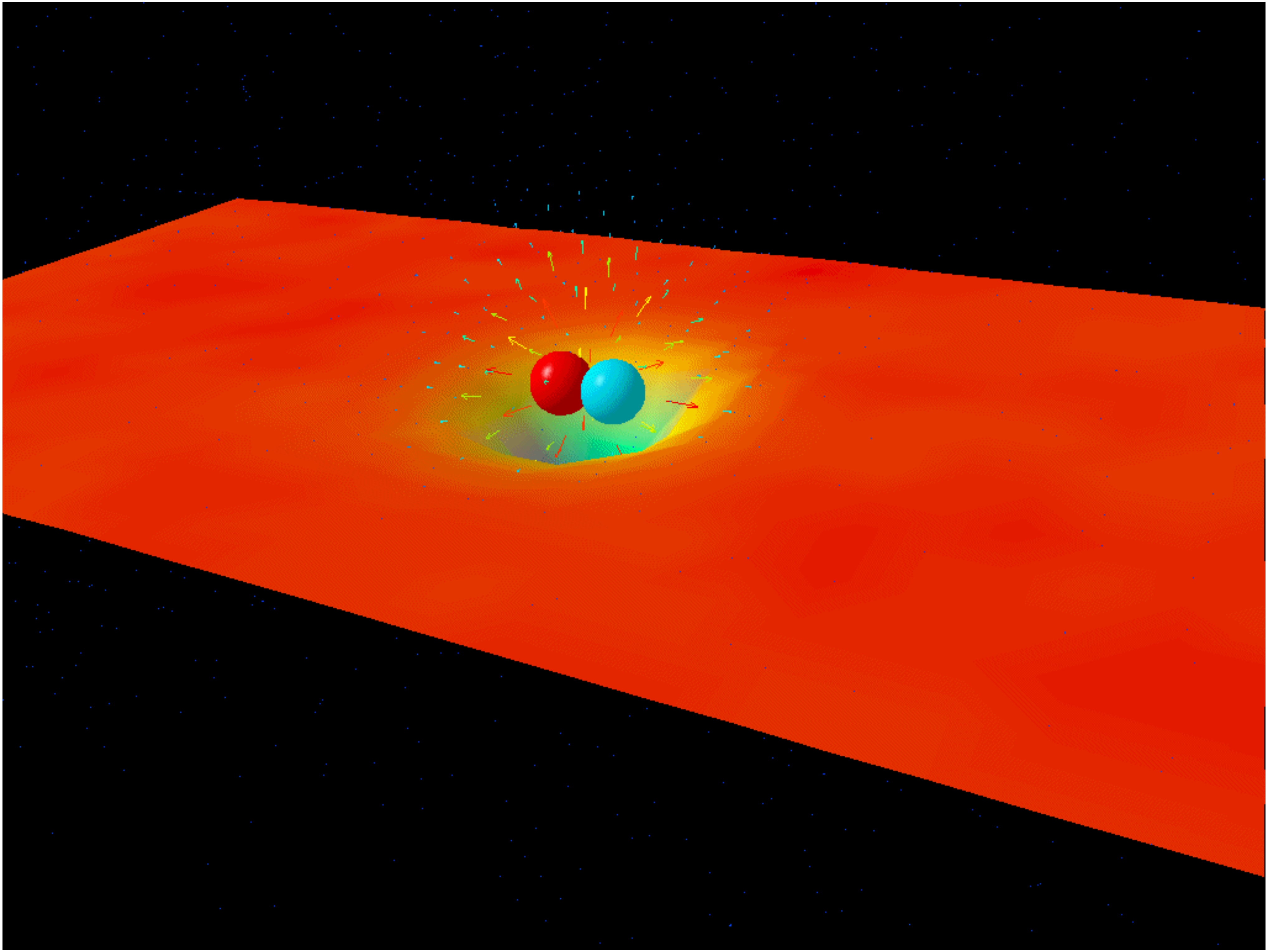
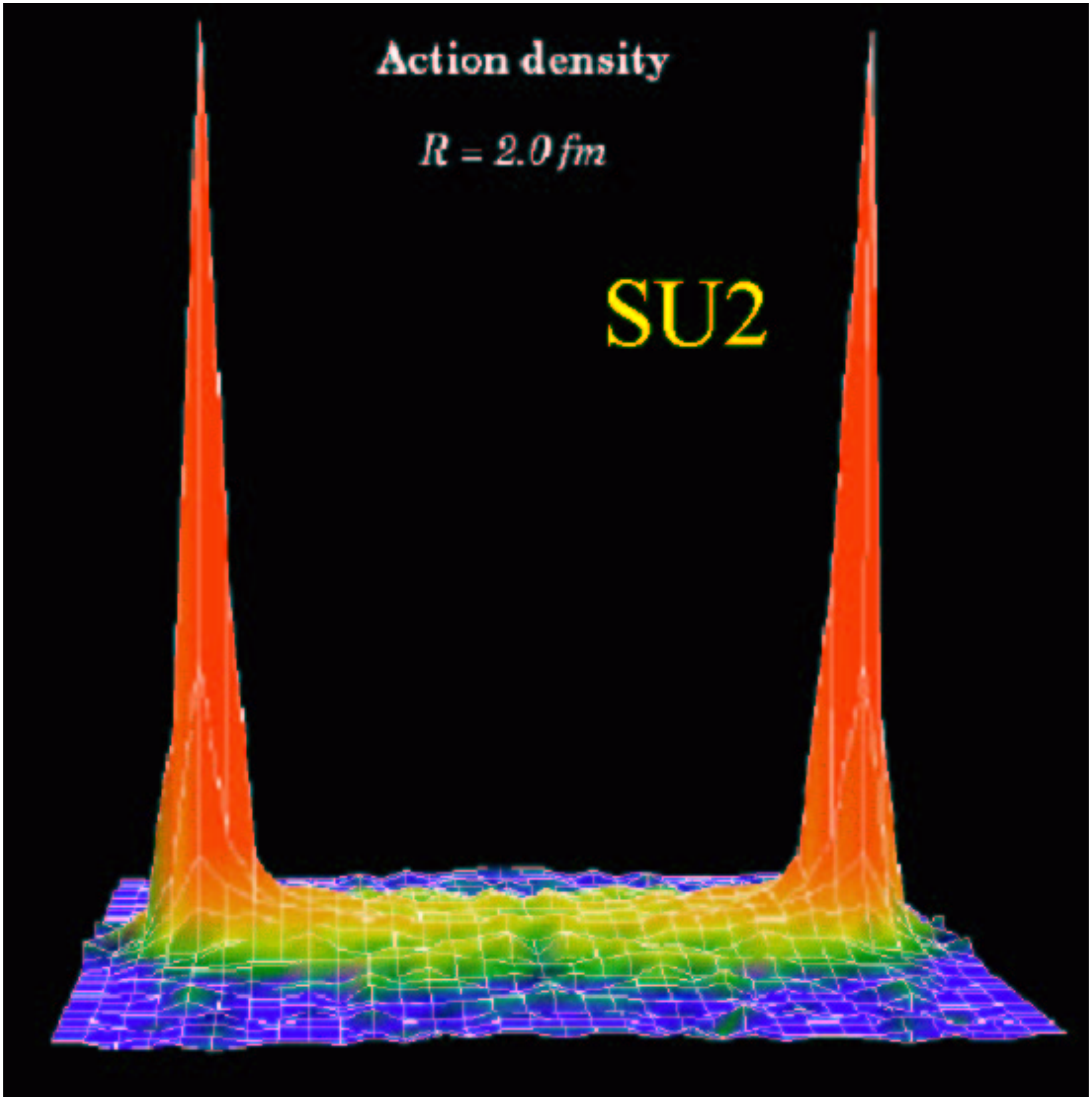
# Part II Algorithms for Multiscale

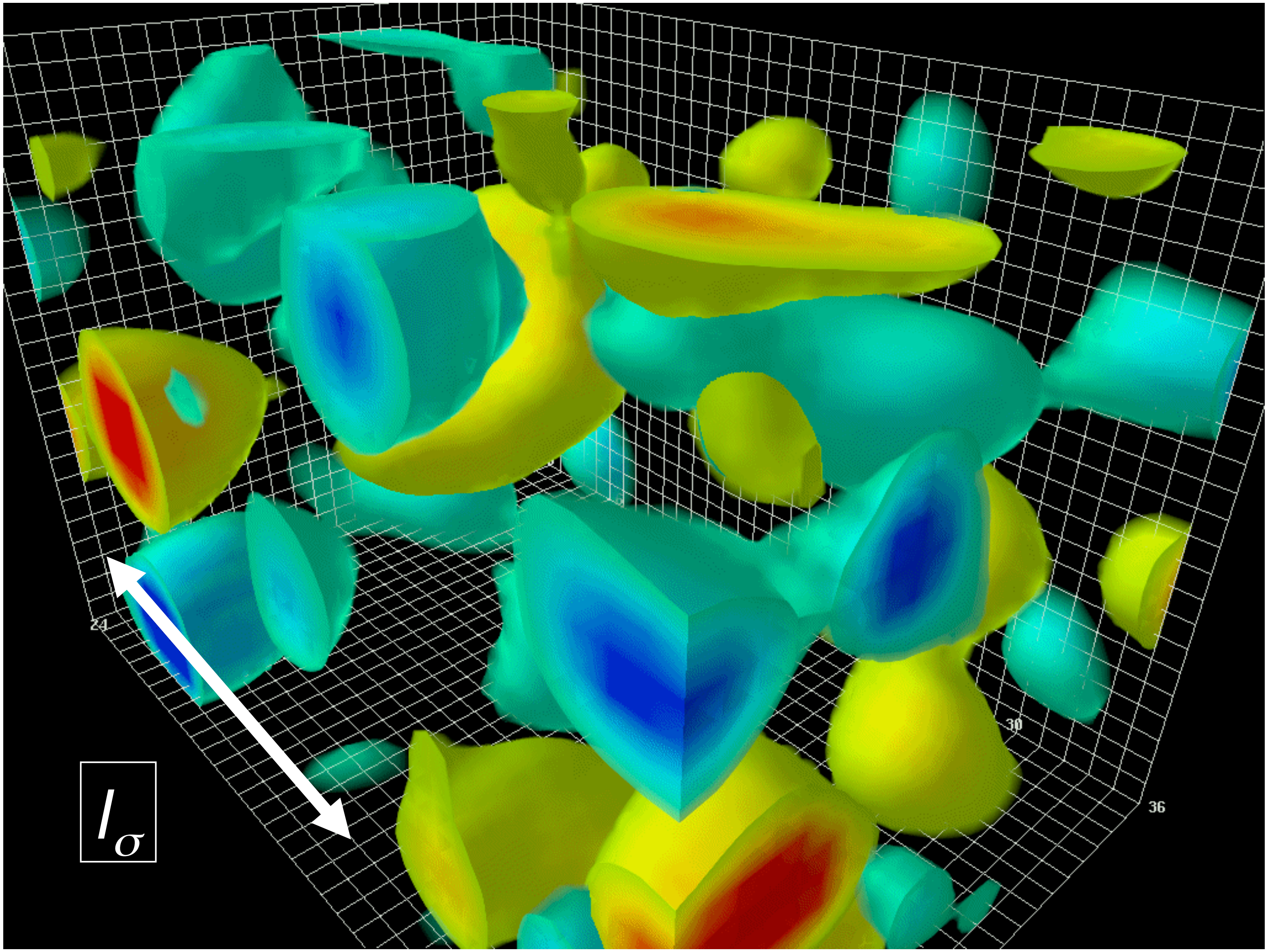
- Fundamental problem is to construct a solver for a elliptic partial derivative operators!





Action density
$R = 2.0\,fm$
SU2

# Instantons, Topological Zero Modes (Atiyah-Singer index) and Confinement length $l_\sigma$



# Introduction

- Physical Pion $\Rightarrow$ critical slowing of Dirac solver

- Finite size effects $\Rightarrow$ large lattices: O($100^4$)
  - a(lattice) << $1/M_{proton}$ << $1/m_\pi$ << L (box)
  - 0.06 fermi << 0.2 fermi << 1.4 fermi << 6.0 fermi
  - 3.3 x 7 x 4.25 $\simeq$ 100

- Solution $\Rightarrow$ Multi-scale algorithms:
  - Eigenvector Deflation (Orginos/Stathopoulos, ..)
  - Schwarz Domain Decomp. (Lüscher)
  - Adaptive Multi-grid (BU/SciDAC/TOPs collab.)

# Symmetries of Naive Discrete $D\,\psi = b$

$$\frac{\gamma_\mu U(x+a\mu,x)\psi(x+a\mu)-\gamma_\mu U(x,x-a\mu)\psi(x-a\mu)}{2a}+m\psi(x)=b(x)$$

- *Hermiticity:* $\gamma_5 D \gamma_5 = D^\dagger$
  $\gamma_5^{\,2} = 1$ and $\gamma_5 = \gamma_1 \gamma_2 \gamma_3 \gamma_4$
- *Gauge :* $U(x,x+\mu) \rightarrow \Omega_x U(x,x+\mu)\, \Omega^\dagger_{x+\mu}$
  *are unitary transformations of A*
- *Chiral:* $D = \exp[i\,\gamma_5\,\theta]\, D\, \exp[\,i\,\gamma_5\,\theta]$ *at* $m=0$

- *Scale: Only quantum fluctuations of glue break scaling at m=0 as a goes to zero. BUT the even/odd modes are decoupled so this is 16 fermions (UGLY: Staggered drops this to 4 copies!??!)*

# The Dirac PDE (for Quarks )

$$\sum_{\mu=1}^{4} \gamma_\mu^{ij} [\frac{\partial}{\partial x_\mu} - iA_\mu^{ab}(x)]\psi_{jb}(x) + m\psi_{ia}(x) = b_{ia}(x)$$

