Scalable parallel implicit solvers for 3D magnetohydrodynamics

L. Chacón

Los Alamos National Laboratory

P.O. Box 1663, Los Alamos, NM 87545

SNL collaborators: J. N. Shadid, R. Pawlowski

Funding: OASCR-AMR program

SCIDAC 2008 Meeting

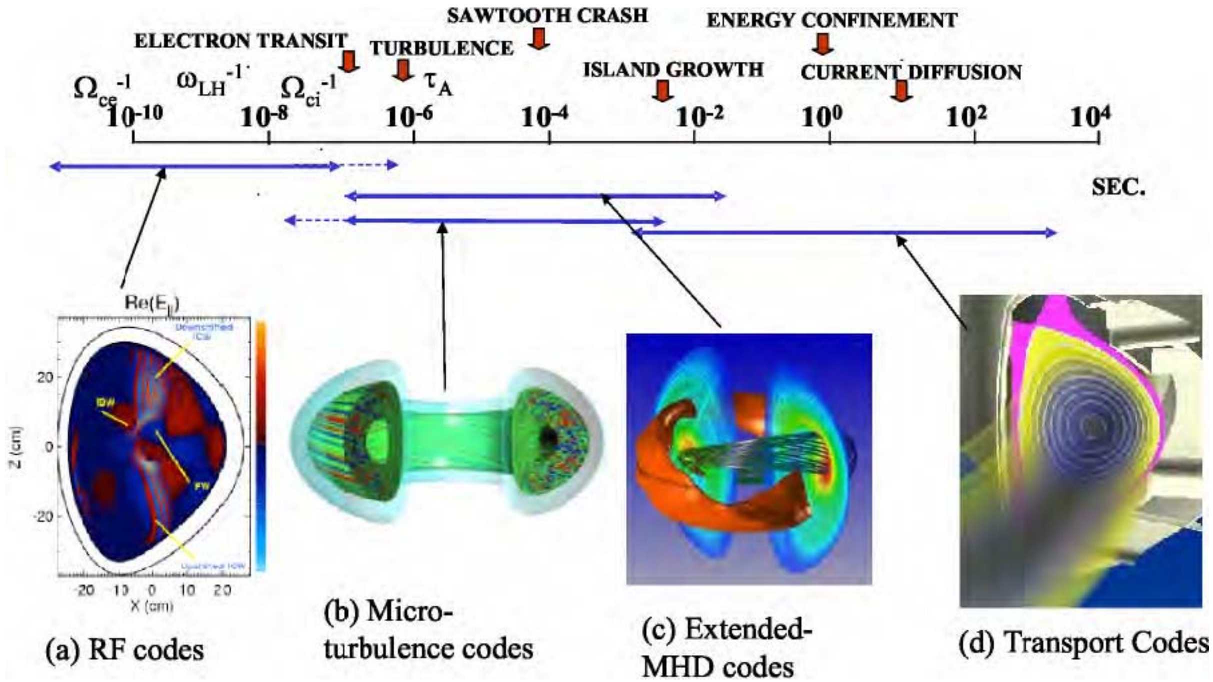
July 13-17, 2008

Seattle, WA

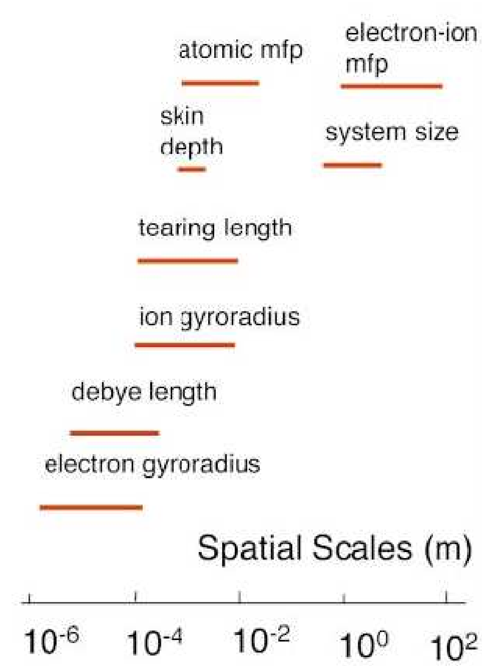
Outline

- Motivation: XMHD and the tyranny of scales
- Parabolization of XMHD: key for SCALABILITY
- 3D compressible resistive MHD
- 3D extended MHD
- Massively parallel performance

"The tyranny of scales"



(a) Time scales in fusion plasmas (FSP report)



(b) Length scales in a typical fusion plasma (Tang, *Phys. Plasmas*, 9 (5), 2002)

"The tyranny of scales will not be simply defeated by building bigger and faster computers" (SBES report, p. 30)

