

Center for Tokamak Transient Simulations
FY2020 Progress Report

E. Held, *Utah State U.*, S. C. Jardin, *PI, PPPL*, J. King, *Tech. X*, L. Lao, *General Atomics*, X. Li, *LBL* R. Samulyak, *Stony Brook U.*, M. Shephard, *RPI*, C. Sovinec, *U. Wisconsin-Madison*, H. Strauss, *HRS Fusion*, S. Williams, *LBL*

Table of Contents

0.0 Project Overview 3

1.0 Neoclassical Tearing Modes 4

 1.1 NTM Studies with NIMROD 4

 1.2 Mode Locking 5

 1.3 Incorporating Kinetic Effects into NTM modeling 5

 1.3.1 Improved computation of the Rosenbluth potentials 5

 1.3.2 Analytic form for Coulomb collision operator for arbitrary flow ordering 6

 1.3.3 Successful implementation and testing of time-implicit moment terms in CEL-DKE for neoclassical transport calculations 6

2.0 Vertical Displacement Events 6

 2.1 Benchmarking Activity 7

 2.2 The Force due to Axisymmetric Halo Currents in ITER 10

 2.3 Calculation of 3D “Sideways Force” in ITER 10

3.0 Disruption Mitigation 10

 3.1 NIMROD SPI Disruption Mitigation Development 10

 3.2 M3D-C1 Disruption Mitigation 11

 3.3 M3D-C1/NIMROD Code verification Benchmarking 12

 3.4 Local Modeling of Pellet ablation and Coupling to MHD codes 12

 3.4.1 Studies of the grad-B drift effect 12

 3.4.2 Influence of radiation model, EOS for mixtures 15

 3.4.3 Simulation with changing radius pellets 16

 3.4.4 Multiscale coupling of LP and tokamak MHD codes 16

4.0 Runaway Electrons 17

 4.1 Reduced Runaway Electron Model in NIMROD 17

4.2 Reduced Runaway Electron Model in M3D-C1.....	18
5.0 Additional Experimental Validation	18
5.1 Validation studies with CTH	18
5.2 Validation studies with JET	19
5.3 Validation studies with MST	20
6.0 Other Physics Studies.....	21
6.1 Peeling-Ballooning Modes and Ideal Stability.....	21
6.2 A New Interpretation of Sawtooth Phenomena in Tokamaks	21
7.0 Physics Code Improvement and Optimization.....	22
7.1 NIMROD Refactoring and Vacuum region basis functions	22
7.2 M3D-C1 code development and Improvements	22
7.2.1 Adapting M3D-C1 for stellarator simulation.....	22
7.2.2 Improving Robustness of Pellet Simulations.....	23
7.2.3 New Capabilities for Toroidal Planes	23
7.3 Improvements to the Lagrangian Particle Code	23
7.3.1 Massively parallel Lagrangian particle pellet / SPI code	23
7.3.2 Cross-code comparison and resolution of disagreements between Frontier and Lagrangian Particle Codes.....	24
7.3.3 New Adaptive Algorithms for LP-based SPI code	25
7.4 3D Sparse LU and triangular solves.....	25
8.0 Solver Improvements.....	25
8.1 Introduction	25
8.2 M3D-C1/NIMROD parameter tuning	26
8.3 One-Sided Communication	27
8.4 GPU Acceleration	28
8.5 Integrating 3D Algorithms.....	30
8.6 Software Releases	30
9.0 Meshing and other M3D-C1 Support.....	30
9.1 Mesh Control for Pellet Simulations	30
9.2 Supporting Mesh Adaptation of the 3D Extruded Meshes.....	31
9.3 Support of a Growing M3D-C1 User Community.....	33
10.0 Publications and Conference Proceedings.....	33

0.0 Project Overview

The Center for Tokamak Transient Simulations is primarily focused on the further development and application of two of the world's most capable extended MHD codes, NIMROD and M3D-C1. These are time-dependent 3D MHD codes that can accurately describe the time evolution of transient event in tokamaks. Some of these transient events such as the seeding of a neoclassical tearing mode (NTM), and its subsequent locking and growing, can lead to a plasma disruption, a potentially catastrophic event for a burning plasma. For this reason, the emphasis of our Center is on better understanding the causes of disruptions, how to avoid them, and how to mitigate them. As part of the mitigation studies, we also seek to predict the consequences of a worst case non-mitigated disruption.

Section 1 describes our efforts in NTMs with the long term goal of better understanding how and when they are formed, under what conditions they will lock and grow, and how this leads to a disruption and how best to prevent it. Since NIMROD and M3D-C1 are fluid codes, and some intrinsic kinetic effects are necessary to fully describe the evolution of a NTM, we have a parallel effort, primarily by Utah State U., to include neoclassical kinetic effects into the fluid codes.

One of the most damaging types of disruption is the vertical displacement event (VDE), discussed in Section 2, where vertical position control is lost and the plasma column firsts moves vertically and then disrupts. We are the first team to attempt to predict the consequences of these VDEs in ITER by using a fully 3D plasma model and a realistic model for the ITER vessel. We are proceeding through steps of first 2D and then 3D modeling and code verification and validation to gain confidence in our results. Again, the goal is to better understand the consequences of VDEs in order to effectively mitigate them.

Section 3 describes our efforts in modeling disruption mitigation by impurity pellet injection. We are approaching this using two methods: stand-alone modeling with NIMROD and/or M3D-C1, and code-coupling modeling where we couple one of these global codes to a code, either Frontier or LPC, that calculates in detail the local ablation physics of the pellet. These are being applied to both solid pellets, and to shattered pellets where you have hundreds of pellet fragments entering the plasma.

It is well known that high energy "runaway" electrons are generated during a disruption. Our efforts to describe these and to couple them to the bulk MHD physics are described in Section 4. We now have relatively crude models of runaway electrons implemented in NIMROD and M3D-C1 but have plans to improve the fidelity of these models by coupling with intrinsic kinetic models.

Our efforts in experimental validation are described in Section 5. This is of utmost importance to test the validity and limitations of our models. None of the projections of our models to ITER will be taken seriously unless we can show that they can describe similar phenomena in today's experiments.

In Section 6 we describe applications of our codes to edge localized modes and to sawteeth. The latter study is somewhat controversial, as it identifies a new mechanism for sawteeth in tokamaks. In section 7-9 we describe our efforts in improving the performance of our codes on some of the largest, most powerful computer in the world. This improvement comes through using improved algorithms, improved data structures, and in finding ways to increase the parallelism of the algorithms that we use. Many of the improvements made by the Computer Science side of our Center will also benefit other applications outside of fusion that have sparse matrix problems similar in form to ours.

1.0 Neoclassical Tearing Modes

1.1 NTM Studies with NIMROD

Tech-X

Typically NTMs are “born” rotating in the plasma fluid frame of reference. Disruptions occur after the NTM breaks and “locks” (magnetic field penetration through the conductor) to the wall. As such, it is of interest to establish both the physics basis of this process on present devices, such as DIII-D, and the scaling of the torques responsible for this locking to planned larger devices such as ITER. We are working with the group at Utah State University to use the drift-kinetic closures to simulate the neoclassical aspect of NTMs. We anticipate that these 5D computations will be computationally expensive and as such we have implemented the heuristic fluid closures [Gianakon et al., Phys. Plasmas 2002] into NIMROD to allow for parameter space exploration.

Work, led by Eric Howell, is underway to explore the dynamics of the NTM seeding on DIII-D shot 174446 in collaboration with Jim Callen, Rob La Haye and Bob Wilcox. This discharge uses the DIII-D ITER baseline parameters and shaping and contains both ELMs and sawteeth. The discharge is terminated by a 2/1 NTM locking and causing a disruption. The 2/1 NTM is present at low amplitude before growth to a large size is triggered by an ELM. Rich multiple mode dynamics are also present with 4/3 activity before the 2/1 grows large and a persistent 3/2 mode throughout the discharge.

While NIMROD is unable to simulate the long time scales associated with these rich dynamics (>1 second) we are able to study the multi-mode dynamics. NIMROD simulations are initialized from the best reconstruction as constrained by magnetics, Thomson scattering and Charge exchange recombination (CER) spectroscopy measurements before the 2/1 begins to grow to large amplitude. The toroidal and poloidal flows based on the CER measurements are included as the island rotation is critical to mode and error field coupling as well as interaction with a resistive wall. In addition, the flow stabilizes any modes associated with the edge pedestal. Because simulation of an ELM crash that triggers and NTM mode is not practical, we apply an $n=1$ edge magnetic perturbation pulse at the boundary. The pulse envelope last for 1ms and is optimized to excite the 2/1 mode by tuning the magnetic coils to achieve the largest vacuum 2/1 response relative to the 3/1 response as measured at the respective resonant surfaces. During and shortly after the pulse the 4/1 and 3/1 modes are driven and the edge magnetic surface are stocasticized. The 2/1 mode is driven to small amplitude and begins to slowly grow. After the edge 3/1 and 4/1 edge modes decay to small amplitude at approximately 5ms

a rapid succession of core modes begins. Initially the $6/5$ mode is driven which leads to a cascade to the $5/4$ then $4/3$ ultimately exciting the $3/2$ mode. At this point the $2/1$ begins to grow rapidly while all other modes except the $3/2$ decay. Examination of the pressure profile shows that the pressure is steepened inside each resonant surface before modes grow. Further examination through consideration of the parallel Ohm's law (which can also be formulated as the radial induction equation) is planned to study the details of the mode coupling. Another complicating factor is the value of the safety factor on axis is raised above one on axis to avoid exciting a large unstable $1/1$ response in the sawtooth discharge.

