Calculating disconnected diagrams with Multigrid

James C. Osborn

Argonne Leadership Computing Facility

Lattice QCD at Zero Temperature

Extreme Computing and its Implications for the Nuclear Physics/Applied Mathematics/Computer Science Interface

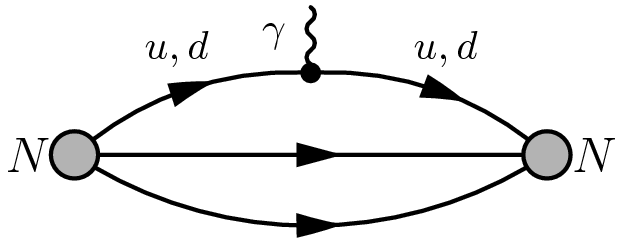
INT, July 5-8, 2011

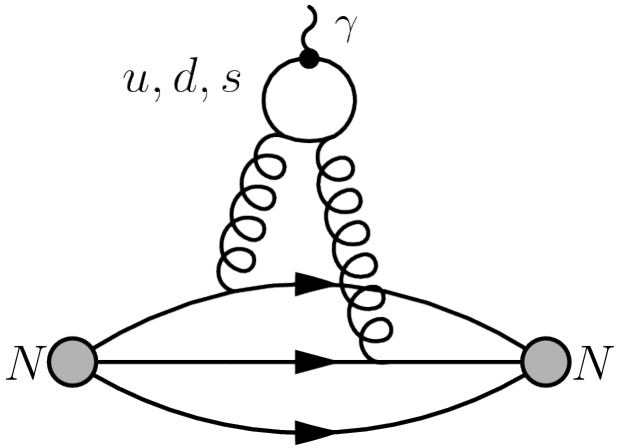
Outline

- Disconnected diagrams
- Adaptive multigrid
- Application to Wilson clover lattice Dirac operator
- The future

Nucleon form factors

Two different types of contribution to form factors





- Connected:
  direct interaction with one of the valence quarks of the nucleon
- u,d quarks contribute
- Only contribution to isovector quantities

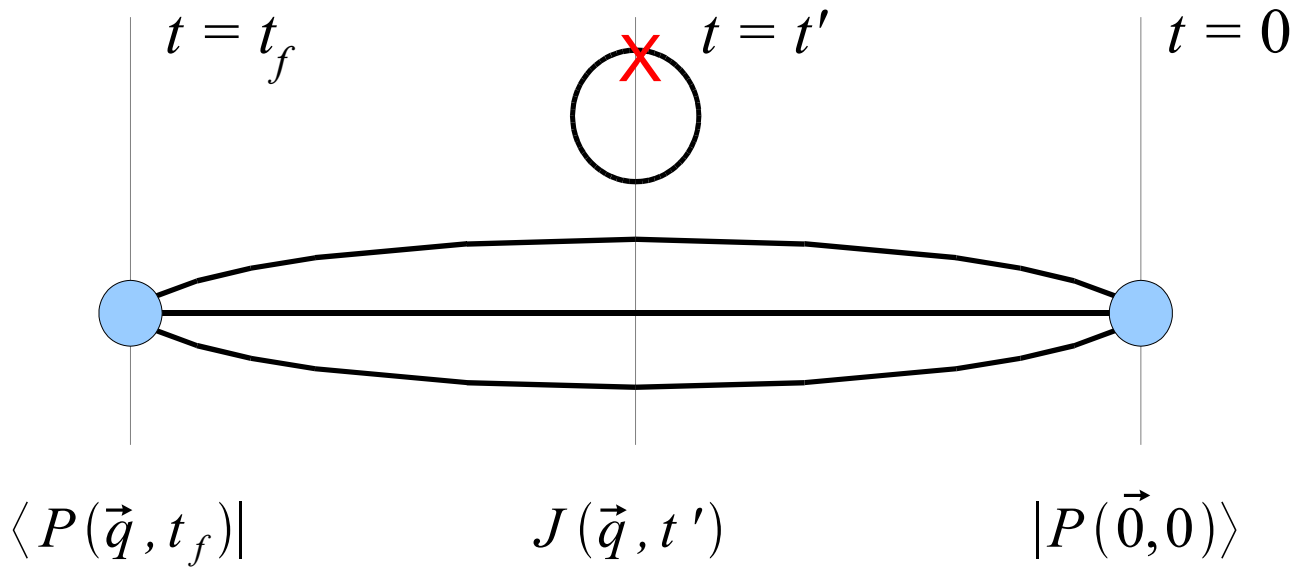
- Disconnected:
  interaction with sea quarks in glue
- Only way strange quarks contribute
- u,d also contribute
- Needed for isoscalar quantities
- Difficult to measure directly in lattice QCD

Disconnected diagrams on the lattice

Want to calculate nucleon matrix elements of the form

$$\langle P(\vec{q}, t_f)| \; J(\vec{q}, t') \; |P(\vec{0}, 0)\rangle$$

$$J(\vec{q}, t') = \sum_{\vec{x}} e^{i\vec{q}\cdot\vec{x}} \, \bar{s}(\vec{x}, t') \, \Gamma \, s(\vec{x}, t')$$



Challenges

- The current gives

$$\sum_{\vec{x}} \bar{s}(\vec{x}, t') \, \Gamma \, s(\vec{x}, t') = \mathrm{Tr}(\Gamma D^{-1})$$

  with trace over color, spin, and space

- Exact trace requires inverting the Dirac operator 12 $N_s{}^3$ times
  - Use approximate methods

- Signal can be very small requiring high statistics:
  correlation between the nucleon 2-point function and the quark loop

$$\langle \mathrm{nucleon} \times \mathrm{trace} \rangle - \langle \mathrm{nucleon} \rangle \langle \mathrm{trace} \rangle$$

Lattice methods for disconnected diagrams

- stochastic sources

$$\mathrm{Tr}\left(\Gamma\, D^{-1}\right) \approx \frac{1}{N}\sum_{i=1}^{N} \eta_i^H\, \Gamma\, D^{-1}\eta_i\ , \quad \left\langle \left(\eta_i^H\right)_x \left(\eta_i\right)_y \right\rangle = \delta_{xy}$$

- options:
  - random source type (Gaussian, Z(N), U(1))
  - dilution (spin, color, space, time)
  - subtraction

