SAND2012-6279C

Towards Scalable Algebraic Multigrid Preconditioners for Parallel Drift-Diffusion and Magnetohydrodynamic Simulations

Modeling and Simulation of Transport Phenomena

Moselle Valley, Germany

July 31, 2012

Paul Lin, John Shadid, Ray Tuminaro, Jonathan Hu
Sandia National Labs

U.S. DEPARTMENT OF ENERGY
National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

Sandia National Laboratories

Motivation and Goals

- High fidelity solutions of transport phenomena for large-scale problems with complex physics
  - Semiconductor simulations
  - CFD/MHD
- Fully implicit Newton-Krylov solution approach
  - robust technique (promising for complex physics and chemistry)
  - but depends on efficiency of sparse linear solver
  - choice of preconditioner critical: robustness, efficiency, scalability
  - large-scale problems: multigrid

Semiconductor Drift-Diffusion Model

(with G. Hennigan, R. Hoekstra, J. Castro, D. Fixel, R. Pawlowski, E. Phipps, L. Musson, T. Smith, Shadid, Lin)

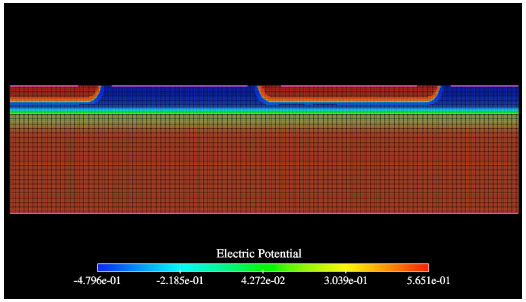
Electric potential $$-\nabla \cdot \epsilon \nabla \psi = q\,(p - n + C)$$

$$\nabla \cdot \mathbf{J}_n - qR = q\frac{\partial n}{\partial t} \qquad \mathbf{J}_n = -qn\mu_n \nabla \psi + qD_n \nabla n$$

$$-\nabla \cdot \mathbf{J}_p - qR = q\frac{\partial p}{\partial t} \qquad \mathbf{J}_p = -qp\mu_p \nabla \psi - qD_p \nabla p$$

- ψ: electric potential
- n: electron concentration
- p: hole concentration
- C: doping profile
- R: generation-recombination term

Defect species: each additional species adds an additional transport-reaction equation

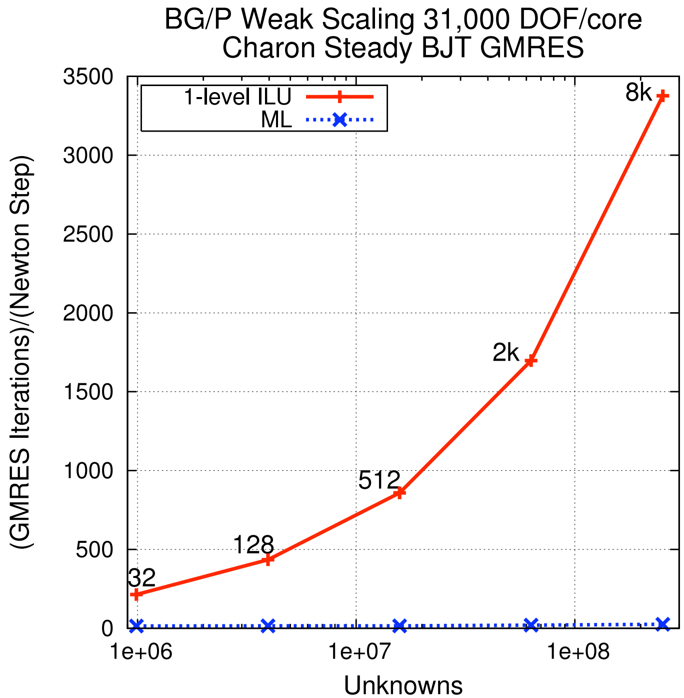


Drift-Diffusion Solution Approach

- Discretization: stabilized FEM or FVM on unstructured meshes
- Fully-implicit solution approach: Newton-Krylov solver
  - Pro: robustness; better for complex physics
  - Con: huge sparse linear systems to solve
- Need efficient solution of large sparse linear systems
- Preconditioning critical for scalability and efficiency
  - Linear system is solved for each Newton step: need to reduce iteration count; need iteration count to scale well
  - Need time/iteration to scale well
- Using solvers in SNL Trilinos library
- Currently MPI-only; one MPI process per core

Need to Reduce Iteration Count

- 1-level preconditioners (e.g. additive Schwarz) do not scale due to lack of global coupling
  - For 2D drift-diffusion GMRES iteration count scales by sqrt(DOF)
- Need methods with global coupling such as multilevel/multigrid



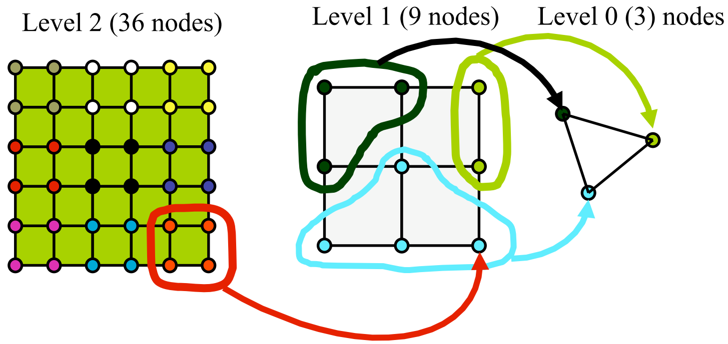
IBM Blue Gene/P

- 36k nodes; each with single 850MHz quadcore PowerPC 450
- 144k cores (147,456)
- 500 TF theoretical peak
- Interconnect: 3D torus +others

Trilinos ML Library: Algebraic Multigrid Preconditioners

(R. Tuminaro, J. Hu, C. Siefert, M. Sala, M. Gee, C. Tong)

- Aggressive coarsening with graph partitioner and pre-specified # of levels
  - Large difference in size between levels
  - Graph partitioner: serial for all levels, parallel for final level



- Aggregates to produce a coarser operator
  - Create graph where vertices are block nonzeros in matrix A_k
  - Edge between vertices i and j added if block $B_k(i,j)$ contains nonzeros
  - Decompose graph into aggregates
- Restriction/prolongation operator
- $A_{k-1} = R_k A_k P_k$

- Petrov-Galerkin smoothed aggregation for nonsymmetric matrices
  - Separate restriction smoothing
  - Local damping parameters

$$P_i = (I - \omega_i D^{-1} A)\hat{P}_i$$
$$R_i = \hat{P}_i^T (I - AD^{-1}\omega_i^{(r)})$$