1.2 Mode Locking

U. Wisconsin-Madison

Professor Ping Zhu's group has developed a new theoretical model on the nonlinear plasma response to external magnetic perturbation without assuming the standard no-slip condition between plasma flow and the island. The model is composed of the equations for the evolution of the width and phase of magnetic islands that are forced by external perturbation and a force-balance equation for the plasma flow. They find that when the island width is much less than the resistive layer width, the island growth is governed by the linear Hahn-Kulsrud-Taylor solution in the presence of time-dependent plasma flow. In the large-island limit, the evolution of both the island width and phase is consistent with Rutherford theory. The island solution is used to construct the quasilinear electromagnetic force, which, together with the viscous one, contributes to the nonlinear variation in plasma flow. The no-slip condition assumed in the conventional error field theory is not imposed, and the island oscillation frequency depends on, but does not necessarily equal, the plasma flow frequency at the rational surface. This work has been published in Ref. [8].

1.3 Incorporating Kinetic Effects into NTM modeling

1.3.1 Improved computation of the Rosenbluth potentials

Last year it was noticed that before proceeding with our work to introduce kinetic effects into simulations of NTMs, improvements were needed to the previous implementation of the Rosenbluth potentials in the field portion of the collision operator. Drs. Andrew Spencer and Held put a great deal of work into constructing an efficient and more accurate method for computing the Rosenbluth potentials. Part of this work entailed finding a robust method of integrating over singularities in the Green's functions of the velocity space Poisson equation. We implemented a numerical splitting technique adapted from the astrophysics community where gravitational potentials from axisymmetric sources are computed. We also were able to derive a series expansion that reduced the dimensionality of a large portion of the calculation. The results of our algorithm were compared to those of a third-party general adaptive integrator to demonstrate convergence to a desired accuracy. Our algorithm was significantly more efficient, and able to quickly evaluate results for cases where the third party software was not able to complete the calculation due to running out of computer memory. In the process of writing our work up for an article for Computer Physics Communications, we found a simplification of the series expansion that may be able to reduce the dimensionality of the entire evaluation of the Rosenbluth potentials. This depends on how efficiently Legendre polynomials can resolve localized

structures of the velocity distribution function in pitch angle, such as arise at trapped/passing boundaries in tokamak plasmas. We are currently exploring this before publishing our work and resuming work on NTMs. If it is found that the expansion can be extended to the whole velocity domain, then kinetic simulations of neoclassical tearing modes will be greatly accelerated.

1.3.2 Analytic form for Coulomb collision operator for arbitrary flow ordering

The major goal of this work was to develop accurate closure relations which can capture runaway electron effects. For an arbitrary relative flow velocity between electrons and ions ($V_{ei} \equiv V_e - V_i$), we have developed formalism to calculate the collisional moments of the Landau (Fokker-Planck) operator. We have derived explicit formulas for the analytic collision coefficients which are expressed in terms of finite series of V_{ei} . Since the moment expansion is performed in each fluid frame with a shifted Maxwellian distribution function, the formulas are exact for arbitrary V_{ei} . The formulas involve only algebraic expressions and can be implemented in symbolic computation programs such as Mathematica and Maple. The theoretical work has been published in [9]. The results may be used in situations where the electron distribution has developed a significant flow moment, on the order of the electron thermal speed, but a relativistic treatment is not yet necessary. Corrections terms in V_{ei} are critical for handling the early-in-time development of a runaway electron (RE) distribution that often accompanies disruptions.

1.3.3 Successful implementation and testing of time-implicit moment terms in CEL-DKE for neoclassical transport calculations

During Dr. Held's sabbatical year at U. of Wisconsin-Madison, he worked closely with his colleague Dr. Chris Hegna and former USU undergrad/now Wisconsin graduate student Joseph Jepson. In the fall of 2019, Dr. Jim Callen also participated in meetings on the predictions of ion CEL-DKE kinetics using the continuum kinetic model in NIMROD. This work led to several improvements in the implementation of time-implicit moment terms, like the anisotropic part of the stress tensor. A time-implicit treatment for moments of the non-Maxwellian distribution function is essential to taking large time steps in hybrid fluid/kinetic simulations of NTMs. The work on the ion CEL-DKE moment terms has paved the way for NTM simulations using an Alfvénic time steps. Joseph is preparing a publication with Drs. Held, Hegna and Callen as co-authors.

2.0 Vertical Displacement Events

A Vertical Displacement Event (VDE) is an off-normal occurrence in a tokamak in which position control of the discharge is lost, and the tokamak plasma moves rapidly upward or downward until it makes contact with the vacuum vessel. The discharge current in ITER will be up to 15 MA. When a plasma with this current makes contact with the vessel, it will induce large currents into the metallic vessel, and these currents will cause large forces. Previous studies commissioned by ITER to calculate these forces assumed that the plasma remained axisymmetric during the VDE to simplify the calculation. However, it is known that the plasma column will deform and produce "sideways forces" in ITER that could potentially damage the machine. Our two flagship MHD codes, NIMROD and M3D-C1 now have the

capability of modeling a fully 3D plasma interacting with a conducting structure. We are using this capability to realistically model a full 3D VDE in ITER and to calculate the expected forces.

2.1 Benchmarking Activity

U. Wisconsin-Madison, PPPL

: The NIMROD and M3D-C1 codes are widely used for modeling macroscopic plasma dynamics in magnetic confinement systems, including disruption applications. As such, verifying their disruption-modeling capabilities is important. During this past year we completed and published a benchmarking study of axisymmetric VDE modeling with the NIMROD, M3D-C1, and JOEK codes [1]. We have also started benchmarking 3D evolution, where asymmetric instability results from contact with the wall. This work is part of the FY2020 theoretical performance target/milestone for Fusion Energy Sciences. [see <http://m3dc1.pppl.gov/2020TheoryMilestone.html>]. The computations are based on an equilibrium fit of NSTX discharge 139536 after feedback stabilization had been partly suppressed to allow vertical instability. While all three codes have been developed for macroscopic stability, they differ in the equations that are solved and in their numerical methods. NIMROD and M3D-C1 solve equations for full-MHD, whereas JOEK [Huysmans, et al., Plasma Phys. Control. Fusion 51, 124012 (2009)] is used to solve reduced-MHD equations. NIMROD solves primitive-field equations for flow velocity and magnetic field, and M3D-C1 and JOEK use potential-field representations. Numerically, NIMROD uses semi-implicit and implicit methods for its temporal advance, together with spectral-elements/Fourier series for its poloidal/toroidal spatial representation. M3D-C1 and JOEK use implicit temporal advances, and M3D-C1's spatial representation uses reduced quintic elements/Hermite cubics, whereas JOEK uses Bezier elements/Fourier series. Thus, this three-code benchmarking compares distinct approaches.

For the 3D comparison, each of the three codes is run in 2D until the last closed flux surface (LCFS) contacts the wall. This phase is similar to the full nonlinear computation of the Krebs, et al. paper, but plasma and wall resistivity values, particle diffusivity, and thermal conductivities are made an order of magnitude larger. This is done so that the subsequent 3D phase occurs more rapidly. After LCFS contact, the computations are restarted in 3D with thermal conductivities increased by two orders of magnitude and particle diffusivity increased by 26. This initiates an artificial thermal quench (TQ), i.e. one that does not occur from 3D instability, which also helps shorten the required computation time. As shown in Fig. 1, after approximately 1 ms into this 3D phase, the safety-factor profile has flattened to a nearly uniform value of 2. An external $m=2$, $n=1$ mode then grows robustly and excites other modes as it saturates, which accelerates the TQ by eroding the outer flux surfaces (Fig. 2). The resolution of the NIMROD computation has been tested by increasing the toroidal Fourier resolution from $n \leq 10$ to $n \leq 21$, starting at 9.3 ms, which is when the mode coupling strengthens. The evolution of thermal energy and magnetic fluctuation energy, shown in Fig. 3, indicates that the computations are already reasonably well resolved with $n \leq 10$. Convergence is facilitated by the relatively low ratio of parallel and perpendicular thermal conductivities of 10^5 . Computed results on net horizontal and vertical forces on the resistive wall are shown in Fig. 4.

The milestone report documents the 3D benchmarking. As an example, Fig. 5 shows the evolution of magnetic fluctuation spectra from NIMROD and JOREK computations with $n \leq 10$ and from an M3D-C1 computation with 16 toroidal planes. While there are quantitative differences in the timing of activity and in energy levels, the three codes reproduce essentially the same dynamics.