Algorithmic challenges in XMHD

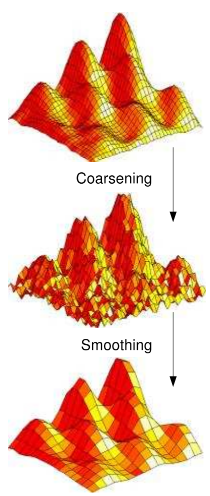
- XMHD has mixed character, with strongly hyperbolic and parabolic components.
- Numerically, XMHD is a nonlinear algebraic system of very stiff equations:
  - Elliptic stiffness (diffusion): $\kappa(J) \sim \frac{\Delta t\, D}{\Delta x^2} \gg 1$
  - Hyperbolic stiffness (linear and dispersive waves): $\kappa(J) \sim \Delta t\, \omega_{fast} \sim \frac{\Delta t}{\Delta t_{CFL}} \gg 1$
- Brute-force algorithms will not be able to cover the span between disparate time/length scales, regardless of computer power (SBES report).
- Key algorithmic requirement: SCALABILITY [$CPU \sim \mathcal{O}(N/n_p)$]!
  - Minimize number of degrees of freedom N: spatial adaptivity.
  - Follow slowest time scales (application dependent): implicit time stepping.
- Scalable implicit methods require MULTILEVEL approaches:

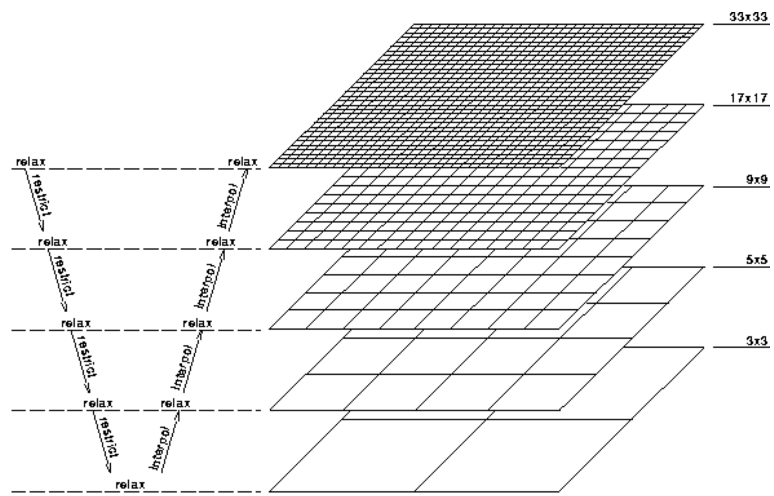
$$CPU \sim \mathcal{O}\left(\frac{N\,\log(N)}{n_p^{\beta}}\right)\,,\ \beta \lesssim 1$$

Why multilevel (multigrid) methods?

- MG employs a divide-and-conquer approach to attack error components in the solution.
  - Oscillatory components of the error are "EASY" to deal with (if a SMOOTHER exists)
  - Smooth components are DIFFICULT.

Idea: coarsen grid to make "smooth" components appear oscillatory, and proceed recursively





- SMOOTHER is make or break of MG!
- In general, smoothers are easy to find for parabolic systems: PARABOLIZATION!

Luis Chacón, chacon@lanl.gov

Parabolization and Schur complement: an example

- PARABOLIZATION EXAMPLE:

$$\partial_t u = \partial_x v \ , \ \partial_t v = \partial_x u.$$

$$u^{n+1} = u^n + \Delta t \partial_x v^{n+1}, v^{n+1} = v^n + \Delta t \partial_x u^{n+1}.$$

$$\boxed{(I - \Delta t^2 \partial_{xx}) u^{n+1} = u^n + \Delta t \partial_x v^n}$$

- PARABOLIZATION via SCHUR COMPLEMENT:

$$\begin{bmatrix} D_1 & U \\ L & D_2 \end{bmatrix} = \begin{bmatrix} I & UD_2^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} D_1 - UD_2^{-1}L & 0 \\ 0 & D_2 \end{bmatrix} \begin{bmatrix} I & 0 \\ D_2^{-1}L & I \end{bmatrix}.$$

Stiff off-diagonal blocks L, U now sit in diagonal via Schur complement $D_1 - UD_2^{-1}L$. The system has been "PARABOLIZED."