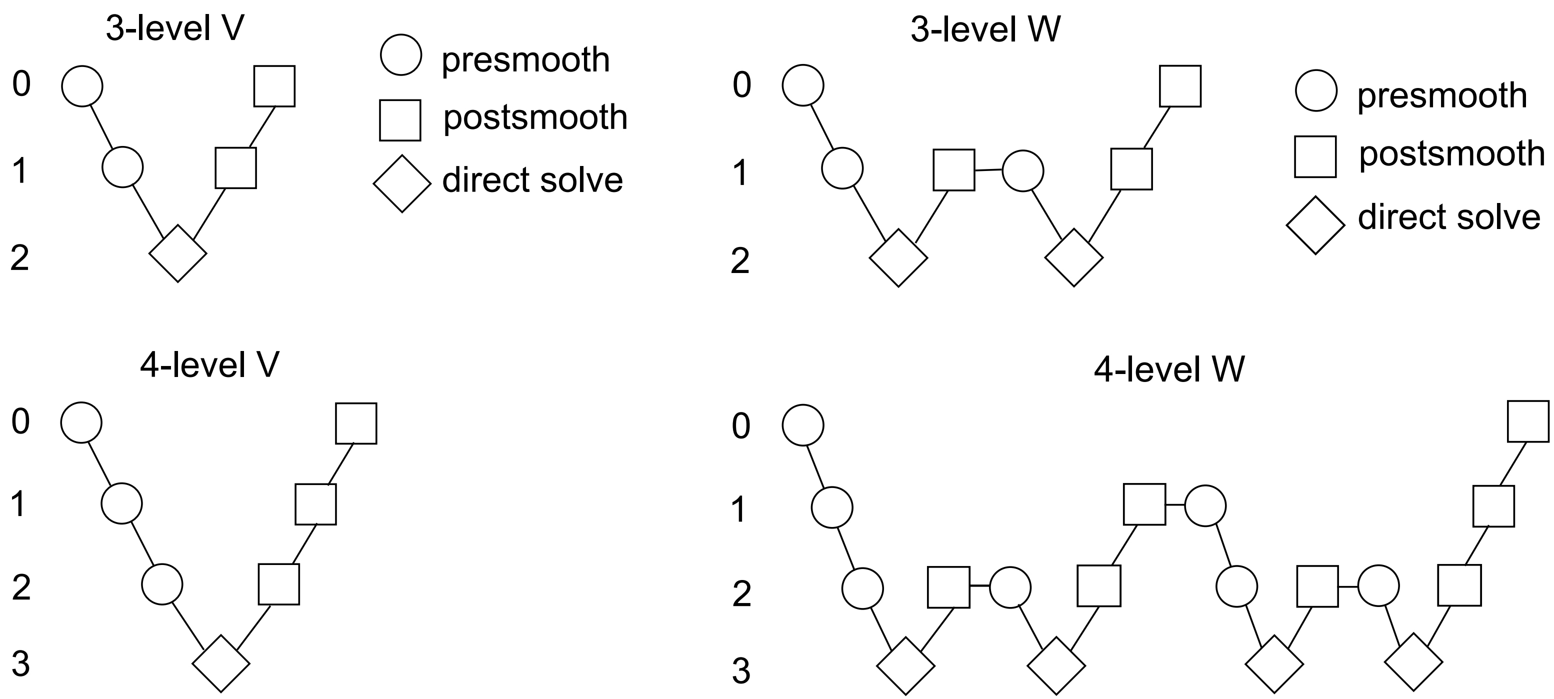
U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

Sandia National Laboratories

Coarsening Schemes

- (Aggressive) coarsening with graph partitioner and coarse matrix repartitioning (Zoltan RCB)
  - METIS for all levels
  - Keep coarsening until matrix is below threshold size
  - More “mesh nodes” on cores; better quality aggregates
- Uncoupled aggregation, coarse matrix repartitioning (Zoltan RCB)
  - Uncoupled: stencil is nearest neighbor, aggregates cannot span processes
  - Keep coarsening until matrix is below threshold size
  - Smaller difference in size between levels, e.g. ~9 for FEM 2D drift-diffusion
  - Better quality aggregates
- Multigrid cycle
  - V-cycle: fewer solves at coarser levels (e.g. 7-lev, 1 KLU apply)
  - W-cycle: more solves at coarser levels (e.g. for 7-levels, apply KLU 32 times)

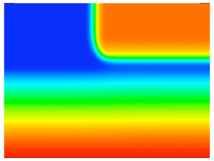
Multigrid Cycles: V-cycle and W-cycle

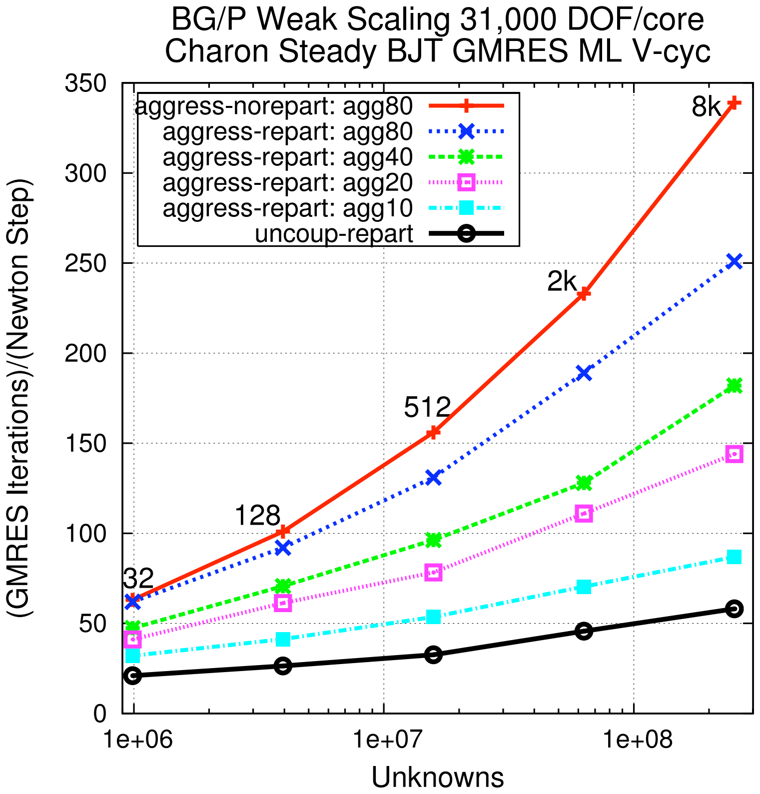


- V-cycle: fewer solves at coarser levels (e.g. for 7-level, one KLU apply)
- W-cycle: more solves at coarser levels
    - 5-, 6- and 7-level W-cycle have 8, 16 and 32 direct solves

