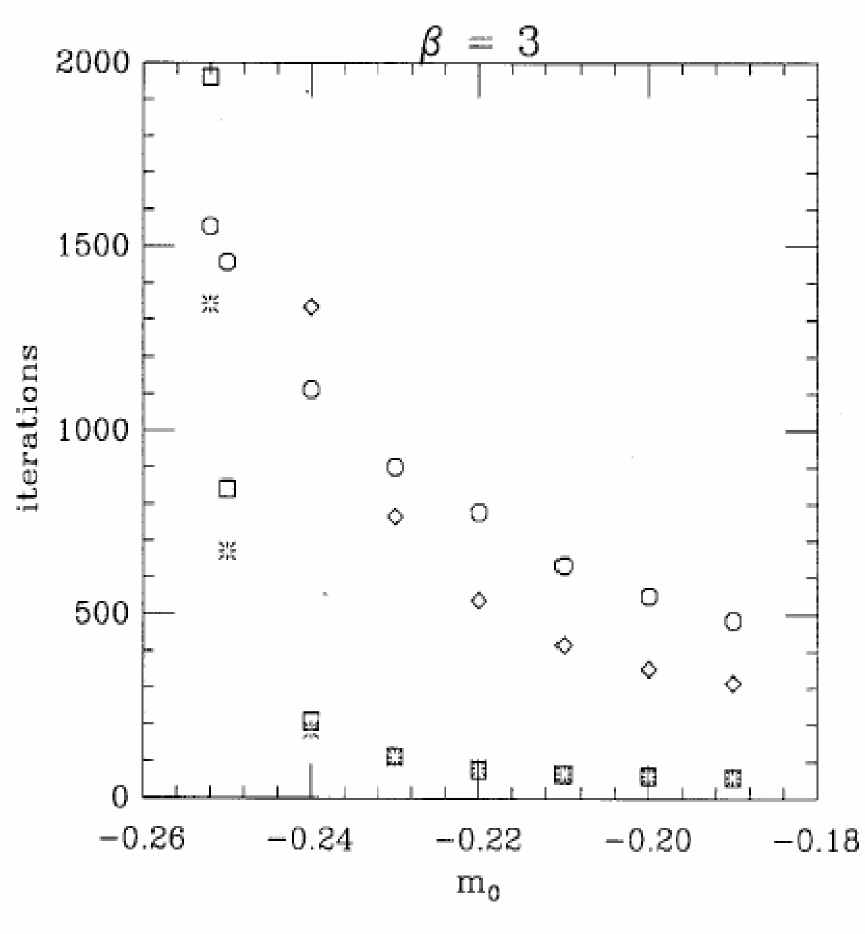


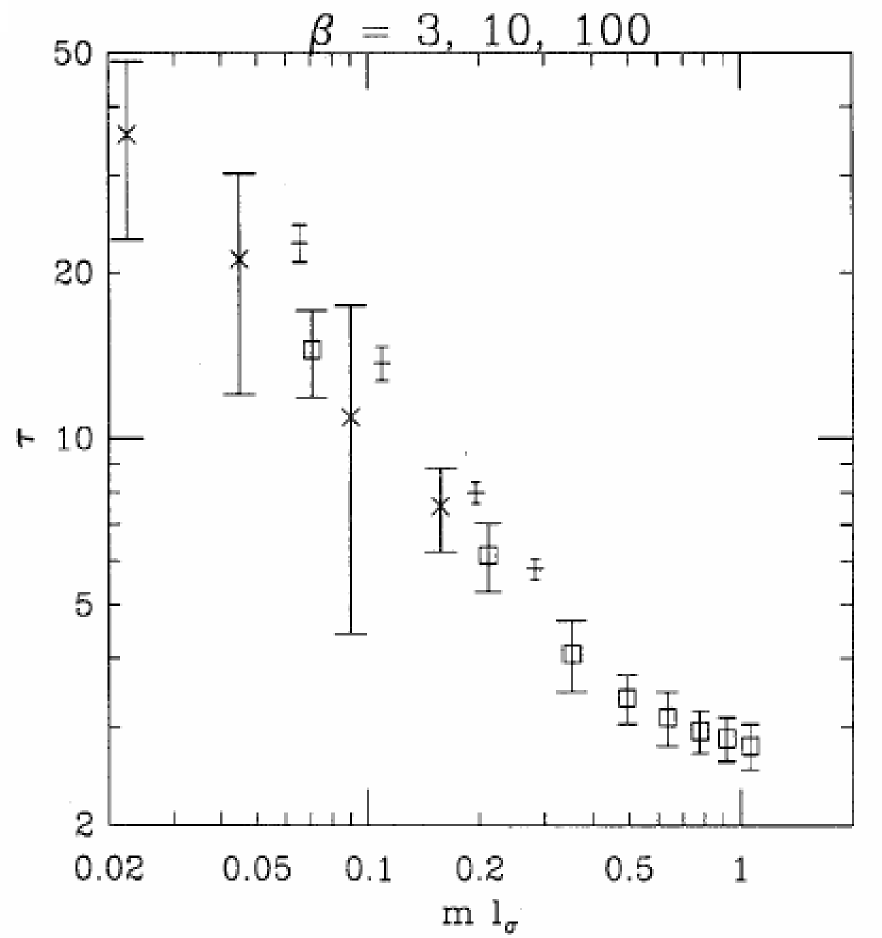


2-d Lattice,
$U_\mu(x)$ on links $\Psi(x)$ on sites

Gauss-Jacobi (Diamond), CG (circle), V cycle (square), W cycle (star)

Universal Autocorrelation: $\tau = F(m\, l_\sigma)$





Gauss-Jacobi (Diamond), CG (circle),
3 level (square & star)

β = 3 (cross) 10(plus) 100(square)

New α Algebraic Multi-grid

New non-trivial smooth aggregation adaptive algebraic Multi-grid !

- Idea is to try naïve iteration
- Start with x = random vector & iterate A x = 0
- Get global “slow mode” in near null space.
- Use cut this vector into blocks (aggregates)
  - Smooth to construct Projector. Precondition with this and see if it is still slow.