Dilution

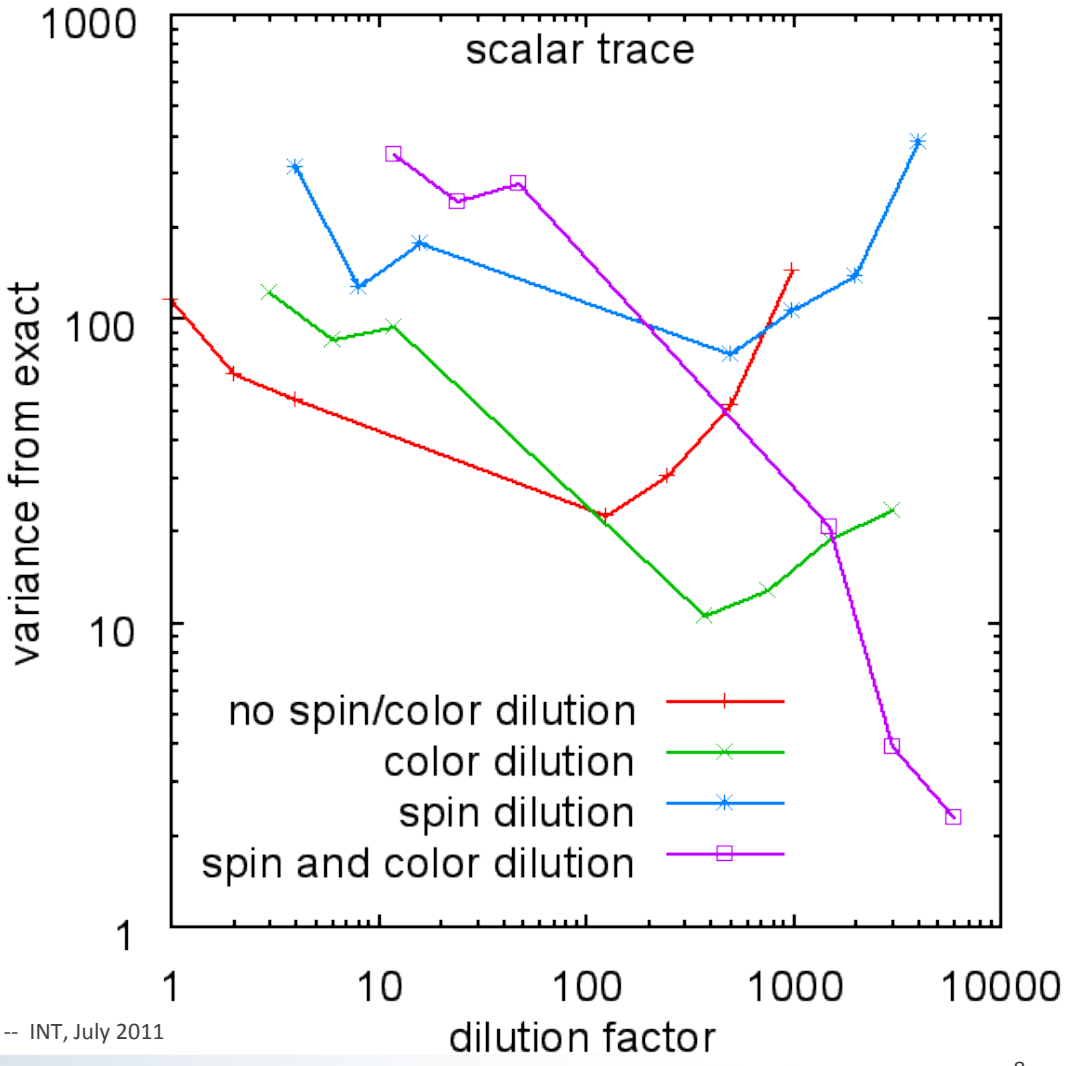
- partition vector indices, use separate sources for each group:
  color, spin, spatial (Wilcox; Foley, et al.)

- spatial dilution:
  - none (dilution factor 1)
  - even/odd (dilution factor 2)
  - cubic diagonal [(0,0,0)(1,1,1)],[(0,0,1)(1,1,0)],...
    (dilution factor 4)
  - inner 5^3 dilution with outer dilution among inner blocks (dilution factors 125, 250, 500)

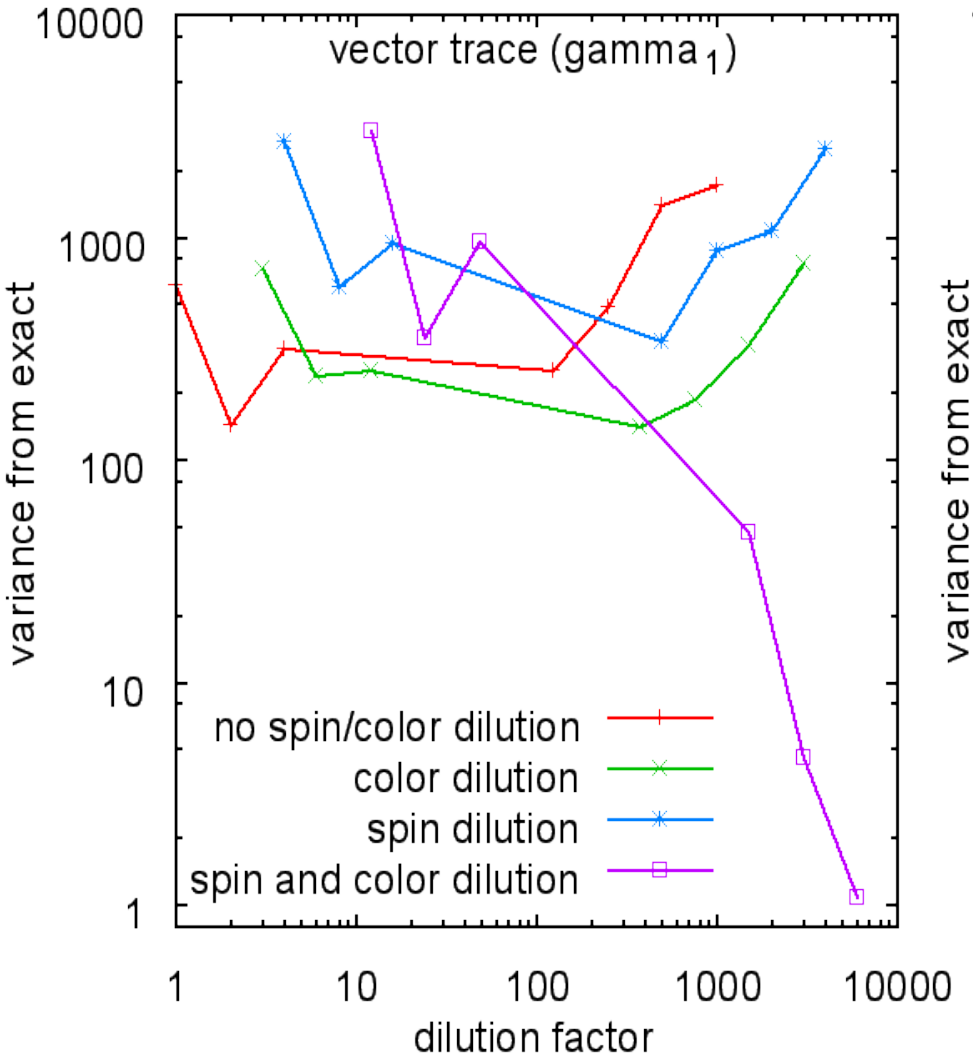
- tests on 10^3x32 quenched lattices with Wilson Dirac matrix
  - trace on single time slice (12,000 components)
  - with/without color/spin dilution

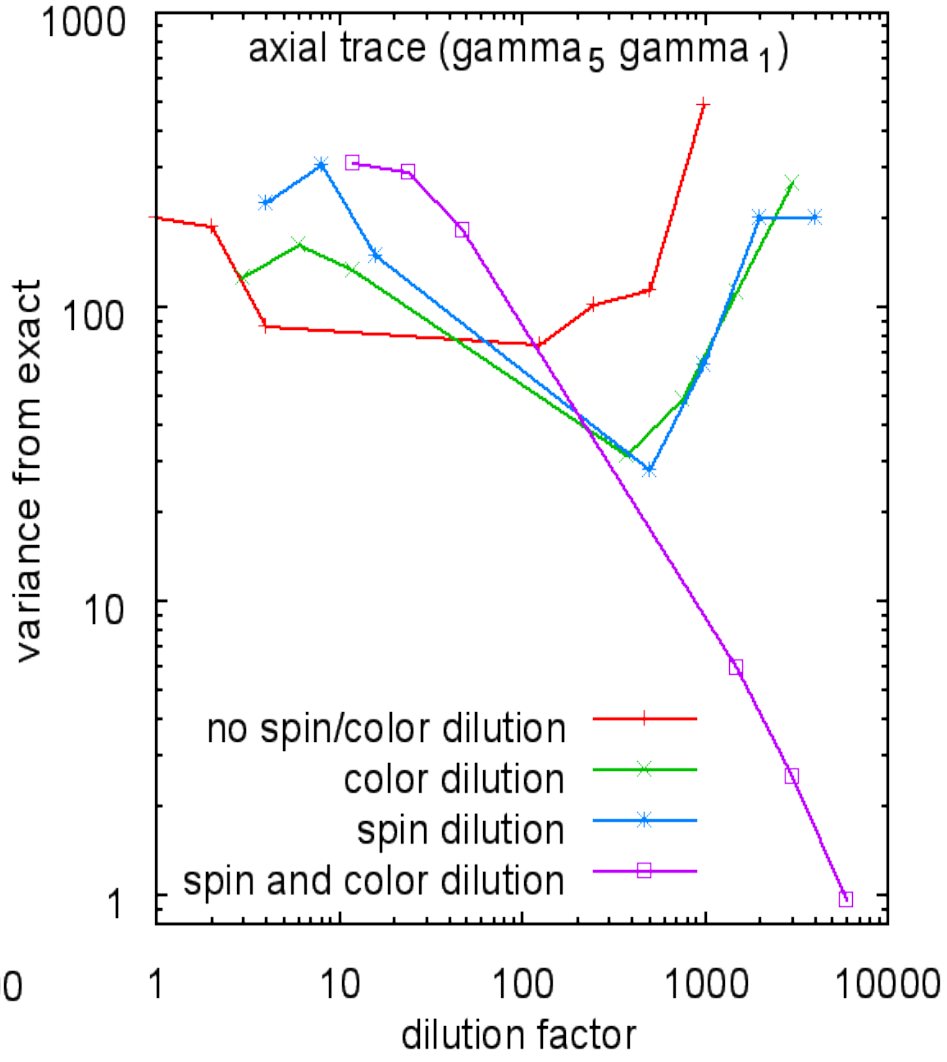
Dilution

- compare dilution (spin,color,spatial) to exact trace
- all points at fixed amount of work
- exact trace: dilution factor = 12,000
- spatial dilution generally helps
- need spin/color dilution for >1000 sources



