Fourier Off-Diagonal Preconditioning and Scaling of Two-Fluid Computations

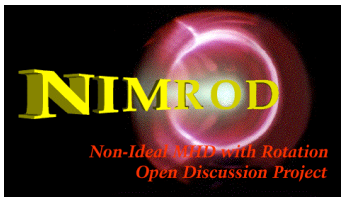
C. R. Sovinec[1] and E. D. Held[2]

[1]*University of Wisconsin-Madison*

[2]*Utah State University*

NIMROD Team Meeting

August 27-28, 2008, San Diego, CA



Outline

- Introduction--NIMROD's physics-based preconditioning and scaling strategy
- Preconditioning
  - Recent work
  - Limited Fourier coupling
  - Results on a small test case
  - Possible improvements
- New fluid-algorithm scaling (with new preconditioner) from Franklin
- Conclusions

Introduction: NIMROD's preconditioning strategy has always been 'physics-based.'

In magnetically confined plasmas with toroidally symmetric boundaries and equilibria, toroidally asymmetric perturbations are small.

- NIMROD's preconditioning is based on this physics.
- No physics in any field-advance of the implicit leapfrog is left out of the preconditioning.
  - Diagonal-in-*n* systems are solved with the parallel SuperLU_DIST library [Li, ACM Trans. Math. Software **29**, 110 (2003)].
- Toroidally gridded systems do not have inherently dominant diagonal blocks for magnetic confinement applications.
  - Chacon's 'parabolization' (PoP **15**, 56103) is a factorization technique for the time-centered system to avoid diagonal sub-dominance. Some flow effects are dropped.
- Here, we explore including more than the toroidally symmetric fields to improve GMRES convergence when advancing **B** in 3D 2-fluid computations.
- We also test our parallel scaling strategy on Franklin.

Hall-MHD in 3D has been problematic because fluctuations do not contribute to the diagonal of the **B**-advance operator, and the whistler is the fastest mode of the system.

- With **A** being a test function and dropping surface terms,

$$-\frac{\Delta t}{2}\int \mathbf{A}\cdot\nabla\times\left(\frac{1}{\mu_0\bar{n}e}\tilde{\mathbf{B}}^{j+1/2}\times\nabla\times\Delta\mathbf{B}\right)dVol =$$

$$-\frac{\Delta t}{2}\int \frac{1}{\mu_0\bar{n}e}(\nabla\times\mathbf{A})\cdot\tilde{\mathbf{B}}^{j+1/2}\times(\nabla\times\Delta\mathbf{B})\,dVol$$

- When the test and trial functions are expanded, the resulting matrix has mixed partials on the diagonal due to the cross product.
- With a Fourier expansion, the first-order toroidal derivatives lead to imaginary terms on the diagonal.
  - The operator is non-Hermitian.
  - It detracts from diagonal dominance when $\tilde{\mathbf{B}}^{j+1/2}$ and Δt are sufficiently large.

Recent work: we have investigated different approaches to bringing Fourier coupling into the preconditioning.

- Polynomial approximation was tested, and results were reported at Sherwood 2008.
- Second-order toroidal differences with toroidally local coefficients and plane-based blocks was coded and tested. [See ICC '08 poster from www.cptc.wisc.edu/sovinec_research for details.]
  - This was expected to complement the standard Fourier-based preconditioning through a multiplicative step.
  - Diagonal challenges of gridded systems also destroy the benefit of this preconditioner at all but extremely small Δt values.
- The latest approach generates specific off-diagonal couplings with respect to Fourier component.
  - Fourier-based block Jacobi or Gauss-Seidel does not add FFTs, and small coupling coefficients make it more effective than using plane-based blocks.
- Electron inertia is also important.

Aside: implicit electron inertia, even at physical m_i/m_e ratios, helps matrix condition numbers.

• The HPD part of our system is increased by adding the $\frac{1}{\varepsilon_0 \omega_e^2}\frac{\partial}{\partial t}\mathbf{J}$ part of electron inertia.

$$\Delta\mathbf{B} - \frac{\Delta t}{2}\nabla\times\left(\mathbf{V}^{j+1}\times\Delta\mathbf{B}\right) + \frac{\Delta t}{2}\nabla\times\frac{1}{\bar{n}e}\left(\mathbf{J}^{j+1/2}\times\Delta\mathbf{B} + \Delta\mathbf{J}\times\mathbf{B}^{j+1/2}\right)$$

$$+\Delta t\,\nabla\times\left(\frac{\eta}{2} + \frac{d_e^2}{\Delta t}\right)\nabla\times\Delta\mathbf{B} - \Delta t\,\kappa_{divb}\nabla\nabla\cdot\Delta\mathbf{B}$$

for the lhs of the **B**-advance, also showing the divergence cleaning term. The electron skin depth is $d_e = c/\omega_e$.

