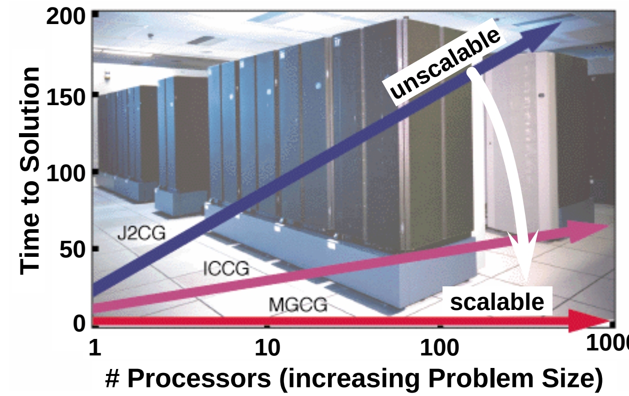
Research Directions in Scalable Algorithms

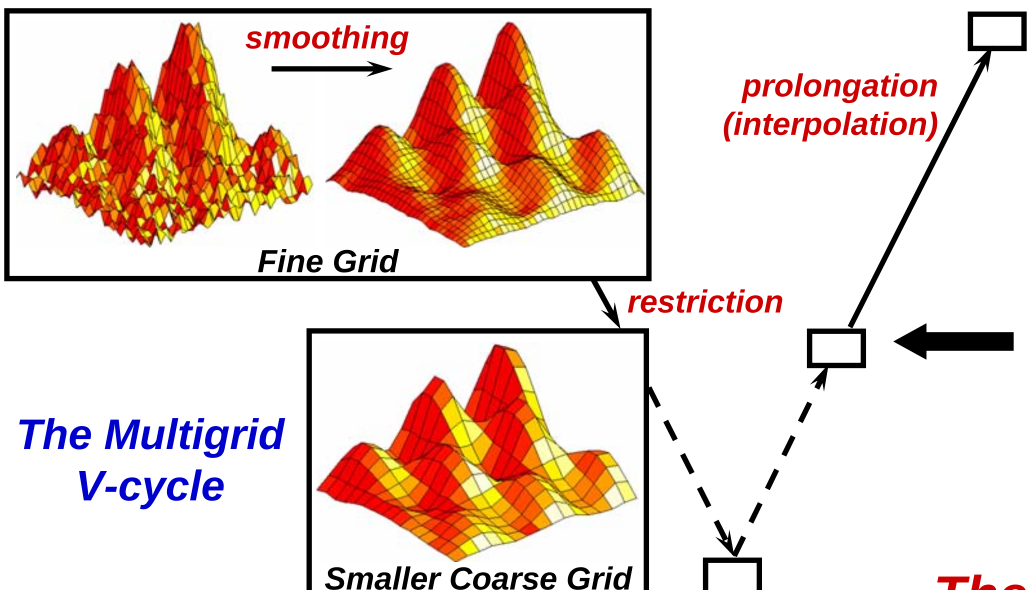
Robert D. Falgout
Center for Applied Scientific Computing
Lawrence Livermore National Laboratory

Panel Discussion
SIAM Conference on CSE
February 20, 2007

The scalable solution of linear systems is crucial in large-scale simulations

- **Multigrid linear solvers are optimal ($O(N)$ operations), and hence have good scaling potential**





The Multigrid V-cycle

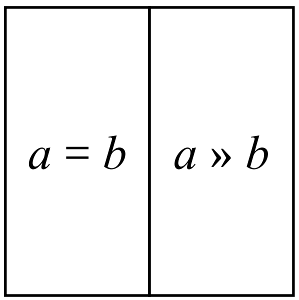
MG uses a sequence of coarse-grid problems to accelerate the solution of the original problem

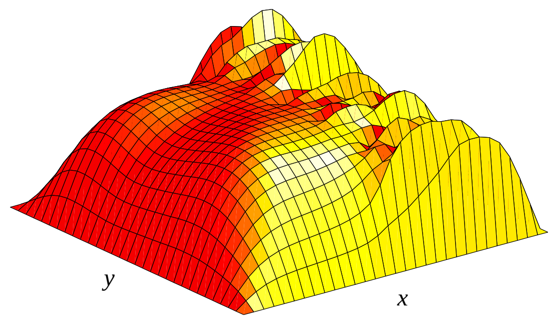
The near null space (kernel) is important!