*4x4 sparse spin matrices: 4 non-zero entries 1,-1, i, -i*

*3x3 color gauge matrices*

*$x_\mu$ = ($x_1$,$x_2$,$x_3$,$x_4$) (space,time)*

## ***On a Hypercubic Lattice** **($x_\mu$ = integer, a = lattice spacing):***

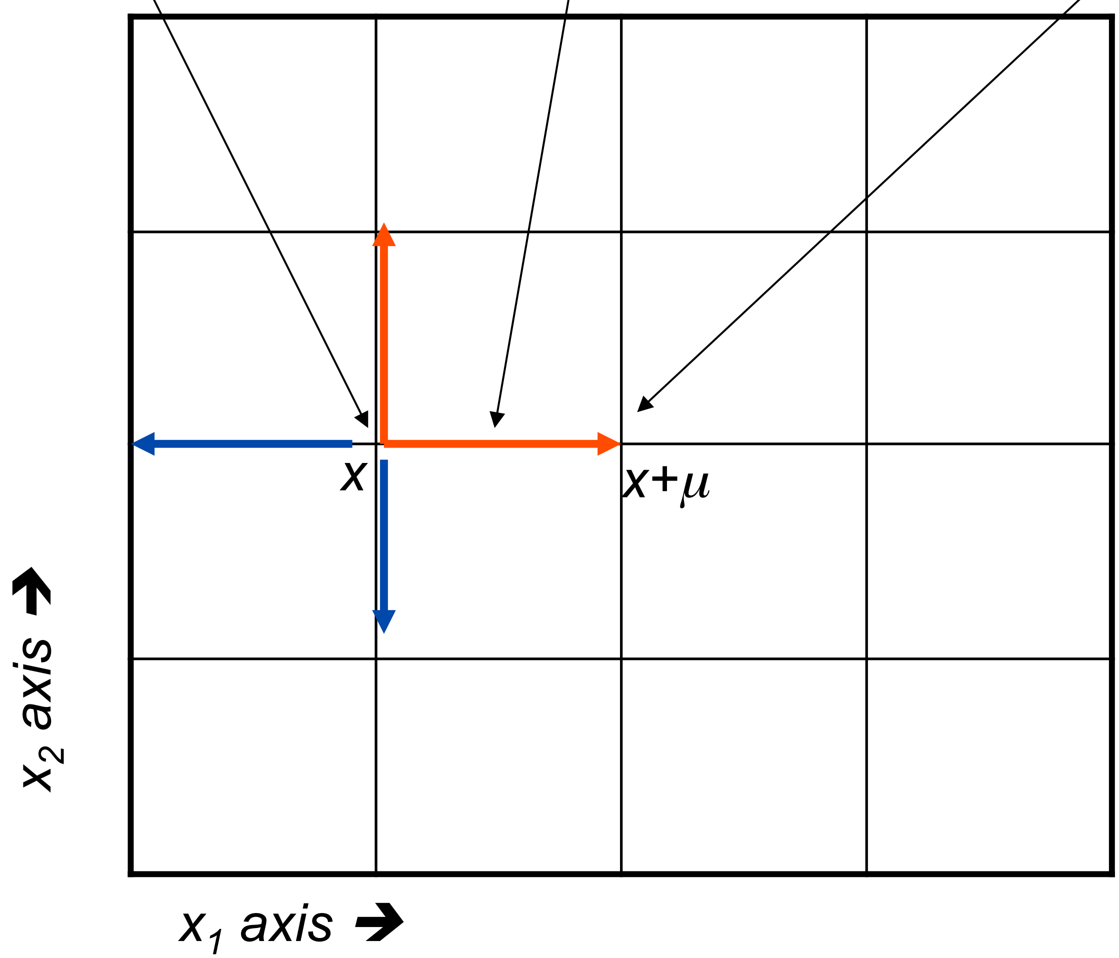
$$\sum_{\pm\mu} \frac{\pm\gamma_\mu - 1}{2} U(x, x\pm\mu)\psi(x\pm\mu) + (am+4)\psi(x) = b(x)$$

*3x3 Unitary : $U(x,x+\mu) = \exp[i\, a\, A_\mu(x)]$ and $U(x,x-\mu) = U^\dagger(x-\mu,x)$*

# Put Wilson Dirac PDE on hypercubic Lattice

$$\bar{\Psi}_{ia}(x) \; \overbrace{\frac{1-\gamma_\mu^{ij}}{2}}^{\textit{Projection Op}} \; U_\mu^{ab}(x) \; \Psi_{jb}(x+\mu)$$

*Dimension:*
*$\mu$=1,2,…,d*




*Color*
*a = 1,2,3*

*Spin*
*i = 1,2,3,4*

# Early QCD attempts in 1990's:

*See Thomas Kalkretuer hep-lat/9409008 review on "MG Methods for Propagators in LGT".*

| group | operator to be inverted | gauge field | lattice sizes |
|---|---|---|---|
| "Israel | $\not{D}+m$ | 2-$d$ $U(1)$ | $\leq 256^2$ |
| [3, 13, and references therein] | staggered fermions | 2-$d$ $SU(2)$ | $\leq 256^2$ |
| 1989–ongoing | | 2-$d$ $SU(3)$ | $\leq 128^2$ |
| "Amsterdam" | $-\not{D}^2+m^2$ | 2-$d$ $SU(2)$ | $\leq 128^2$ |
| [14, and references therein] | staggered fermions | | |
| 1990–1992 | staggered fermions | 2-$d$ $SU(2)$ | $\leq 128^2$ |
| | and Wilson fermions | | |
| "Boston" | $-\Delta+m^2$ | 2-$d$ $U(1)$ | $\leq 64^2$ |
| [7, and references therein] | | 4-$d$ $U(1)$ | $\leq 16^4$ |
| 1990–1991 | | 2-$d$ $SU(2)$ | $\leq 32^2$ |
| | $(\gamma_\mu+1)D_\mu+m$ | 2-$d$ $U(1)$ | $64^2$ |
| | Wilson fermions | | |
| [29] | $(\gamma_\mu+1)D_\mu+m$ | 2-$d$ $U(1)$ | $64^2$ |
| 1990–1992 | Wilson fermions | 4-$d$ $SU(3)$ | $16^4$ |
| "Hamburg" | $-\Delta+m^2$ | 2-$d$ $SU(2)$ | $\leq 128^2$ |
| [21, 18, 22, 23, 1, 17, 19, 20, 2, 24] | | 4-$d$ $SU(2)$ | $\leq 18^4$ |
| 1990–ongoing | $-\not{D}^2+m^2$ | 2-$d$ $SU(2)$ | $\leq 162^2$ |
| | staggered fermions | 4-$d$ $SU(2)$ | $\leq 18^4$ |

Table 1: *Overview of works on MG methods for propagators in lattice gauge theories.*

*Israel:* Ben-Av, M. Harmatz, P.G. Lauwers & S.Solomon

*Boston:* Brower, Edwards, Rebbi & Vicari

*Amsterdam:* A. Hulsebos, J Smit J. C. Vick

*Hamburg:* T. Kalkreuter, G. Mack & M. Speh

受入
91-5-116
高工研図書室

SUPERCOMPUTER
COMPUTATIONS
RESEARCH INSTITUTE

**PROJECTIVE MULTIGRID FOR WILSON FERMIONS**

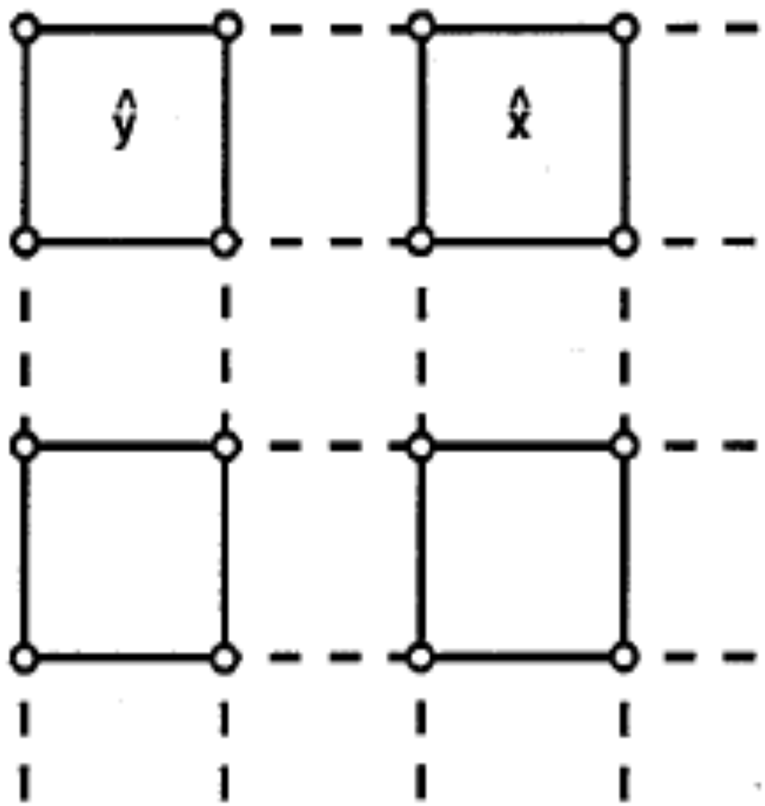
by

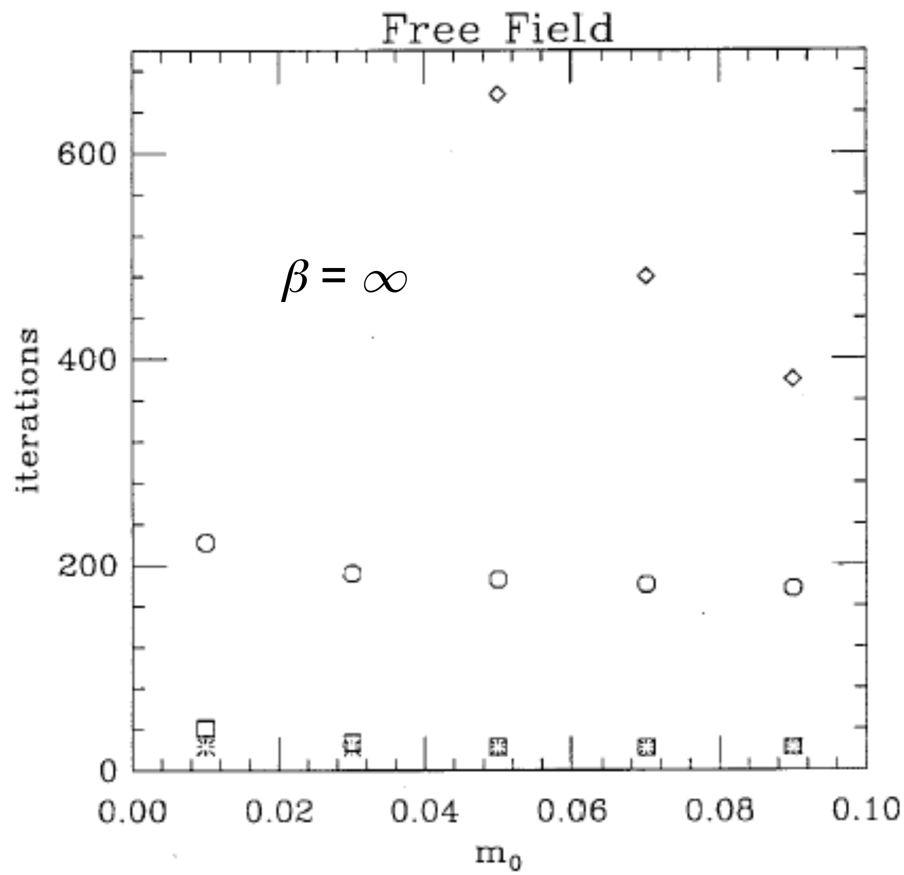
Richard C. Brower, Robert G. Edwards,
Claudio Rebbi, and Ettore Vicari

**FSU-SCRI-91-54**

*† R. C. Brower, R. Edwards, C.Rebbi,and E. Vicari,*
*"Projective multigrid for Wilson fermions", Nucl. Phys.B366 (1991) 689*
*(aka Spectral AMG, Tim Chatier, 2000)*

# 2x2 Blocks for U(1) Dirac

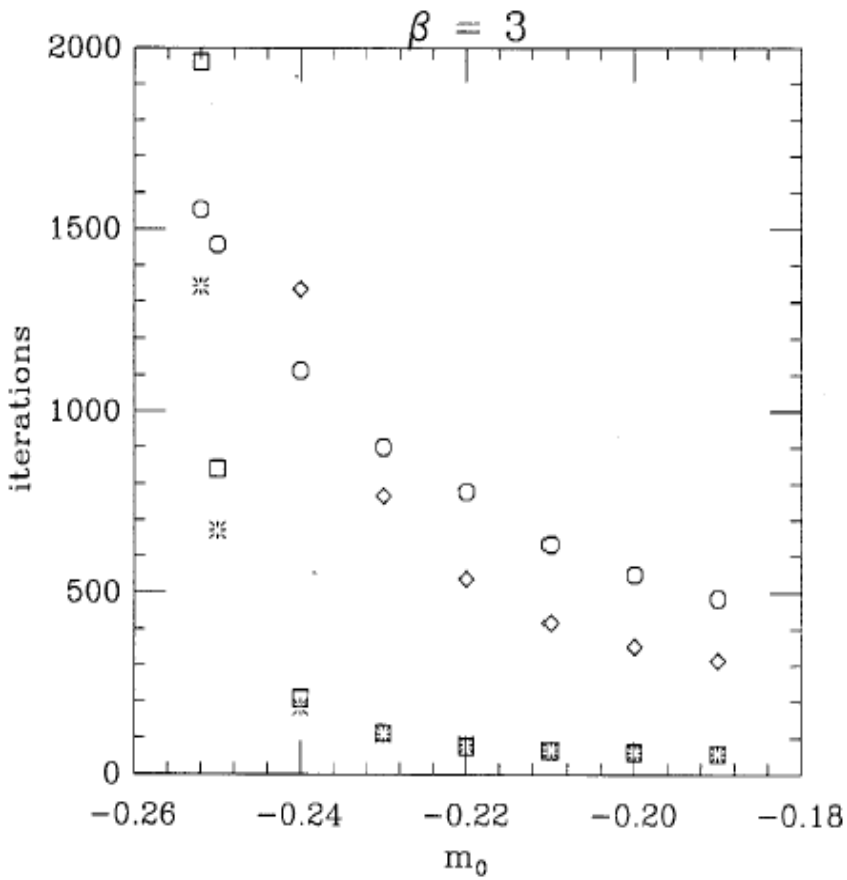


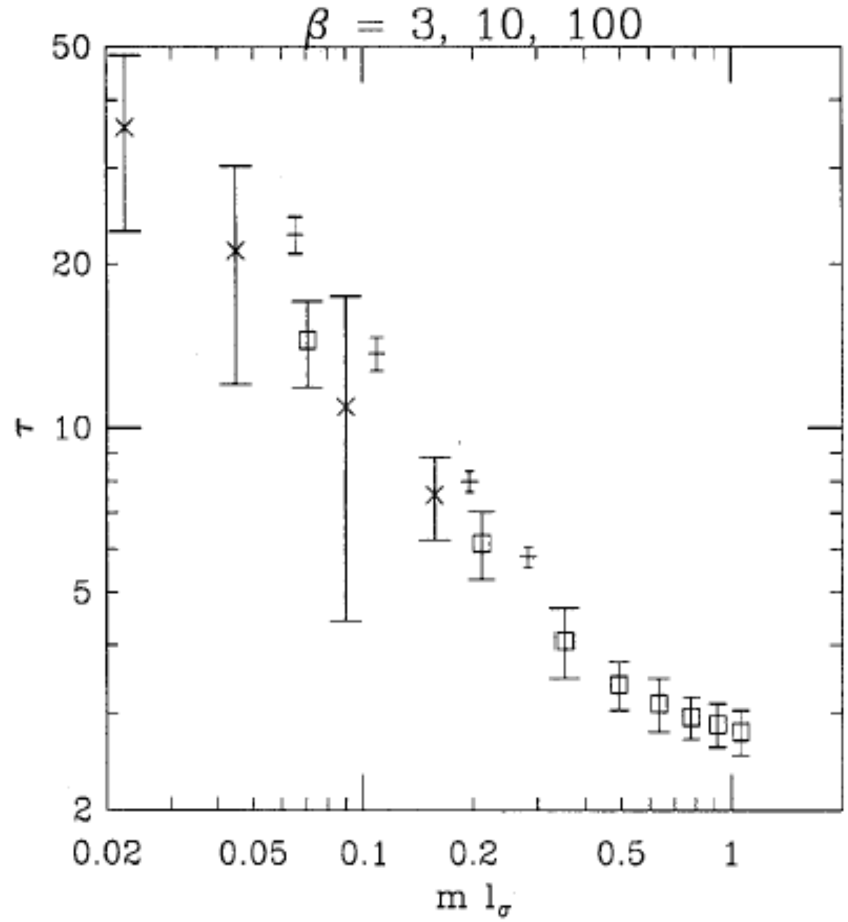

*2-d Lattice,*
*$U_\mu(x)$ on links $\Psi(x)$ on sites*

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

# Universal critical slowing: $\tau = F(m\, l_\sigma)$




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




*$\beta$ = 3 (cross) 10(plus) 100( square)*

# Failure & Success of MG attempts in 1990's : Why?

- Partial success (RG) weak coupling
- Maintain Gauge invariance
- Maintain $\gamma_5$ Hermiticity
- Local adaptive blocking
- Learn from Failure