• Physically, electron inertia leads to the electron cyclotron resonance, which keeps the *R*-mode phase speed from growing indefinitely as k increases.

• This helps limit stiffness, hence condition numbers, in two-fluid computations.

The limited off-diagonal Fourier coupling has several advantages.

• For block Gauss-Seidel preconditioning, the Fourier representation helps keep the spectral radius of $(D+L)^{-1}U$ less than 1 at large time-step. [Here, the notation refers to block-based splitting $A=L+D+U$.]

• With limited couplings, matrix elements can be generated, and matrix-vector products are fast relative to full matrix-free product operations.

  • The required coding for these matrices is a generalization of existing code for the diagonal-in-Fourier systems.

  • When used in iteration, these matrices are not factored.

• Generating matrix elements is computationally intensive but scales well in parallel. [Generation of full convolution matrices is not practical, however, even in parallel.]

• The extra communication during Jacobi/Gauss-Seidel iteration (preconditioner looping) is point-to-point.

Implementation has been a nontrivial but straightforward extension of existing NIMROD coding.

- The preconditioner loop over Fourier index in iter_3d_cg.f now includes block off-diagonal matrix-vector multiplications.
  - It can take multiple passes.
  - There is new point-to-point communication (next slide).
- Extra matrices (but not factors) are generated from the management routine for the advance (just adv_b_3dnsym for now).
- Regularity conditions have been generalized to handle matrices where the row Fourier index is not the same as the column Fourier index.
- The new b_hmhd_cpop routine is based on b_hmhd_op, but wavenumbers for the column and row differ, and coefficients use complex $n>1$ components. [It is called separately for each of a user-selected range of couplings.]

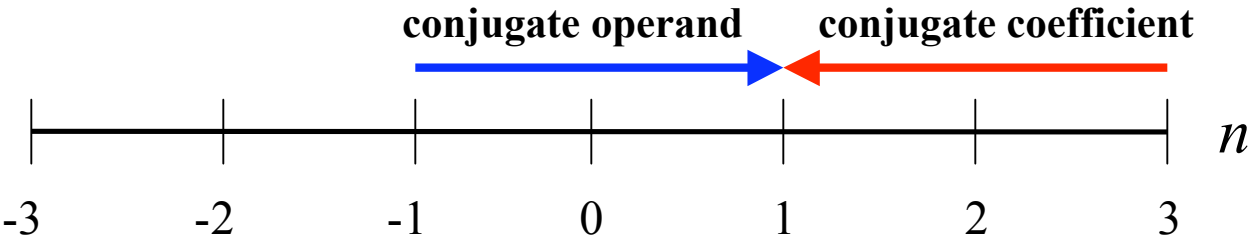
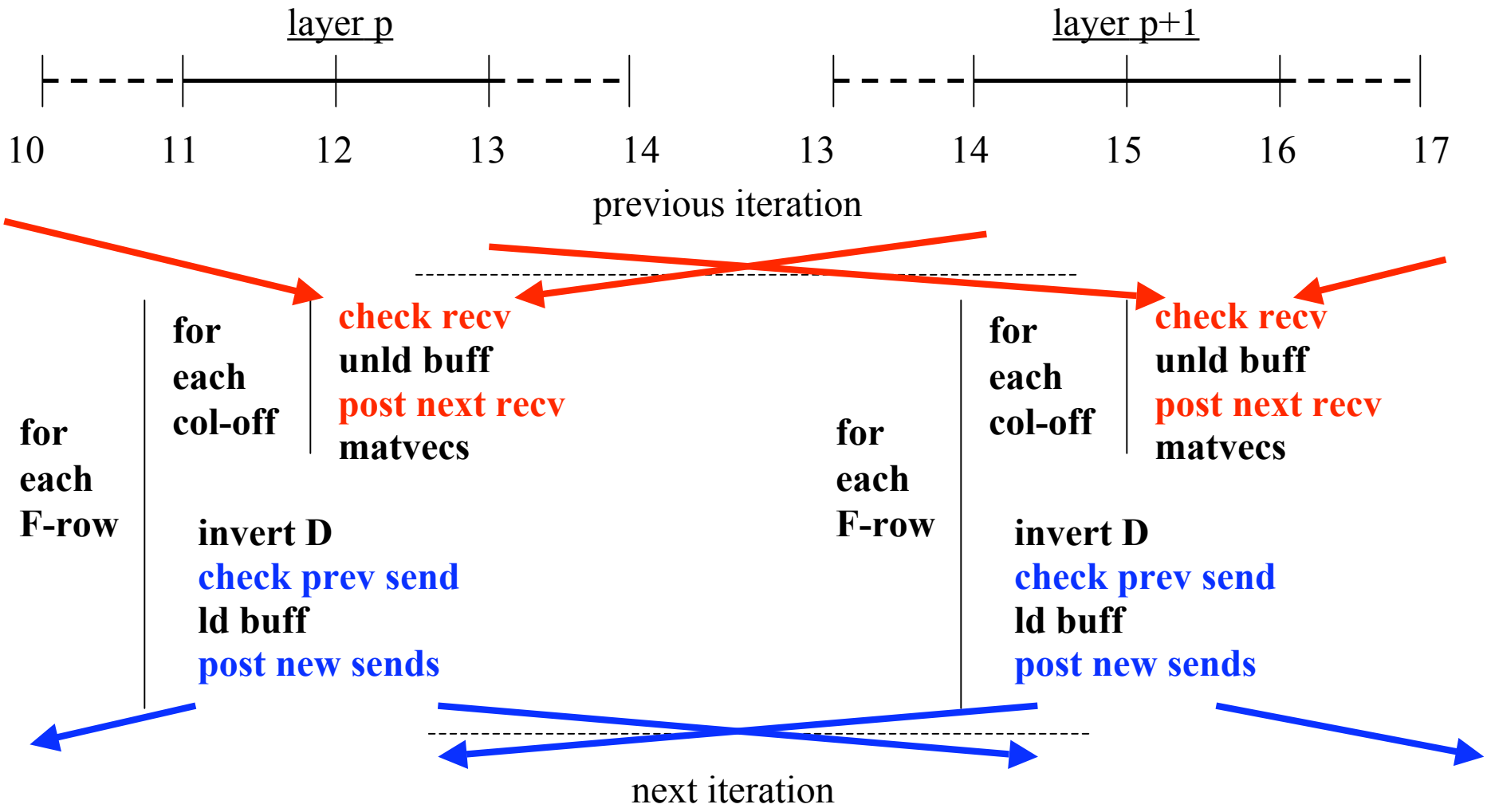