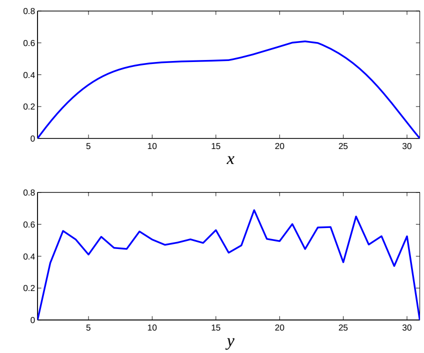
Error left by relaxation can be geometrically oscillatory

- **7 GS sweeps on**

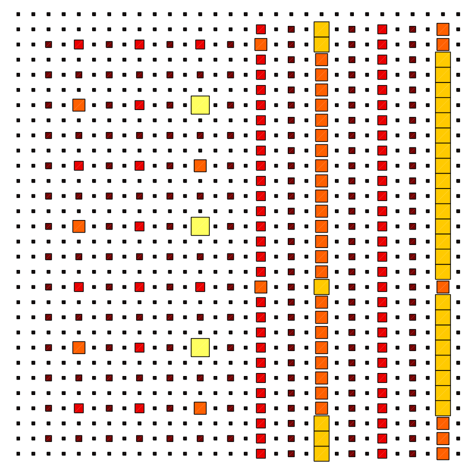
$$-au_{xx} - bu_{yy} = f$$







- **AMG automatically coarsens grids — can "follow physics"**
- **This example still targets geometric smoothness and pointwise smoothers**
- **Not sufficient for some problems!**



AMG coarse grids

Electromagnetic problems have huge near null spaces that are geometrically oscillatory

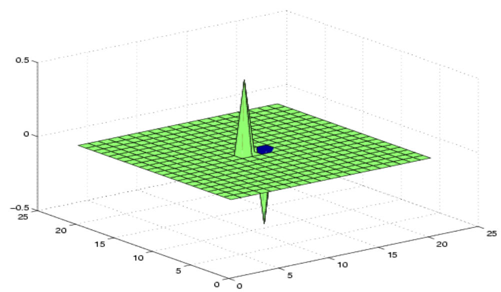
- **Three classes of PDEs:**

$$\nabla \times \alpha \nabla \times \mathbf{E} + \beta \mathbf{E} = f$$ – Definite Maxwell (α, $\beta > 0$)

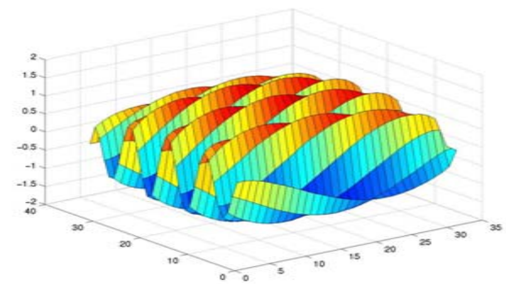
$$\nabla \times \alpha \nabla \times \mathbf{E} - k^2 \mathbf{E} = f$$ – Indefinite Maxwell ($\alpha > 0$)

$$-\nabla^2 u - k^2 u = f$$ – Helmholtz

- **Requires specialized smoothers and coarse grids**



Local: specialized relaxation
(Definite Maxwell, Indefinite Maxwell)



Global: specialized coarse grids
(Helmholtz, Indefinite Maxwell)