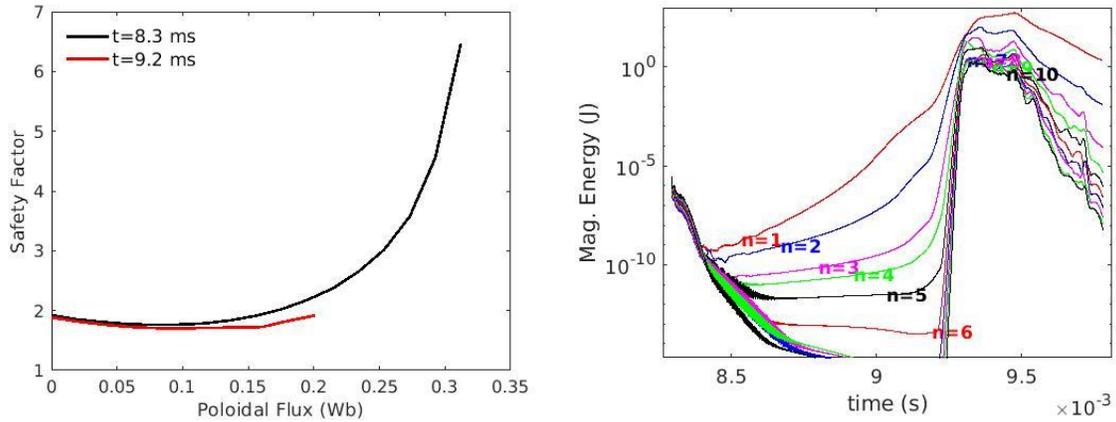


Figure 1. Closed-flux safety-factor profiles at the beginning of the 3D phase and at the onset of (2,1) activity (left), and evolution of magnetic fluctuation energies from the 3D NIMROD computation with toroidal resolution of $n \leq 10$ (right).

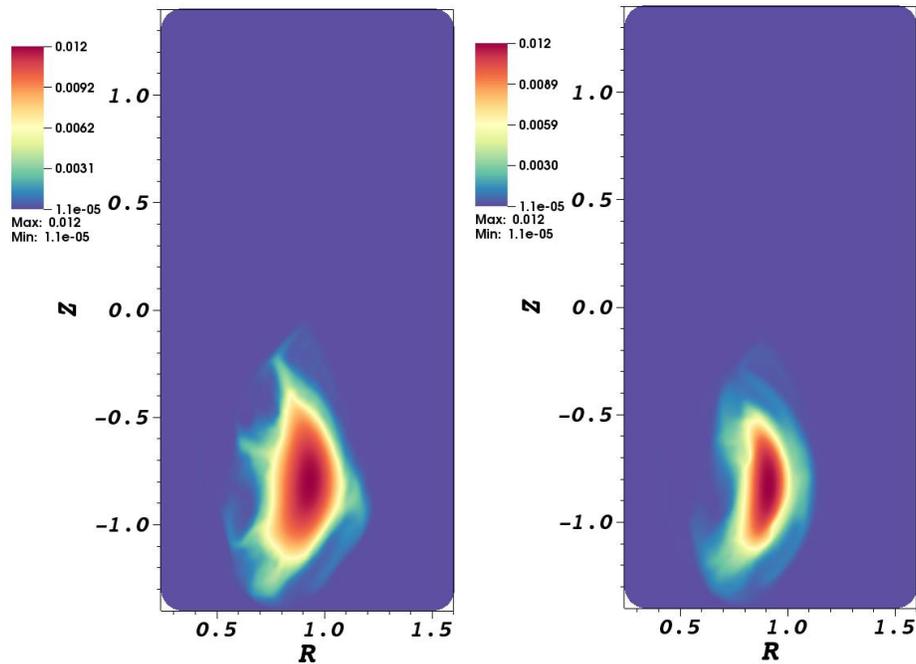


Figure 2. Color contours of plasma pressure over a constant toroidal-angle plane at times 9.35 ms (left) and 9.4 ms (right) of the NIMROD computation with $n \leq 21$.

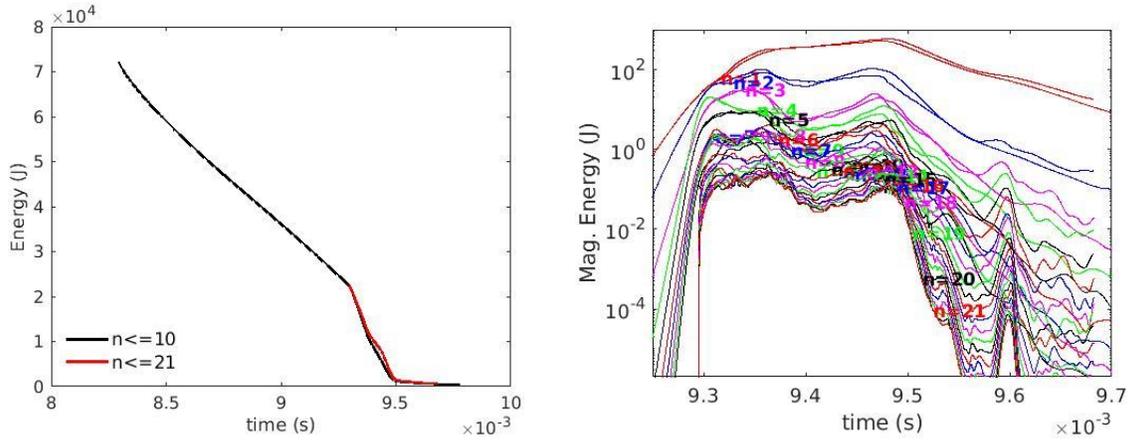


Figure 3. Evolution of thermal energy (left) and of magnetic fluctuation energies (right) from the two NIMROD computations with differing toroidal resolution.

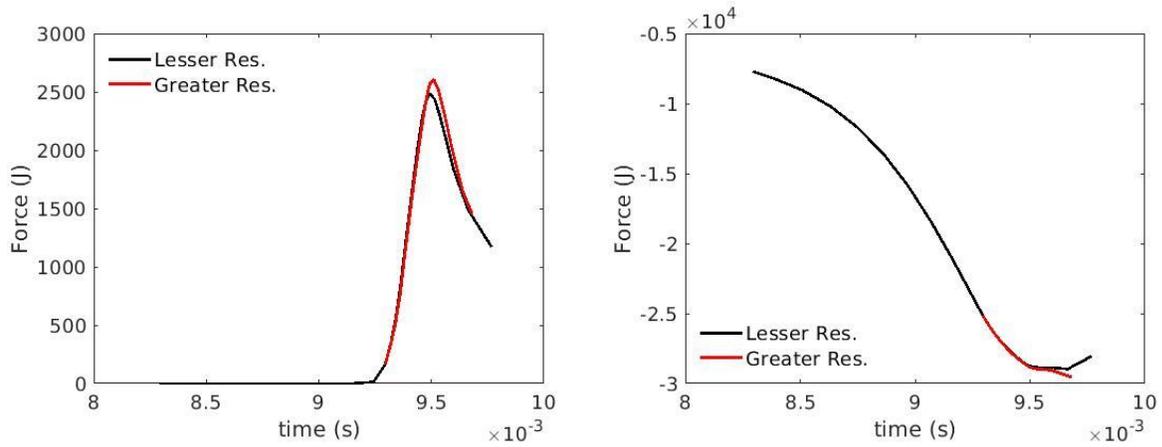


Figure 4. Evolution of horizontal (left) and vertical (right) forces on the resistive wall, as computed in the 3D NIMROD computations with $n \leq 10$ and with $n \leq 21$.

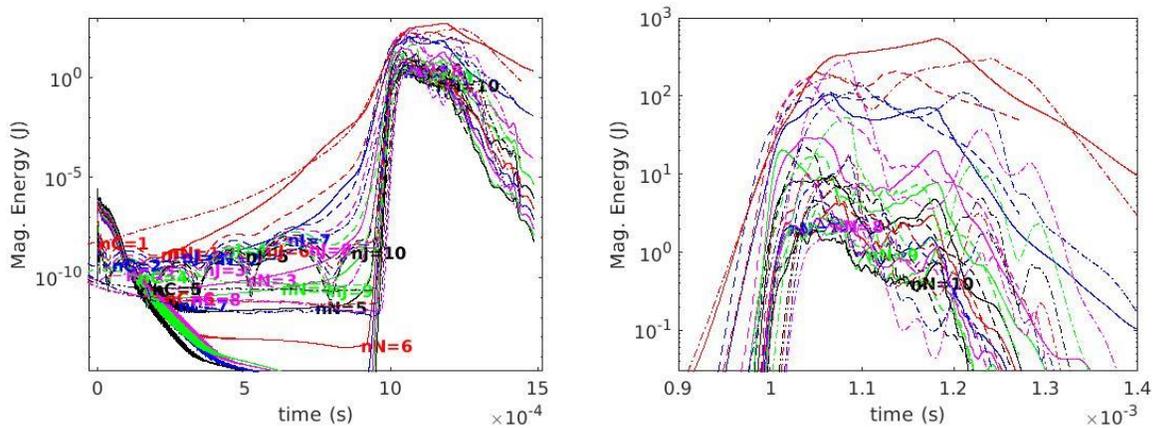


Figure 5. Comparison of magnetic fluctuation energies from NIMROD (solid traces), JOEKE (dashed), and M3D-C1 (dash-dot) over the 3D evolution. Here, computation times are shifted to match the start of the rapid-growth phase with $t = 0$ being the start of the 3D run segment. [JOEKE data courtesy of F. J. Artola.]