Illustration of ±2 column offsets to n=1 row. Here, the n=-1 operand ($\Delta\mathbf{B}$ iterate) is evaluated from the complex conjugate of n=+1.

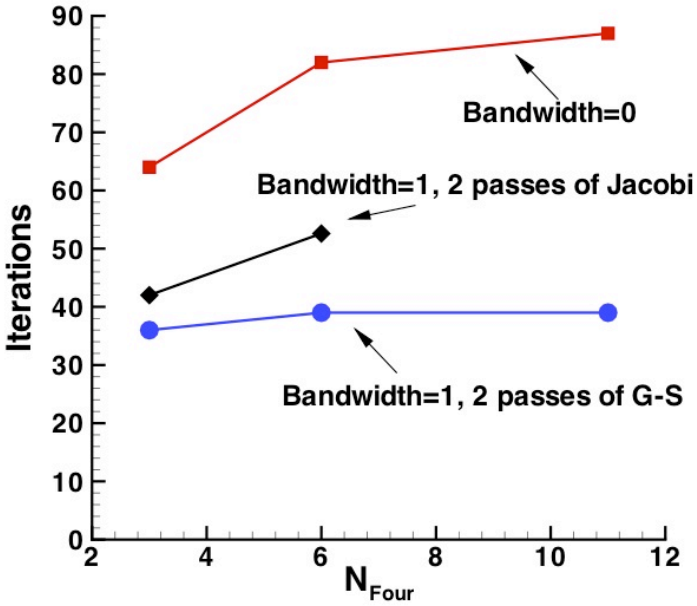
New layer-to-layer communication has been coded to overlap with computation during Jacobi/GS iteration.