$$D_1 - UD_2^{-1}L = (I - \Delta t^2 \partial_{xx})$$

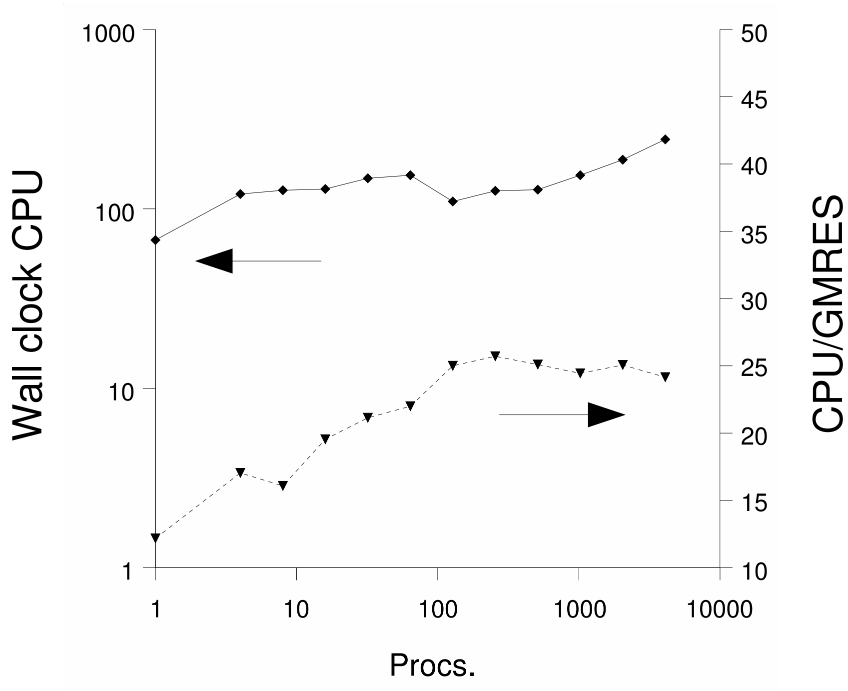
Our approach to a successful fully implicit algorithm for XMHD

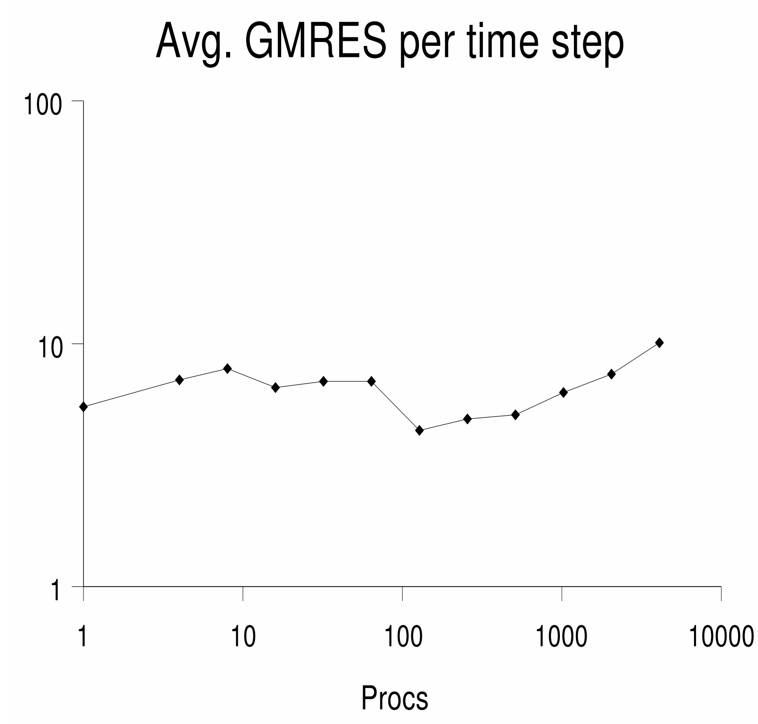
- Even if a smoother exists, MG is remarkably temperamental.
- Combination of Krylov methods and MG preconditioning is optimal:
  - MG preconditioning provides scalability
  - Krylov solver provides robustness

> We seek to develop a successful algorithm for XMHD based on Newton-Krylov-MG

- Rest of the talk focuses on preconditioning development!
- Prove the concept in resistive MHD, and then move to XMHD.

Massively parallel performance of 3D resistive MHD with PETSc toolkit





Jacobian-Free Newton-Krylov Methods

- Objective: solve nonlinear system $\vec{G}(\vec{x}^{n+1}) = \vec{0}$ efficiently (scalably).
- Converge nonlinear couplings using Newton-Raphson method: $\left.\frac{\partial \vec{G}}{\partial \vec{x}}\right|_k \delta \vec{x}_k = -\vec{G}(\vec{x}_k)$.
- Jacobian-free implementation: $\left(\frac{\partial \vec{G}}{\partial \vec{x}}\right)_k \vec{y} = J_k \vec{y} = \lim_{\epsilon \to 0} \frac{\vec{G}(\vec{x}_k + \epsilon \vec{y}) - \vec{G}(\vec{x}_k)}{\epsilon}$
- Krylov method of choice: GMRES (nonsymmetric systems).
- Right preconditioning: solve equivalent Jacobian system for $\delta y = P_k \delta \vec{x}$:

$$J_k P_k^{-1} \underbrace{P_k \delta \vec{x}}_{\delta \vec{y}} = -\vec{G}_k$$

APPROXIMATIONS IN PRECONDITIONER DO NOT AFFECT ACCURACY OF CONVERGED SOLUTION; THEY ONLY AFFECT EFFICIENCY!

