Parallel Scaling of the NIMROD Code to 10K Processor Cores[1]

Carl Sovinec, Univ. of Wisconsin-Madison, and the NIMROD Team

September 8, 2009

The NIMROD code (http://nimrodteam.org) is a flexible simulation tool for modeling nonlinear MHD-like behavior in magnetized plasma. It is presently being used in a number of studies of laboratory and natural plasmas including edge-localized modes (ELMs), resonant magnetic perturbation, internal kink/giant sawteeth, disruption mitigation, runaway electron generation, magnetic tearing, and island evolution in tokamaks; MHD/Hall dynamo activity and pressure-driven modes in reversed-field pinches; MHD effects on confinement and merging of spheromaks; non-inductive startup in spherical tori; translation and rotating-magnetic-field generation of field-reversed configurations; the onset of magnetotail substorms due to ballooning and reconnection; and the stability of astrophysical jets.[2] Numerical resolution for the large ranges of spatial and temporal scales that are inherent for many of these applications requires efficient parallel computation. This report describes a recent scaling study performed on the Cray XT-4 ("Franklin") at the National Energy Research Scientific Computing Center (NERSC, http://www.nersc.gov) and highlights improvements that allow NIMROD to run efficiently on more than 10,000 processor cores.

The NIMROD algorithm uses implicit and semi-implicit time-advance methods to handle the temporal stiffness of nonlinear evolution over times that are many orders of magnitude longer than it takes for waves to propagate globally. Two-dimensional spectral finite elements combined with finite Fourier series provide accurate spatial representation for extreme anisotropy, which is characteristic of magnetized plasma, when nonlinear evolution changes the magnetic topology.[3] Each time-step of the implicit advance requires the solution of large ill-conditioned matrices. The Fourier representation of the periodic coordinate leads to dense submatrices from the convolutions. However, NIMROD takes advantage of the fact that the symmetric parts of the magnetic field, number density, and pressure are much larger than their asymmetric components. It uses matrix-free conjugate gradient (CG) and generalized minimal residual (GMRES) iteration with fast Fourier transforms (FFTs) and block-based preconditioning, where each block represents a distinct Fourier component. With the Fourier representation, diagonal blocks hold the largest matrix elements and are inverted with the parallel sparse direct solver, SuperLU_DIST.[4] The preconditioning has recently been improved by generating matrix elements for a limited number of off-diagonal blocks for use in block-Gauss-Seidel-like steps with asynchronous communication overlapped with on-processor computation. Further improvement to parallel scaling has been obtained by reordering data and loops to allow fewer, larger collective communications, which occur before and after computations at collocation points.

Our new scaling uses a nonlinear computation of the tokamak internal kink mode with fast reconnection effects, increasing resolution with the number of processors for a weak scaling. The physical parameters include the Lundquist number of 10^6, two-fluid Ohm's law with sound gyroradius of $\rho_s=0.015a$, where a is the minor radius, and a realistic electron-ion mass ratio. The computation has been performed in cylindrical geometry for convenience of the scaling study, but similar NIMROD computations have been performed in toroidal geometry. The poloidal representation uses two-dimensional basis functions of polynomial degree eight, with nonuniformly packed meshes ranging from 512 to 2048 elements. The number of Fourier components ranges from 11 to 342. Each processor core is allocated 16 high-order elements and at most three Fourier components; de-aliasing influences the Fourier component range, and in some cases, a minority of processor cores have two Fourier components. The

[1] Work supported by U.S. Dept. of Energy through the Center for Extended Magnetohydrodynamics (http://w3.pppl.gov/cemm) and through the Plasma Science and Innovation Center (http://www.psicenter.org).

[2] See http://nimrodteam.org/NIMROD_pubs.html for a publication list.

[3] C. R. Sovinec, A. H. Glasser, T. A. Gianakon, D. C. Barnes, R. A. Nebel, S. E. Kruger, D. D. Schnack, S. J. Plimpton, A. Tarditi, M. S. Chu, and the NIMROD Team, J. Comput. Phys. **195**, 355 (2004).