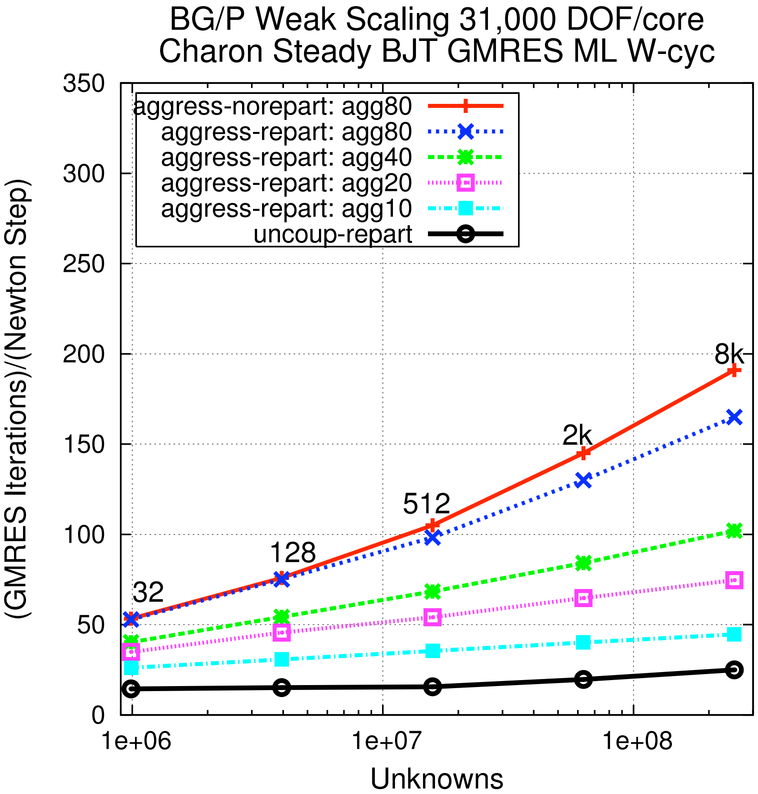
Weak Scaling: Iteration Count for Different Coarsening Schemes

- 2D BJT steady-state drift-diffusion
- 4-level aggressive coarsening: METIS/METIS/ParMETIS
- METIS, coarse matrix repart; Uncoupled, coarse matrix repart
- Scaled to 8192 cores and 250 million DOF; 31000 DOF/core





- V-cycle



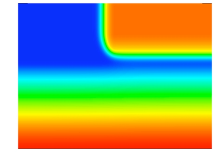
- W-cycle

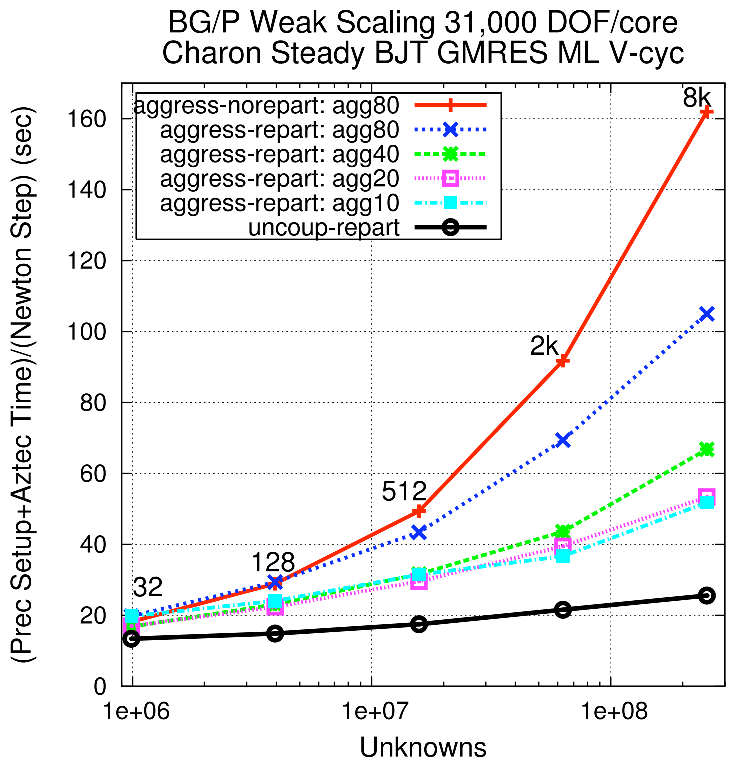
- GMRES

U.S. DEPARTMENT OF ENERGY
NNSA National Nuclear Security Administration
Sandia National Laboratories

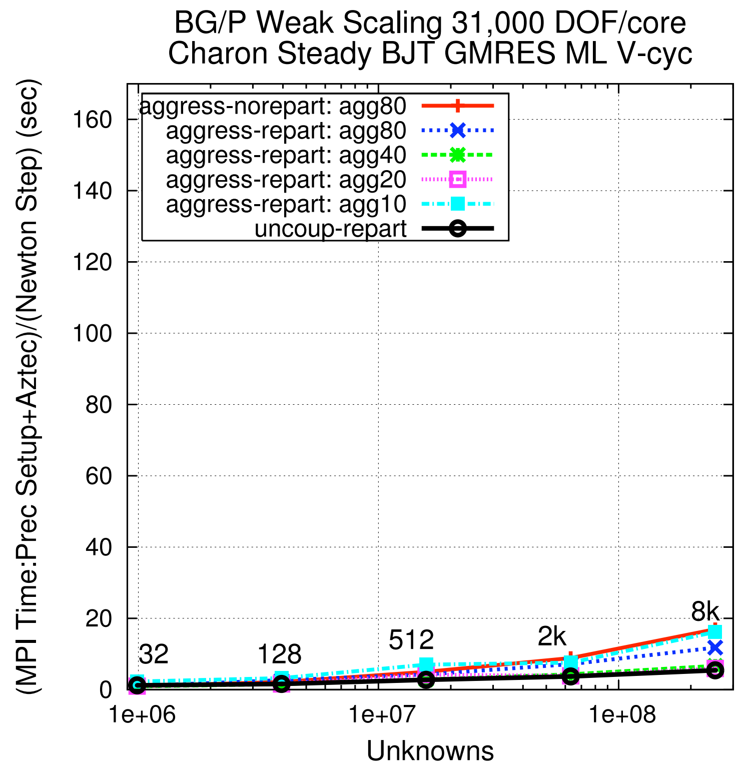
Weak Scaling: Solution Time for Different Coarsening Schemes