- The rest of the talk will discuss the development of suitable preconditioners P_k!

Los Alamos NATIONAL LABORATORY EST. 1943 Office of Science ASCR

Luis Chacón, chacon@lanl.gov

3D *resistive* MHD implicit solver

L. Chacón, Phys. Plasmas 15, 056103 (2008)

Resistive MHD model equations

$$
\begin{aligned}
\frac{\partial \rho}{\partial t} &+ \nabla \cdot (\rho \vec{v}) = 0, \\
\frac{\partial \vec{B}}{\partial t} &+ \nabla \times \vec{E} = 0, \\
\frac{\partial (\rho \vec{v})}{\partial t} + \nabla \cdot \Big[\rho \vec{v}\vec{v} - \vec{B}\vec{B} &- \rho \nu \nabla \vec{v} + \overleftrightarrow{I}(p + \frac{B^2}{2})\Big] = 0, \\
\frac{\partial T}{\partial t} + \vec{v} \cdot \nabla T &+ (\gamma - 1) T \nabla \cdot \vec{v} = 0,
\end{aligned}
$$

- Plasma is assumed polytropic $p \propto n^{\gamma}$.
- Resistive Ohm's law:

$$\vec{E} = -\vec{v} \times \vec{B} + \eta \nabla \times \vec{B}$$

Resistive MHD Jacobian block structure

- The linearized resistive MHD model has the following couplings:

$$
\begin{aligned}
\delta\rho &= L_\rho(\delta\rho, \delta\vec{v}) \\
\delta T &= L_T(\delta T, \delta\vec{v}) \\
\delta\vec{B} &= L_B(\delta\vec{B}, \delta\vec{v}) \\
\delta\vec{v} &= L_v(\delta\vec{v}, \delta\vec{B}, \delta\rho, \delta T)
\end{aligned}
$$

- Therefore, the Jacobian of the resistive MHD model has the following coupling structure:

$$
J\delta\vec{x} = \begin{bmatrix} D_\rho & 0 & 0 & U_{v\rho} \\ 0 & D_T & 0 & U_{vT} \\ 0 & 0 & D_B & U_{vB} \\ L_{\rho v} & L_{Tv} & L_{Bv} & D_v \end{bmatrix} \begin{pmatrix} \delta\rho \\ \delta T \\ \delta\vec{B} \\ \delta\vec{v} \end{pmatrix}
$$

- Diagonal blocks contain advection-diffusion contributions, and are "easy" to invert using MG techniques. Off diagonal blocks L and U contain all hyperbolic couplings.

PARABOLIZATION: Schur complement formulation

- We consider the block structure:

$$J\delta\vec{x} = \begin{bmatrix} M & U \\ L & D_v \end{bmatrix} \begin{pmatrix} \delta\vec{y} \\ \delta\vec{v} \end{pmatrix} \; ; \; \delta\vec{y} = \begin{pmatrix} \delta\rho \\ \delta T \\ \delta\vec{B} \end{pmatrix} \; ; \; M = \begin{pmatrix} D_\rho & 0 & 0 \\ 0 & D_T & 0 \\ 0 & 0 & D_B \end{pmatrix}$$

- M is "easy" to invert (advection-diffusion, MG-friendly).

Schur complement analysis of 2x2 block J yields:

$$\begin{bmatrix} M & U \\ L & D_v \end{bmatrix}^{-1} = \begin{bmatrix} I & 0 \\ -LM^{-1} & I \end{bmatrix} \begin{bmatrix} M^{-1} & 0 \\ 0 & P_{Schur}^{-1} \end{bmatrix} \begin{bmatrix} I & -M^{-1}U \\ 0 & I \end{bmatrix},$$

$$P_{Schur} = D_v - LM^{-1}U.$$

- EXACT Jacobian inverse only requires M^{-1} and P_{Schur}^{-1}.
- Schur complement formulation is fundamentally unchanged in Hall MHD!

Physics-based preconditioner (I): small-flow approximation

- The Schur complement analysis translates into the following 3-step EXACT inversion algorithm:

$$
\begin{aligned}
\text{Predictor} &: \quad \delta\vec{y}^* = -M^{-1}G_y \\
\text{Velocity update} &: \quad \delta\vec{v} = P_{Schur}^{-1}[-G_v - L\delta\vec{y}^*], \; P_{Schur} = D_v - LM^{-1}U \\
\text{Corrector} &: \quad \delta\vec{y} = \delta\vec{y}^* - M^{-1}U\delta\vec{v}
\end{aligned}
$$

- MG treatment of P_{Schur} is impractical due to M^{-1}.