2.2 The Force due to Axisymmetric Halo Currents in ITER

PPPL

Vertical displacement events (VDEs) can occur in elongated tokamaks causing large currents to flow in the vessel and other adjacent metallic structures. To better understand the potential magnitude of the associated forces and the role of the so called ‘halo currents’ on them, we have used the M3D-C¹ code in Ref. [2] to simulate potential VDEs in ITER. We used actual values for the vessel resistivity and pre-quench temperatures and, unlike most of the previous studies, the halo region is naturally formed by triggering the thermal quench with an increase in the plasma thermal conductivity. We used the 2D non-linear version of the code and vary the post- thermal quench thermal conductivity profile as well as the boundary temperature in order to generate a wide range of possible cases that could occur in the experiment. We also show that, for a similar condition, increasing the halo current does not increase the total force on the wall since it is offset by a decrease in the toroidal contribution.

2.3 Calculation of 3D “Sideways Force” in ITER

PPPL

The report obtainable as Ref. [12] describes the first ever 3D MHD simulation of a vertical displacement event disruption in ITER using a model for the ITER vacuum vessel with a realistic time constant and a multi-region conductivity model. The vertical vessel forces that we calculate are consistent with those obtained previously with 2D simulations [1]. The horizontal (or sideways) forces calculated are substantially smaller than what one would obtain from a straightforward scaling of JET results to larger values of the plasma current and magnetic field. This result was obtained with M3D-C1. If it can be independently verified, it will remove a large uncertainty regarding the effects of disruptions on the ITER vessel.

3.0 Disruption Mitigation

3.1 NIMROD SPI Disruption Mitigation Development

General Atomics

NIMROD SPI simulations are concluding a scan of several parameters, including viscosity and toroidal deposition to study their impact on the quench dynamics. As expected, lower viscosity results in shorter thermal quench time due to a stronger linear response resulting in faster onset of (2,1) and (3,2) tearing modes. The final thermal collapse is usually dominated by a (1,1) core mode. These instabilities disrupt

the plasma and are accompanied by a bright burst of rapid thermal collapse. Due to the instability driven thermal collapse, radiation efficiency is at best ~50%

Dependence on the toroidal deposition is less obvious, but typically narrow depositions leads to a stronger instability drive. However, the strong nonlinear nature of these simulations complicates a straight linear interpretation. Although onset is earlier, the saturation amplitudes may be smaller than for the broader deposition. Also, the higher concentration of a narrower toroidal deposition results in less overall ablation and assimilation. Analysis continues on the toroidal dependence.

For SPI thermal quenches, the current spike is observed to occur several 100 μ s after the plasma thermal quench. Symmetric multi-injector SPI simulations show that the current spike can be suppressed if the final dominant (1,1) core mode can be suppressed. This is achieved in NIMROD with dual SPI injectors placed 180° apart (similar to KSTAR's SPI system). Simultaneous dual SPI injectors can completely suppress odd n-modes and result in a benign thermal quench (absence of radiation spikes and current spikes) with high radiation efficiency > 80%. Analysis continues on multi-injector SPI.

These parameter scan results are being analyzed and compiled for publication and submitted for an invited presentation at the upcoming IAEA meeting.

U. Wisconsin-Madison

Prof. Ping Zhu's groups at the University of Science and Technology of China and at Huazhong University of Science and Technology investigate a range of disruption-related macroscopic dynamics. Their new study of disruption mitigation investigates why a (2,1) MHD mode is almost universally observed before the TQ and why an internal cold bubble structure is observed. Their recent NIMROD computations with massive gas injection (MGI) modeling [Izzo, Nucl. Fusion 46, 541 (2006)] find that the impurity density penetration and radiation cooling lead to the formation of an X-point reconnection site and then stochastic magnetic field at the interface between impurity and main-species plasma. As this interface moves with changing magnetic topology, the impurity distribution spreads inside the q=1 surface through the O-point of an inner (2,1) structure and forms a cold bubble that leads to the final phase of the TQ.

3.2 M3D-C1 Disruption Mitigation

General Atomics

Making use of new physics and numerical techniques to improve robustness of the simulations, we performed initial M3D-C1 validation with DIII-D pellet-injection experiments. Simulations with different pellet compositions were performed showing a faster thermal quench for pure-neon pellets as compared to mixed-neon-deuterium pellets. The pure-neon pellet showed increased radiation and also drove a strong $n = 1$ instability, while the mixed-neon-deuterium pellet resulted in a quiescent radiation-dominated thermal quench. This result qualitatively agrees with both experiment and NIMROD simulations. Additional modeling showed that more poloidally localized impurity sources result in early MHD instability.

M3D-C1 has also been upgraded to allow for an arbitrary number of pellet impurity sources. This has been used to perform initial modeling of dual pellet injection and will be used to simulate plumes of shattered pellet fragments.

3.3 M3D-C1/NIMROD Code verification Benchmarking

General Atomics

Building off a highly successful 2D benchmark, 3D simulations of the same scenario with axisymmetric, on-axis impurity deposition were performed by both M3D-C1 and NIMROD. Both codes found that the plasma remained axisymmetric deep into the thermal quench. Eventually, the current sheet that forms goes non-axisymmetrically unstable, resulting in a current spike followed by a fast current quench. The M3D-C1 simulations showed a very pronounced current spike of 10% of the total current. Further M3D-C1 simulations were carried out, showing the existence of the current spike independent of time step, toroidal resolution, or hyper-diffusivity.

In addition, a benchmark between M3D-C1 and NIMROD is underway for an externally injected pellet in DIII-D. Initial results show excellent agreement, but M3C1 simulations have been stymied due to the formation of negative-temperature regions. Work on this is in progress.

3.4 Local Modeling of Pellet ablation and Coupling to MHD codes

Stony Brook U. and General Atomics

3.4.1 Studies of the grad-B drift effect

In a tokamak magnetic field, the pellet ablation cloud propagates along magnetic field lines. In addition, a rapid movement of the ablated material towards the outward major radius R direction has been observed. This motion has been attributed to a vertical curvature and grad-B drift current induced inside the ionized ablated material by the $1/R$ toroidal field variation. The uncompensated vertical drift current inside the cloud causes charge separation at the boundary. The resulting electrostatic field induces the $E \times B$ drift to the large-R side of the torus. This transverse motion, which we call the grad-B drift effect, established the pellet shielding length of the ablation cloud.

In the axially symmetric approximation, grad-B drift cannot be resolved. Without the transverse motion, the ablated material would extend along magnetic field lines, increasing the pellet shielding and eventually stopping the pellet ablation. Therefore, a finite shielding length of 16 cm (a theoretical estimate) was imposed in both codes for axisymmetric pellet ablation in magnetic fields. Axisymmetric simulations predict a strong reduction of the ablation rate in magnetic field of increasing strength for both deuterium and neon pellets (see Figure below).

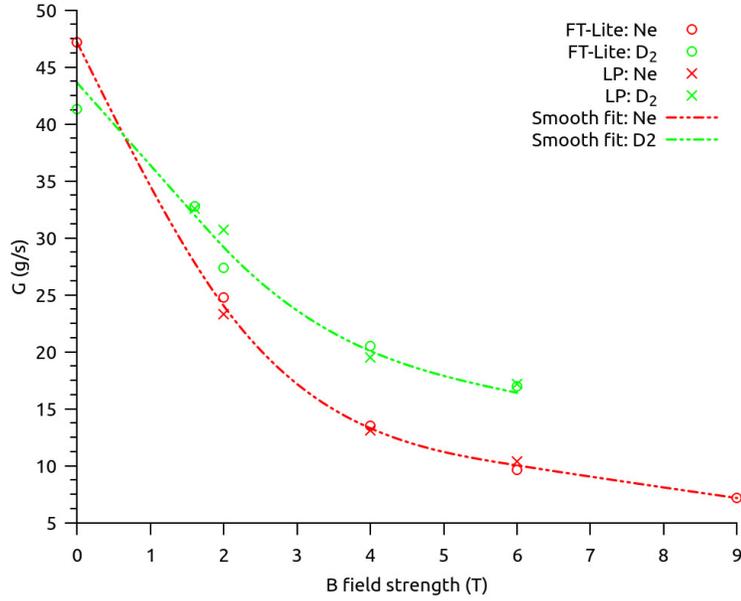


Figure 6. Ablation rate as a function of B for neon (red) and deuterium (green) pellets computed from the FronTier and Lagrangian particle pellet codes in axially symmetric approximation. The shielding length was set as 16 cm for all values of the magnetic field.

We have performed extensive studies of the grad-B drift effect using the Lagrangian particle code. The drift velocity v_D is implemented as

$$\frac{dv_D}{dt} = \frac{2}{R\langle\rho\rangle} \left\langle P \left(1 + \frac{M^2}{2} \right) - P_\infty \right\rangle - \frac{v_D}{\langle\rho\rangle} \frac{2B^2}{\mu_0 v_A}$$

where R is the tokamak major radius, P and ρ are the ablation cloud pressure and density, P_∞ is the ambient plasma pressure, M is the Mach number of the ablation flow, v_A is the Alfvén velocity, and $\langle A \rangle$ denotes the integral of quantity A along magnetic field lines. A neon pellet cloud cross-section showing the shielding length obtained self-consistently via the grad-B drift is shown in Figure 7 below.