  - If so find another near null in orthogonal space if not go to the next level.
- Use Multi-grid at every stage to adaptively find the next better MG operator!

A basic adaptive AMG algorithm ...

Given B^1, select x_1

1 V–cycle on $A_1 x_1 = 0$; $\bar{B}^1 \leftarrow [B^1, x_1]$

$P_2^1 B^2 = \bar{B}^1$; $x_2 \leftarrow$ Last col of B^2

$A_2 = (P_2^1)^T A_1 P_2^1$

2 V–cycle on $A_2 x_2 = 0$; $\bar{B}^2 \leftarrow [B^2, x_2]$

$P_3^2 B^3 = \bar{B}^2$; $x_3 \leftarrow$ Last col of B^3

$A_3 = (P_3^2)^T A_2 P_3^2$

$L-1$

L

Update x_1

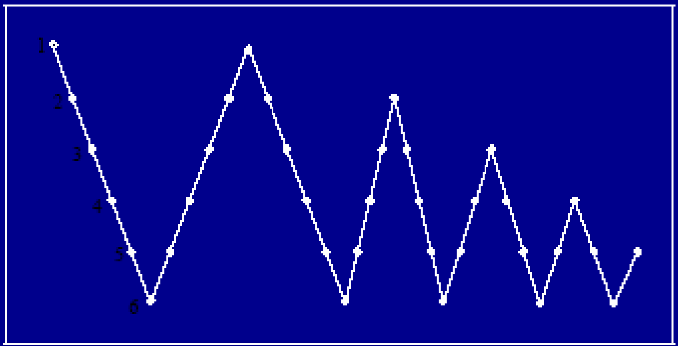
$x^1 \leftarrow P_2^1 P_3^2 \ldots P_{L-2}^{L-1} x_{L-1}$