Need suitable simplifications (SEMI-IMPLICIT)!

- We consider the small-flow-limit case: $M^{-1} \approx \Delta t$
- This approximation is equivalent to splitting flow in original equations.

Physics-based preconditioner (II)

- Small flow approximation: $M^{-1} \approx \Delta t$ in steps 2 & 3 of Schur algorithm:

$$\begin{aligned} \delta\vec{y}^* &= -M^{-1} G_y \\ \delta\vec{v} &\approx P_{SI}^{-1}\left[-G_v - L\delta\vec{y}^*\right] \; ; \; P_{SI} = D_v - \Delta t L U \\ \delta\vec{y} &\approx \delta\vec{y}^* - \Delta t U \delta\vec{v} \end{aligned}$$

where:

$$\boxed{P_{SI} = \rho^n \left[\overleftrightarrow{I}/\Delta t + \theta(\vec{v}_0 \cdot \nabla \overleftrightarrow{I} + \overleftrightarrow{I} \cdot \nabla \vec{v}_0 - \nu^n \nabla^2 \overleftrightarrow{I})\right] + \Delta t \theta^2 W(\vec{B}_0, p_0)}$$

$$W(\vec{B}_0, p_0) = \vec{B}_0 \times \nabla \times \nabla \times [\overleftrightarrow{I} \times \vec{B}_0] - \vec{j}_0 \times \nabla \times [\overleftrightarrow{I} \times \vec{B}_0] - \nabla[\overleftrightarrow{I} \cdot \nabla p_0 + \gamma p_0 \nabla \cdot \overleftrightarrow{I}]$$

- P_{SI} is block diagonally dominant by construction!
- We employ multigrid methods (MG) to approximately invert P_{SI} and M: 1 V(4,4) cycle

Luis Chacón, chacon@lanl.gov

Serial performance (2D tearing mode)

Δt convergence study (128x128)

Δt	GMRES/Δt	CPU_{exp}/CPU	$\Delta t/\Delta t_{CFL}$
0.5	8.0	8.0	380
0.75	9.5	10.0	570
1.0	11.2	12.7	760
1.5	14.6	14.6	1140

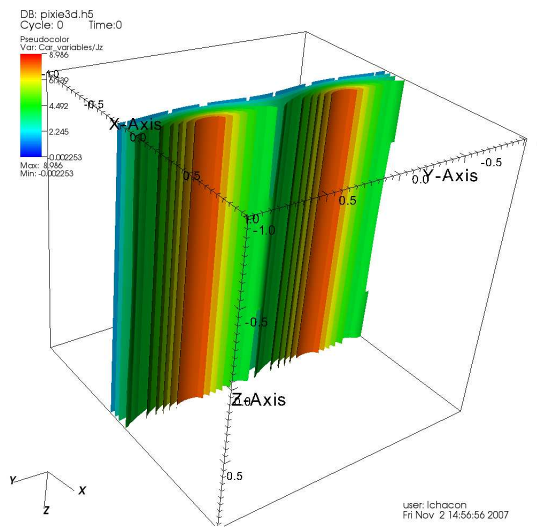
Grid convergence study ($\Delta t = 1200\Delta t_{CFL}$)

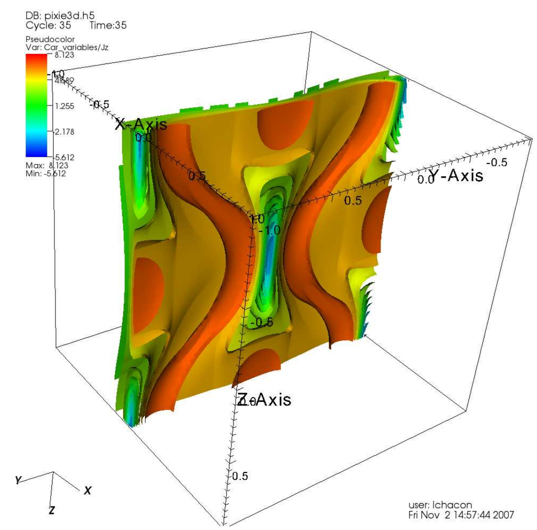
N	GMRES/Δt	CPU_{exp}/CPU	$\Delta t/\Delta t_{CFL}$
32x32	14	2.43	159
64x64	11.8	5.8	322
128x128	11.2	13.3	667
256x256	11.4	28.5	1429

$CPU \sim \mathcal{O}(N)$ OPTIMAL SCALING!