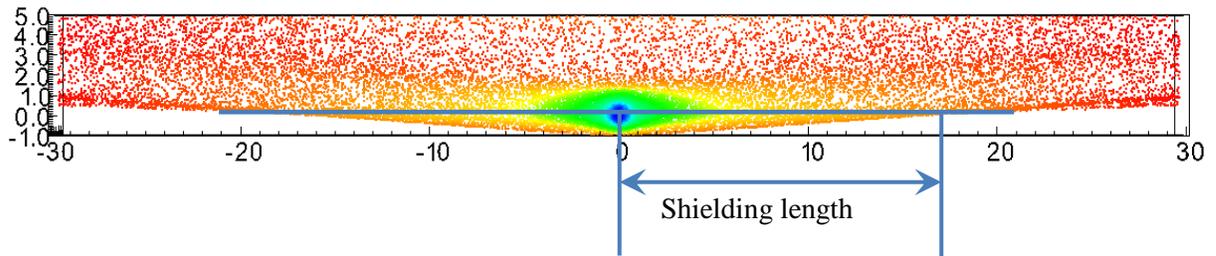


Figure 7. Cross-section of neon pellet ablation cloud showing the shielding length obtained self-consistently via the grad-B drift.

We performed studies of the influence of various terms in the grad-B drift formula and showed that the Mach number term provides an essential contribution while the last term, describing the Alfvén wave drag, is negligibly small. With grad-B drift, we observe a smaller reduction of the ablation rate in magnetic field compared to simulations with fixed shielding length. This is due to a combined effect of shorter shielding lengths and slightly changed hydrodynamic states in the ablation cloud. Grad-B drift in DIII-D ($R_0 = 1.6$ m) is stronger compared to ITER ($R_0 = 6.2$ m), all other factors assumed equal.

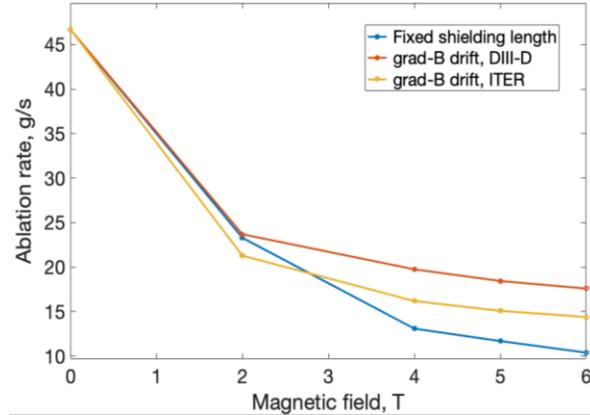


Figure 8. Ablation rate of neon pellets in magnetic fields with resolution of grad-B drift.

Tables below shows values of the ablation rate and the shielding length for the DIII-D and ITER major radii.

B (T) DIII-D	Shielding length, cm	G(g/s)		B(T) ITER	Shielding length, cm	G(g/s)
2	18	23.7		2	35.5	21.3
4	14	20.0		4	30	16.2
5	13	18.5		5	27.5	15.1
6	12	17.6		6	25.5	14.4

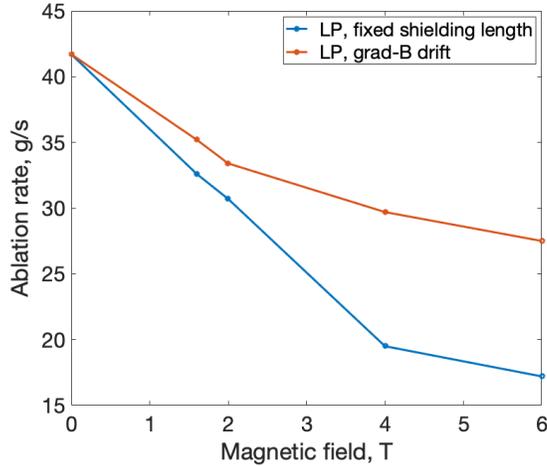


Figure 9. Ablation rate of deuterium pellets in magnetic fields with resolution of grad-B drift.

This study shows that any empirical $G(B)$ fitting function should be dependent on the tokamak major radius. Since B/R does not change significantly for various tokamaks ($B[\text{Tesla}]/R[\text{m}] \sim 1$), we are building a pellet ablation simulation database assuming the constant $B/R=1$ ratio

3.4.2 Influence of radiation model, EOS for mixtures

All simulation results for neon pellets presented above used a thin optical limit radiation model and a database precomputed with the CRETIN code. Additional analysis shows that re-absorption of radiation could be important. Using the PrismsPECT code, E. Hollmann and P. Parks computed an emissivity table by considering the full neon radiation spectrum and the radiation absorption in a density – temperature thermodynamic domain typical for neon pellet ablation clouds. Simulations with the new reduced radiation model resulted in significantly increased temperatures. In turn, the increased temperature caused more intense radiation. Preliminary studies showed small influence of the radiation model change on the pellet ablation rates, but this work will continue in the future with improved radiation tables.

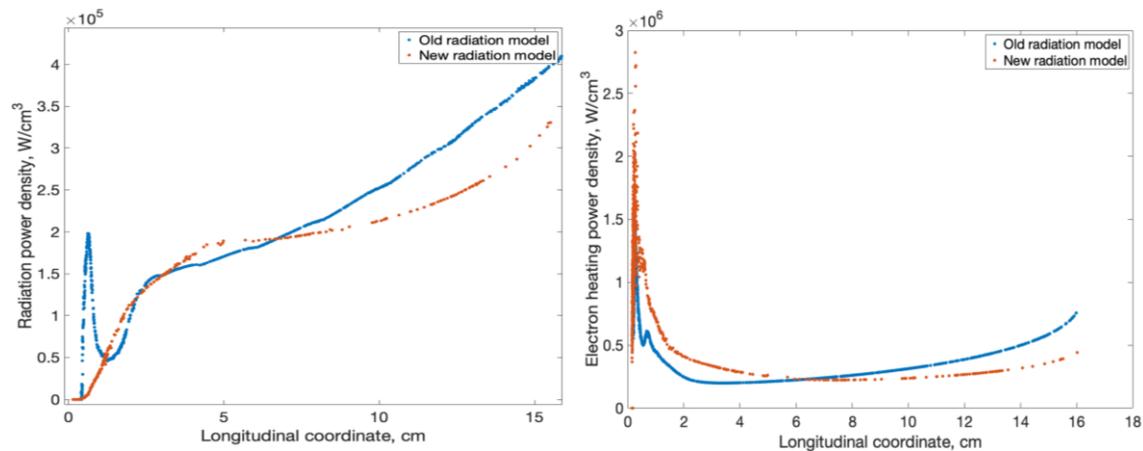


Figure 10. Comparison of the radiation and electron heating power densities for the old (thin optical limit) and new radiation models. With both radiation models, the electron heat flux is significantly higher than radiation close to the pellet surface. With the new radiation model, radiation is much close to the electron heat flux in the far field.

We have also developed an equation of state model for neon – deuterium mixtures in the approximation of local thermodynamic equilibrium (Saha model). The model was implemented in a stand along code for building tabulated EOS data sets for the Lagrangian particle code. The new EOS model will be used in the future for the simulation of composite pellets.

3.4.3 Simulation with changing radius pellets

All simulation results presented above were obtained with a fixed-radius pellet in order to obtain steady-state ablation rates. In this section, we summarize simulations with decreasing radius pellets due to the ablation process. This study established a very important result: for simulations in magnetic fields (for both settings of the shielding length: the fixed shielding length and the grad-B drift-defined shielding length), the ablation rate still scales mainly as $G \sim r_p^{4/3}$, just as in the case of the theoretical scaling law for spherically-symmetric ablation.

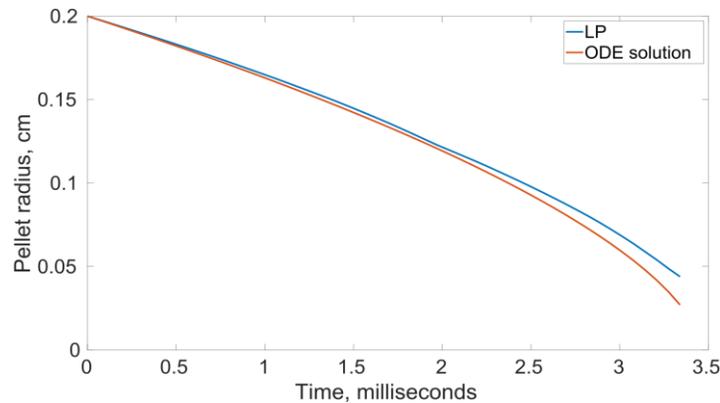


Figure 11. Simulation of a neon pellet with decreasing radius due to the ablation process in 2 T magnetic field with the shielding length established by the grad-B drift. Blue lines depict LP simulation result. Red line corresponds to the ODE solution based on the analytic scaling law. The initial ablation rate for the scaling law was obtained from LP simulations.

3.4.4 Multiscale coupling of LP and tokamak MHD codes

General Atomics

We worked with colleagues at Stony Brook University to couple M3D-C1 and NIMROD to the Lagrangian-particle (LP) code being developed to simulate the detailed physics of pellet ablation more accurately. A file-based loose coupling was performed. A script was created such that the LP code could read needed plasma information (e.g., density, temperature, magnetic field strength) from M3D-C1 output. The LP code was then advanced to steady-state to simulate the ablation, ionization, and advection of the impurities around the pellet. A file was created specifying the location and configuration of each LP that has drifted out of the code’s simulation domain. A smooth impurity source for each charge state was created in M3D-C1 by projecting the LPs as delta functions onto the M3D-C1 finite elements. The next step will be to iterate between the codes for a dynamic simulation.