- **Good recent progress for Definite Maxwell!**

Adaptive AMG employs the idea of: *using the method to improve the method*

- **Requires no a-priori knowledge of near null-space**
- **Idea:** uncover slowly-converging error components by applying the "current method" to the system $Ax = 0$, then use these to adapt (improve) the method

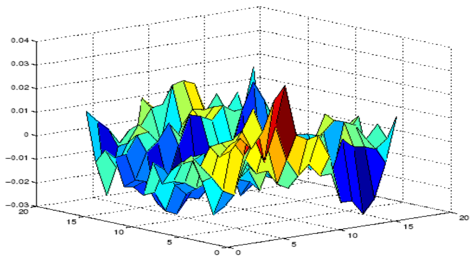
- **PCG can be viewed as an adaptive method**
  - Not optimal because it uses a global view
  - The key is to view slow-to-converge components as "representatives" of locally smooth error

- **Two methods:** α*AMG* and α*SA* (SISC pubs)
- **Prolongation in α*SA* formed by**
  - "chopping up" the representatives, then
  - smoothing to lower the overall energy

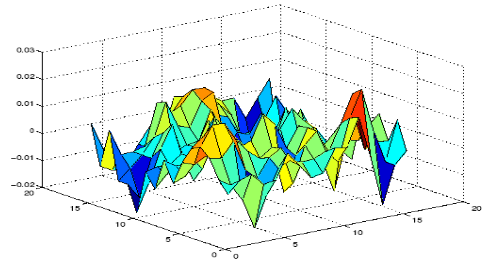
We are applying our adaptive *AMG* methods to QCD

- **Quantum Chromodynamics (QCD) is the theory of strong forces in Standard Model of particle physics**
- **Challenges:**
  - **The system is complex and indefinite**
  - **The system can be extremely ill-conditioned**
  - **Near null space is unknown and oscillatory!**

Real part



Imaginary part

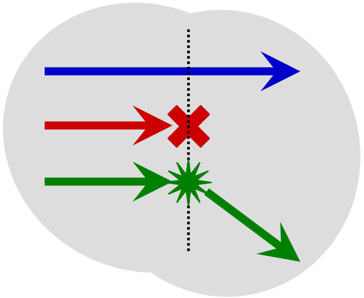


- **Uniform convergence of α*SA* in 2D (first such result)**
- **Extending to 4D**

Scalable, robust simulation of transport is a major issue in many codes

- **Transport plays a crucial role in many applications**
  - Stockpile stewardship, astrophysics, ICF
- **High dimensionality makes it a challenging problem**
  - 6D phase space (space, angle, energy) + time
- **Mono-energetic Boltzmann equation is a key kernel in radiative transfer and neutron transport**

$$\frac{1}{\nu}\frac{\partial\psi}{\partial t}+\underbrace{\Omega\cdot\nabla\psi}_{\text{Streaming}}+\underbrace{\sigma(r)\psi}_{\text{Absorption}}=\underbrace{\sigma_s(r)\int_{S^2}\psi(r,\Omega',t)\,d\Omega'}_{\text{(In)Scattering}}+q$$

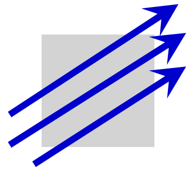


Underlying nature of transport equation changes in different parameter regimes

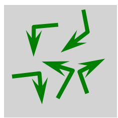
- Discretizion in *angle* (S_N discrete ordinates) and *space* (Petrov-Galerkin, corner balance) leads to

$$\left(\begin{array}{ccc|c} H_1 & & & -\Sigma_1 \\ & \ddots & & \vdots \\ & & H_n & -\Sigma_n \\ \hline -S_1 & \cdots & -S_n & I \end{array}\right) \left(\begin{array}{c} \Psi_1 \\ \vdots \\ \Psi_n \\ \hline \Phi \end{array}\right) = \left(\begin{array}{c} Q_1 \\ \vdots \\ Q_n \\ \hline 0 \end{array}\right)$$