# *MG-QCD Applied Math/Physics Collaboration!*

## Many different people (TOPS, QCD) and institutions involved in the collaboration

- CU Boulder
  - Tom Manteuffel
  - Steve McCormick
  - Marian Brezina
  - John Ruge
  - James Brannick
  - Christian Ketelsen
  - Scott MacLachlan
- Lawrence Livermore
  - Rob Falgout
- Columbia
  - David Keyes
- Boston University
  - Rich Brower
  - Claudio Rebbi
  - Mike Clark
  - James Osborn
- Penn State
  - James Brannick
  - Ludmil Zikatanov
- Tufts
  - Scott MacLachlan
- Argonne
  - James Osborn

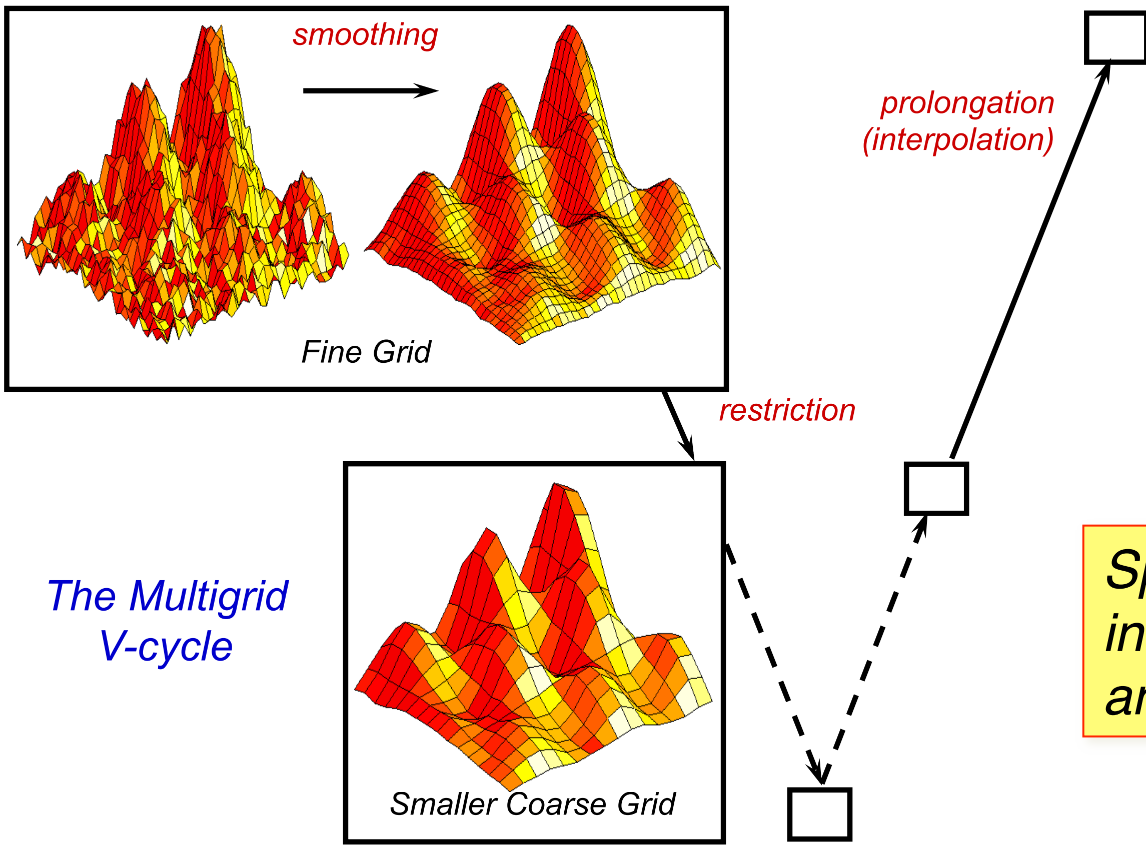
*Harvard U*
*Mike Clark*

# *Adaptive Smooth Aggregations Algebraic MultiGrid*

Slow convergence of Dirac solver is due small eigenvalues for vectors in near null subspace: $\mathcal{S}$.

$$D: \mathcal{S} \simeq 0$$

*The Multigrid V-cycle*




*Spilt the vector space into near null space $\mathcal{S}$ and the complement $\mathcal{S}_\perp$*

# 3 approaches to near null space

1. "Deflation": $N_\nu$ exact eigenvector projection

$$D_{cc} = \sum_{|\lambda|<\epsilon} \widetilde{\psi}^*_{\lambda,x} D_{x,y}\, \psi_{\lambda',y} = \sum_{|\lambda|<\epsilon} |\widetilde{\lambda}\rangle \lambda \langle\lambda|$$

$$D_{cf} = D_{fc} = 0 \quad , \quad P^\dagger = \sum_{\lambda<\epsilon} |\widetilde{\lambda}\rangle\langle\lambda|$$

2. "Inexact deflation" plus Schwarz (Lüscher)
5. Multi-grid preconditioning
   - 2 & 3 use the same splitting $\mathcal{S}$ and $\mathcal{S}_\perp$

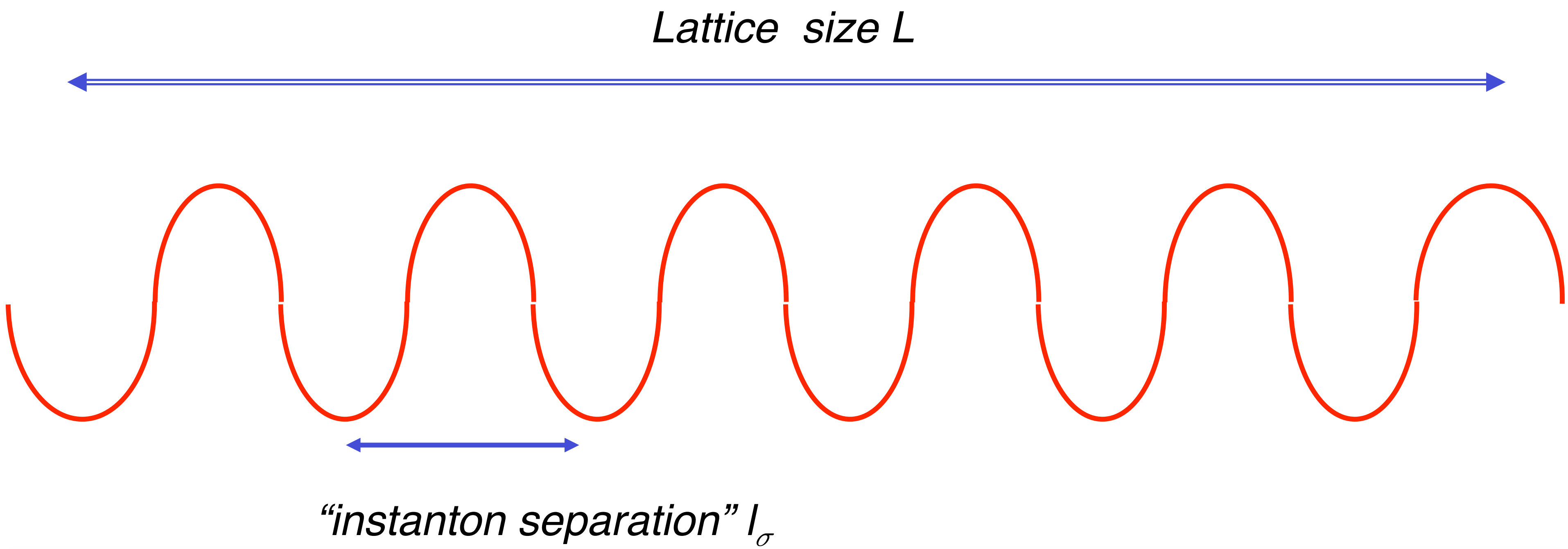
Little Dirac: $D_{cc} \equiv \hat{D} = P^\dagger D P$

Zero on coarse space

$$D_{Schur} = [D - DP\frac{1}{P^\dagger DP}P^\dagger D]_{ff}$$

# Eigenvector Deflation vs MG

## (Gedanken Toy Model with N-minima)




**EXPONENTIAL GROWTH OF NEAR NULL SPACE**

*N wells* $\Rightarrow$ *N near null states basis vectors*

$\Rightarrow$ *Size of null space* $\mathcal{S}$ = *const V where* $N \sim L^4/l_\sigma^4$

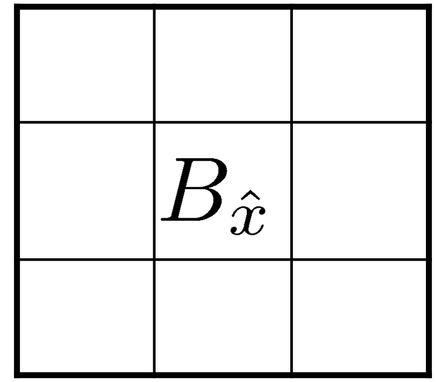
## *Choosing the Prolongator (P) and Restrictor (R = P$^\dagger$) ?*

*Relax from random vector to find near null vectors.*

$$D_{yx}\psi_x^{(s)} \simeq 0 \quad s = 1, \cdots, N_\nu$$

*Cut up on sublattice in Blocks of size 4d*

$$P_{x,\hat{x}s} = \begin{cases} \psi_x^{(s)} & x \in B_{\hat{x}} \\ 0 & \text{otherwise} \end{cases}$$




*d=2*

---

*for d=1,*
*s =1*

$$P = \left| \begin{array}{cccc} \psi_1 & 0 & 0 & \cdots \\ \psi_2 & 0 & 0 & \\ \psi_3 & 0 & 0 & \\ \psi_4 & 0 & 0 & \\ 0 & \psi_5 & 0 & \cdots \\ 0 & \psi_6 & 0 & \\ 0 & \psi_7 & 0 & \\ 0 & \psi_8 & 0 & \\ \cdots & \cdots & & \end{array} \right|$$

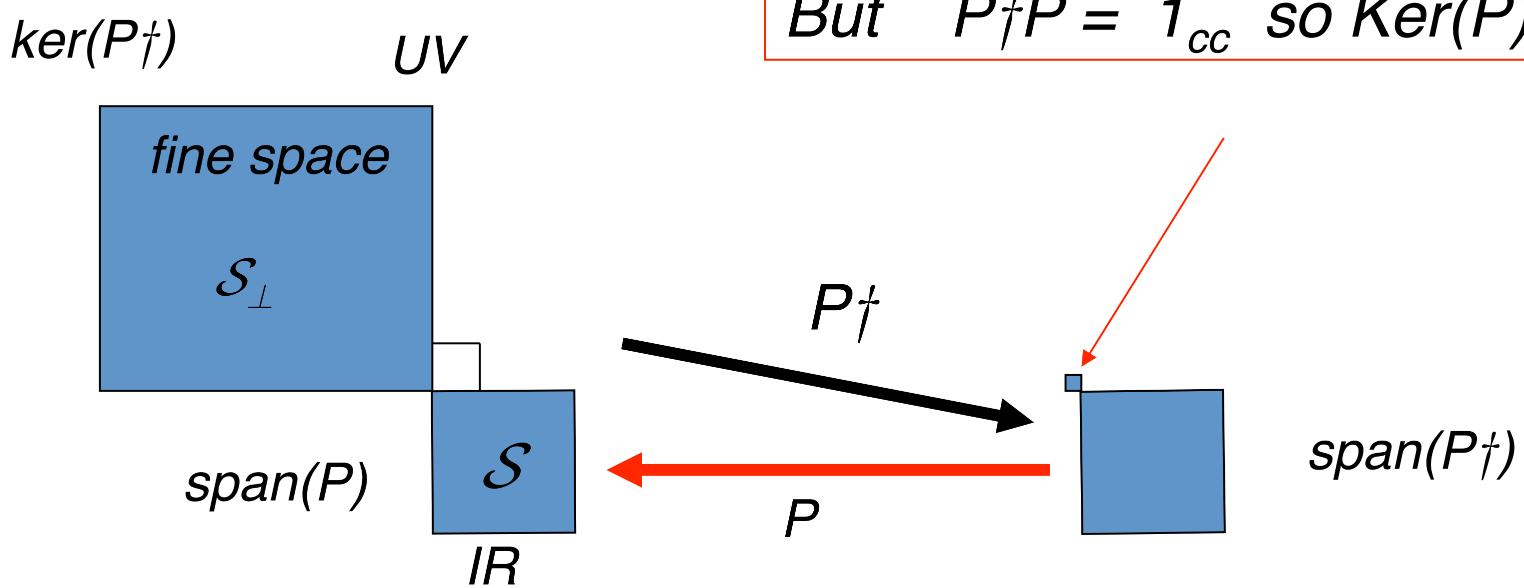
Coarse approximation:

$$D_{cc} \equiv \hat{D} = P^\dagger D P$$

# $P^\dagger$:fine $\rightarrow$ coarse ( non-square matrix[†])

*(fine lattice vector space)* *(coarse lattice vector space)*

*But* $P^\dagger P = 1_{cc}$ *so* $Ker(P) = 0$




$\mathcal{S}$ = span(P) = Image(P†)

rank(P) = rank(P†) = dim($\mathcal{S}$) = $N_\nu N_B = 2N_\nu L4/44$

† *See Front cover of Gilbert Strang's undergraduate text !*

# Oblique Projector  Algebra of splitting

$$\hat{D} = D_{cc} = P^\dagger D P \qquad D_{Schur} = [D - DP\frac{1}{P^\dagger D P}P^\dagger D]$$

$$P^\dagger D_{Schur} = D_{Schur} P = 0$$

But P2 ≠ P is not a "proper projection operator" --

The  projectors operator (Π^2 = Π ) are:

$$\Pi_L^\dagger = DP\frac{1}{P^\dagger D P}P^\dagger \quad , \quad \Pi_R = P\frac{1}{P^\dagger D P}P^\dagger D$$

$$D_{Schur} = D - \Pi_L^\dagger D \Pi_R = (1 - \Pi_L^\dagger) D (1 - \Pi_R)$$

Lüscher's  "oblique"  projectors are:  $P_L$ = 1- $\Pi\dagger_L$ and  $P_R$ = 1 - $\Pi_R$