$B^1 \leftarrow [B^1, x_1]$

Run setup

$\Downarrow$

New V–cycle



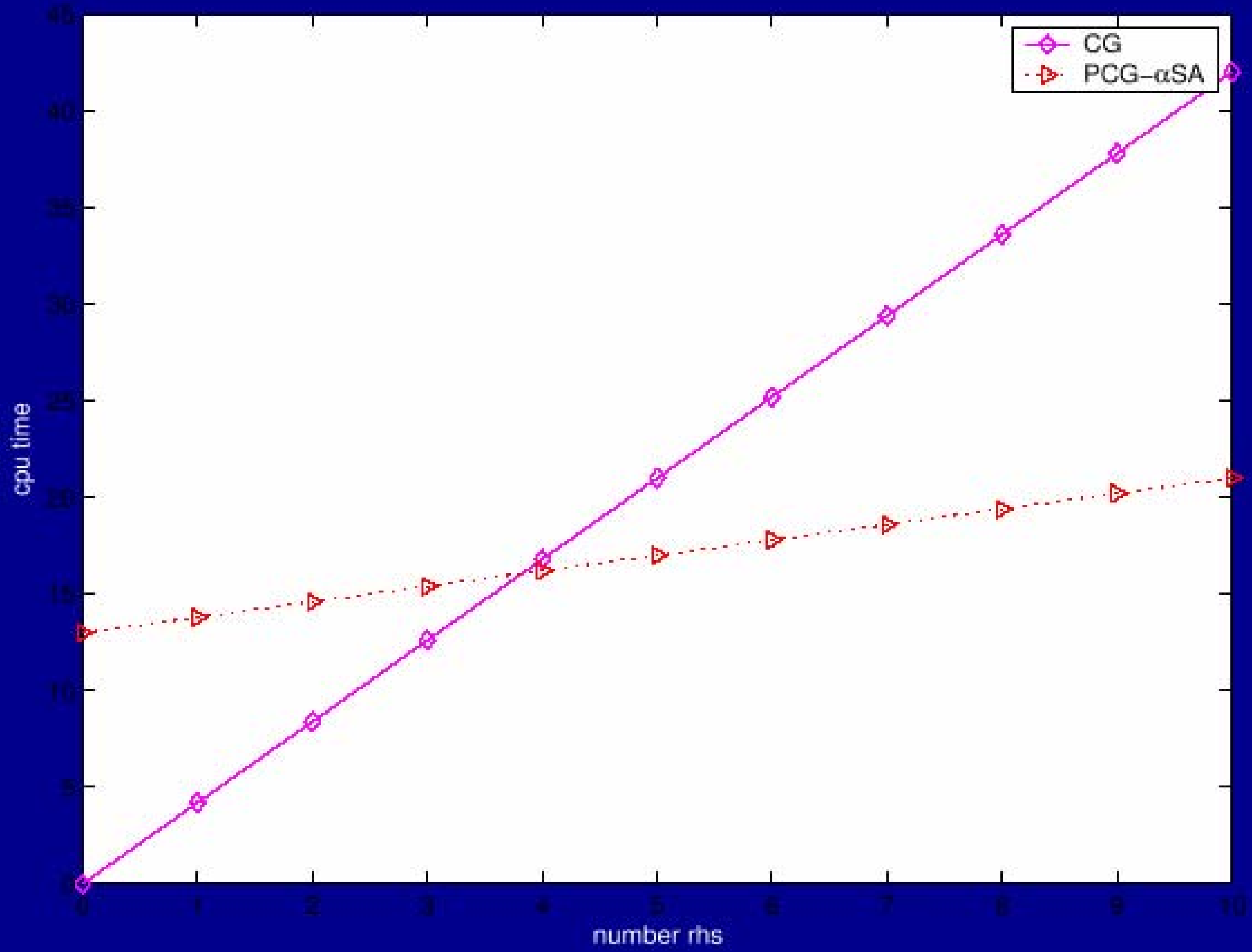
Numerical results: general case

Test on 1 flavor 2-d Schwinger Model with Wilson Fermions

β / m_g	.001	.01	.05	.1	.3
2	.37 / .99	.33 /.99	.31 /.96	.31 /.94	.31 /.85
3	.50 / .99	.42 /.98	.42 /.97	.40 /.93	.31 /.86
5	.38 / .99	.31 /.99	.31 /.96	.29 /.92	.28 /.83

Table 1: αSA-PCG / CG, $Levels = 3$, $r = 8$, $n = 16384$, $m_g := \rho - \rho_{cr}$.

Numerical results: general case



αSA-PCG, $Levels = 3$, $r = 8$, for $\beta = 3$, $N = 64$, $m_g = .01$.

3. Disconnected Diagrams

- Pseudo Fermions with low eigenvalue projection
  (Duncan,Eichetn and Yoo)
- Noisy estimators: Gaussian vs Z_2 (Dong, K.F.Liu)
- Unbiased Subtraction (Mathur and Dong)
- Solution to pollution is dilution!
- Schwarz Methods & Multigrid Methods

Q: How to take a Trace?
A: Pseudo Fermion Monte Carlo

$$Tr[\mathcal{O} D^{-1}] \sim \int d\phi \quad [D^\dagger \mathcal{O}]_{xy} \phi_y \phi_x^* \quad e^{-\phi_x^* [D^\dagger D]_{xy} \phi_y}$$

- *Can do "standard" Monte Carlo with low eigenvalue subtraction on H = γ_5 D*
- *Or "perfect" Monte Carlo – Gaussian η_x*

choose $\quad \eta_x \in Prob[\eta] \sim e^{-\eta_x^* \eta_x} \quad$ solve $\quad \phi = D^{-1} \eta$