[4] X. S. Li and J. W. Demmel, ACM Trans. Math. Software **29**, 110 (2003).

tests represent a short segment of the nonlinear evolution when the magnetic perturbation reaches approximately half of its peak amplitude prior to the first reconnection event. Computations with very high resolution performed for this scaling include modes that are well into the dissipation range; resolution of the fine reconnection scale is achieved with 86 Fourier components. However, the resolution afforded by parallel scaling can be exploited for modeling higher temperature or for ELM evolution, for example, where very short wavelengths are generated nonlinearly.

Results of the scaling analysis are summarized in Figure 1. Scaling with an increasing number of Fourier components is separated into continuous traces with 512, 1024, or 2048 high-order elements. Computations with 512 and 1024 elements use four cores per node on the Franklin computer. Additional memory per core is needed for the two largest computations with 2048 elements, and the traces with red symbols show results with two cores per node. Two smaller 2048-element computations with four cores per node are shown for comparison. The "Loop Time" traces indicate the wall-clock time for 50 time-steps without the relatively short set-up and completion times. The "FFT Time" traces show that the amount of time spent in collective communication for the collocation computations does not impede scaling to 10,000 processor cores with the improved preconditioning and organization of data. The former also leads to the nearly constant number of GMRES iterations required for a 2-norm relative tolerance of 10^{-9} for the two-fluid magnetic-field advance as the number of Fourier components is increased, as shown in Figure 2.

The increase in total time with the number of spectral elements results from the increasing number of operations required for direct solution of the diagonal blocks. While direct solves address the dominant stiffness of the time-advance, they may not be required as part of a preconditioning step. We are presently working with LBNL's Xiaoye Li, the author of SuperLU_DIST, to test threshold-based ILU as a possible alternative with potential for better scaling.

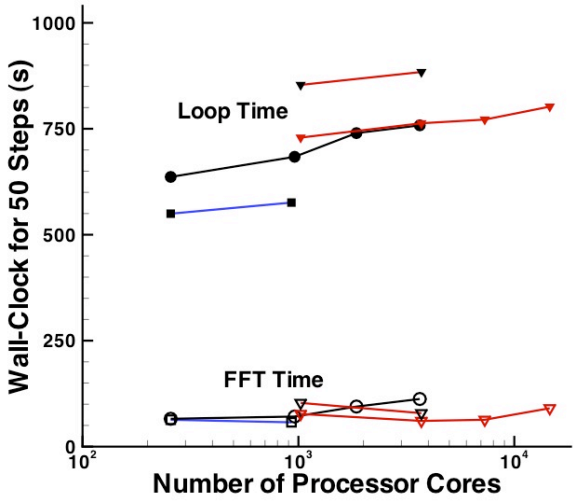

Figure 1. Scaling results with 512 elements (blue), 1024 elements (black), and 2048 elements (red) for total time and FFT/collective communication time for 50 steps. The red-symbol traces show results with two cores per node. Traces with black symbols show results with four cores per node.

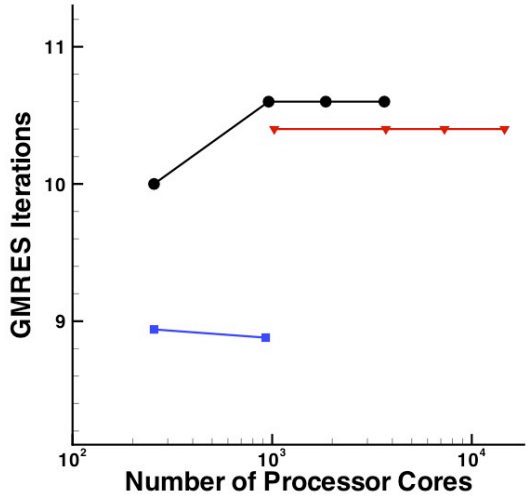

Figure 2. GMRES iterations per step for the 2-fluid magnetic-field advance.