# 2-level Multigrid Cycle (simplified)

- *Smooth:* $x' = (1 - D)\, x + b \quad \Rightarrow r' = (1 - D)\, r$
- *Project:* $D_c = P^\dagger D P \quad \& \quad r_c = P^\dagger r$
- *Solve:* $A_c e_c = r \quad \Rightarrow \quad e_c = A^{-1}{}_c P^\dagger r$
- *Prolongate* $e = P\, e\_c$
- *Update* $x' = x + e \quad \Rightarrow r' = b - D(x + e)$
  $= [\, 1 - D\, P\, (P^\dagger D\, P)^{-1} P^\dagger]\, r$

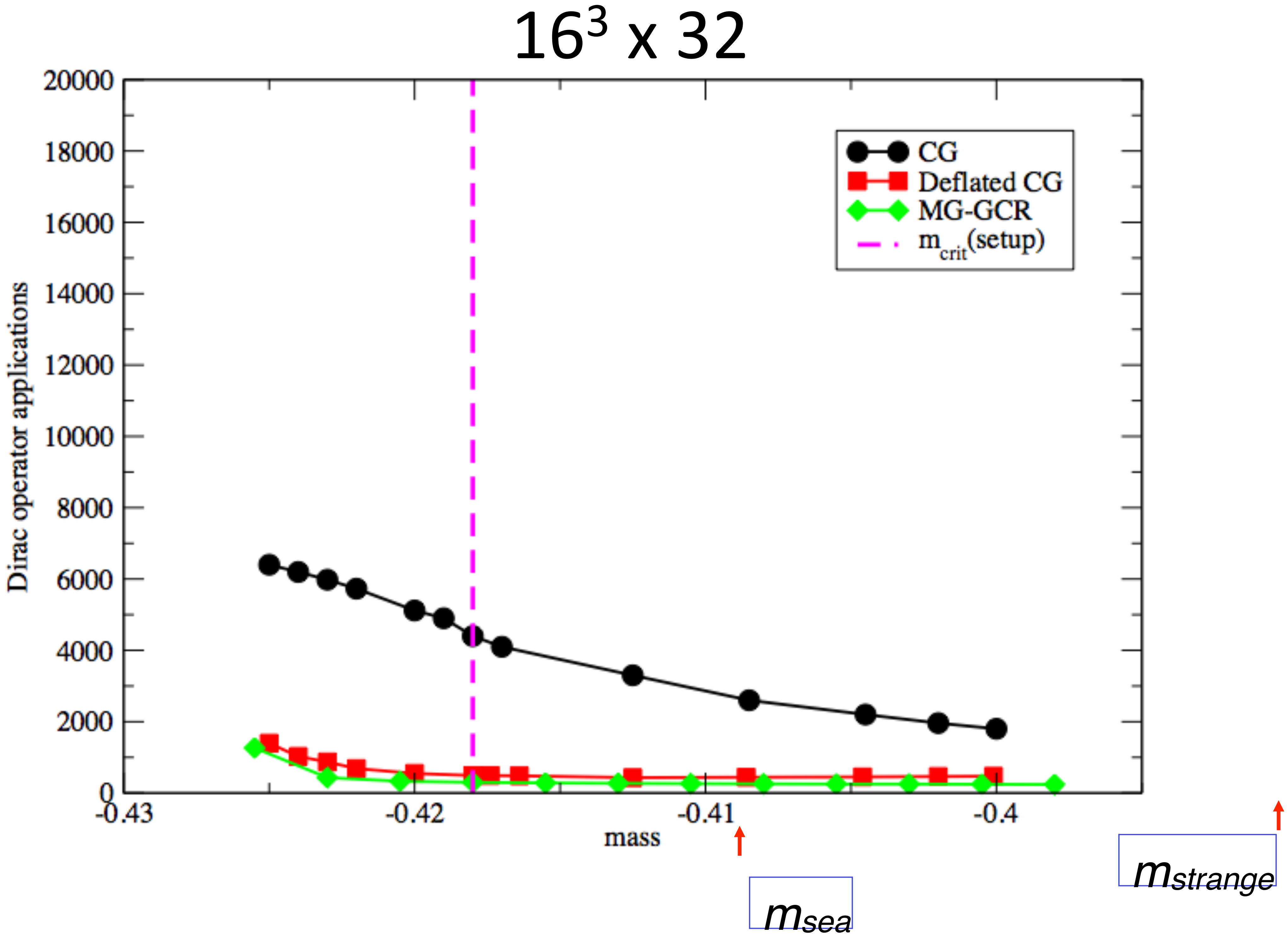
*RESULT: D is preconditioned by M = P (P† D P)-1 P†*

$M\text{-}1\, D\, x = M\text{-}1\, b \quad \Rightarrow r' = (\, 1 - D\, M\text{-}1\, )$
$r$

Note since $P\dagger\, r' = 0 \ \Rightarrow$ full (exact) deflation on $\mathcal{S}$

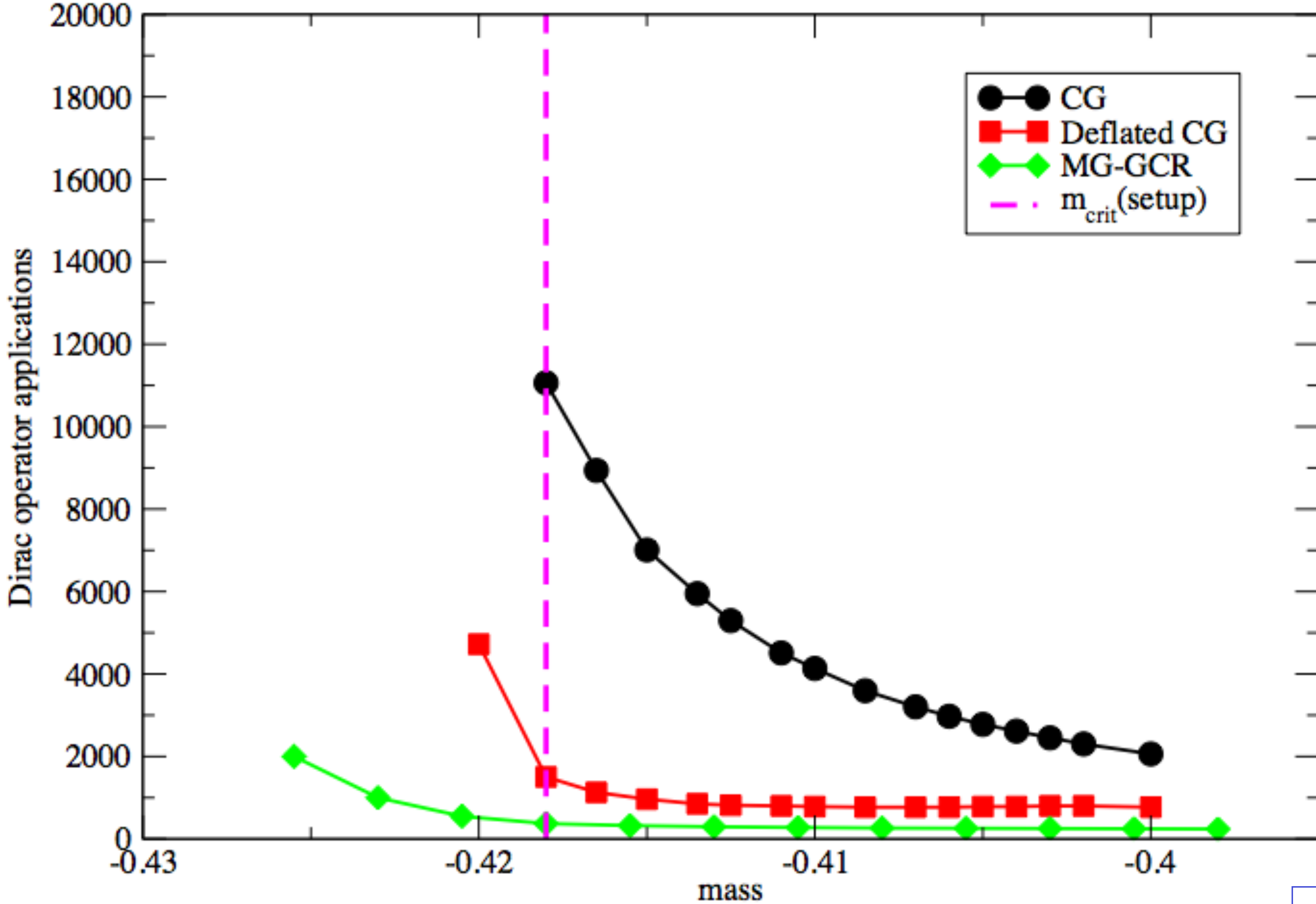
*Test of MG algorithm in 4-d Wilson*

- Devil is in the details!
  - Rigorous MG proofs for normal equation ($D^\dagger D \psi = b$)
  - But must block D to avoid higher complexity.
  - Multigrid is recursive to multi-levels.
  - Must preserves Gauge invariance and $\gamma_5$ ( $[\gamma_5, P] = 0$)
- Current benchmarks for Wilson-Dirac Operator:
  - Asym $V = 16^3 \times 64$, $24^3 \times 64$, $32^3 \times 96$ ($N_f = 2$, 400MeV pion)
  - $N_\nu = 20$ null vectors ${\psi^s}_x$ with 4th order MR with subsetrefinement.
  - MG Blocks = $4^4 \times N_c \times 2$ and 3 level V MG cycle with smoothing etc,etc
  - pre and post-smoothing is done by Minimum Residual.
  - Entire cycle is used as preconditioner in CG



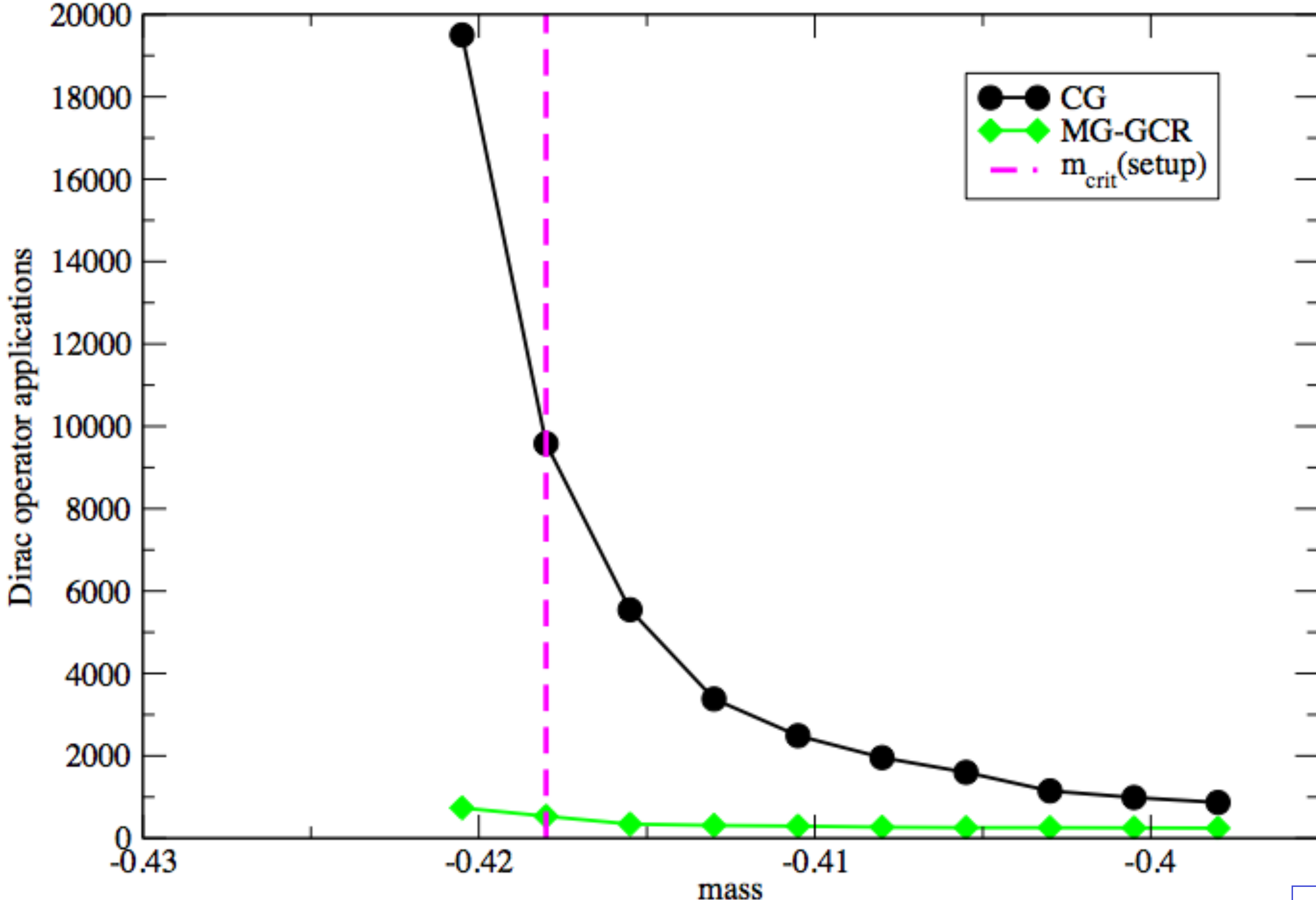
16³ x 32
Dirac operator applications
mass
CG
Deflated CG
MG-GCR
m_crit(setup)
m_sea
m_strange

# $24^3$ x 64




$m_{strange}$

# $32^3$ x 96




$m_{strange}$

# Lack of Critical slowing down:

CG iteration count is insensitive to quark mass and lattice volume!