- 2D BJT steady-state drift-diffusion
- 4-level aggressive coarsening: METIS/METIS/ParMETIS
- METIS, coarse matrix repart; Uncoupled, coarse matrix repart
- Scaled to 8192 cores and 250 million DOF; 31000 DOF/core





- V-cycle: prec setup+Aztec



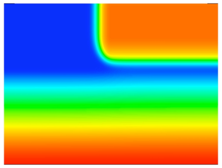
- V-cycle: MPI time

- GMRES

U.S. DEPARTMENT OF ENERGY
NNSA National Nuclear Security Administration
Sandia National Laboratories

Weak Scaling Study: 1-level vs. Multigrid

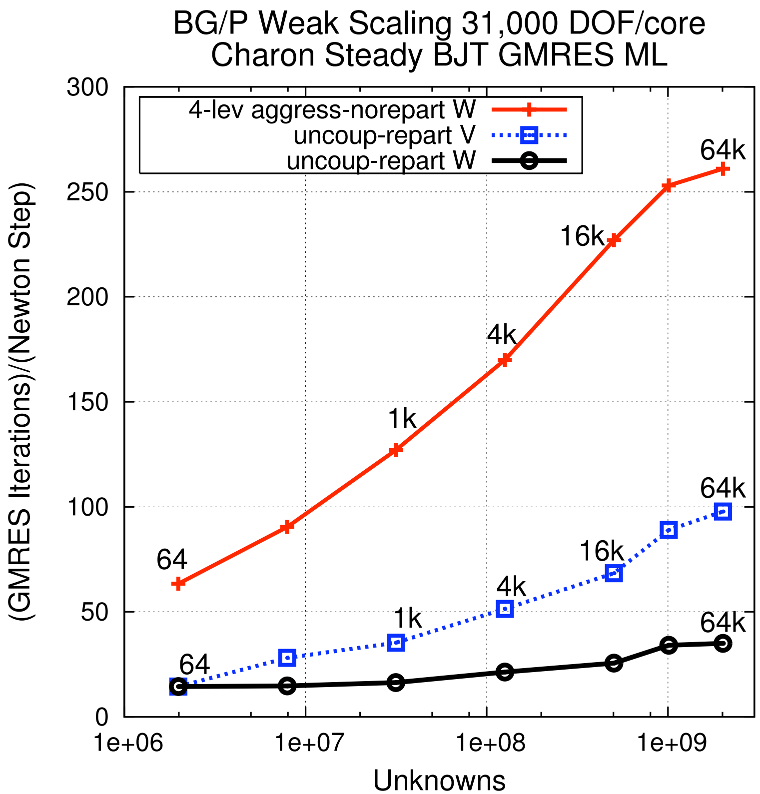
- 2D BJT steady-state drift-diffusion
- Uncoupled aggregation with coarse matrix repartitioning
- Problem scaled to 8192 cores and 252 million DOF
- GMRES Krylov solver

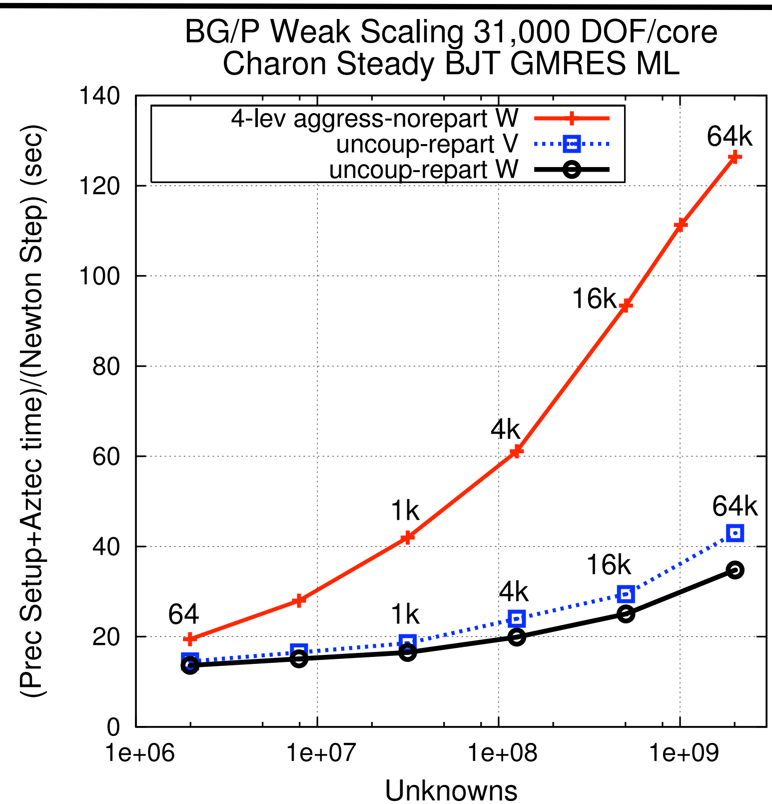


core	fine grid unknowns	1-level ILU		Uncoupled V-cyc		Uncoupled W-cyc	
		ave its per Newt step	time per Newt (s)	ave its per Newt step	time per Newt(s)	ave its per Newt step	time per Newt (s)
32	988533	214	55	21	13.5	14	13.1
128	3.95E+06	435	192	26	14.9	15	13.6
512	1.58E+07	859	697	33	17.5	16	15.6
2048	6.31E+07	1697	2634	46	21.6	20	18.1
8192	2.52E+08	3377	10559	58	25.6	25	22.6

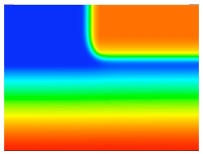
- Compared with 1-level preconditioner for 8192-core, 252 million DOF case
  - Uncoupled agg V-cyc reduces iterations by 182x, time by 412x
  - Uncoupled agg W-cyc reduces iterations by 422x, time by 467x