Dilution


vector trace (gamma 1)
variance from exact
10000
1000
100
10
1
1
10
100
1000
10000
dilution factor
no spin/color dilution
color dilution
spin dilution
spin and color dilution


axial trace (gamma 5 gamma 1)
variance from exact
1000
100
10
1
1
10
100
1000
10000
dilution factor
no spin/color dilution
color dilution
spin dilution
spin and color dilution

Variance reduction

- unbiased subtraction

$$\mathrm{Tr}\left(\Gamma\, D^{-1}\right) \approx \left\langle \eta^{H}\, \Gamma \left(D^{-1} - O_{tr}\right) \eta \right\rangle_{\eta} \quad , \quad O_{tr} = O - \Gamma^{-1} \frac{1}{N} \mathrm{Tr}\, \Gamma\, O$$

  - hopping parameter expansion

$$O = 2\kappa\left(1 + \kappa M + \kappa^{2} M^{2} + \ldots\right)$$

  - eigenvalue projection

$$O = P_{ev}\, D^{-1}\, P_{ev} - D^{-1}$$

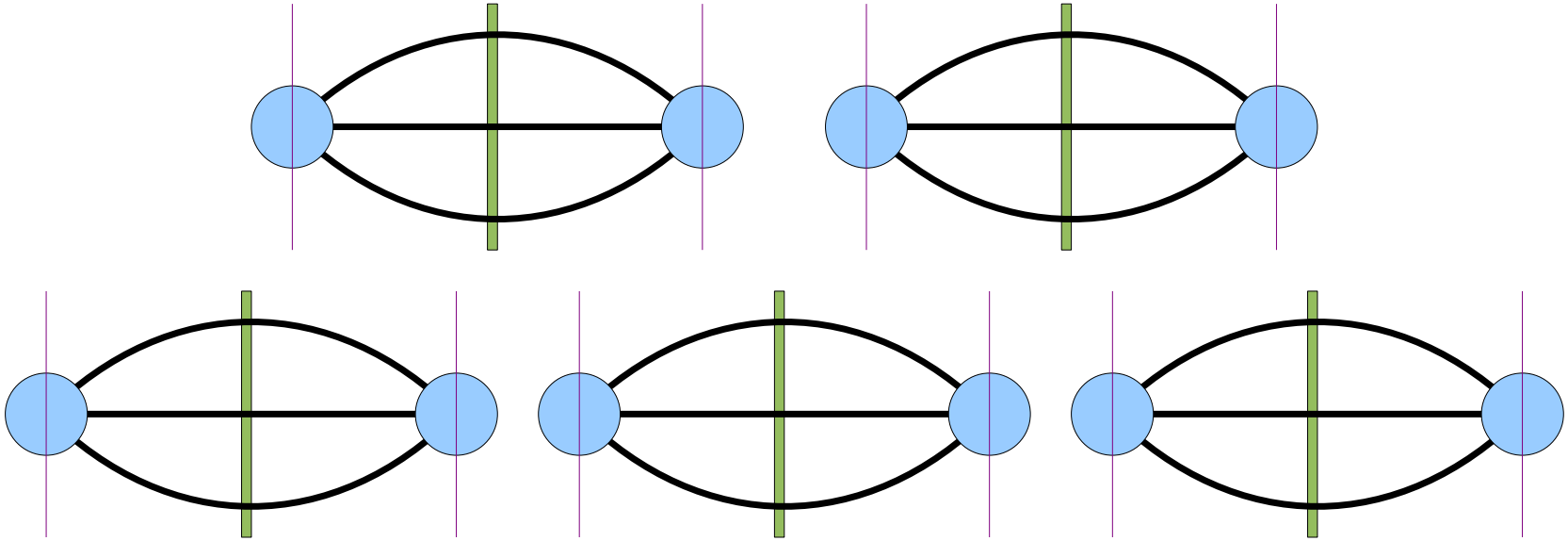
  - multigrid subtraction

$$O = P_{fc}\, D_{c}^{-1}\, P_{cf}$$

$$\mathrm{Tr}\, \Gamma\, P_{fc}\, D_{c}^{-1}\, P_{cf} = \mathrm{Tr}\, P_{cf}\, \Gamma\, P_{fc}\, D_{c}^{-1}$$

Gauge noise

- gauge noise is significant
  - prefer many lattices: $O(1000)$
  - make best use of existing lattices (multiple timeslices per lattice)



Disconnected diagrams on the lattice

- Disconnected diagrams are hard
- Many methods for improving errors have been developed and used
  - Finding best method for a given problem is also a challenging problem
- High precision calculations will require large number of Dirac equation solves

MULTIGRID

QCD-MG collaboration

- Argonne
  - James Osborn
- Boston University
  - Ron Babich
  - Rich Brower
  - Claudio Rebbi
- Colorado U., Boulder
  - Marian Brezina
  - Christian Ketelsen
  - Tom Manteuffel
  - Steve McCormick
  - John Ruge
- Harvard
  - Mike Clark
- KAUST
  - David Keyes
- LLNL
  - Rob Falgout
- Penn State
  - James Brannick
  - Ludmil Zikatanov
- Tufts
  - Scott MacLachlan
- Washington, University of
  - Saul Cohen

The problem

- Lattice QCD requires repeated solution of Dirac equation

$$[D(U) + m]\psi = \eta$$

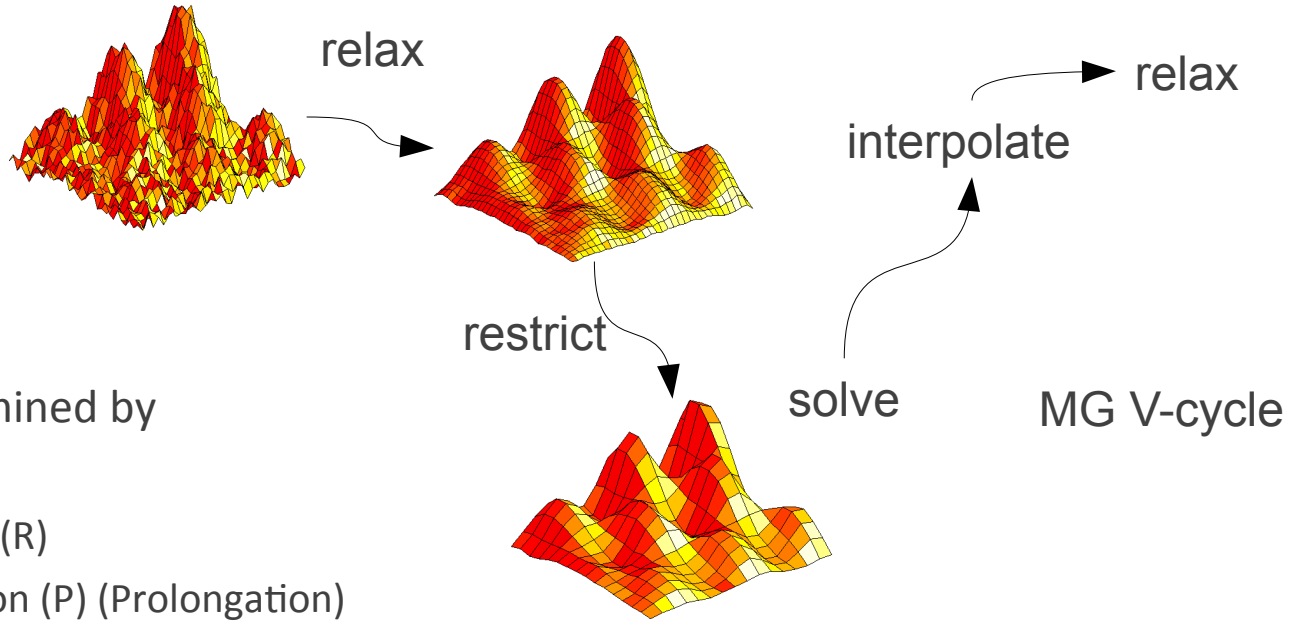
- Much of the work goes into solution
  - Usually over 90% for analysis
  - Typically from 50-90% for gluon configuration generation