$m_s$ (- 0.38922)

Lattice volumes

| Mass: | $16^3 \times 64$ | $24^3 \times 64$ | $32^3 \times 96$ |
|---|---|---|---|
| -.3980 | 40 | 40 | 41 |
| -.4005 | 41 | 41 | 42 |
| -.4030 | 42 | 42 | 43 |
| -.4055 | 42 | 43 | 43 |
| -.4080 | 43 | 44 | 45 |
| -.4105 | 44 | 46 | 49 |
| -.4130 | 45 | 49 | 52 |
| -.4155 | 47 | 54 | 57 |
| -.4180 | 50 | 62 | 89 |

*Small increase is probably not significant?*

physical $m2_\pi$

Chiral limit: $m2_\pi = 0$

# Implementation Details

- Mixed precision
  - Outer GCR solver on fine level in double precision
  - MG preconditioner and all levels below in single precision
  - Comparison to mixed precision Krylov methods (iterative refinement)

- Implemented in DOE SciDAC Lattice QCD libraries
  - QDP/C QCD Data Parallel library
  - Multi-lattice support and improved arbitrary size dense matrix support
  - Optimized for BG/P, x86

# Fine and coarse operators

- MG normally done on Hermitian positive definite systems ($D^\dagger D$)
  - Coarse operator constructed from Galerkin prescription $R = P^\dagger$, $A_c = P^\dagger A P$
  - Increases complexity of coarse operator (has 2-hop corner terms)
- Instead using just D
  - Want R to be rich in low left-modes
  - For $\gamma_5$-Hermitian operator can set $R = P^\dagger \gamma_5$
- Solving Wilson-clover operator
  - Using even-odd preconditioning on fine system
  - $D\,x = b \rightarrow D_p\, x_p = b \rightarrow D_r\, x_r = b_r$
  - Construct coarse operator from $D_p$ then construct reduced operator
  - $D_p$ no longer $\gamma_5$-Hermitian, but use same R anyway

$$D = \begin{pmatrix} D_{ee} & D_{eo} \\ D_{oe} & D_{oo} \end{pmatrix}$$

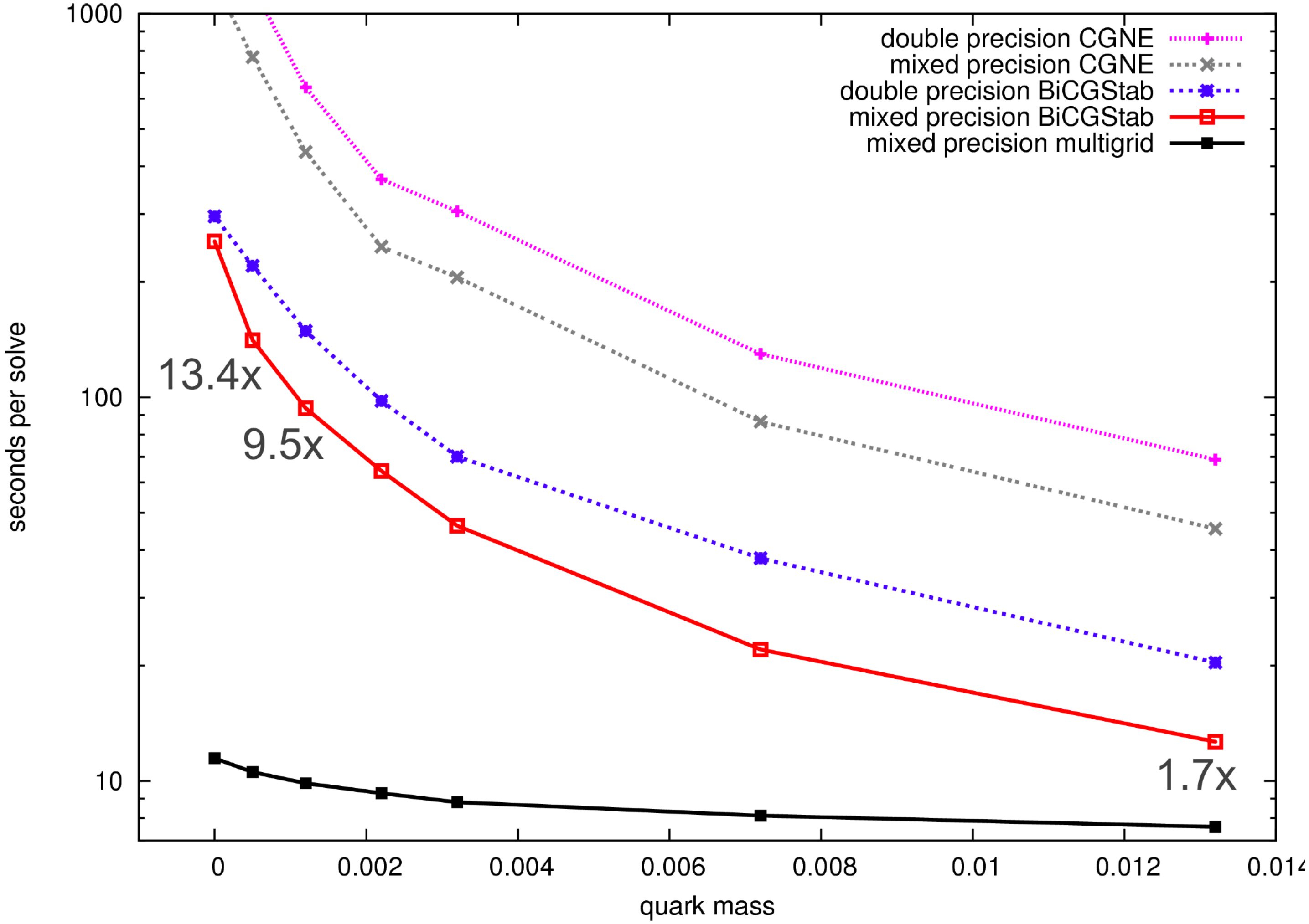
$$D_p = \begin{pmatrix} 1 & D_{eo} D_{oo}^{-1} \\ D_{oe} D_{ee}^{-1} & 1 \end{pmatrix}$$

$$D_r = 1 - D_{eo} D_{oo}^{-1} D_{oe} D_{ee}^{-1}$$

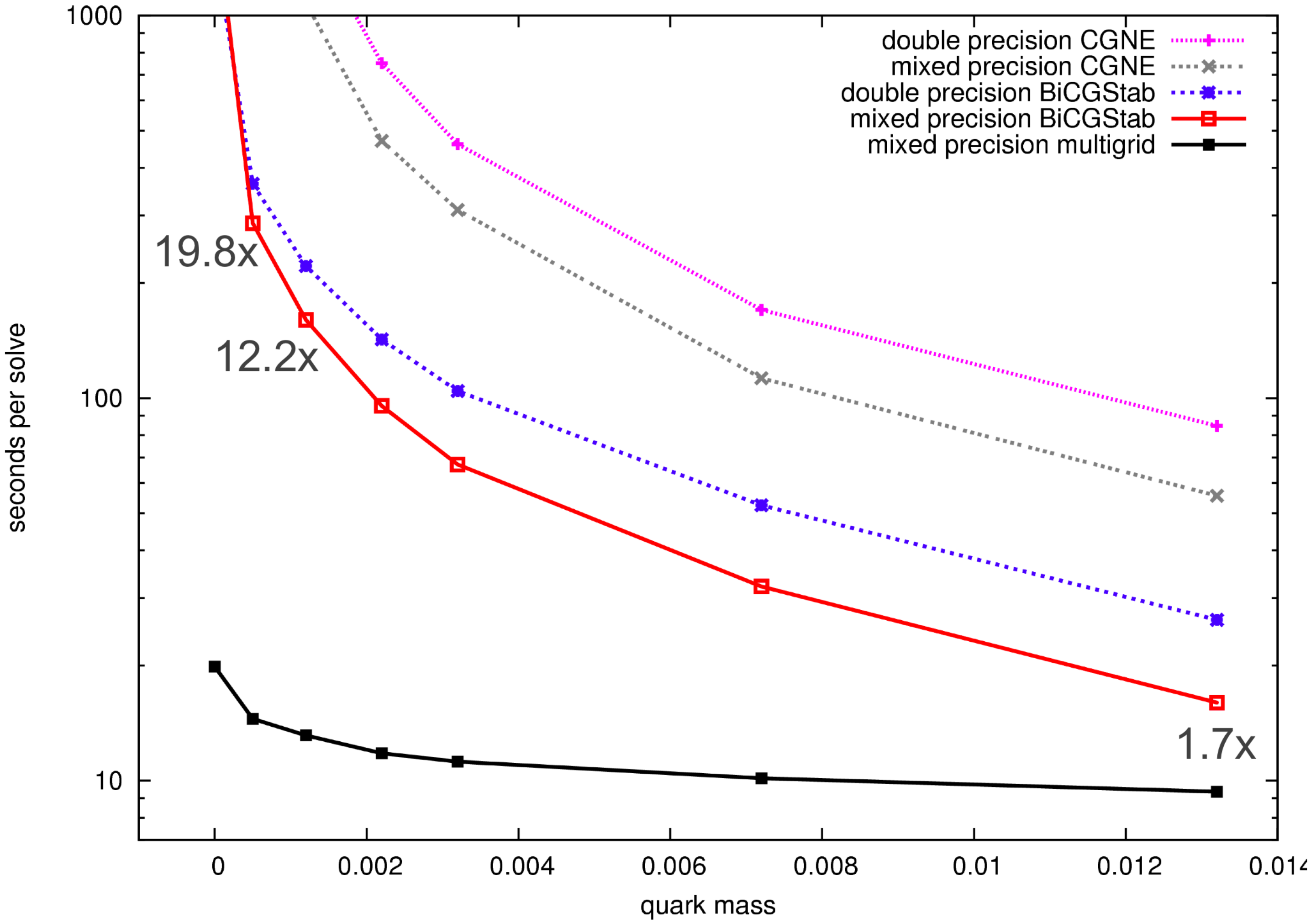
# Numerical results

- Using gluon configurations from (Jlab) Hadron Spectrum Collaboration
  - Anisotropic: $a_s \approx 0.12$ fm, $a_t \approx 0.035$ fm
  - $24^3$x128 and $32^3$x256
  - Up, down quark masses ~ 2.7x physical
- Results obtained on BG/P
  - 256 cores for $24^3$x128
    - 1st coarse lattice: $8^3$x16 with 24 vectors
    - 2nd coarse lattice: $4^3$x4 with 32 vectors
  - 1024 cores for $32^3$x256
    - 1st coarse lattice: 16x8x8x32 with 24 vectors
    - 2nd coarse lattice: 4x4x4x16 with 32 vectors