Reducing Iteration Count: Improved Aggregation



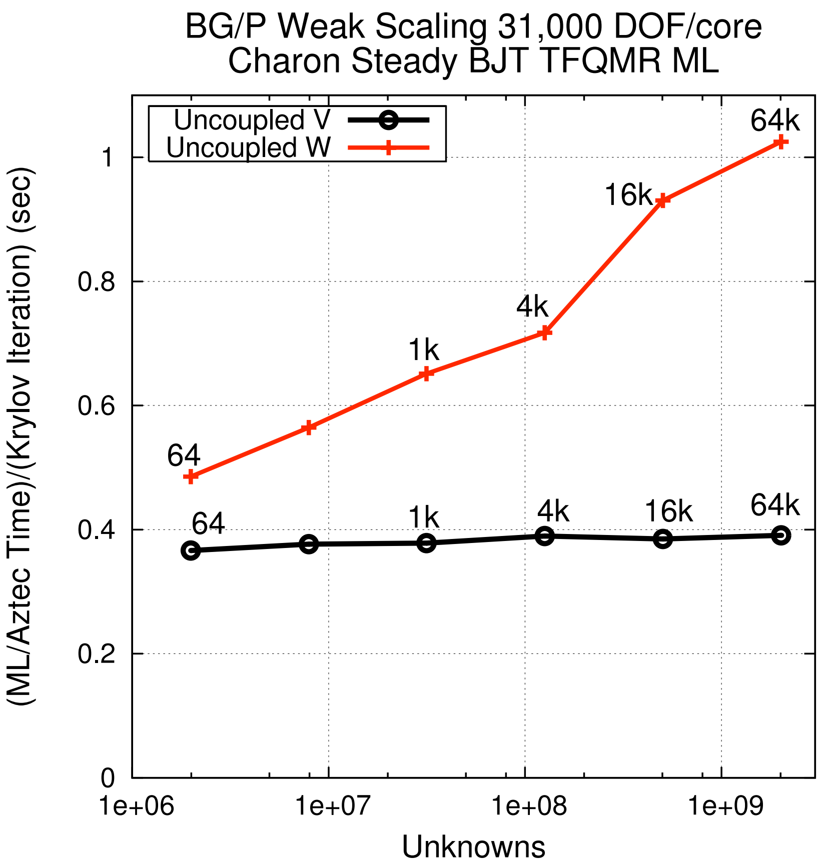


- Uncoupled with matrix repartitioning
  - more levels (up to 7); better aggregates
- Significantly reduces iterations: W-cyc by ~8x, V-cyc of ~3x for 2 billion DOF for 64k
- Time reduction: W-cyc 3.6x 2 billion DOF, 64k



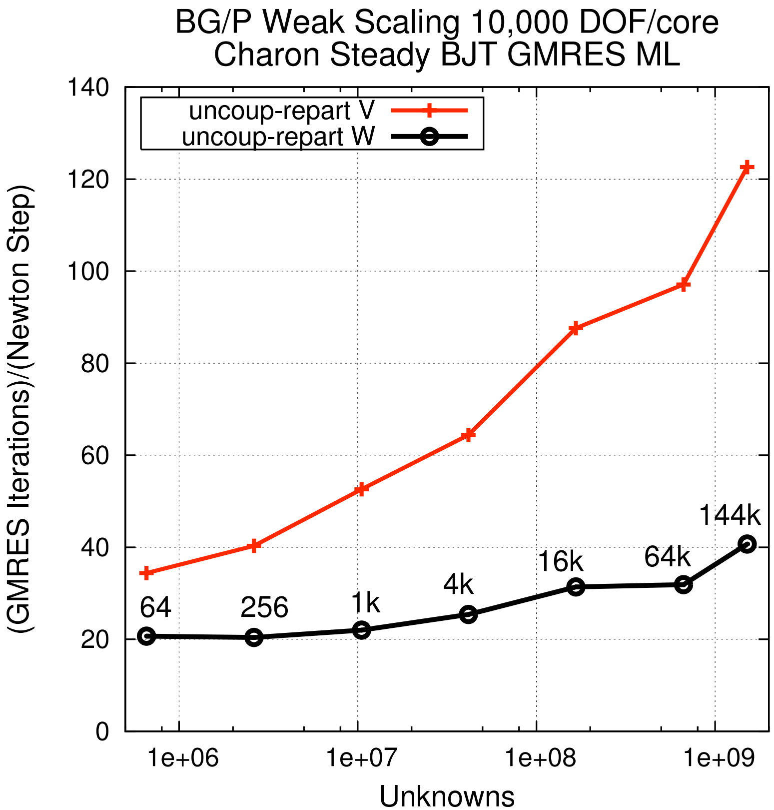
U.S. DEPARTMENT OF ENERGY · NNSA National Nuclear Security Administration · Sandia National Laboratories

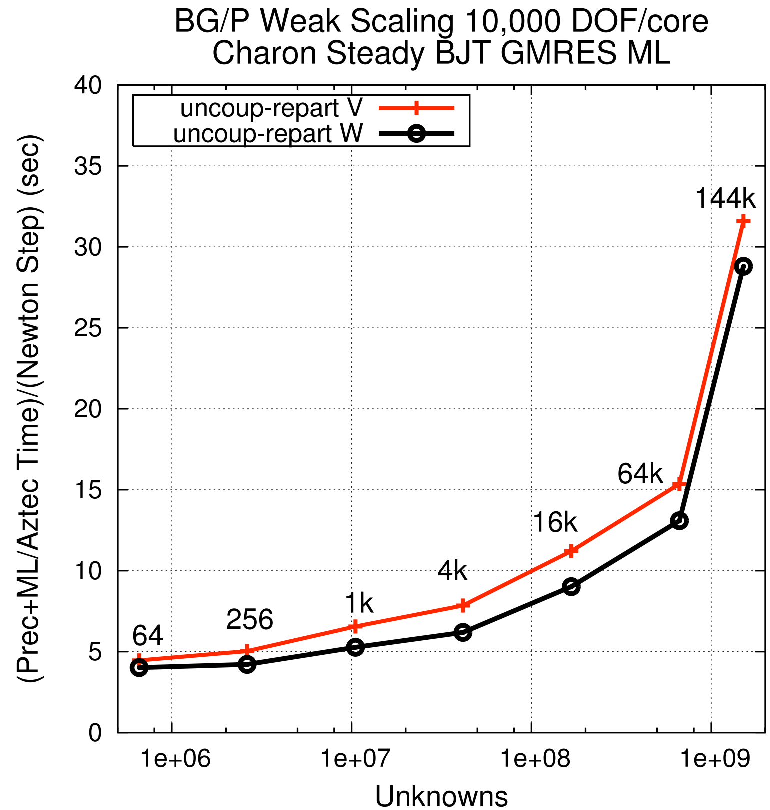
Uncoupled Aggregation: Time/Iteration



- TFQMR: can look at time/ iteration
- V-cyc time/iteration flat from 64 to 64k cores
- W-cyc time/iteration not doing well due to significant increase in work on coarse levels
- V-cyc and W-cyc require about the same total time, even though W-cyc had fewer iterations/Newton step