Standard Deviation

$$\mathcal{A}(\eta) = \eta_x^* A_{xy} \eta_y \Rightarrow Tr[A] \simeq \overline{\mathcal{A}} \pm \frac{\sigma}{\sqrt{N_{sample}}}$$

where $\langle \mathcal{A}(\eta) \rangle_\eta = \overline{\mathcal{A}}$

Gaussian Noise: $\sigma^2_{Gaussian} = \overline{\mathcal{A}^2} - \overline{\mathcal{A}}^2 = \sum_{xy} A^*_{xy} A_{xy}$

Z_2 Noise: $\sigma^2_{Z_2} = \sum_{x \neq y} A^*_{xy} A_{xy}$

(with $\eta_x = (\pm 1 \pm i)/\sqrt{2} \rightarrow \eta_x^* \eta_x = 1$)

Shur Trace Decomposition

$$Tr[\mathcal{O}D^{-1}] = Tr[\mathcal{O}_{rr}D^{-1}_{rr}] - Tr[\mathcal{O}_{rr}D^{-1}D_{br}D^{-1}_{rr}] + (r \leftrightarrow b)$$

Exact Trace ?

Stochastic estimate

Can rearrange trace

$$Tr[\mathcal{O}_{rr}D^{-1}D_{br}D^{-1}_{rr}] = Tr[D_{br}D^{-1}_{rr}\mathcal{O}_{rr}D^{-1}] = Tr[\mathcal{O}^{eff}_{eo}D^{-1}]$$

Solving $D\psi = \eta_e$ gives $< \eta_b^\dagger \mathcal{O}^{eff}_{br} \psi >_\eta = < \eta_b^\dagger \mathcal{O}^{eff}_{br} D^{-1} \eta_b >_\eta$.