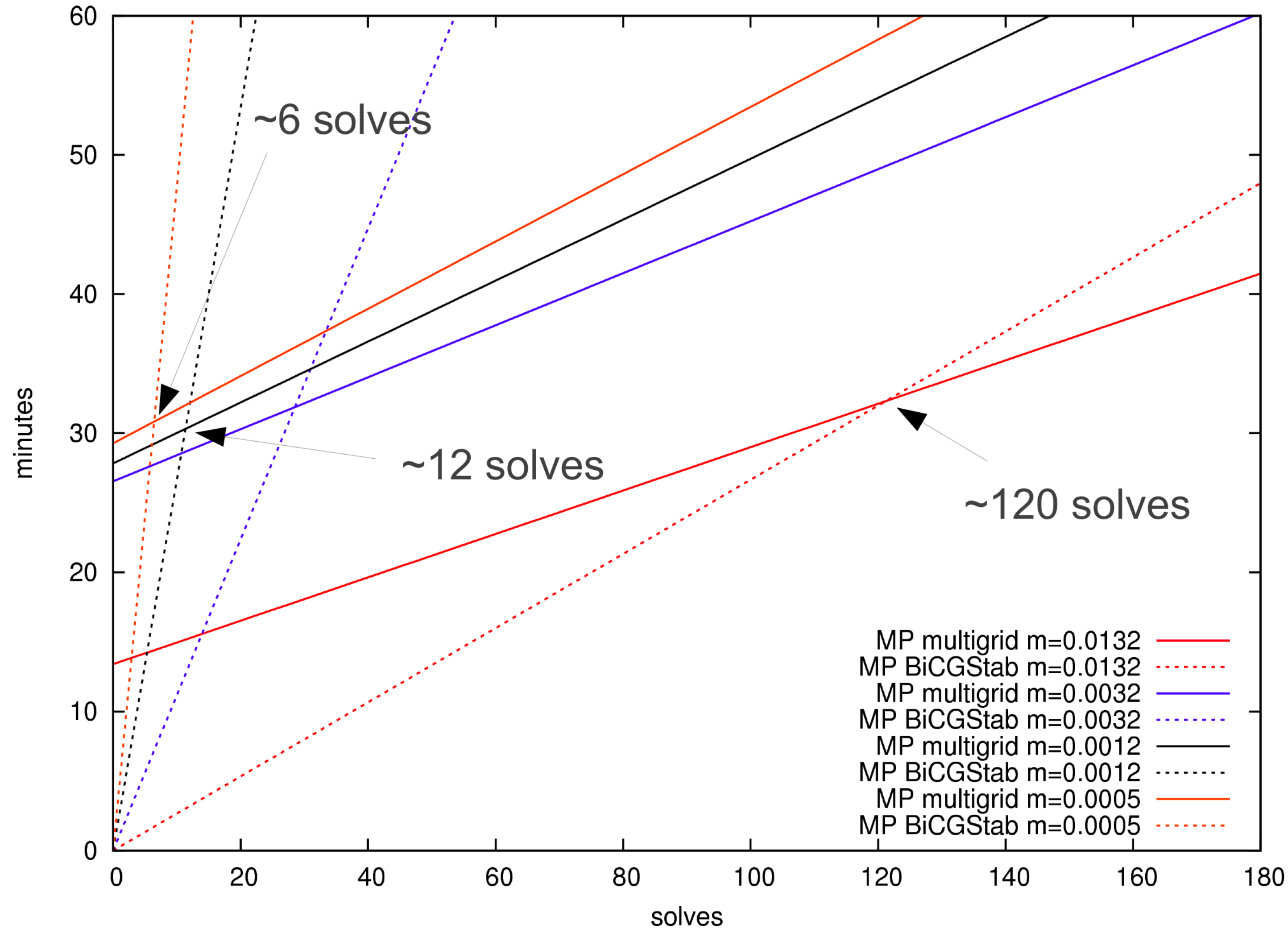
# Results $24^3$x128



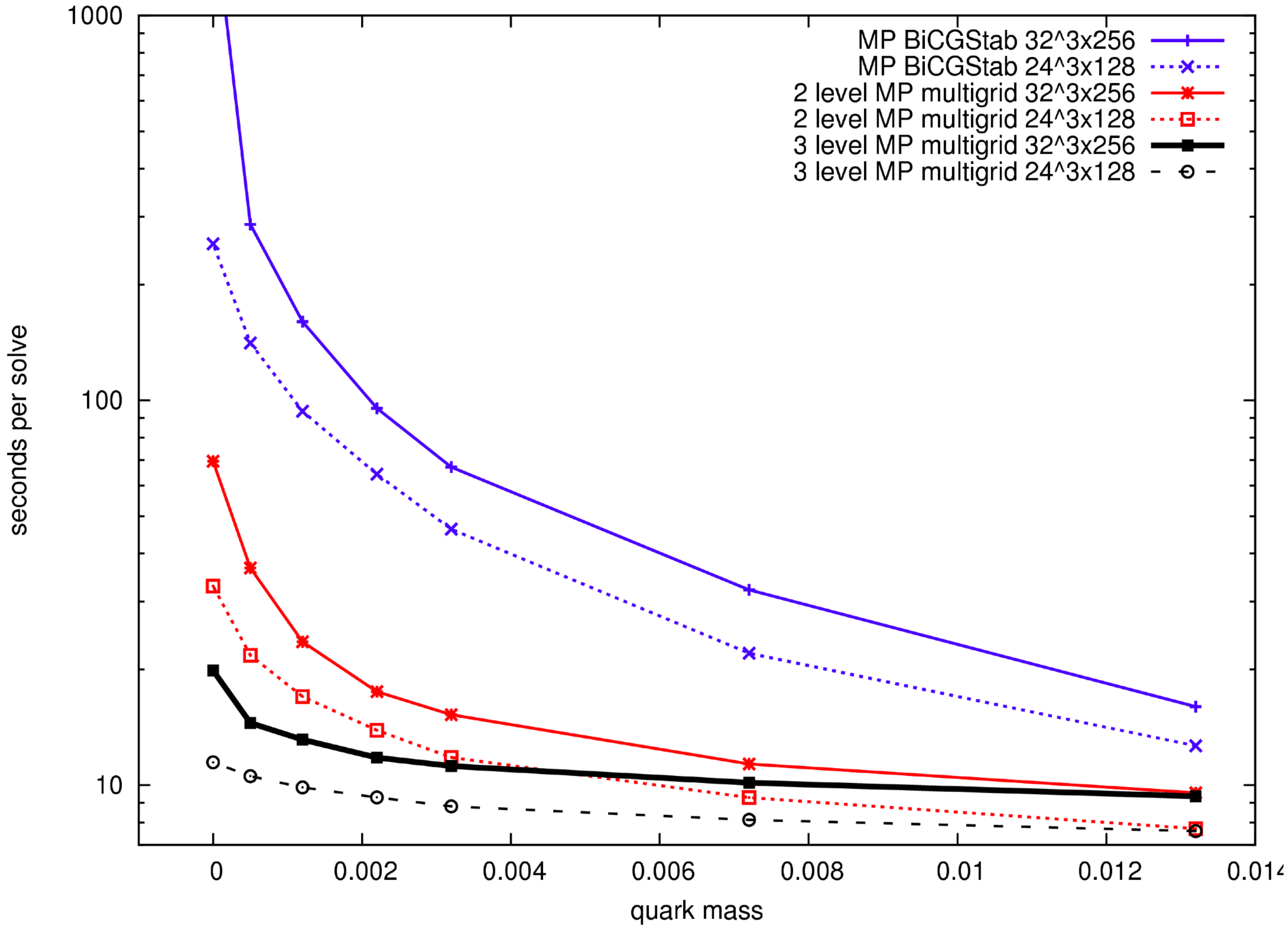
# Results $32^3$x256



# Total cost $32^3$x256

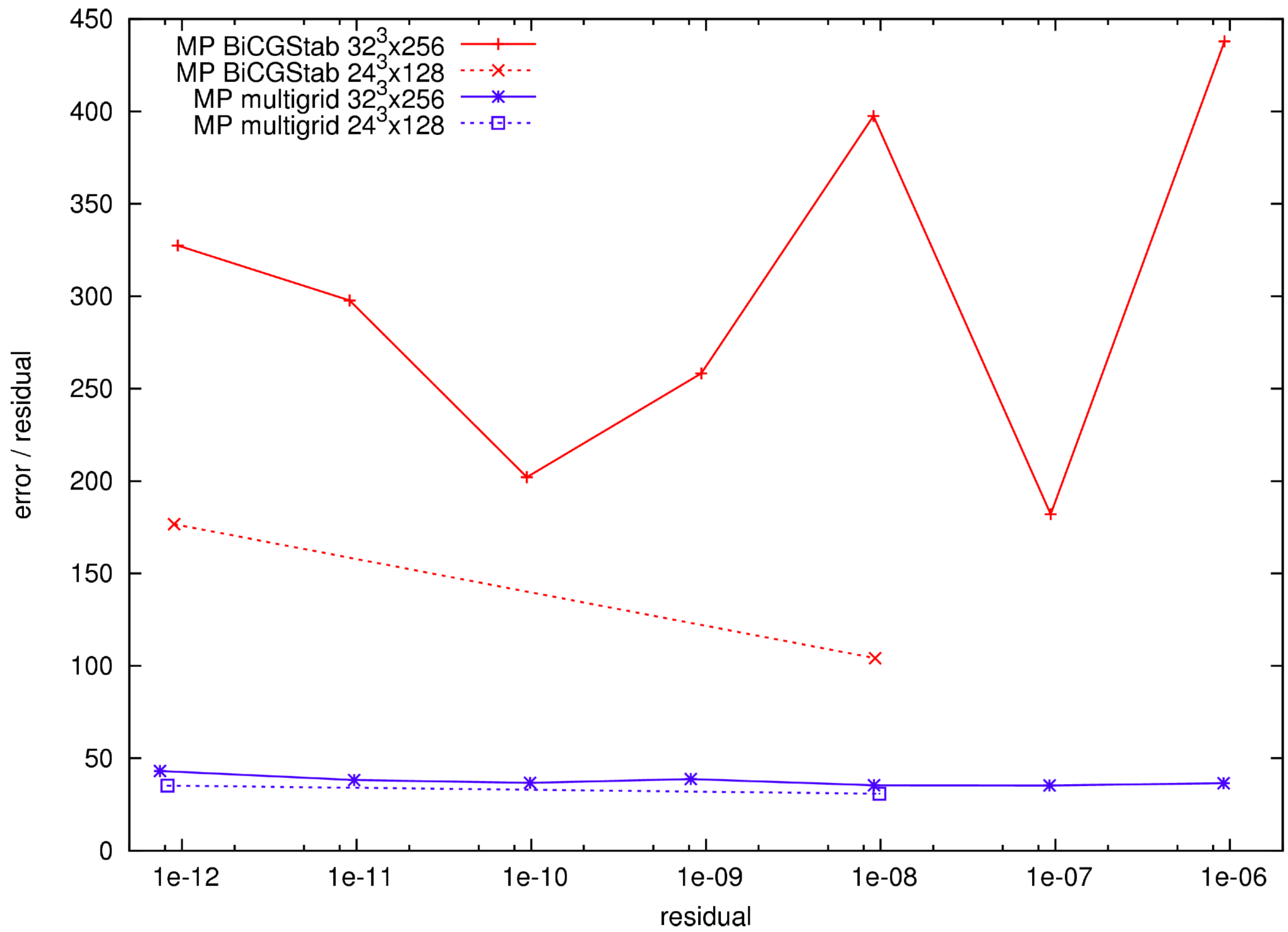


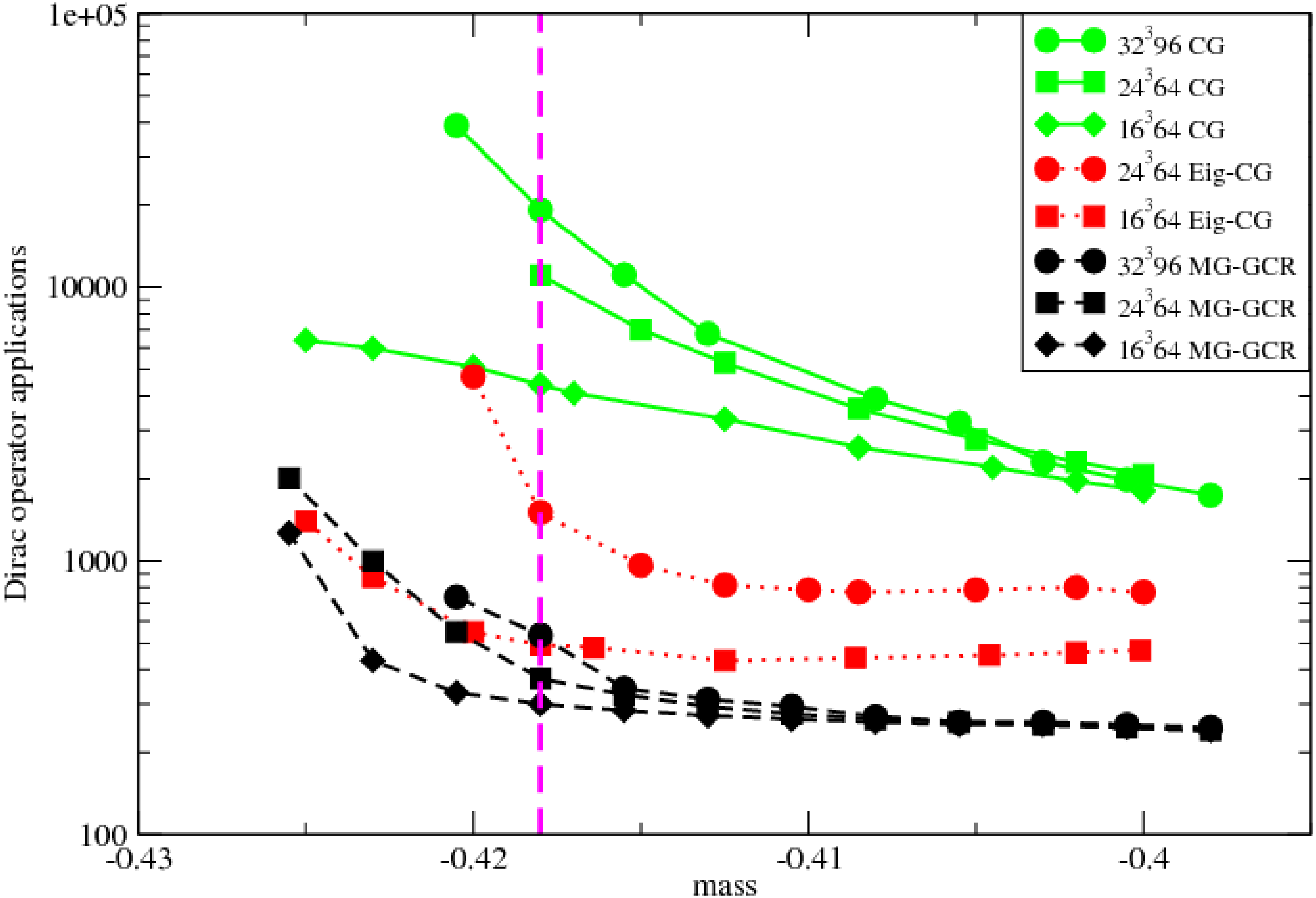
# 2 level vs 3 level



# Error vs residual

- Error:
  e = x* - x
- Residual:
  r = b – A x
     = A e
- Residual not as sensitive to low modes

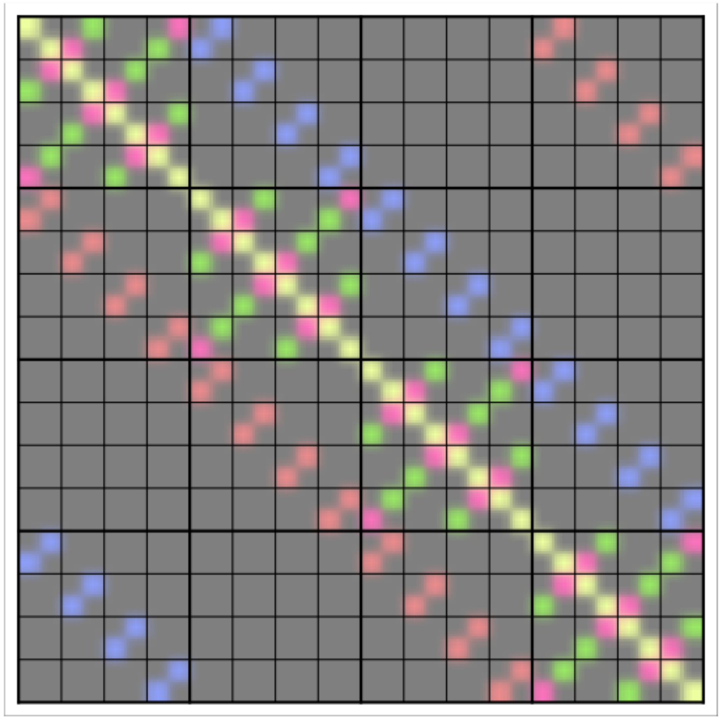


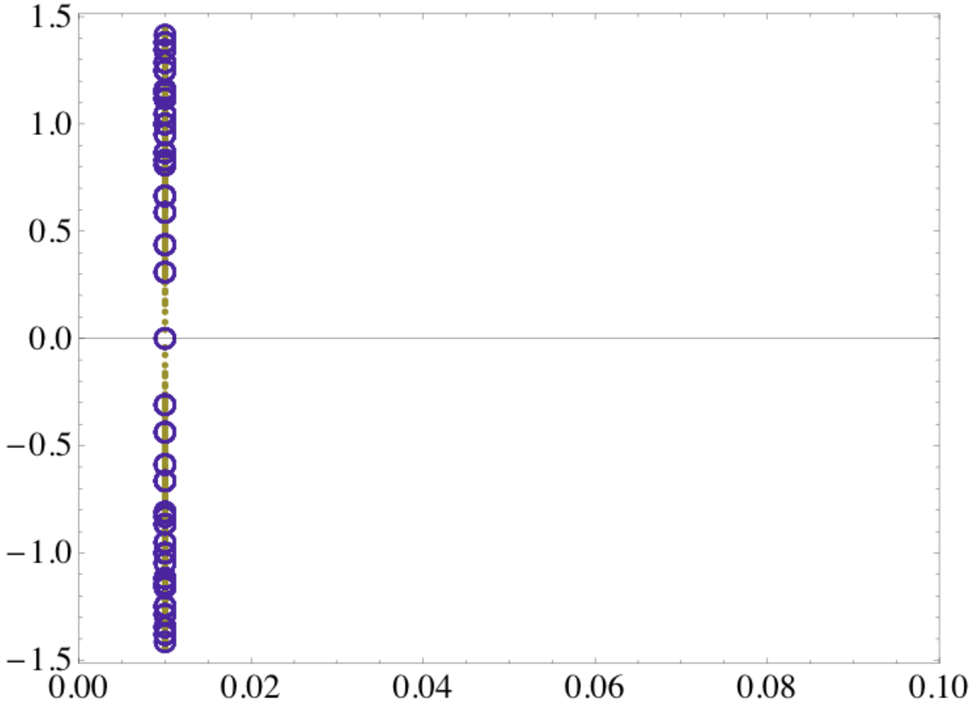


1e+05
10000
1000
100
-0.43
-0.42
-0.41
-0.4
mass
Dirac operator applications
32³96 CG
24³64 CG
16³64 CG
24³64 Eig-CG
16³64 Eig-CG
32³96 MG-GCR
24³64 MG-GCR
16³64 MG-GCR

# Naive Operator

$$D^{\text{naive}}_{x,x'} = m\,\delta_{x,x'} - \sum_{\mu} \frac{1}{2}\left[\gamma_\mu U_{x,\mu}\delta_{x+\hat{\mu},x'} - \gamma_\mu U^\dagger_{x,\mu}\delta_{x,x'+\hat{\mu}}\right]$$

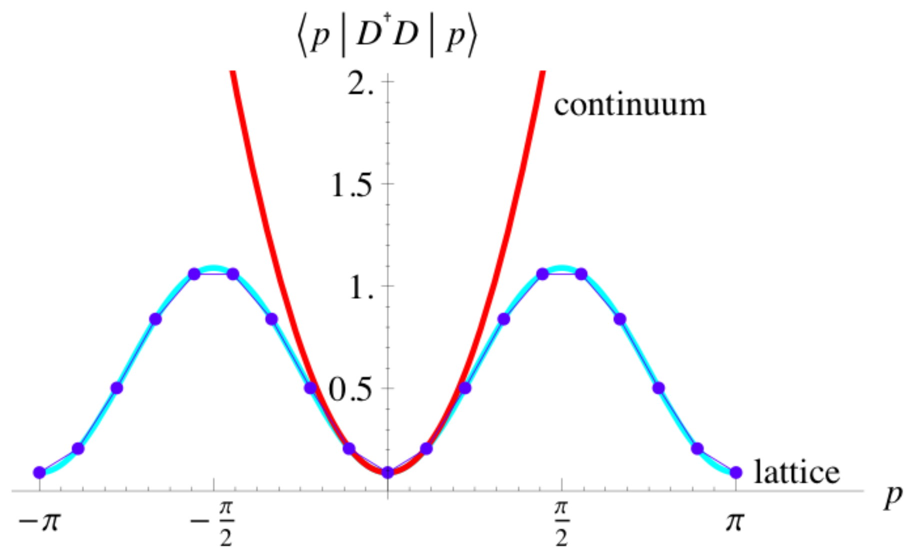




# Naive Operator

$$D^{\text{naive}}_{x,x'} = m\,\delta_{x,x'} - \sum_{\mu} \frac{1}{2} \left[ \gamma_\mu U_{x,\mu} \delta_{x+\hat{\mu},x'} - \gamma_\mu U^{\dagger}_{x,\mu} \delta_{x,x'+\hat{\mu}} \right]$$