- Exhibits critical slowing down
  - Condition number diverges as mass decreases ($\kappa \propto 1/m$)
  - Standard Krylov solvers (CG, BiCGStab, …) become inefficient as condition number grows
  - Difficult to simulate at physical light (up, down) quark masses

- Multigrid methods have been very successful in beating this in other fields

Multigrid

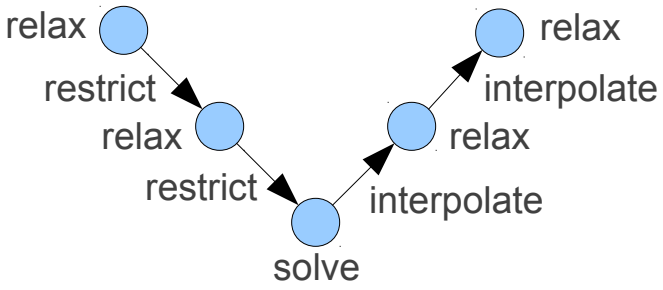
- Standard solvers (stationary, Krylov) good at reducing high frequency error components, not good with low frequency errors
- MG projects error onto coarse grid, solves, then interpolates correction back to fine grid

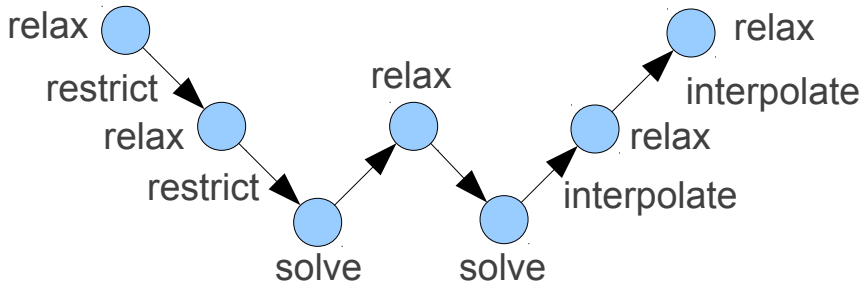


- V-cycle determined by
  - Relaxation
  - Restriction (R)
  - Interpolation (P) (Prolongation)
  - Coarse operator

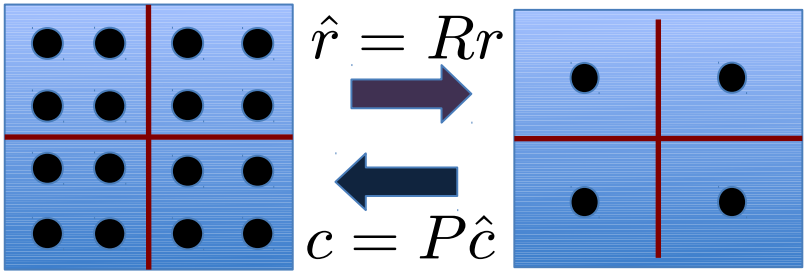
Multigrid

- MG V-cycle typically used as preconditioner for outer solver
  - Here using GCR (Generalized Conjugate Residuals)
- Used recursively: MG cycle used to solve on coarse grid, …
- Choice of cycle:
  - V-cycle, W-cycle, …
  - Here using GCR solver for coarse system with MG preconditioner





Choosing P & R



- Coarse grid solve:

$$P A_c^{-1} R\, r$$

$$A_c = RAP$$

- Algebraic MG: P & R formed from elements of A (or approximation to)
- Adaptive MG: P & R formed from slow-to-converge modes of A
  - Want P to preserve (right) low modes of A
  - Form P from representative low modes chopped into blocks (aggregates)
  - R from left low modes of A

$$P = \begin{pmatrix} v_1 & v_2 & & & \\ v_1 & v_2 & & & \\ \vdots & \vdots & & & \\ v_1 & v_2 & & & \\ & & v_1 & v_2 & \\ & & v_1 & v_2 & \\ & & \vdots & \vdots & \\ & & v_1 & v_2 & \\ & & & & \ddots \end{pmatrix}$$

Setup methods

- Repeated relaxation (inverse iteration) on random vectors
  - Simple (don't need to construct coarse operator)
  - Can vary number of iterations/cycles
  - Vectors may be locally redundant

- Adaptive smooth aggregation (αSA) (Brezina, et al., 2004)
  - Construct new MG cycle with current vectors, use to find new vector
  - Requires construction of coarse operator
  - New vectors should give new important components

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 γ_5-Hermitian operator can set $R = P^\dagger \gamma_5$

- Also keeping chirality independent of blocking
  - Treat $(1 \pm \gamma_5)P$ as separate vectors for prolongation/restriction
  - Helps alleviate problems due to indefinite operator ($|P^\dagger \gamma_5 D P| \approx 0$)

Fine and coarse operators

- Solving Wilson-clover operator
  - Using even-odd preconditioning on fine system
  - $D\ x = b \rightarrow (D\ D_d^{-1})\ (D_d\ x) = b \rightarrow D_p\ x_p = b$
  - $D_r\ x_{p,e} = b_e - D_{eo}\ D_{oo}^{-1}\ b_o \rightarrow D_r\ x_r = b_r$
  - Construct coarse operator from D_p then construct reduced operator
  - D_p no longer γ_5-Hermitian, but use same R (= $P^\dagger \gamma_5$) anyway