4.0 Runaway Electrons

High energy runaway electrons will be generated during a disruption. There is another SciDAC, SCREAM, which is primarily concerned with describing the generation of runaway electrons. However, our disruption modeling codes NIMROD and M3D-C1 need some model for the runaway electrons and how the effect the MHD evolution of the disrupting plasma. We describe relatively crude models of runaway electron dynamics that we have coupled with the MHD codes to start to understand their effects on the MHD. The longer term plans is to couple our codes with more complete kinetic models being developed within the SCREAM SciDAC.

4.1 Reduced Runaway Electron Model in NIMROD

U. Wisconsin-Madison

When the thermal quench of a disruption cools the electron species, the resulting large electric field can exceed conditions for collisional electron force-balance along the magnetic field. This accelerates the most energetic, least collisional part of the distribution, and the ensuing relativistic runaway beam of electrons then carries current. Besides the damage that this beam can cause, it lengthens the timescale of the current quench and affects the current density distribution. Most theoretical studies of REs focus on their kinetic physics, yet the interaction with the overall discharge evolution and the influence on macroscopic instabilities is also important. Fully kinetic descriptions of REs for macroscopic evolution are desirable, but they will be computationally challenging, particularly for spatially 3D modeling. As a practical step, we are implementing a variant of a fluid model that has been used by Helander, et al., [Phys. Plasmas 14, 122102 (2007)] and by Cai and Fu [Nucl. Fusion 55, 022001 (2015)] for studying macroscopic stability with RE effects. A similar model has also been implemented in the reduced-MHD codes EXTREM [Matsuyama, et al., Nucl. Fusion 57, 066038 (2017)] and JOREK [see Bandaru, et al., Phys. Rev. E 99, 063317 (2019)]. The REs are described by cold-beam drift kinetics, where the large parallel speed models relativistic propagation, and perpendicular flow is governed by drifting. The REs are then described by a fluid continuity equation, and their electrical current is coupled into the Ohm's law for macroscopic dynamics by considering the RE population to be free of electrical resistivity. The rest of the system of equations is the standard extended-MHD system, except that the parallel force density on the runaway electrons appears in the flow velocity equation. This model is summarized by Eqs. (1-3).

$$\frac{\partial n_r}{\partial t} + \nabla \cdot \left(n_r v_{r,\parallel} \hat{\mathbf{b}} + n_r \mathbf{v}_{r,\perp} \right) = S_r \quad (1)$$

$$\mathbf{E} \cong -\mathbf{V} \times \mathbf{B} + \eta \left(\mathbf{J} + e n_r v_{r,\parallel} \hat{\mathbf{b}} \right) \quad (2)$$

$$\rho \left(\frac{\partial}{\partial t} \mathbf{V} + \mathbf{V} \cdot \nabla \mathbf{V} \right) = e n_r E_{\parallel} \hat{\mathbf{b}} + \mathbf{J} \times \mathbf{B} - \nabla p - \nabla \cdot \Pi \quad (3)$$

Postdoctoral associate Dr. Ge Wang started implementing this model in NIMROD during the previous reporting period. He used a least-squares approach for the flux terms in Eq. (1) to improve numerical stability when computing with large parallel RE speeds. Before his term ended in December, he had also

started to implement the source term in Eq. (1), which is described in the Matsuyama and Bandaru papers. Graduate student Alex Sainterme has since taken the lead on this development effort. He has completed the implementation of the source term and has developed an analytical solution for spatially decoupled evolution, which he is using for verification. He has also implemented an outer iteration loop to converge the advance of magnetic field and the evolution of RE density with consistent temporal centering.

4.2 Reduced Runaway Electron Model in M3D-C1

PPPL

Chen Zhao and Chang Liu have led the effort to incorporate the runaway electron model given by Eq. (1)-(3) into M3D-C1, which has led to two publications this year.[4-5]. In these papers, they consider the interactions of runaways on low mode number tearing modes in tokamaks. The fluid runaway electron model has been implemented to the M3D-C1 code. To benchmark, they reproduced the MHD linear tearing mode results (with runaway electrons) in a circular cylinder presented in previous analytic studies [Helander, P., et al, Phys. Plasmas 14 144102 (2007)] and have extended them with a numerical eigenvalue calculation. It was found that the low mode number tearing mode has a rotation caused by the MHD - runaways interaction, and the perturbed toroidal current scale length is much smaller with runaways than without and decreases as the runaway speed increases. This model has since been extended to include the runaway source term [J. Connor and R. Hastie 1975 *Nucl. Fusion* **15** 415] and is being applied to model some DIII-D shots in which significant runaways were generated.

5.0 Additional Experimental Validation

Validation of our models with existing experiments is an essential step in model development and application to ITER. We list three areas in which we have made progress in validating certain aspects of our models.

5.1 Validation studies with CTH

Tech- X

We are collaborating with professors Jim Hanson and Dave Maurer at Auburn University in using NIMROD to simulate disruptions in the Compact Toroidal Hybrid experiment (CTH). Experimentally they observe that the application of sufficient external 3D magnetic fields allows operation with a edge safety factor less than 2 (a hard operation limit in axisymmetric tokamaks). However, in these low safety factor (low-q) discharges with small 3D external fields, disruptions often result during the current decay after peak current. Disruptions are suppressed in low-q discharges with a high degree of 3D shaping. Eric Howell, is helping advise postdoctoral researcher Omar Ortiz, at Auburn University on using NIMROD to investigate the dynamics of these low-q disruptions.

5.2 Validation studies with JET

HRS Fusion

M3D-C1 simulations were carried out of asymmetric wall force in JET asymmetric vertical displacement event (AVDE) disruptions. The main physics result is the wall force resulting from an AVDE is quenched when the current quench time τ_{CQ} is less than the magnetic wall penetration time τ_{vde} . The simulations are being done with several values of τ_{CQ}/τ_{vde} . Simulations with the experimental $\tau_{CQ}/\tau_{vde} = 0.005s$ are in progress. Shorter values of τ_{CQ}/τ_{vde} were used to allow shorter runs, and to see whether there is a τ_{CQ}/τ_{vde} scaling of the wall force, where γ is the effective growth rate of the MHD instability causing the asymmetry. The runs were initialized with data from shot 71985 which experienced an AVDE disruption. The current in the simulations was driven with a time dependent electric field, to give control over the current quench time, while keeping γ fixed.

The wall force can be approximated with the Noll force F_N , where B is the magnetic field strength, I is the toroidal current, Z is the vertical displacement, and δI is the rms amplitude of the toroidal variation of I . This can be compared in the JET data and simulations, which are in good agreement. This is shown in Fig. 11.

A main new result is that the wall force is produced by a (1,1) external kink mode, which is destabilized when the vertical displacement event (VDE) causes magnetic flux to be scraped off at the edge of the plasma. The causes q to drop at the plasma edge. When q drops to $q = 1$, the instability is triggered. It was found that when the current quench time $\tau_{CQ} < \tau_{vde}$, q does not drop to 1. This causes a weaker instability, which produces a weaker wall force. This is seen in Fig.12.

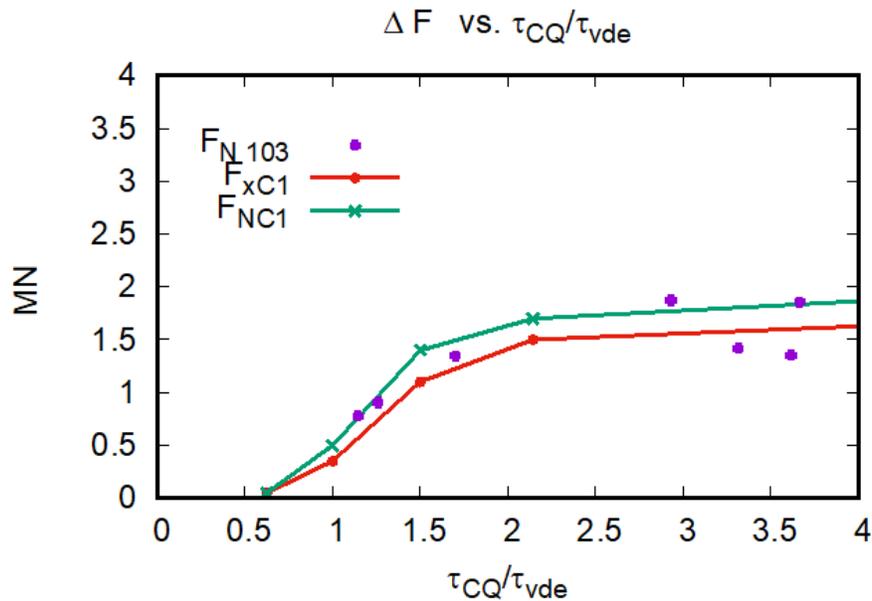


Fig.11 JET data and M3D-C1 simulations of Noll force., which are in good agreement. When the current quench time is small, the force vanishes. The CQ time is normalized to the VDE time. Also shown is M3D-C1 wall force, which agrees well with the Noll force.