- Straightforward discretization of continuum operator
- Results in $2^d$ doubler modes



# Wilson Operator

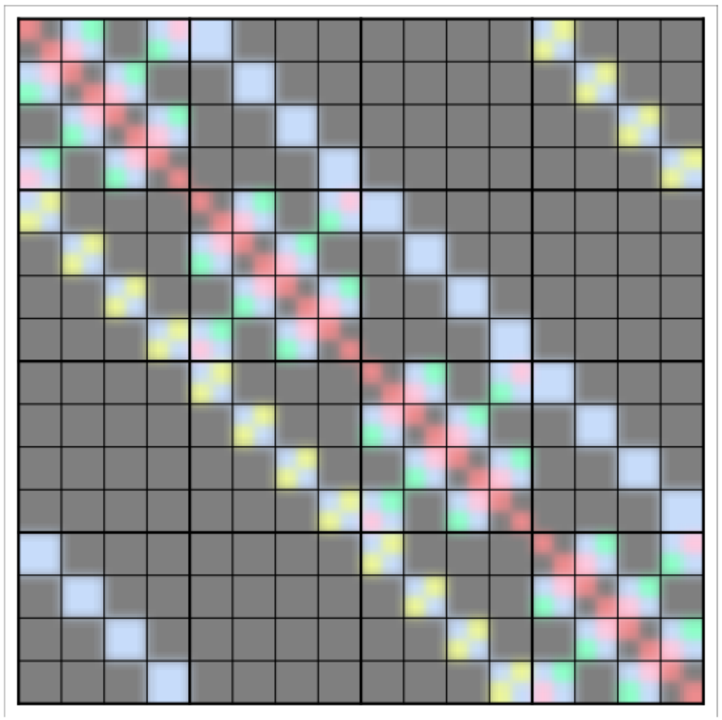
$$D^{\text{Wilson}}_{x,x'} = (m+d)\,\delta_{x,x'} - \sum_{\mu} \frac{1}{2}\left[(1+\gamma_\mu)U_{x,\mu}\delta_{x+\hat{\mu},x'} + (1-\gamma_\mu)U^{\dagger}_{x,\mu}\delta_{x,x'+\hat{\mu}}\right]$$

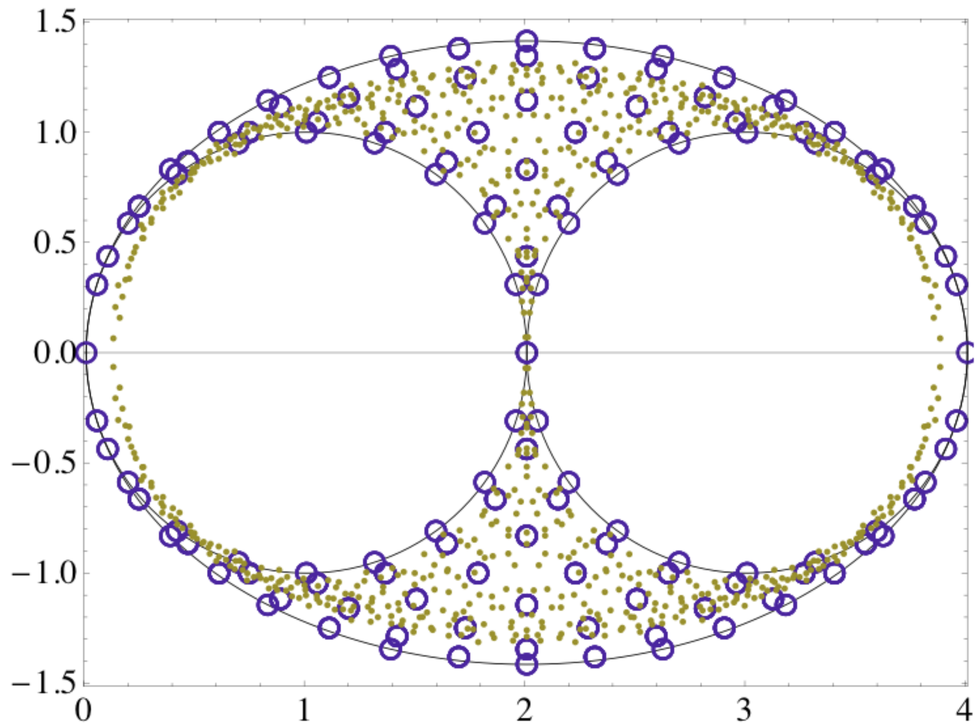
- Suppresses doublers with Wilson term; gives them a mass like $1/a$
- Explicitly breaks chiral symmetry; large additive renormalizations (e.g. to quark mass)

# Wilson Operator

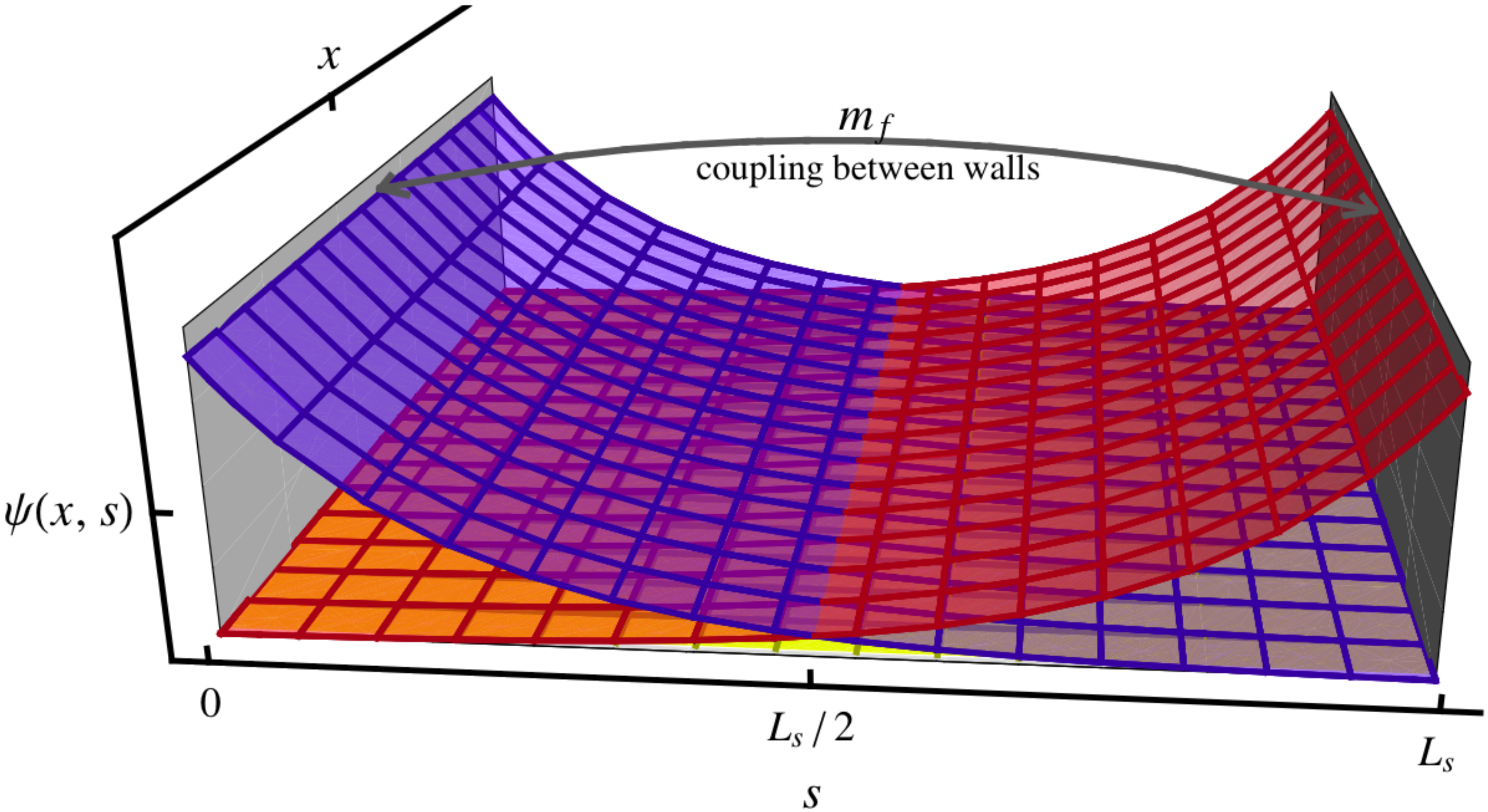
$$D^{\text{Wilson}}_{x,x'} =$$

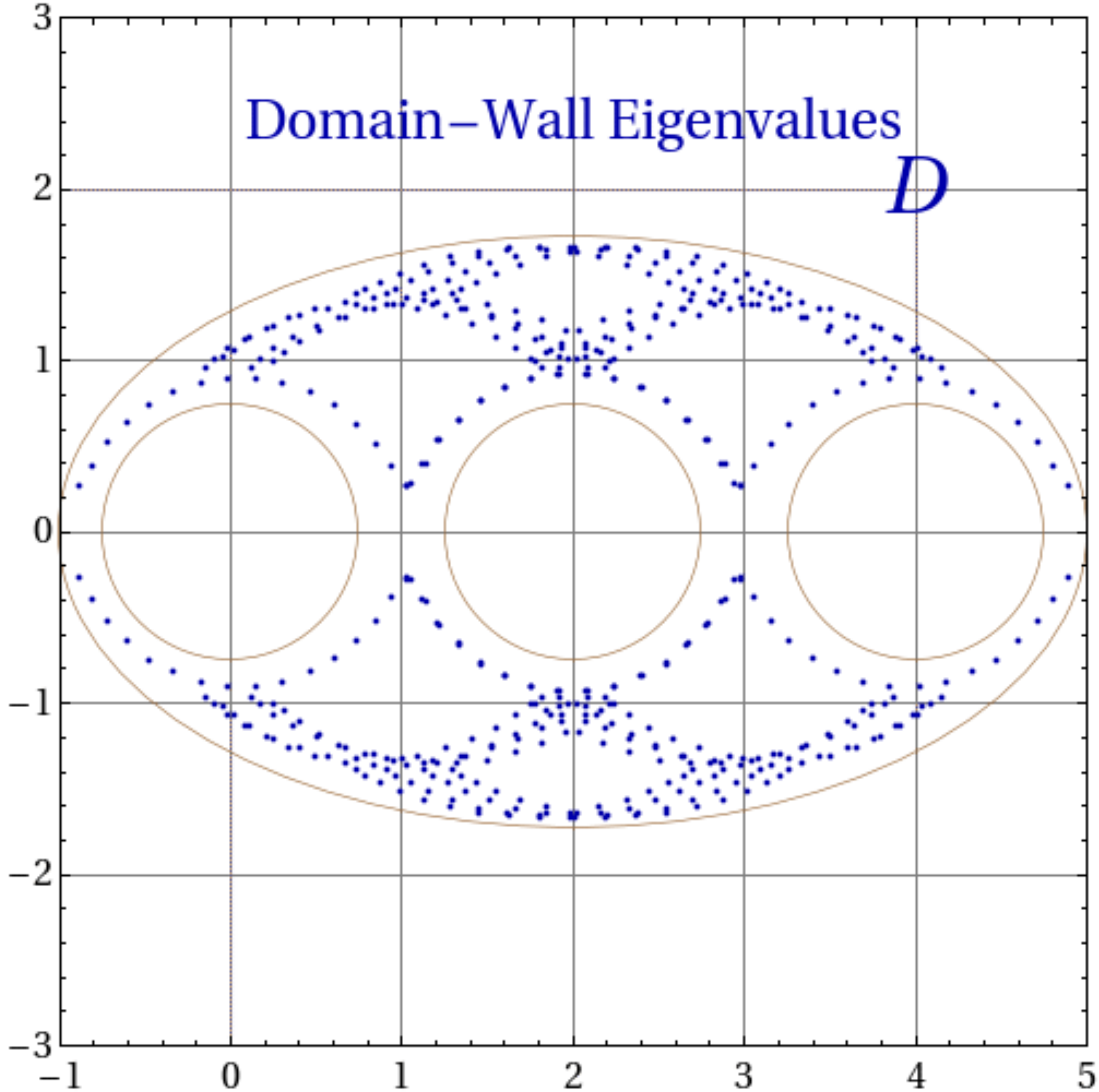
$$(m+d)\,\delta_{x,x'} - \sum_{\mu} \frac{1}{2}\left[(1+\gamma_\mu)U_{x,\mu}\delta_{x+\hat{\mu},x'} + (1-\gamma_\mu)U^\dagger_{x,\mu}\delta_{x,x'+\hat{\mu}}\right]$$





# A Chiral Fermion





Domain–Wall Eigenvalues
$D$
3
2
1
0
−1
−2
−3
−1
0
1
2
3
4
5

# Domain-Wall Operator

$$D^{\text{dwf}}_{x,s;x',s'} = \delta_{s,s'} D^{\parallel}_{x,x'} + \delta_{x,x'} D^{\perp}_{s,s'}$$