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

$$D_d = \begin{pmatrix} D_{ee} & 0 \\ 0 & D_{oo} \end{pmatrix}$$

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

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

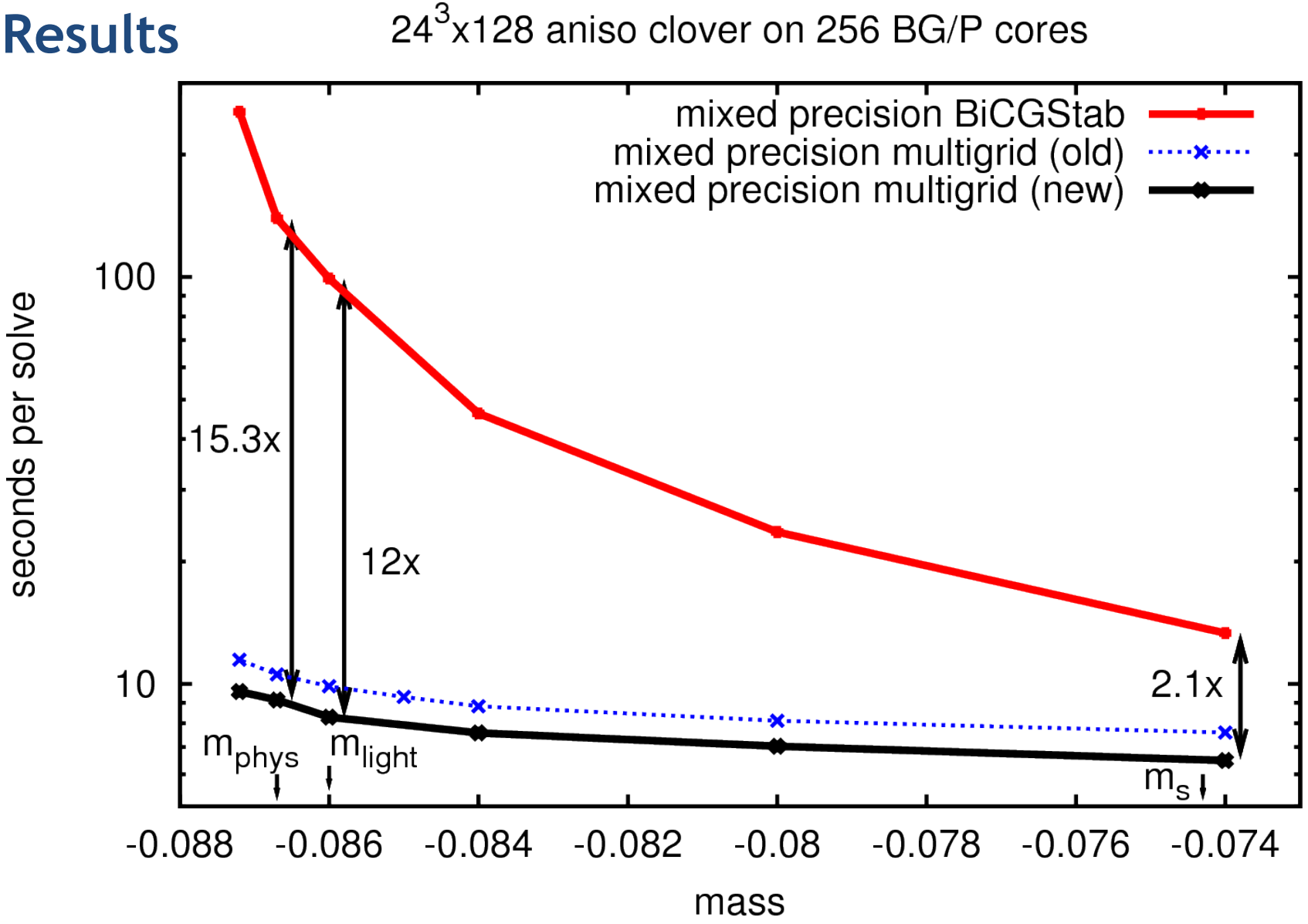
Implementation Details

- (0,~4) V-cycle
  - No pre-relaxation, ~4 steps GCR post-relaxation
- 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

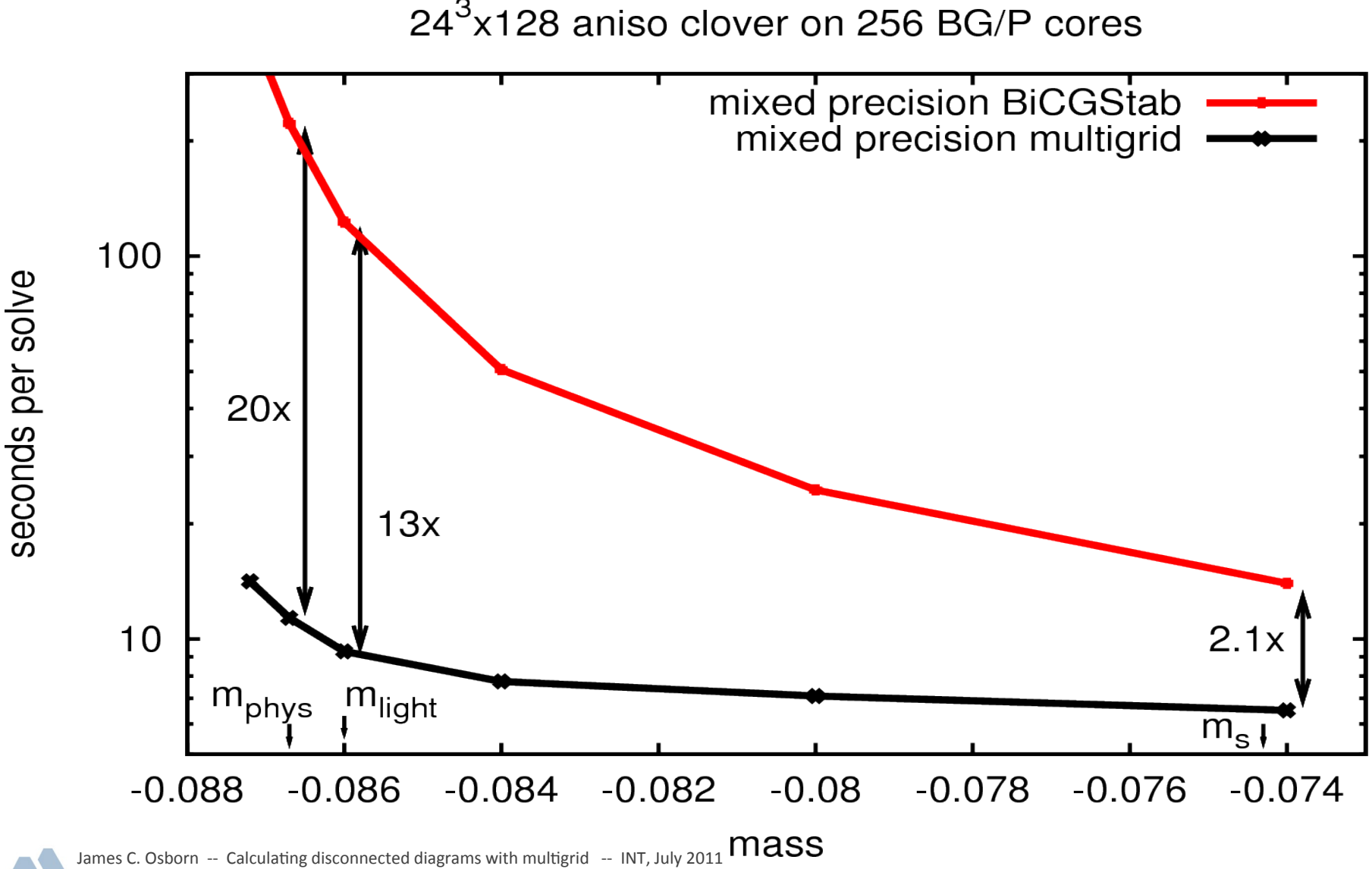
Numerical results

- Using gauge configurations from Hadron Spectrum Collaboration
  - Anisotropic: $a_s \approx 0.12$ fm, $a_t \approx 0.035$ fm
  - $24^3 x128$ and $32^3 x256$
  - Dynamical $m_\pi \approx 220$ MeV ($m = -0.086$)
- Results obtained on BG/P
  - 256 cores for $24^3 x128$
    - 1^{st} coarse lattice: $8^3 x16$ with 24 vectors
    - 2^{nd} coarse lattice: $4^3 x4$ with 32 vectors
  - 1024 cores for $32^3 x256$
    - 1^{st} coarse lattice: 16x8x8x32 with 24 vectors
    - 2^{nd} coarse lattice: 4x4x4x16 with 32 vectors