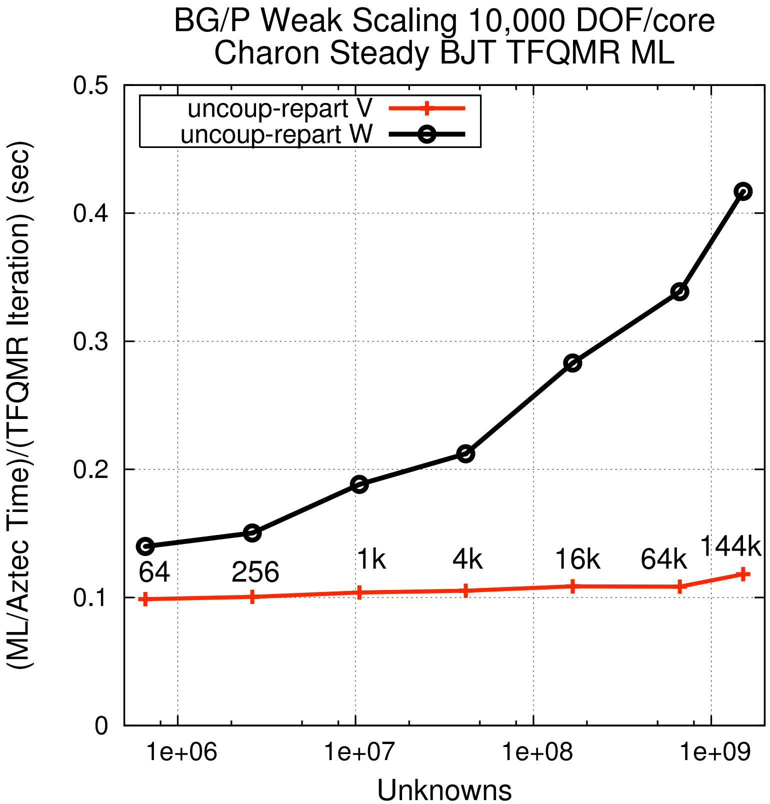
Weak Scaling to 147,000 Cores





- 10,000 DOF/core; 1.47 billion DOF at 147,000 cores
- GMRES; Uncoupled aggregation with coarse matrix repartitioning
- Problem size increased 2304x: W-cycle iter increased 2.0x; (prec +Aztec) time increased 8.8x (prec setup time is double Aztec)

Weak Scaling to 147,000 Cores: Time/Iteration

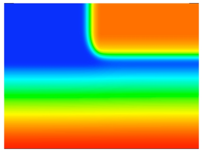


- 10,000 DOF/core; 1.47 billion DOF at 147,000 cores
- TFQMR uncoupled aggregation
- V-cyc: time/iteration
  - 64 to 64k increases 28%
  - 64k to 144k increases 12%
- W-cyc time rapidly increases due to larger amount of work at coarser levels

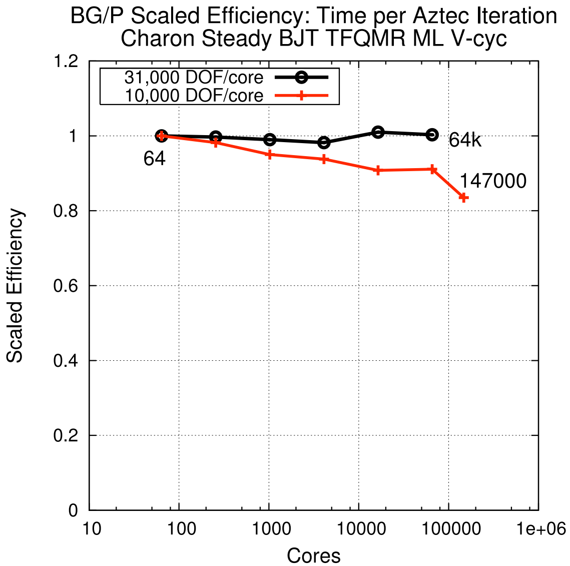
U.S. DEPARTMENT OF ENERGY NNSA National Nuclear Security Administration

Sandia National Laboratories

Overall Performance: Still Have Work To Do



- Time per iteration scales well for V-cycle
- Next challenge to improve overall performance
    - Improve preconditioner setup time
    - Improved repartitioning to minimize data movement in traversing mesh hierarchy and application of preconditioner
    - Eliminate re-computation of symbolic graph algorithms for projection and for matrix graphs (static meshes)
    - Work to obtain true h-independent iteration counts



Time per Krylov Iteration

- V-cyc overall performance:
- 64 to 64k: prec+Aztec 3x slower
    - 35% due to prec, 65% due to Aztec inc
    - Prec 1.6x slower
    - Aztec 5.7x slower (4.3x iter inc)
- 64 to 147000: prec+Aztec 7.1x slower
    - 60% due to prec, 40% due to Aztec inc
    - Prec 8.3x slower
    - Aztec 5.7x slower (3.6x iter inc)

Resistive MHD Model

(J. Shadid, R. Pawlowski, E. Cyr, L. Chacon)

Navier-Stokes + Electromagnetics

$$\rho \frac{\partial \mathbf{u}}{\partial t} + \rho(\mathbf{u} \cdot \nabla \mathbf{u}) - \nabla \cdot (\mathbf{T} + \boxed{\mathbf{T_M}}) - \rho \mathbf{g} = 0$$

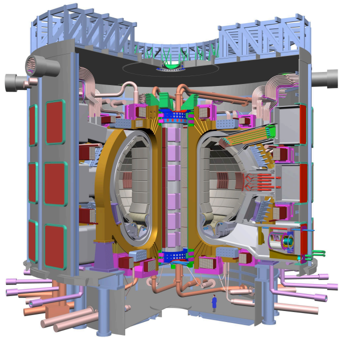
$$\mathbf{T} = -(P + \frac{2}{3}\mu(\nabla \cdot \mathbf{u}))\mathbf{I} + \mu[\nabla \mathbf{u} + \nabla \mathbf{u}^T]$$

$$\mathbf{T_M} = \frac{1}{\mu_0}\mathbf{B} \otimes \mathbf{B} - \frac{1}{2\mu_0}\|\mathbf{B}\|^2 \mathbf{I}$$