---

$$D^{\parallel}_{x,x'} =$$

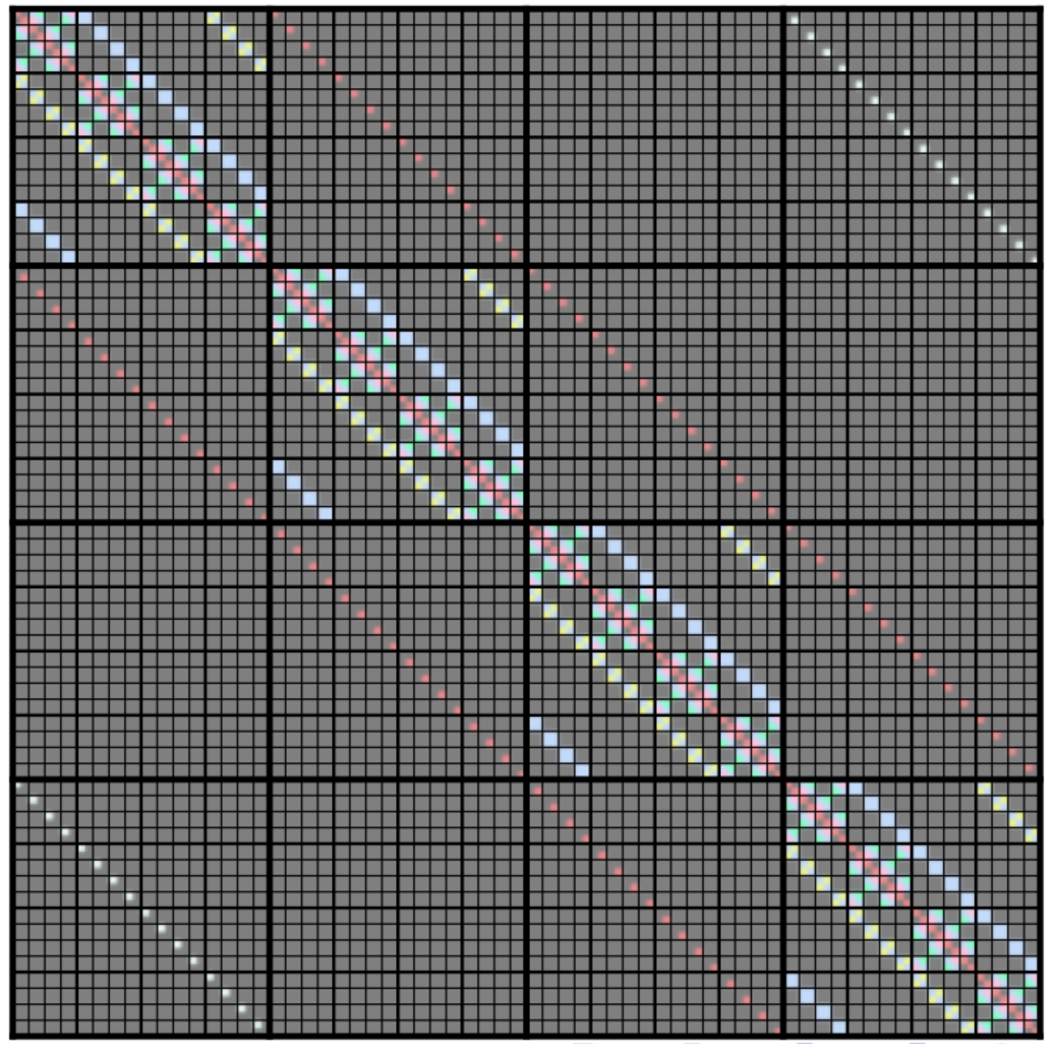
$$(M_5 - d)\delta_{x,x'} + \frac{1}{2}\sum_{\mu}\left[(1-\gamma_\mu)U_{x,\mu}\delta_{x+\hat{\mu},x'} + (1+\gamma_\mu)U^{\dagger}_{x',\mu}\delta_{x-\hat{\mu},x'}\right]$$

$$D^{\perp}_{s,s'} = \frac{1}{2}\left[(1-\gamma_5)\delta_{s+1,s'} + (1+\gamma_5)\delta_{s-1,s'} - 2\delta_{s,s'}\right] - \frac{m}{2}\left[(1-\gamma_5)\delta_{s,L_s-1}\delta_{0,s'} + (1+\gamma_5)\delta_{s,0}\delta_{L_s-1,s'}\right]$$

# Domain-Wall Operator

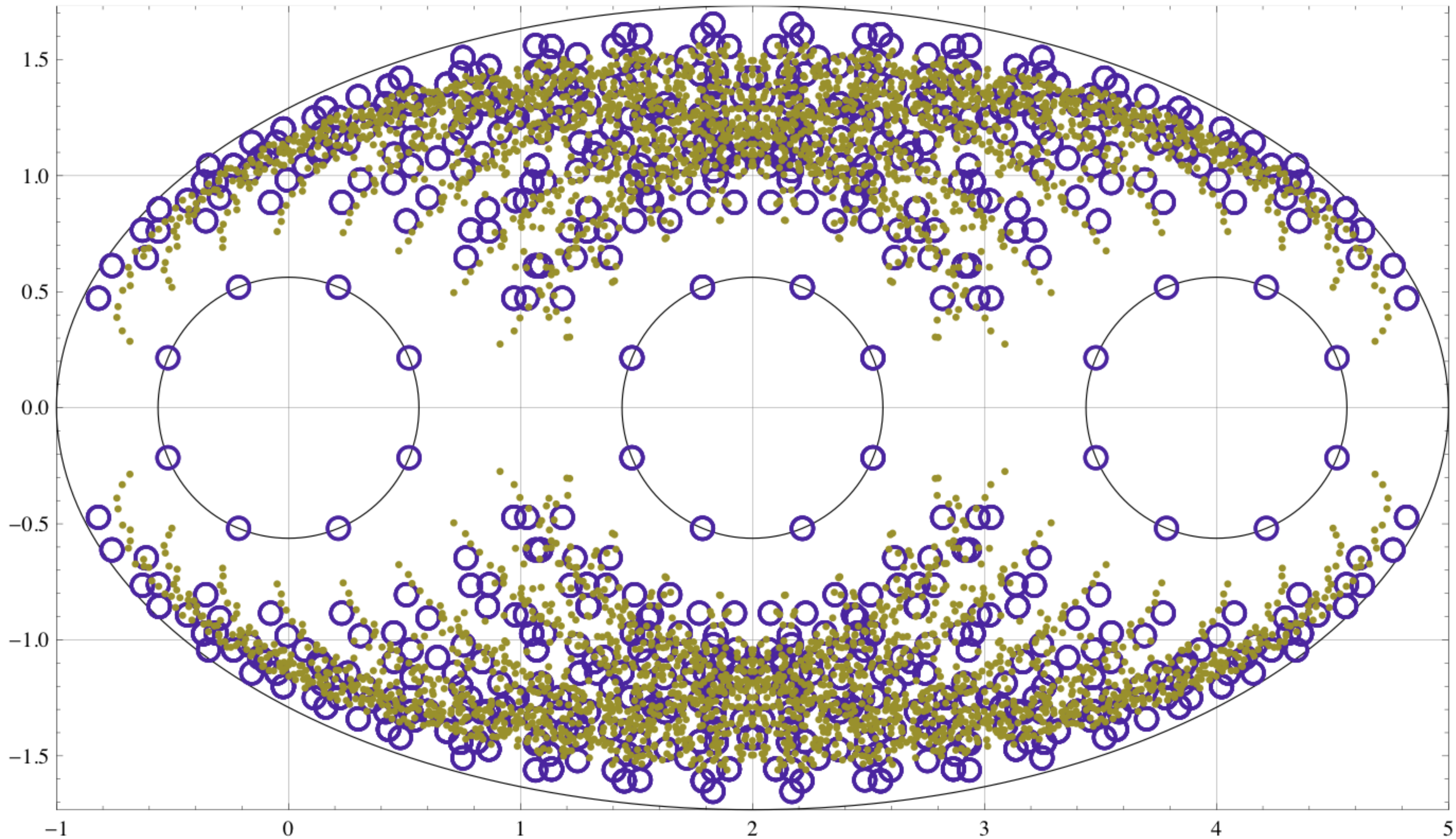
## Structure

$$D^{\mathrm{dwf}}_{x,s;x',s'} =$$
$$\delta_{s,s'} D^{\parallel}_{x,x'} + \delta_{x,x'} D^{\perp}_{s,s'}$$



# Domain-Wall Operator

## Spectrum

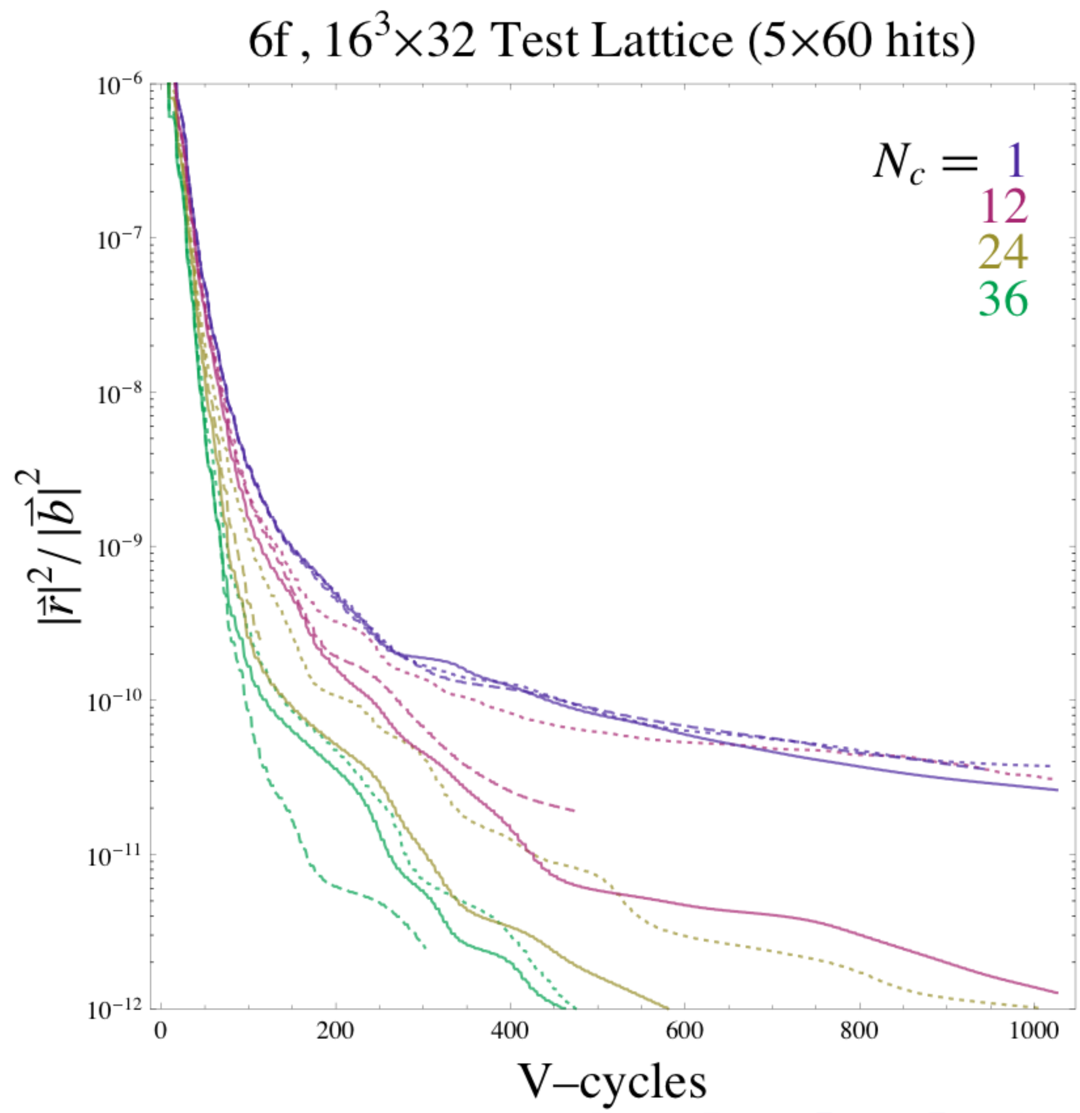


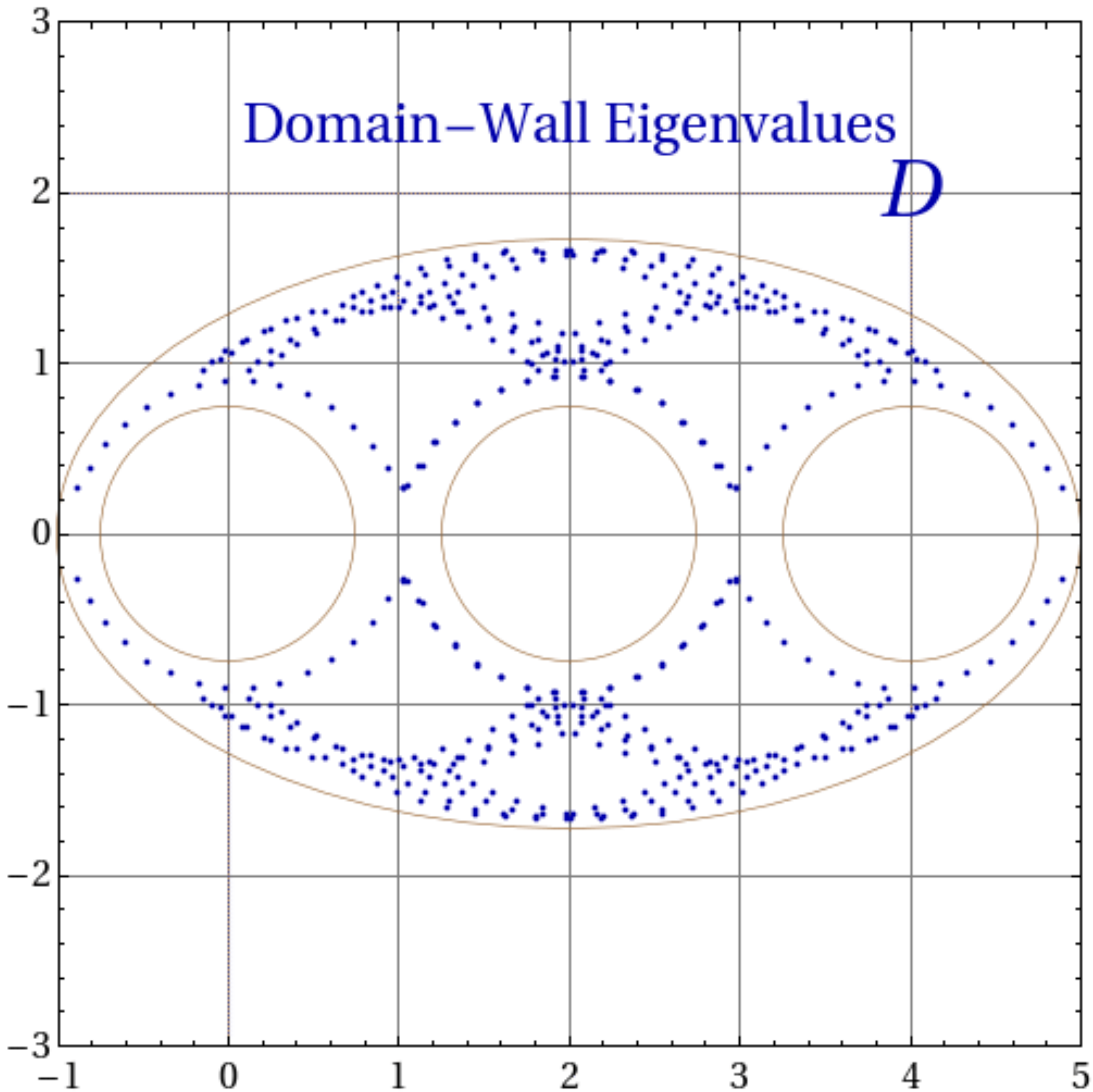
# Applied ASAM on DWF Normal Equation

Test on 4d SU(3) lattices

Using QDP multigrid capabilities

Examining dependence on quality of $V$






*Non Normal*
*Non Hermitian*
*Non Pos. Def.*

$$D = U\sqrt{D^\dagger D}$$

$$D^\dagger D = H^2$$

$$H = \Gamma_5 D$$

$$(\Gamma_5)^n D$$

# Part III: GPUs

- The problem of programming
  is making a Parallel problem Serial!