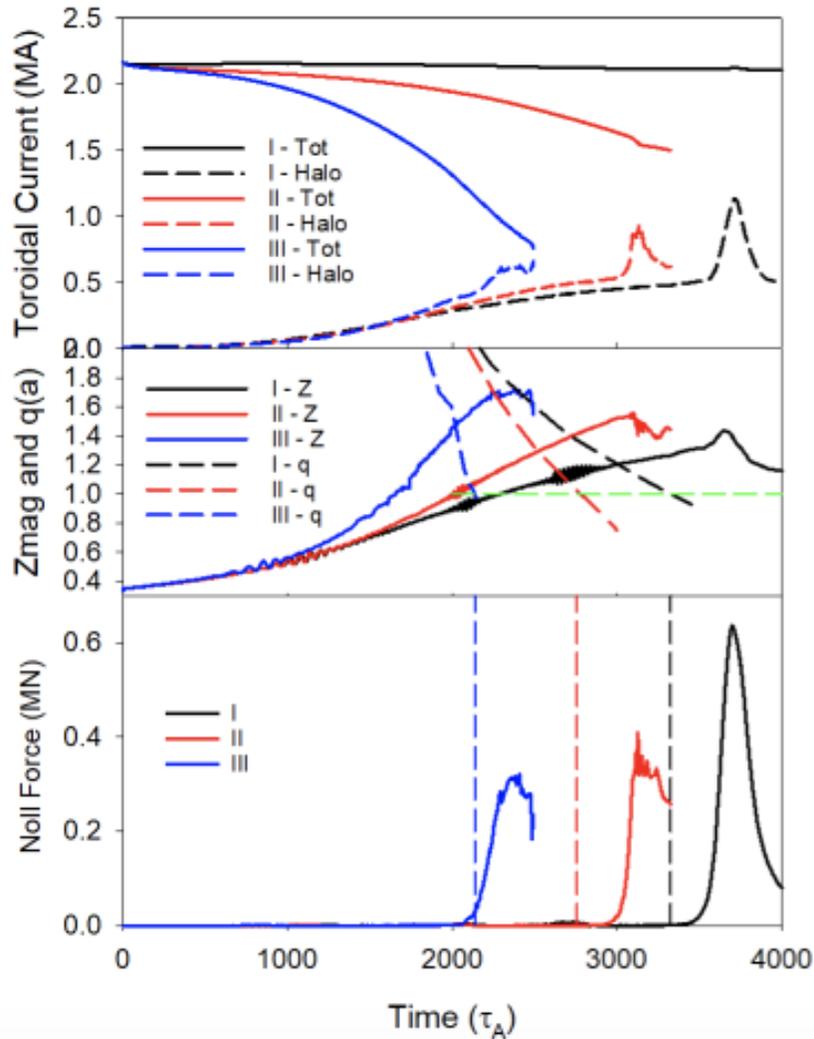


Fig. 12 shows the relation between the edge q and the Noll force for three cases with different CQ times. The upper panel shown the toroidal current and halo current as a function of time for the three cases. The second panel shows the vertical displacement and q . When $q = 1$, the Noll force is produced, shown in the bottom panel. The Noll force and halo current are smaller, the faster the current quench.

5.3 Validation studies with MST

U. Wisconsin-Madison

Graduate student Brian Cornille is simulating tokamak disruption physics in the Madison Symmetric Torus (MST), and he is conducting development work that will contribute to NIMROD disruption modeling. The simulation effort is in coordination with MST experimental research. Cornille had previously applied 0-beta NIMROD simulations to an MST tokamak campaign, which investigated the effects of resonant magnetic perturbation (RMP) on RE confinement. The effects of RMP on magnetic topology shown by Cornille's simulations provide a plausible explanation for the experimental observations of RE de-confinement, and this is an important part of a recent publication [6]. Cornille is now working with Matt Beidler of Oak Ridge National Laboratory to improve the coupling of NIMROD

with the KORC code [Carbajal, et al., Phys. Plasmas 25, 042512 (2017)]. Cornille will use KORC computations to more thoroughly investigate RE de-confinement in MST. He is also conducting finite-beta simulations to support a new MST campaign on tokamak disruption onset.

6.0 Other Physics Studies

6.1 Peeling-Ballooning Modes and Ideal Stability

U. Wisconsin-Madison

Professor Ping Zhu's group previously conducted stability analysis of NSTX equilibria with lithium-conditioning and found that enhanced resistivity due to the increased impurity level can provide a stabilizing effect on low-n edge localized modes [Banerjee, et al., Nucl. Fusion 24, 054501 (2017)]. Recent work considers the intermediate-n peeling-ballooning (PB) instabilities with linear stability analysis of EAST high-confinement mode equilibria using NIMROD two-fluid calculations [7]. The new modeling provides better insight into the physical mechanism behind the beneficial effects of impurities on pedestal stability, but it also shows that the stabilizing effect on PB instabilities in the EAST tokamak is weaker than that found in NSTX.

Professor Zhu's group has also performed an evaluation of linear ideal-MHD stability of the China Fusion Engineering Test Reactor (CFETR) base-line scenario for various first-wall locations. The initial-value code NIMROD and eigenvalue code AEGIS are employed in this analysis. Despite the distinctly different approaches in modeling the scrape-off-layer (SOL) region, the dominant growth in each of the low-n ($n = 1-10$) modes is consistent. The higher-n modes are dominated by ballooning modes and are localized in the pedestal region, while the lower-n modes have more prominent external-kink components and broader mode profiles. The influences of plasma profiles and wall shape are examined using NIMROD. The low-n ideal MHD instabilities in the presence of a resistive wall are studied using AEGIS. For the designed first-wall location, the $n = 1$ resistive wall mode (RWM) is found to be unstable, but it can be stabilized by uniform toroidal rotation above 2.9% of the core Alfvén speed.

6.2 A New Interpretation of Sawtooth Phenomena in Tokamaks

PPPL

The ubiquitous sawtooth phenomena in tokamaks are so named because the central temperature rises slowly and falls rapidly, similar to the blades of a saw. First discovered in 1974, it has so far eluded a theoretical explanation that is widely accepted and consistent with experimental observations. In Ref. [3] we propose a new theory for the sawtooth phenomena in auxiliary heated tokamaks, which is motivated by our recent understanding of "magnetic flux pumping." In this theory, the role of the $(m, n) = (1, 1)$ mode is to generate a dynamo voltage, which keeps the central safety factor, q_0 , just above 1.0 with low central magnetic shear. When central heating is present, the temperature on axis will increase until at some point, the configuration abruptly becomes unstable to ideal MHD interchange modes with equal poloidal and toroidal mode numbers, $m = n > 1$. It is these higher order modes and the localized magnetic stochasticity they produce that cause the sudden crash of the temperature profile, not

magnetic reconnection. Long time 3D MHD simulations with M3D-C1 demonstrate these phenomena, which appear to be consistent with many experimental observations. This work was also presented as an invited talk at the 2019 APS-DPP meeting.

7.0 Physics Code Improvement and Optimization

In addition to these improvements in the communication and sparse matrix routines, Section 6.1 is an effort to reduce communication on the NIMROD side, Sections 6.2-6.4 describe improvements in the local ablation codes, and Section 6.8 is an effort to make the M3D-C1 solver use fewer iterations.

7.1 NIMROD Refactoring and Vacuum region basis functions

Tech-X

A refactoring of the NIMROD code using the abstraction concepts from Fortran 2008/2013 is progressing. The original Fortran 90 code was written in an abstract way minimizing the need for a complete rewrite. The new code is more flexible and will enable such CTTS-related features such as better performance through mixed parallelism. The abstract NIMROD branch has been moved to a repository on gitlab which enables modern development practices suitable for a distributed team such as continuous integration testing, code review and merge requests. Work is underway, lead by Jacob King, to understand the best programming paradigm for using GPUs. The use of OpenACC with NIMROD finite-element assembly kernels has shown the potential for good GPU acceleration with the PGI compiler and Nvidia hardware. This work used simple data structures and in the future we plan to understand how to use OpenACC with the full NIMROD data structures.

U. Wisconsin-Madison

Brian Cornille is investigating different numerical basis functions for the meshed vacuum region that is used in NIMROD VDE computations. He has implemented an expansion of a scalar magnetic potential and will consider the use of edge finite elements for vector potential. If successful, this aspect of his work will improve the code's treatment of boundary corners and allow elimination of most of the degrees of freedom in the meshed vacuum region. The latter is expected to improve the efficiency of the magnetic advance with coupled subdomains for inside and outside the resistive wall.

7.2 M3D-C1 code development and Improvements

7.2.1 Adapting M3D-C1 for stellarator simulation

PPPL and HRS Fusion

Tokamaks are inherently at risk of disruptions unless parameters are carefully kept away from stability boundaries. Stellarators on the other hand appear disruption free. It is worthwhile to be able compare tokamaks and stellarators using the same code. The M3D-C1 equations are expressed in a cylindrical coordinate system (r, θ, z) . A stellarator coordinate system can be given in VMEC coordinates (ρ, ψ, θ) , where ρ , ψ , and θ . It appears possible to modify the M3DC1 finite element discretization to use element based on VMEC

coordinates, while retaining the formulation of the equations in cylindrical coordinates. The implementation is being carried out by Yao Zhou

7.2.2 Improving Robustness of Pellet Simulations

General Atomics

A variety of upgrades have been made to M3D-C1 in order to improve the robustness of pellet-injection simulations, particularly the tendency for negative temperature regions to form. We added minimum densities and temperatures for which the KPRAD impurity routines are advanced. Internal evolution of the density and temperature was also added to the impurity/radiation evolution, along with adaptation of the KPRAD time step with large changes in the thermal energy, in order to avoid temperature overshoots. Finally, a new temperature-dependent thermal-conductivity model was added.