Luis Chacón, chacon@lanl.gov

Serial performance (3D island coalescence)



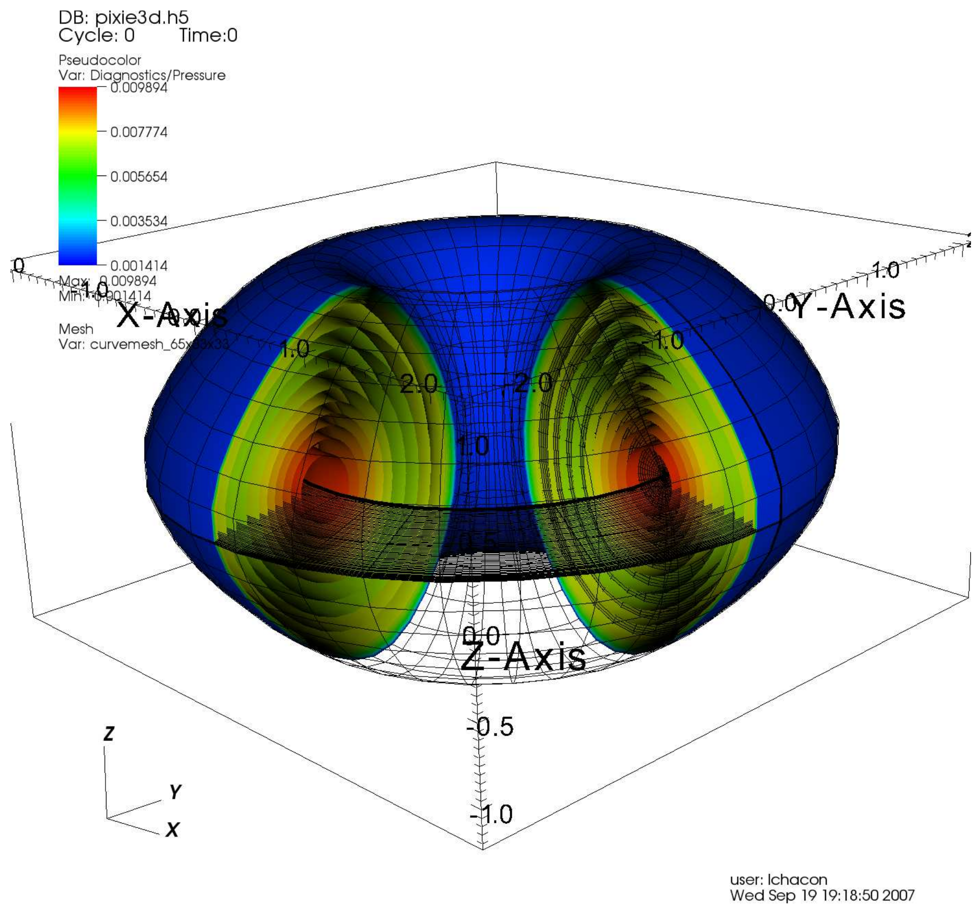


10 time steps, $\Delta t = 0.1$, V(3,3) cycles, mg_tol=1e-2

Grid	GMRES/Δt	CPU
16^3	5.5	81
32^3	7.9	1176
64^3	7.0	11135

Application: m=n=1 kink mode in 3D CDXU Tokamak

64x32x32 mesh with 16 processors



Luis Chacón, chacon@lanl.gov

3D *extended* MHD implicit solver

Extended MHD model equations

$$\begin{aligned}
\frac{\partial \rho}{\partial t} &+ \nabla \cdot (\rho \vec{v}) = 0, \\
\frac{\partial \vec{B}}{\partial t} &+ \nabla \times \vec{E} = 0, \\
\frac{\partial (\rho \vec{v})}{\partial t} + \nabla \cdot \left[\rho \vec{v}\vec{v} - \vec{B}\vec{B} \right. &- \left. \rho \nu \nabla \vec{v} + \overleftrightarrow{I} (p + \frac{B^2}{2}) \right] = 0, \\
\frac{\partial T_e}{\partial t} + \vec{v} \cdot \nabla T_e &+ (\gamma - 1) T_e \nabla \cdot \vec{v} = 0,
\end{aligned}$$

- Plasma is assumed polytropic $p \propto n^{\gamma}$.
- We assume cold ion limit: $T_i \ll T_e \Rightarrow \boxed{p \approx p_e}$.
- Generalized Ohm's law:

$$\boxed{\vec{E} = -\vec{v} \times \vec{B} + \eta \nabla \times \vec{B} - \frac{d_i}{\rho}(\vec{j} \times \vec{B} - \nabla p_e)}$$