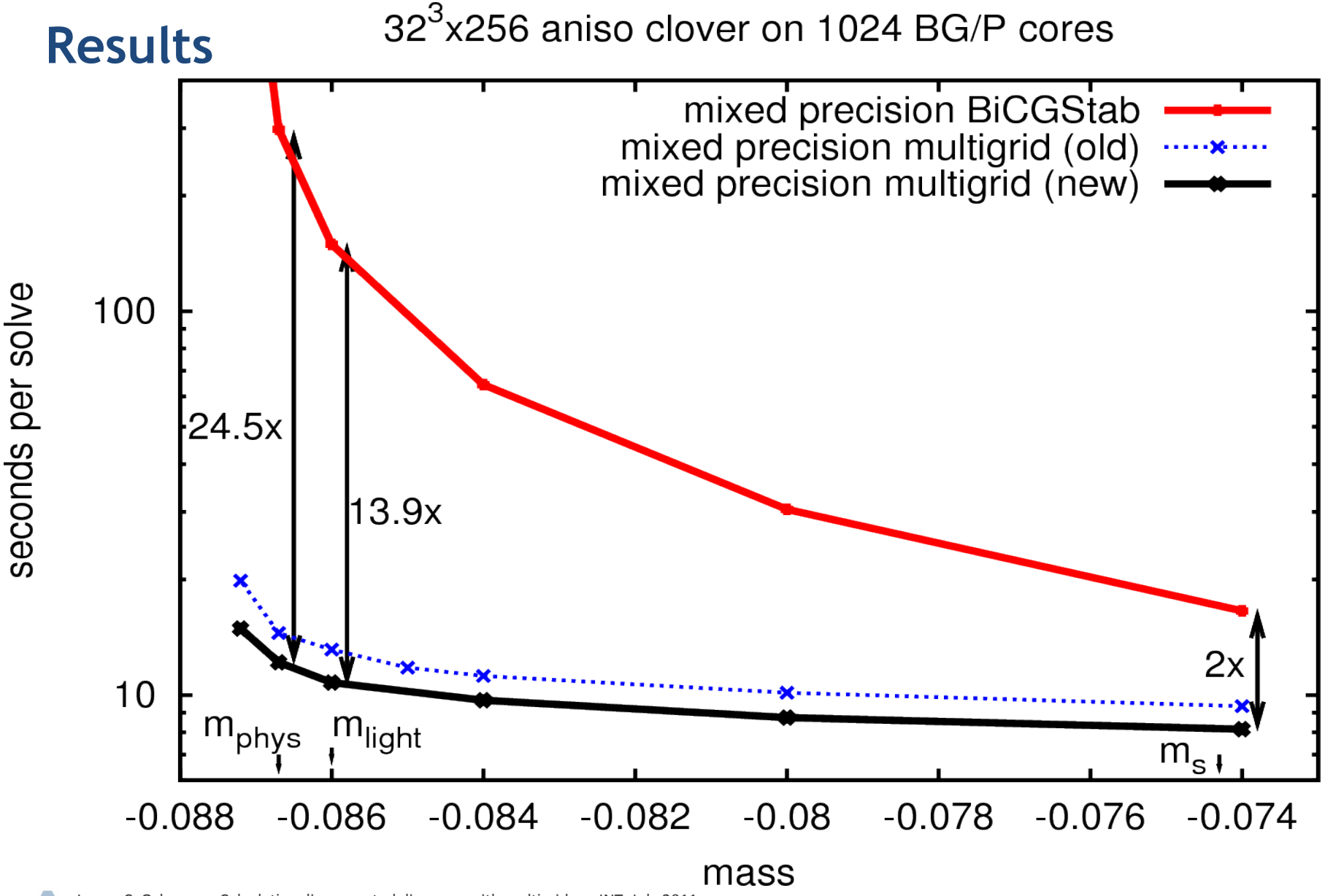
Results


24^3x128 aniso clover on 256 BG/P cores
mixed precision BiCGStab
mixed precision multigrid (old)
mixed precision multigrid (new)
seconds per solve
100
10
15.3x
12x
2.1x
m_{phys}
m_{light}
m_s
-0.088
-0.086
-0.084
-0.082
-0.08
-0.078
-0.076
-0.074
mass

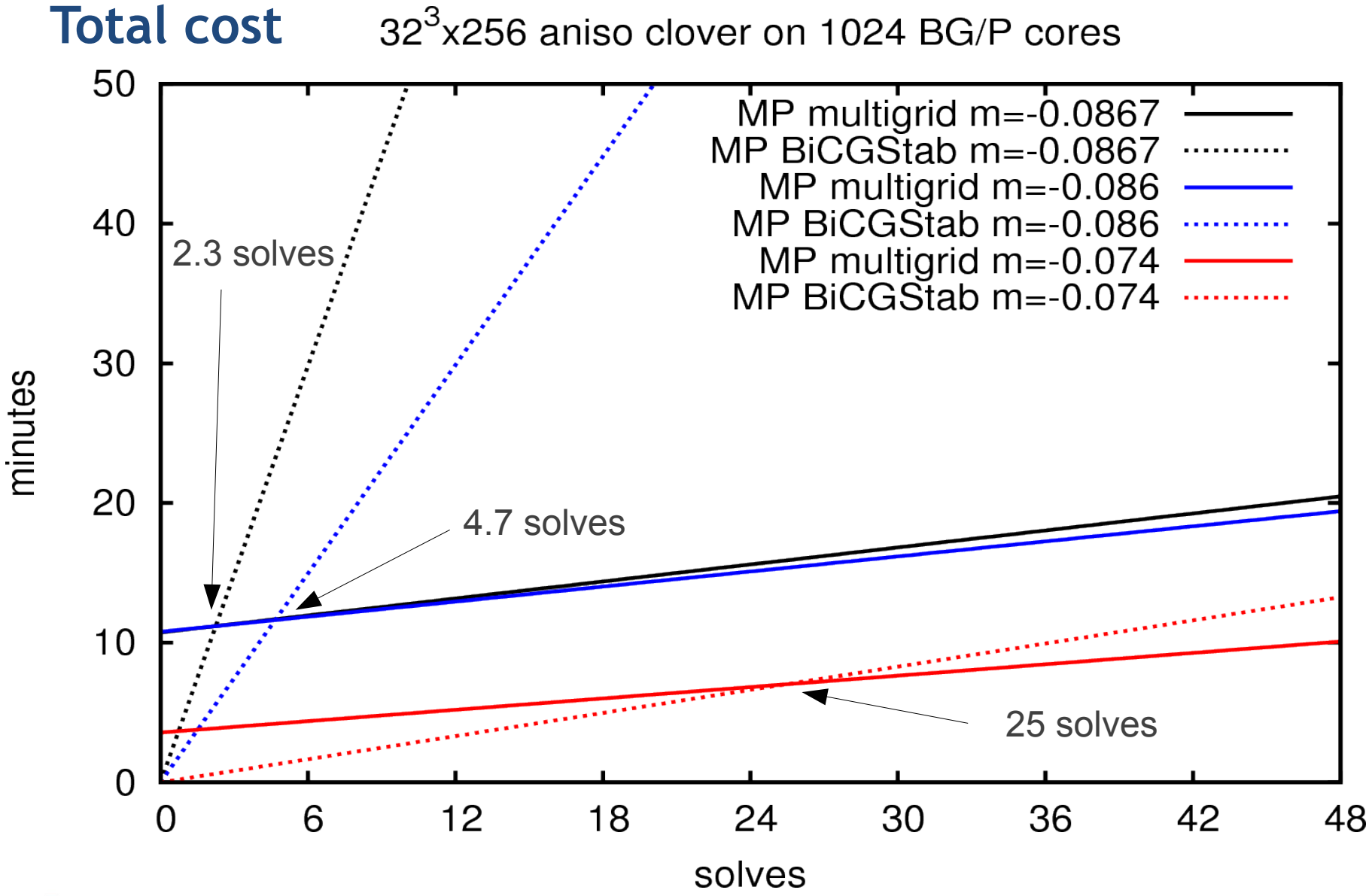
“Exceptional lattice”