❑**This looks good since η_e is restricted to bndy of Λ_o**

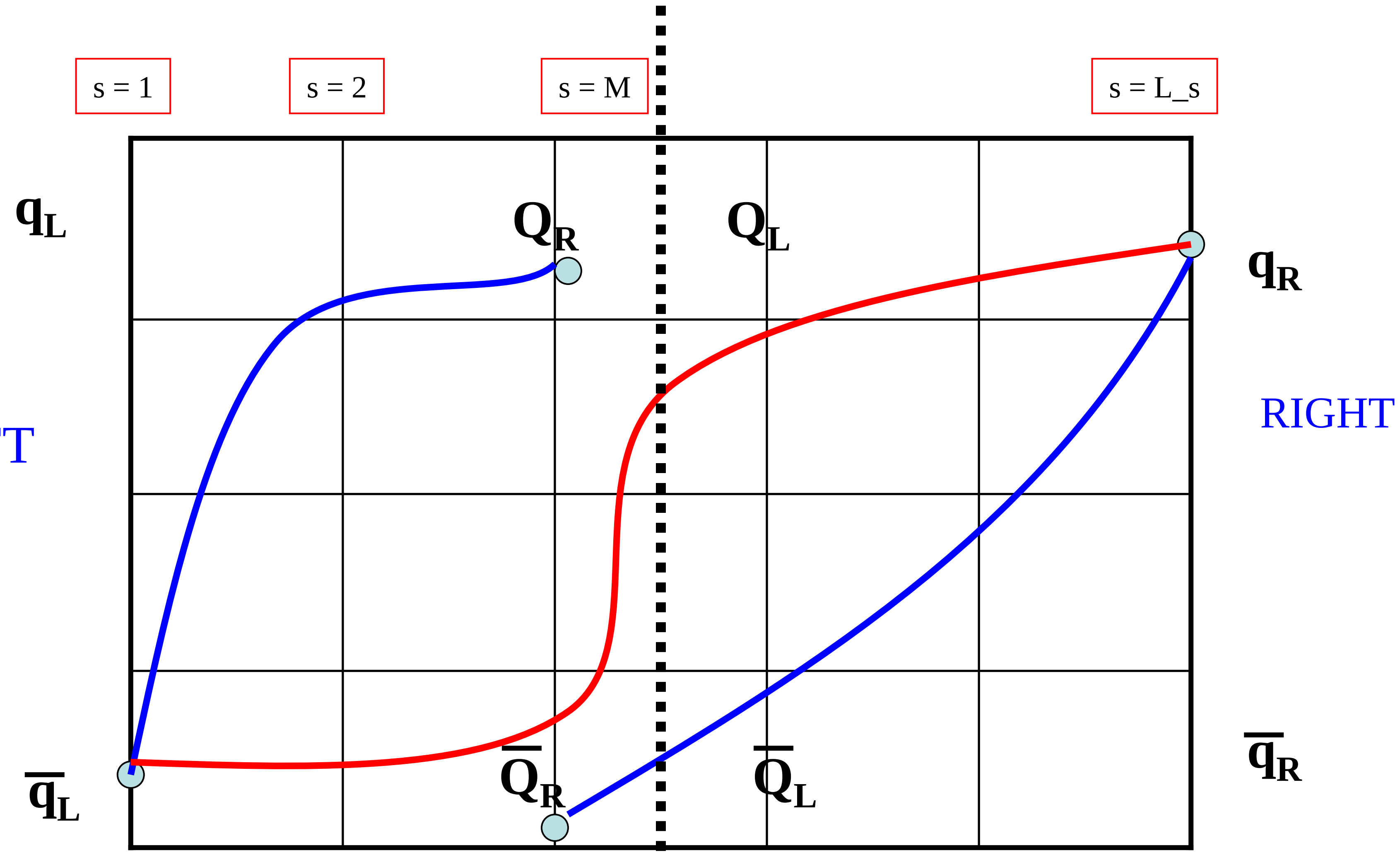
Multi-grid Trace Project

Everything can work together

BUT it is not Simple to design pre-conditioner and code efficiently!

- MG Speed up Inverse
- Amortize Pre-conditioner with multiple RHS.
- MG variance reduction at long distances.
- Unbiased subtraction at short distance.
- Low eigenvalue projection.
- Dilution.

4. Taking the 5th Dimension Seriously



$$\langle Q_x \bar{q}_y \rangle$$

$$\langle q_x \bar{Q}_y \rangle$$

$$\langle q_x \bar{q}_y \rangle = [D_{ov}^{-1}(m)]_{xy}$$

5-d Vector Current → 4-d Vector/Axial Current

$$\Delta_\mu \mathcal{J}^a_\mu(x,s) + \Delta_5 \mathcal{J}^a_5(x,s) = 0 \Rightarrow$$

Vector:
$$\Delta_\mu V^{a,DW}_\mu(x) = \sum_s \mathcal{J}^a_\mu(x,s) = 0$$

Axial:
$$\begin{aligned}\Delta_\mu A^{a,DW}_\mu(x) &= \sum_s^{L_s/2} [\mathcal{J}^a_\mu(s,x) - \mathcal{J}^a_\mu(L_s - s, x)] \\ &= -2m\,\bar{q}_x \lambda^a \gamma_5 q_x + 2\bar{Q}_x \gamma_5 \lambda^a Q_x\end{aligned}$$

Define Overlap Axial by the decent relation:

$$\langle A^{ov}_\mu(x) \psi_y \bar{\psi}_z \rangle_c \equiv \langle A^{DW}_\mu(x) q_y \bar{q}_z \rangle_c$$

What is best use of 5th Dimension?

- **Let glue be a true 5-d Gauge theory?** Improved isolation of Left and Right domain walls by "localization"?

- **Quantum Links uses replaces U_μ by fermionic bilinears.**
  (R.Brower, S.Chandrasekharan, S.Riederer, U.-J.Wiese D-Theory: Field Quantization by Dimensional Reduction of Discrete Variables **hep-lat/0309182)**

- **Should the 5-d theory be SUSY broken by domain walls boundaries ?**

- What is hadronic content of 5-d DW QCD?
  Hadronic AdS^5/CFT works pretty well. Why?

“QCD and a Holographic Model of Hadrons” Erlich, Katz, Son, Stephanov, hep-ph/05011

Observable	Measured (MeV)	Model A (MeV)	Model B (MeV)
m_π	139.6±0.0004	139.6*	140
m_ρ	775.8±0.5	775.8*	793
m_{a_1}	1230±40	1363	1256
f_π	92.4±0.35	92.4*	86.5
$F_\rho^{1/2}$	345±8	329	337
$F_{a_1}^{1/2}$	433±13	452	449
$g_{\rho\pi\pi}$	6.03±0.07	5.43	6.05

$$S = \int d^4x \int_{-L_s/2}^{L_s} ds\Big[\frac{1}{4\sqrt{f(s)}} F_{\mu\nu}F_{\mu\nu} + \frac{r^4\sqrt{f(s)}}{2R^4} F_{\mu 5}F_{\mu 5} + m_q(\ldots)\Big]$$

where $\Sigma(x) = P\,\exp[i\int_{-\infty}^{\infty} \lambda^a A_5^a(x,s)ds]$ obey Chiral L

The 5-d flavor Vector Field A has even/odd in s pieces for vector/axial hadrons, with Goldstone modes at the boundaries! (Brower, Guralnik and Tan)

FINI