Extended MHD Jacobian block structure

- The linearized extended MHD model has the following couplings:

$$\begin{aligned}
\delta\rho &= L_\rho(\delta\rho, \delta\vec{v}) \\
\delta T &= L_T(\delta T, \delta\vec{v}) \\
\delta\vec{B} &= L_B(\delta\vec{B}, \delta\vec{v}, \delta\rho, \delta T) \\
\delta\vec{v} &= L_v(\delta\vec{v}, \delta\vec{B}, \delta\rho, \delta T)
\end{aligned}$$

- Jacobian coupling structure:

$$J\delta\vec{x} = \begin{bmatrix} D_\rho & 0 & 0 & U_{v\rho} \\ 0 & D_T & 0 & U_{vT} \\ L_{\rho B} & L_{TB} & D_B & U_{vB} \\ L_{\rho v} & L_{Tv} & L_{Bv} & D_v \end{bmatrix} \begin{pmatrix} \delta\rho \\ \delta T \\ \delta\vec{B} \\ \delta\vec{v} \end{pmatrix}$$

- We have added off-diagonal couplings.

Extended MHD Jacobian block structure (cont.)

- The coupling structure can be substantially simplified if we note $(p \approx p_e)$:

$$\frac{1}{\rho}(\vec{j} \times \vec{B} - \nabla p_e) \approx \frac{D\vec{v}}{Dt}$$

and therefore:

$$\vec{E} \approx -\vec{v} \times \vec{B} + \eta \nabla \times \vec{B} - d_i \frac{D\vec{v}}{Dt}$$

- This transforms jacobian coupling structure to:

$$J\delta\vec{x} \approx \begin{bmatrix} D_\rho & 0 & 0 & U_{v\rho} \\ 0 & D_T & 0 & U_{vT} \\ 0 & 0 & D_B & U_{vB}^R + U_{vB}^H \\ L_{\rho v} & L_{Tv} & L_{Bv} & D_v \end{bmatrix} \begin{pmatrix} \delta\rho \\ \delta T \\ \delta\vec{B} \\ \delta\vec{v} \end{pmatrix}$$

We can therefore reuse ALL resistive MHD PC framework!

Extended MHD preconditioner

- Use same Schur complement approach.
- M block contains ion scales only! Approximation $M^{-1} \approx \Delta t$ is very good in extended MHD (ion scales do NOT contribute to numerical stiffness).
- Additional block U_{vB}^{H} results, after the Schur complement treatment, in new term in Schur complement:

$$P_{SI} = \rho^n \left[\overleftrightarrow{I} / \Delta t + \theta (\vec{v}_0 \cdot \nabla \overleftrightarrow{I} + \overleftrightarrow{I} \cdot \nabla \vec{v}_0 - \nu^n \nabla^2 \overleftrightarrow{I}) \right] + \Delta t \theta^2 W(\vec{B}_0, p_0)$$

$$W(\vec{B}_0, p_0) = \vec{B}_0 \times \nabla \times \nabla \times [\overleftrightarrow{I} \times \vec{B}_0 - \frac{d_i}{\theta \Delta t} \overleftrightarrow{I}] - \vec{j}_0 \times \nabla \times [\overleftrightarrow{I} \times \vec{B}_0] - \nabla [\overleftrightarrow{I} \cdot \nabla p_0 + \gamma p_0 \nabla \cdot \overleftrightarrow{I}]$$

- This system supports dispersive waves $\omega \sim k^2$!
- We have shown analytically that additional term is amenable to damped JB smoothing!

We can use classical MG!

Luis Chacón, chacon@lanl.gov

Preliminary serial performance results (2D tearing mode)

$d_i = 0.05$

1 time step, $\Delta t = 1.0$, V(3,3) cycles, mg_tol=1e-2

Grid	GMRES/Δt	CPU_{exp}/CPU	$\Delta t/\Delta t_{exp}$
32x32	32	0.9	110
64x64	9	9.3	384
128x128	8	37.9	1436
256x256	10	117.0	5660