- Traditional source iteration (SI) = block Gauss-Seidel
- Thin limit (little scattering): nearly block lower triangular and SI converges rapidly

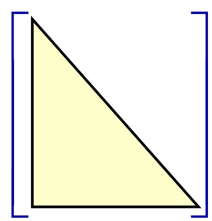


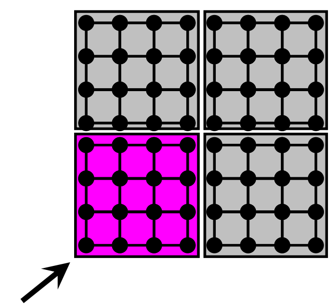
- Thick limit (high scattering): the system for the scalar flux Φ behaves like diffusion
  - SI converges slowly
  - DSA / TSA used to accelerate convergence



Very little work has been done on MG for the Boltzmann transport equations

- **MG developed mainly for 2nd order elliptic problems**
- **Challenges:** not elliptic, not symmetric, involves 1st order terms & integral terms
- **Many methods require so-called sweeps to invert the triangular streaming operators H_i**

$H_i =$ 



- **Current parallelization techniques may be sufficient even for BG/L → $O(dP^{1/d} + M)$**
  - Sweeping many directions M delays effect of P term
- **Parallel MG alternative to sweeps an open problem**

True scalability will require parallel multilevel methods in time

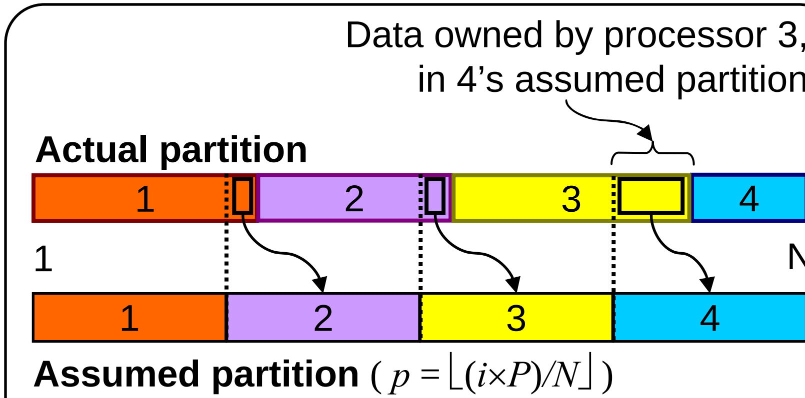
- As we refine the mesh, we also refine the time step
- To date, have relied on increases in processor speed
- This “solution” probably won’t work indefinitely

- Doing concurrent work in time is not a natural concept (we live our lives sequentially in time)
- It is possible, however, though not trivial

- Related to the sweep problem in transport
- Some work has been done on this already (e.g., Stefan Vandewalle at Leuven, Belgium)
- Still a very open (and interesting) problem!

New assumed partition (AP) algorithm enables scaling to 100K+ processors

- **Answering global data distribution queries previously required $O(P)$ storage and computations**
- **On BG/L, storing $O(P)$ data is not always practical or possible**
  - e.g., no `MPI_AllGather()`
- **New algorithm employs an assumed partition to answer queries through a kind of rendezvous algorithm**
- **Reduces storage to $O(1)$ and computations to $O(\log P)$!**



Actual partition info is sent to the assumed partition processors → distributed directory