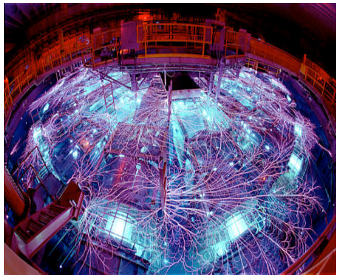
$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{u}) = 0$$

$$\rho C_p \left[\frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T \right] + \nabla \cdot \mathbf{q} - \boxed{\eta \|\mathbf{J}\|^2} = 0$$

$$\boxed{\frac{\partial \mathbf{B}}{\partial t} - \nabla \times (\mathbf{u} \times \mathbf{B}) + \nabla \times (\frac{\eta}{\mu_0} \nabla \times \mathbf{B}) = \mathbf{0}}$$



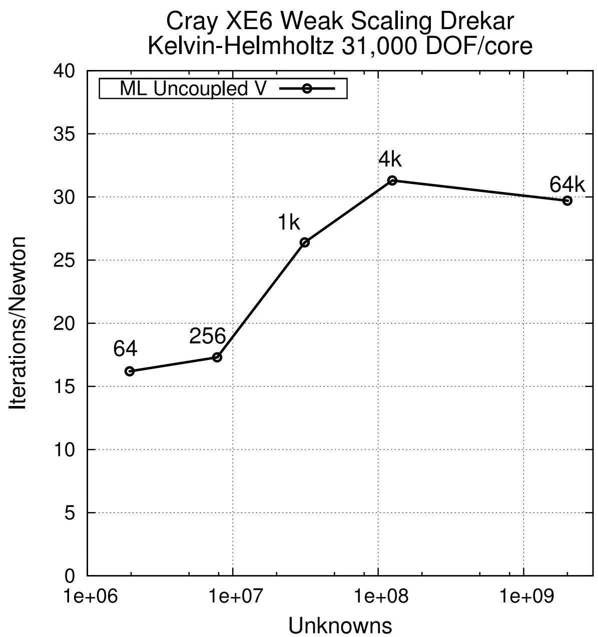
ITER

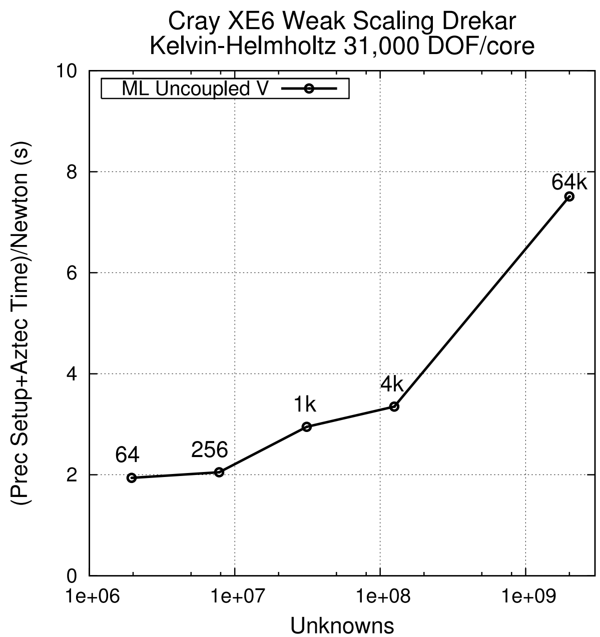


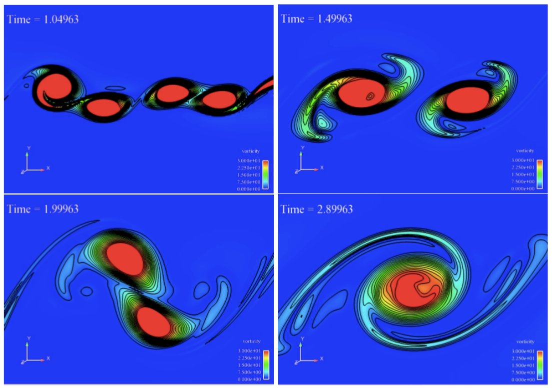
SNL Z-machine

Preliminary Weak Scaling: Kelvin-Helmholtz

Drekar (CFD/MHD): R. Pawlowski, E. Cyr, J. Shadid, P. Gabel



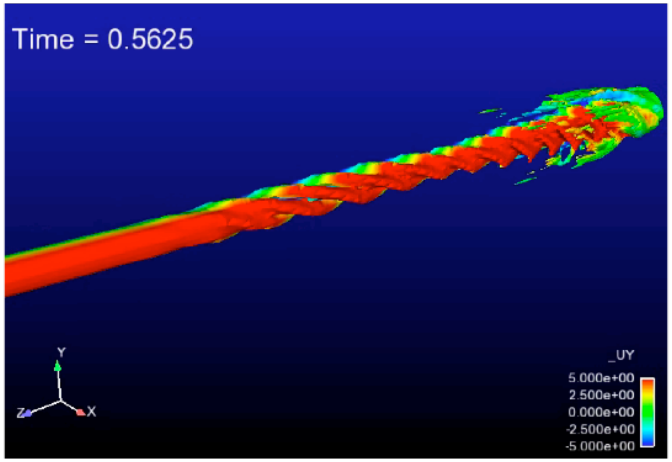




- 2D Transient Kelvin-Helmholtz instability
- Re=5000 shear layer, CFL ~2
- Cray XE6

Preliminary Weak Scaling: 3D CFD (Fixed CFL)

Drekar (CFD/MHD): R. Pawlowski, E. Cyr, J. Shadid

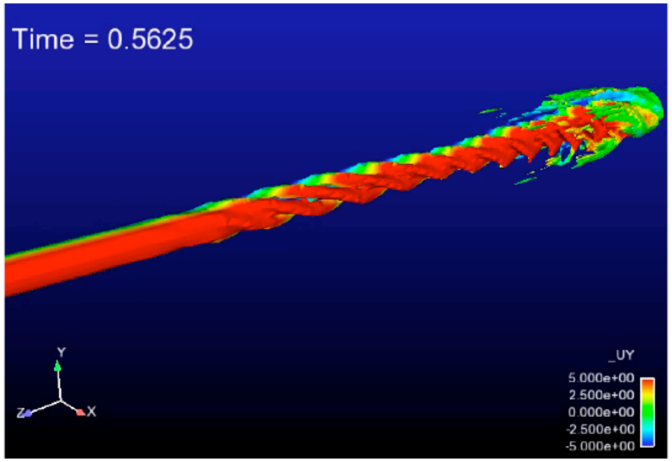


- 3D transient swirling jet, fixed CFL (CFL~1)
- GMRES, ML prec
- Cray XE6

cores	Newt/dt	Iter/Newt step	Time/Newton step (sec)	
			Prec	Aztec
256	3.7	14	1.3	1.0
2048	4.0	20	1.8	1.6
16384	4.0	30	2.8	3.0
131072	3.8	34	5.4	3.5