Luis Chacón, chacon@lanl.gov

Massively parallel performance

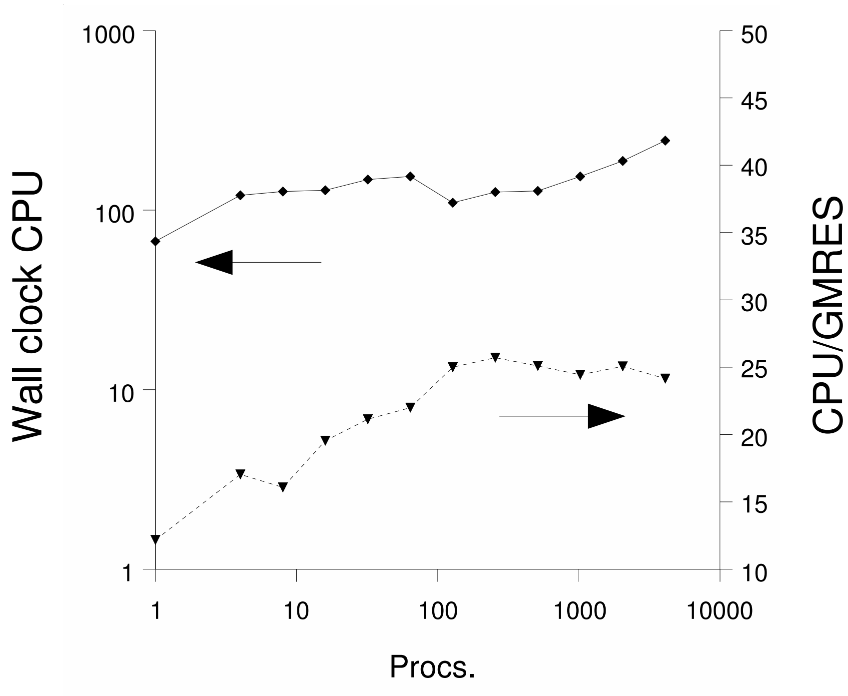
L. Chacón, Phys. Plasmas 15, 056103 (2008)

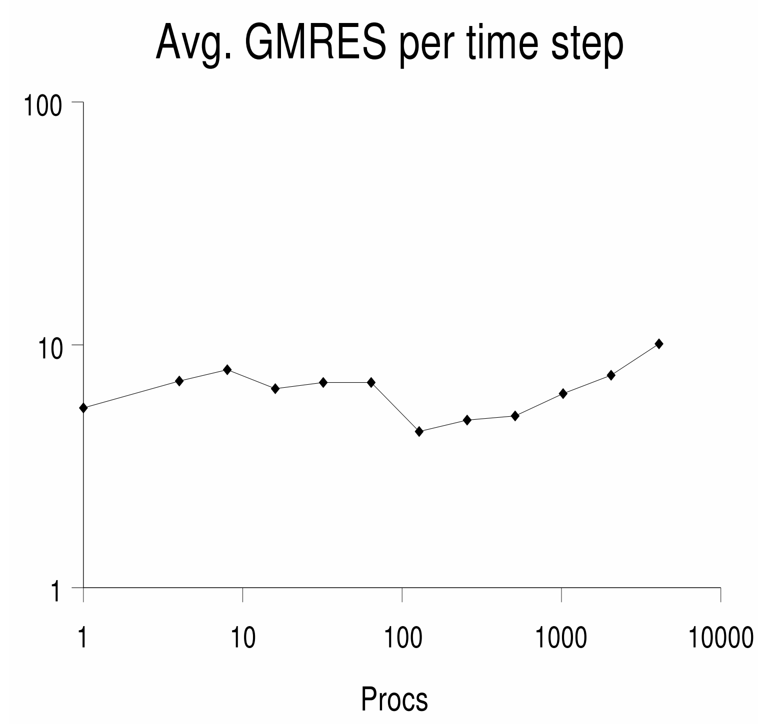
Approach to parallelization

- We employ the PETSc toolkit (Parallel Extensible Toolkit for Scientific computing)
- In our preconditioner, we have parallelized our own multigrid solver (using PETSc's distributed arrays constructs)
  - We implement a matrix-light multigrid, where only diagonals are stored; residuals are calculated matrix-free.
  - Operator coarsening is done via rediscretization (instead of Galerkin procedure); this avoids forming a matrix at any grid level: avoids communication of matrix elements.
  - Prolongation and restriction are performed with second-order splines locally at each processor: no communication required.
- As a result, our MG implementation features excellent parallel scalability.
- We do not feature a coarse-solve beyond the processor skeleton grid.
  - This eventually degrades algorithmic scalability (only shows at > 1000-processor level).

Luis Chacón, chacon@lanl.gov

Massively parallel performance with PETSc toolkit (3D island coalescence, 16^3 grid points per processor, on Cray XT (Franklin) at NERSC)





Luis Chacón, chacon@lanl.gov

Conclusions

- Developed a scalable, multilevel-based, fully implicit NK-MG solver for XMHD.

  Key algorithmic breakthrough: PARABOLIZATION + MG.

- Equivalence between parabolization and the Schur decomposition:
  - Provides a rigorous foundation for the parabolization step.
  - Provides a path to generalize approach when more complete XMHD models are considered.
- Demonstrated excellent algorithmic performance under grid refinement and with time step.
- Demonstrated excellent parallel performance in massively parallel environments.
- Future work:
  - Bring Hall MHD to production stage (high-order dissipation required).
  - Add MG coarse-grid solve to avoid algorithmic performance loss with thousands of processors.
  - Extend MG functionality to singular-point coordinate systems.

Luis Chacón, chacon@lanl.gov