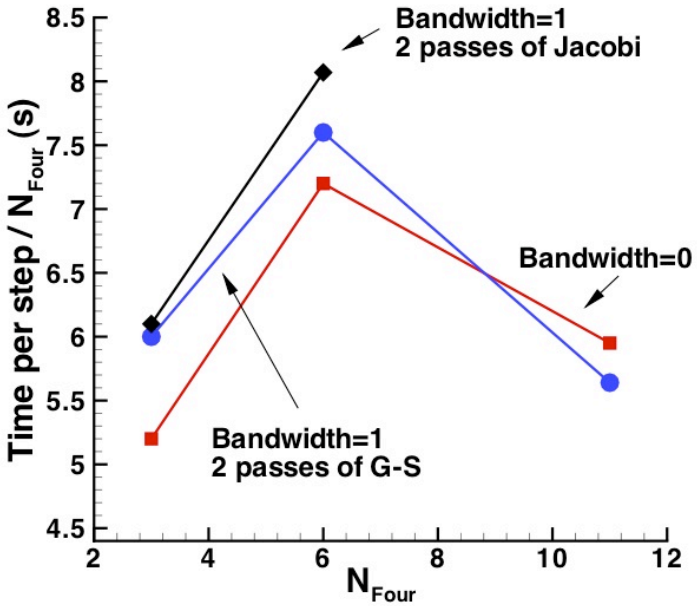
- Fourier-row looping goes up and down in index (~SSOR).
- Selecting Gauss-Seidel iteration-type uses the on-layer part of L only.

The limited Fourier coupling shows promise for large timestep two-fluid computations.

- The test case is the two-fluid 1/1 cylindrical kink. [The physics is helically symmetric, but NIMROD treats it as a 3D computation.]
- Realistic electron inertia helps convergence at very large timestep.
- Here, the representation is 9×9 biquintic, and $\Delta t/\tau_{Hp} = 2$.



G-S passes makes the iteration nearly independent of Fourier comps in this case.



Serial CPU time is shown per Fourier comp. Data at N_{Four}=11 is from Bassi unlike others.

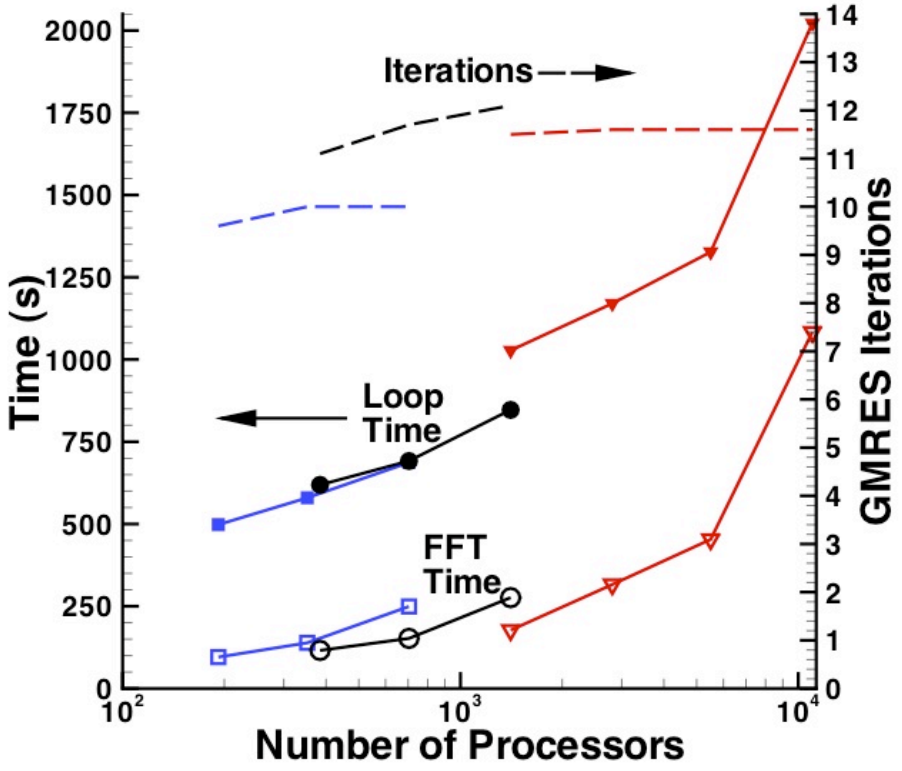
- The near-constant iteration count with increasing n_{max} is important.
- This case is dominated by n=1; others may be more challenging.

Further refinements of the scheme are possible if necessary.

- Present scheme is helping limit **B** iteration in two 3D two-fluid **production** studies. [Sovinec, King, and Murphy--applications]
  - Off-diagonal coupling of ±1 is usually satisfactory.
  - Relaxation parameter choice of 0.5 - 1 depends on size of perturbations and timestep.
- Refinement 1: Fourier couplings can be chosen according to their physical importance and do not need to be adjacent in Fourier index.
- Refinement 2 (more speculative): An incomplete factorization with limited Fourier coupling may be possible within each layer.
  - This would approximate a direct solve over all Fourier components in a layer.
  - The 'incomplete' part means dropping any loss of sparseness across the poloidal plane when forming the Schur complements.
  - Off-layer coupling can still be left to iteration so that all layers can be done simultaneously.