Preliminary Weak Scaling: 3D CFD (Fixed dt)

Drekar (CFD/MHD): R. Pawlowski, E. Cyr, J. Shadid

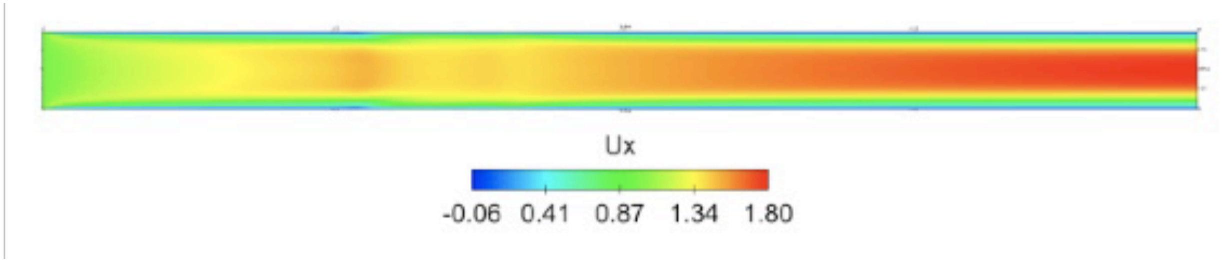


- 3D transient swirling jet, fixed dt=0.001
- GMRES, ML prec
- Cray XE6

cores	Newt/dt	Iter/Newt step	Time/Newton step (sec)		TFQMR time/iter
			Prec	Aztec	
256	Failed to converge				
2048	2.3	22	1.9	1.0	0.15
16384	3.6	27	2.2	2.4	0.18
131072	3.8	34	5.4	3.5	0.19

Preliminary Weak Scaling: MHD Generator

Drekar (CFD/MHD): R. Pawlowski, E. Cyr, J. Shadid



- 3D Steady-state MHD Generator
- Inlet V=1, permanent magnet supplies nonzero By
- Cray XE6

cores	Iter/Newt step	Time per Newton step (sec)		
		Prec	Aztec	Prec+Aztec
32	10	15.4	2.2	17.6
256	14	16.1	3.2	19.3
2048	24	17	5.6	22.6
16384	38	20.5	9.9	30.4

Concluding Remarks and Future Work

- Newton-Krylov/AMG methods are promising for large-scale simulations (semiconductor drift-diffusion, CFD/MHD)
- Scalable linear solvers critical to scalability and efficiency for large-scale simulation
- Massively parallel simulations on up to 147,000 cores
  - AMG V-cycle: time per iteration scales well
  - Need to improve preconditioner setup and iteration count
- Issues
  - Strong convection effects, hyperbolic systems
  - Highly non-uniform FE aspect ratios
- Need to worry about hybrid (MPI/threading); depend on Trilinos

Thanks For Your Attention!

Paul Lin (ptlin@sandia.gov)

Acknowledgments:

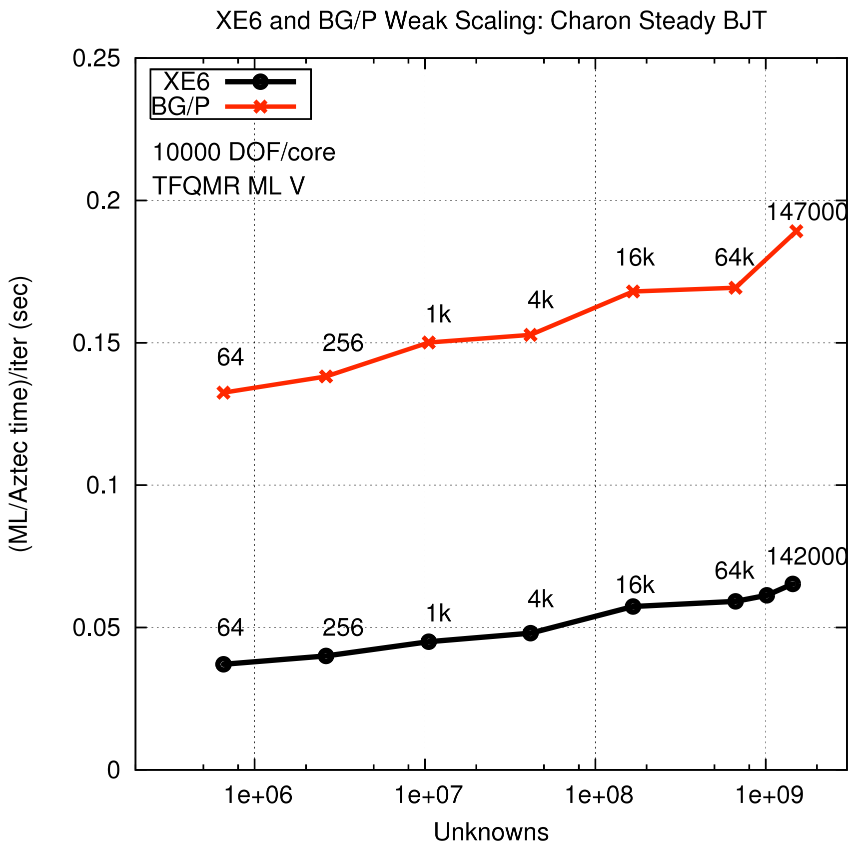
- LLNL BG/P team
  - John Gyllenhaal: fighting with XL compiler bugs
  - Tom Spelce, Scott Futral, Adam Bertsch, Dave Fox
- LLNL computing support: Sheila Faulkner, Tim Fahey

Funding Acknowledgment:
The authors gratefully acknowledge funding from the DOE NNSA Advanced Simulation & Computing (ASC) program and DOE Office of Science ASCR Applied Math Research program

U.S. DEPARTMENT OF ENERGY National Nuclear Security Administration Sandia National Laboratories

Extra Slides

Cray XE6 and IBM BG/P Weak Scaling



- Steady-state drift-diffusion BJT
- TFQMR time per iteration
- Cray XE6 2.4GHz 8-core Magny-Cours
- IBM Blue Gene/P 850 MHz quadcore PowerPC
- 10,000 DOF/core

U.S. DEPARTMENT OF ENERGY National Nuclear Security Administration

Sandia National Laboratories