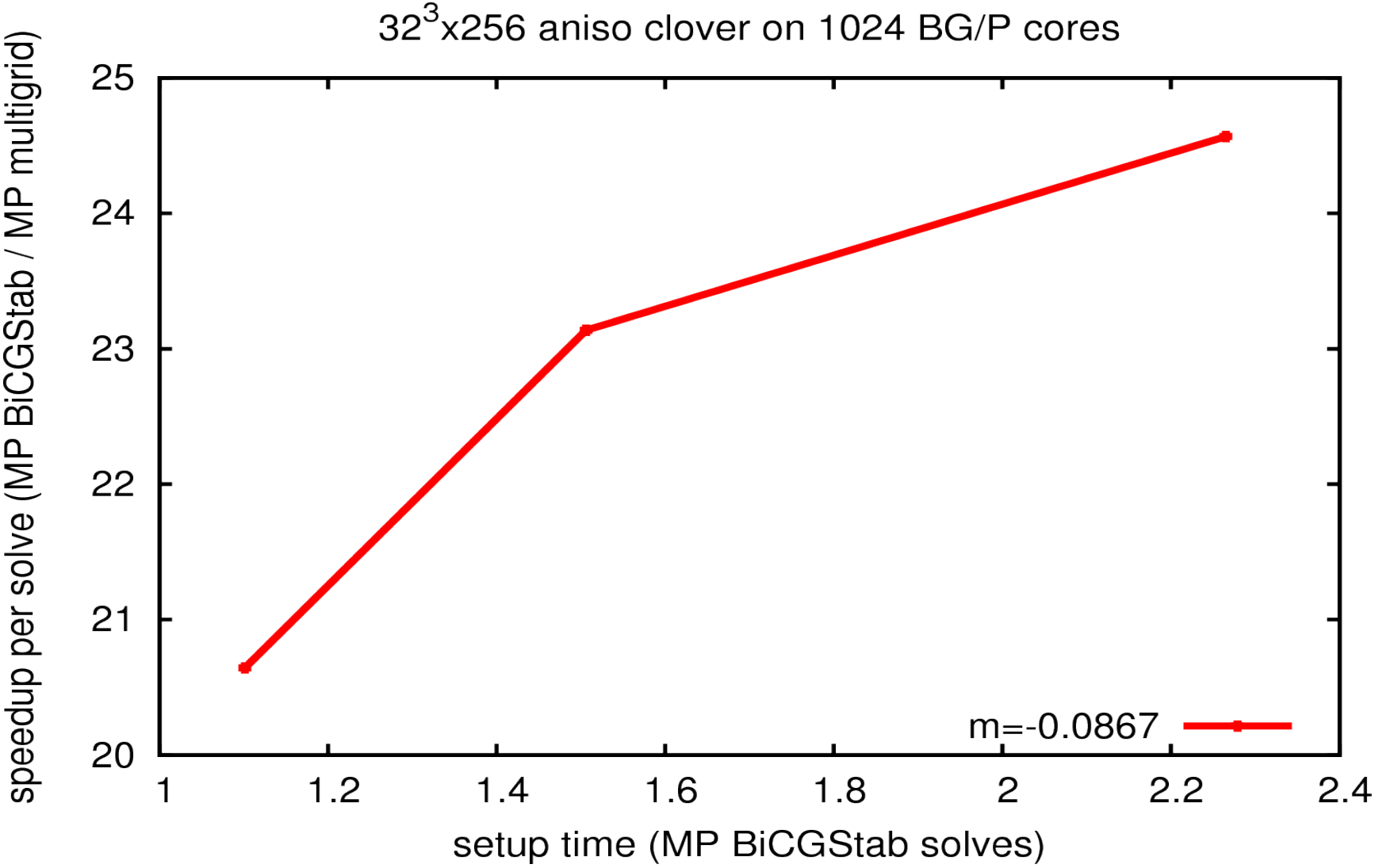
Results


32^3x256 aniso clover on 1024 BG/P cores
mixed precision BiCGStab
mixed precision multigrid (old)
mixed precision multigrid (new)
seconds per solve
100
10
24.5x
13.9x
2x
m_{phys}
m_{light}
m_s
-0.088
-0.086
-0.084
-0.082
-0.08
-0.078
-0.076
-0.074
mass

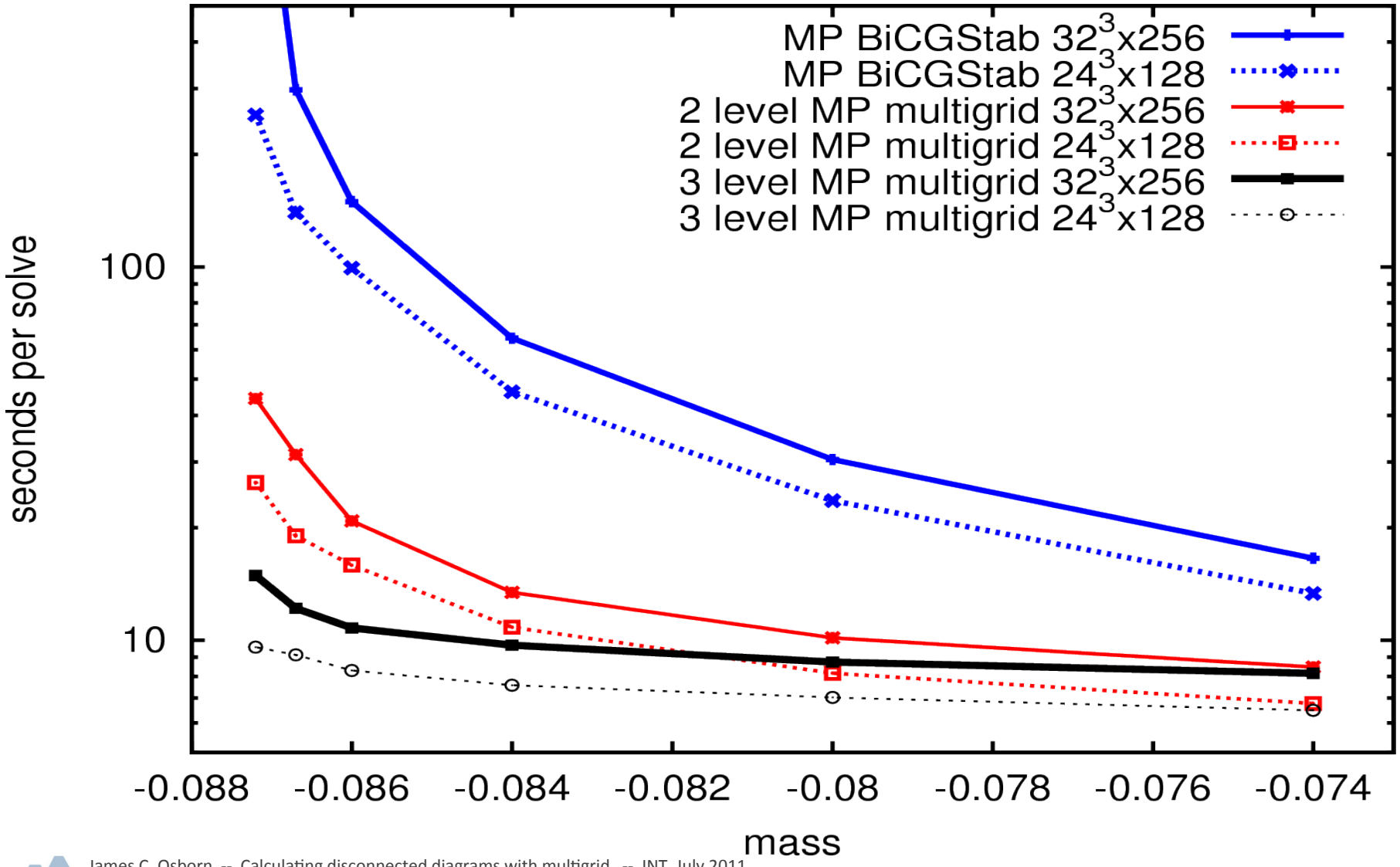
Total cost


32^3x256 aniso clover on 1024 BG/P cores
MP multigrid m=-0.0867
MP BiCGStab m=-0.0867
MP multigrid m=-0.086
MP BiCGStab m=-0.086
MP multigrid m=-0.074
MP BiCGStab m=-0.074
2.3 solves
4.7 solves
25 solves
minutes
solves
0
10
20
30
40
50
0
6
12
18
24
30
36
42
48