New scaling study from Franklin: shows effectiveness of Fourier toroidal representation with new preconditioner.

- Application is the 1/1 (3D for nimrod) two-fluid kink in nonlinear stage.
- Parameters provide a weak scaling of a production computation.
- Largest has 64×32 poly_degree=8, N_{Four}=171, 11,008 processors, 86 layers.
- Δt is 0.1 τ_A.
- As with the small test, iteration count does not increase with N_{Four}.
  - Bandwidth=1, 2 passes.
- Overall efficiency in going from 192 to 5504 processors is 38%.
- Loss of scaling is from FFT at large N_{Four}.
  - Present looping does separate FFTs for each quadrature-point in each element, starving FFTs.
  - **It can be addressed by sending all quad-pts in fewer FFTs.**



Blue: 32 blocks; Black: 64 blocks; Red: 128 blocks. Within color shows increasing N_{Four}. Times are for 50 steps, so ~ 40s/step at most.

Conclusions

- Improvements to computational linear algebra are essential for studying two-fluid macroscopic dynamics numerically.
- Implicit electron inertia is HPD and limits stiffness from the *R*-mode.
- A small test case and a weak scaling study show that the limited off-diagonal coupling can keep the **B**-advance iteration count fixed as the number of Fourier components is increased.
- Communication for the new preconditioner is hidden by computation.
- The two decomposition types (grid block and layer) work reasonably well together.
    - Practical scaling to 10k processors (totaling 10s of TFlop/s peak) is within reach with some FE-integration loop modifications for the FFTs.

Solver basics: the generalized minimal residual (GMRES) method finds an orthonormal bases for a Krylov space, starting from the initial residual, then finds the linear combination of bases that minimizes the residual.

- The required Krylov space K_m is $\left(r_0, Ar_0, AAr_0, \ldots, A^{m-1}r_0\right)$

  solving $Ax = b, \quad r_0 \equiv b - Ax_0 \Rightarrow x = x_0 + \delta x, \quad A\delta x = r_0$

- The Arnoldi algorithm generates the next basis vector (MGrSchm) as soon as the next Krylov vector is computed. [from Saad's book]

- For A of rank n, the first m+1 basis vectors define an $n \times m$+1 matrix V_{m+1} and the m+1$\times m$ Hessenberg matrix that satisfy

  $AV_m = V_{m+1}H_m$ typically $m << n$

- The best possible solution, $\delta x = V_m y_m$, is found from a relatively small least-squares problem.

$$\min\|r_0 - A\delta x\|_2 \Rightarrow \min\|r_0 - AV_m y\|_2 = \min\|\beta v_1 - V_{m+1}H_m y\|_2 = \min\|V_{m+1}(\beta e_1 - H_m y)\|_2$$

where $\beta \equiv \|r_0\|$ and Givens rotations provide the remaining error after each step.

- With right preconditioning, we solve $AM^{-1}u = r_0, \ \delta x = M^{-1}u, \ M \cong A$

For reference, we show the preconditioned Arnoldi algorithms with modified Gram-Schmidt orthogonalization. [Yousef Saad, Iterative Methods for Sparse Linear Systems, 2nd ed. SIAM.]

Choose v_1, $\|v_1\|_2 = 1$
for $j = 1, \ldots m$ do
  $w_j = M^{-1} A v_j$
  for $i = 1, \ldots j$ do (MGS)
    $h_{ij} = \left(w_j, v_i\right)$
    $w_j = w_j - h_{ij} v_i$
  end
  $h_{j+1,j} = \|w_j\|_2$
  $v_{j+1} = w_j / h_{j+1,j}$
end

to construct the basis set that spans

$$\left\{v_1, M^{-1}Av_1, \left(M^{-1}A\right)^2 v_1, \ldots\right\}$$

left preconditioning
[with $v_1 \sim r_0 = M^{-1}(b - Ax_0)$]

Choose v_1, $\|v_1\|_2 = 1$
for $j = 1, \ldots m$ do
  $w_j = AM^{-1} v_j$
  for $i = 1, \ldots j$ do (MGS)
    $h_{ij} = \left(w_j, v_i\right)$
    $w_j = w_j - h_{ij} v_i$
  end
  $h_{j+1,j} = \|w_j\|_2$
  $v_{j+1} = w_j / h_{j+1,j}$
end

to construct the basis set that spans

$$\left\{v_1, AM^{-1}v_1, \left(AM^{-1}\right)^2 v_1, \ldots\right\}$$

right preconditioning
[with $v_1 \sim r_0 = b - Ax_0$]