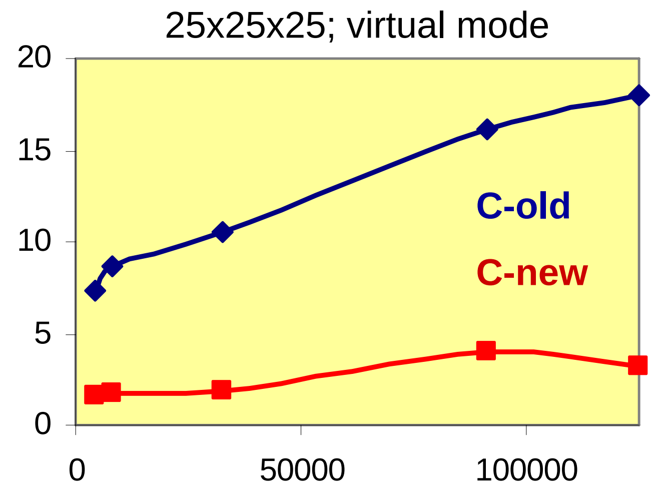
- **Now available in *hypre***
- **AP idea has general applicability beyond *hypre***

AMG is 16x faster and uses less memory with new AP and coarsening algorithms on BG/L

	global partition (old)		assumed partition (new)	
# of procs	C-old	C-new	C-old	C-new
4,096	12.42	3.06	12.32	2.86
64,000	67.19	10.45	**19.85**	**4.23**

7pt 3D Laplacian; 30x30x30 unknowns per processor; co-processor mode;
BoomerAMG-CG; total times in seconds; coarsening algorithms C-old & C-new

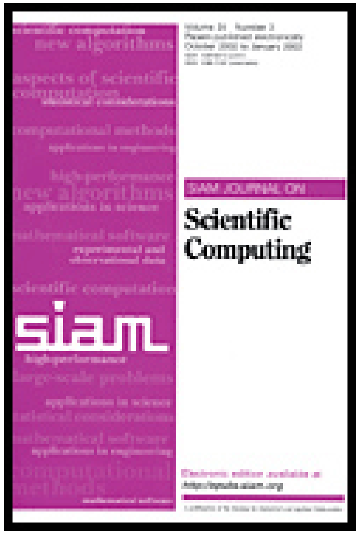
- **15x overall speedup on 64K procs!**
- **2 billions unknowns on 125K procs!**



SIAM J. Scientific Computing

Special Issue on CS&E

http://www.siam.org/journals/sisc.php



Guest Editors-in-Chief:

Chris Johnson

David Keyes

Ulrich Ruede

Call for papers:

- Modeling techniques
- Simulation techniques
- Analysis techniques
- Tools for realistic problems

Deadline for submissions: April 30, 2007

Guest Editorial Board: Gyan Bhanot, IBM Watson Research Center, Yorktown Hights; Rupak Biswas, NASA Ames Research Center, Moffett Field, CA; Edmond Chow, D. E. Shaw Research, LLC; Phil Colella, Lawrence Berkeley National Laboratory; Yuefan Deng, Stony Brook University, Stony Brook, NY Brookhaven National Lab; Lori Freitag Diachin, Lawrence Livermore National Laboratory; Omar Ghattas, The University of Texas at Austin, Laurence Halpern, Univ. Paris XIII; Robert Harrison, Oak Ridge National Laboratory; Bruce Hendrickson, Sandia National Laboratories; Kirk Jordan, IBM, Cambridge MA; Tammy Kolda, Sandia National Laboratories; Louis Komzsik, UGS Corp., Cypress, CA, USA; Ulrich Langer, Johann Radon Institute for Computational and Applied Mathematics (RICAM) Linz; Hans Petter Langtangen, Simula Research Laboratory, Oslo; Steven Lee, Lawrence Livermore National Laboratory; Kengo Nakajima, University of Tokyo; Aiichiro Nakano, University of Southern California, Los Angeles; Esmong G. Ng, Lawrence Berkeley National Laboratory; Stefan Turek, University of Dortmund, Germany; Andy Wathen, Oxford University; Margaret Wright, New York University; David P. Young, Boeing.

siam Society for Industrial and Applied Mathematics

Thank You!

This work was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under contract no. W-7405-Eng-48.

UCRL-PRES-228198