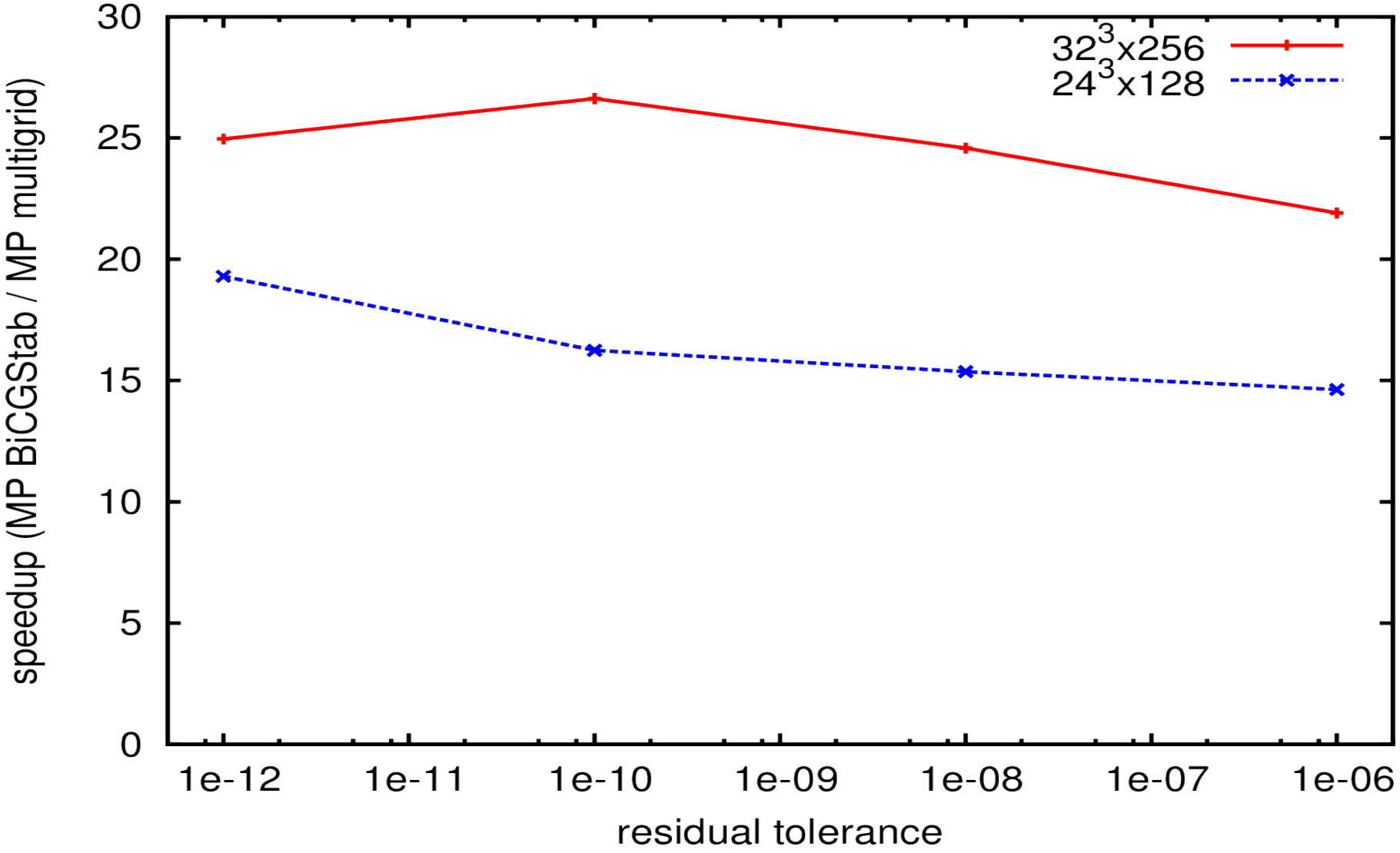
Setup cost vs speedup (physical quark mass)



2 level vs 3 level

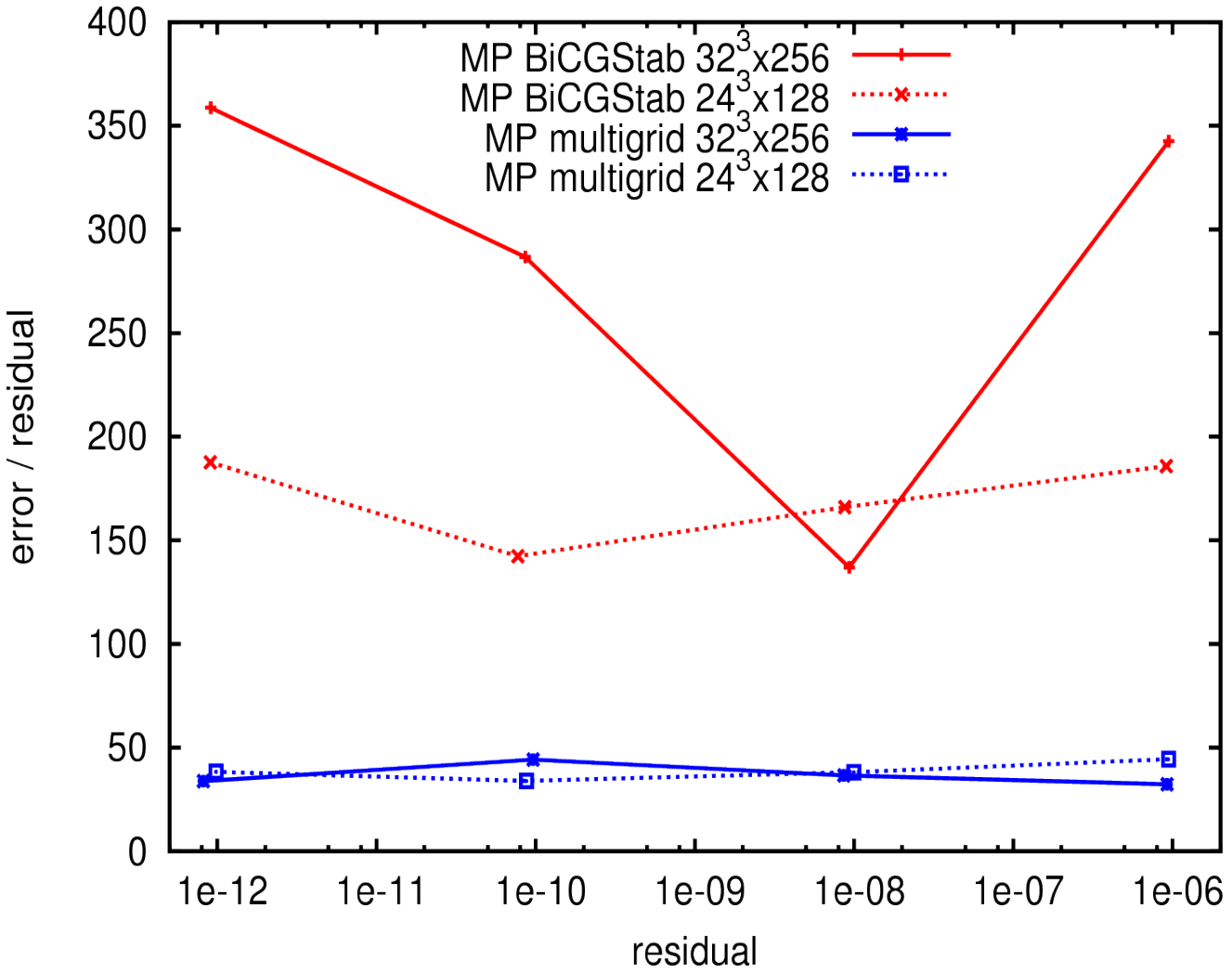


Speedup vs residual (physical mass)



Error vs residual

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



Obligatory exascale reference

- Multigrid reduces time to solution by reducing problem size
  - Coarsest lattice has 1 site/core
- Analysis jobs are trivially parallel over gauge configurations
  - Run on as small a partition as possible
  - Can still consume a large (Leadership Class) number of flops per project, though individual jobs are not large (Leadership Class) themselves
- Configuration generation needs to scale to large machines to evolve gauge field quickly – major challenge for exascale

Obligatory exascale reference

- Remaining challenges:
  - Scale to large number of cores
    (while retaining similar quality of solver)
  - Integrate with HMC
  - Update low modes directly (Lüscher's DD-HMC)
  - Implement other Dirac operators
    (Domain Wall: Saul, Improved Staggered)
  - Port to other architectures (GPUs)

Summary

- Disconnected diagrams are hard (require many solves)
- Multigrid can reduce cost of solves by 20-25x
  - Error very stable and relatively small to Krylov methods
  - Speedup (and relative error) improves for larger lattices
  - Less sensitive to “exceptional” configurations
  - Makes projects requiring many solves at light masses feasible
- Requires more work to efficiently scale, but still useful even if not running at the optimal configuration