Collectively, these changes allowed M3D-C1 pellet-injection simulations to proceed farther than ever before, well into the thermal quench.

7.2.3 New Capabilities for Toroidal Planes

General Atomics, RPI, PPPL

We began to exploit the ability for M3D-C1 to have non-uniform toroidal planes in order to perform simulations with more toroidally localized impurities. The Fourier-mode diagnostic was upgraded in order to account for the possibility of non-uniform planes, permitting accurate tracking of the kinetic and magnetic energy harmonics in such simulations. In addition, a new capability was added to change the number of planes in a restarted M3D-C1 run. This allows for great computational saving by running with relatively few planes until instability onsets, then increasing the number of planes to resolve the MHD activity accurately. General

7.3 Improvements to the Lagrangian Particle Code

Stony Brook U.

7.3.1 Massively parallel Lagrangian particle pellet / SPI code

The development of massively parallel Lagrangian particle (LP) code for the ablation of pellets and SPI fragments has been fully completed. The work started in 2019, and the LP code data structures were redesigned to make them compatible with P4EST (Parallel Forest of Octrees) software library, developed in the past within an ASCR-funded project of Omar Ghattas. P4EST library enables a dynamic management of a collection of adaptive octrees on distributed memory supercomputers and scales to hundreds of thousands of processor cores. The construction and search of quad- and octrees is one of the most time-consuming tasks in the LP code, and the improvement of our previous Open-MP-parallelized libraries by P4EST led to the increase in parallel performance and accuracy. In 2020, we added all pellet ablation physics models to the new LP code and performed extensive verification simulations. For typical production simulations of pellet ablation using two hundred cores, we achieved 72% of ideal scalability. The new massively parallel LP code is now used for all physics simulations.

7.3.2 Cross-code comparison and resolution of disagreements between Frontier and Lagrangian Particle Codes

Simulation of the ablation of cryogenic pellets in tokamaks with the resolution of all relevant physics processes is a challenging numerical task due to sensitivity to small changes. The pellet ablation rate, computed as the amount of energy of hot electrons reaching the pellet surface divided by the heat of vaporization, is a fraction of two small quantities: both of these values are very small compared to the energy absorbed by the cloud. Since small changes in the ablation cloud can significantly change the ablation rate, cross-code comparison of FronTier (FT) and Lagrangian particle (LP) codes is a critical step in our verification program.

In 2019, we observed significant disagreements between of FronTier and Lagrangian particle codes for simulations of pellets in magnetic fields. While the results agreed reasonably well in zero magnetic field, properties of ablation channels and the ablation rates became different in both codes with the increase of the magnetic field. All these disagreements were completely resolved in the 1st half of 2020. Main improvements in FronTier involved algorithms for the explicit tracking of the ablation cloud – ambient plasma interface. In the Lagrangian particle code, the implementation of adaptive algorithms for kinetic models were significantly improved as well as some aspects of solvers. Now both codes are in full agreement for the entire parameter space of neon and deuterium pellets. Figure below shows the agreement between FronTier and Lagrangian particle codes in simulations of neon pellets.

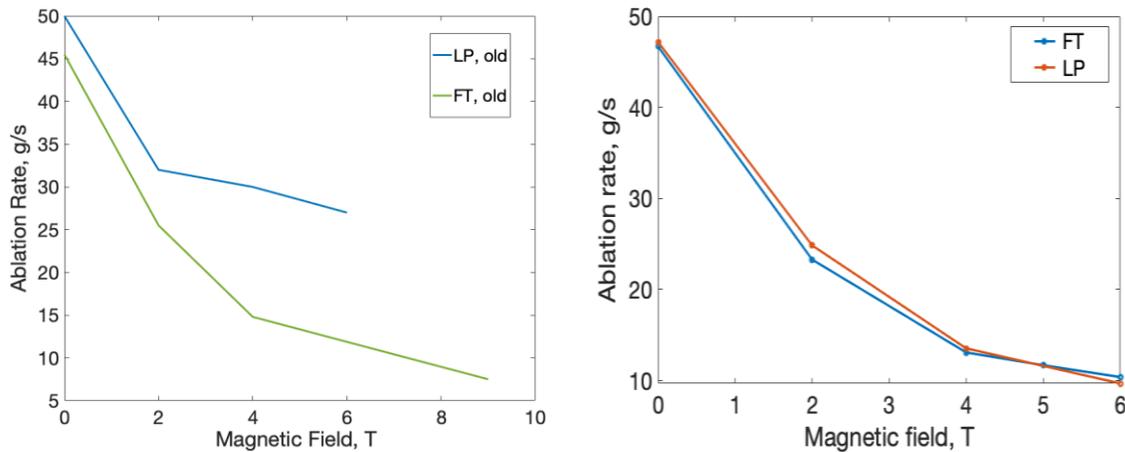


Figure 13. Simulation of neon pellet ablation rates in magnetic fields. Left plot: disagreement between 2D axisymmetric FronTier code and 3D Lagrangian particle code observed in 2019. Right plot: current versions of the codes are in complete agreement.

In addition, both codes have been extensively tested in the spherically symmetric approximation. The 1D FronTier code is in good agreement with the 3D LP code with spherically symmetric initial conditions and both codes agree with analytic scaling laws.

After the successful completion of the verification program, the Lagrangina particle code is being used for most of physics results since it resolves important 3D effects (such as grad-B drift) not available in the axisymmetric FronTier pellet code.

7.3.3 New Adaptive Algorithms for LP-based SPI code

We have implemented grad-B drift algorithms for multiple SPI fragments in the LP code. Current work focuses on implementation and verification of a plasma cooling model that enables SPI simulation fully within the LP code (the plasma cooling model is not needed in a coupled simulation where the plasma temperature is obtained from a tokamak MHD code NIMROD or M3D-C).

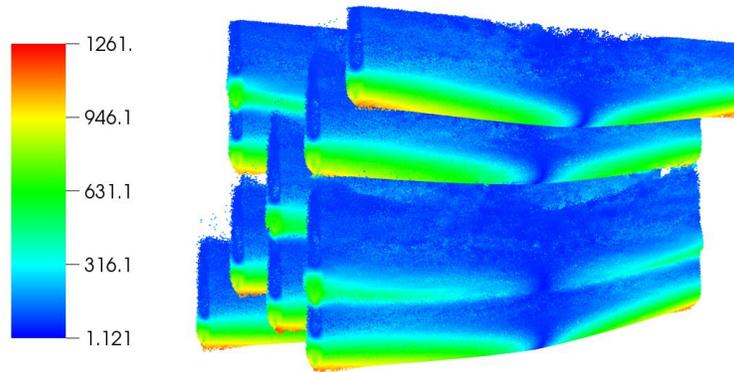


Figure 14. Fragment of SPI simulation with grad-B drift effect.

7.4 3D Sparse LU and triangular solves

Tech-X and LBL

Jacob King is working with Sherry Li and Yang Liu of LBL we have coupled NIMROD to the beta “version 7” of SuperLU_DIST. This version enables additional parallelism through 3D factorization of the sparse matrices used in as block Jacobi preconditioner for each Fourier mode during NIMROD’s 3D nonlinear solves. Note that 3D factorization here refers to the processor grid layout while 3D refers to three spatial dimensions with regards to NIMROD.

We are presently working with Yang Liu to apply a machine learning algorithm to automatically tune the best parameters for the 3D process grid decomposition.

8.0 Solver Improvements

LBL

8.1 Introduction

Samuel Williams (LBL PI), Sherry Li, Nan Ding, and Yang Liu are improving the sparse matrix solvers used in CTTS. M3D-C1 and NIMROD solve ill-conditioned systems of equations using SuperLU preconditioned GMRES. In the worst case, the system of equations is structured as block diagonal and factored for each GMRES solve. Then, on each GMRES iteration, the factored blocks are used as a preconditioner necessitating a pair of sparse triangular solves (SpTRSV) for the upper and lower triangular matrices.

Although one can partially amortize factorization time by factoring once every N solves, one cannot amortize triangular solve time. Thus, SpLU and SpTRSV performance and scalability is critical to M3D-C1 and NIMROD performance and scalability. To that end, in FY20, we explored four techniques to improve SpLU and SpTRSV performance. Concurrent with these activities we collaborated with both the M3D-C1 and NIMROD teams to 1) ensure we remained cognizant and responsive to their computational needs, 2) integrated and evaluate our solutions in production environments, and 3) aide the application developers in using GPUs and planning for the future. This was accomplished via monthly application-specific telecons.

8.2 M3D-C1/NIMROD parameter tuning

Both M3D-C1 and NIMROD have multiple runtime parameters that can significantly affect the solve time and memory usage. Manually tuning these parameters quickly become untrackable as each simulation requires long runtime or multiple compute nodes. Here we adopted a machine-learning-based autotuning framework called GPTune [GPTUNE] to auto detect the best performing parameters with only a handful of expensive simulations. For M3D-C1, the parameters considered are [ROWPERM, COLPERM, p_r , NSUP, NREL, nblock]. Here ROWPERM, COLPERM, p_r , NSUP, NREL are SuperLU_DIST parameters representing the type of row and column permutations, 2D process decomposition, the maximum supernode size and the relaxed supernode size, and nblock is a M3D-C1 parameter representing the number of Jacobi blocks. For NIMROD, the parameters are [ROWPERM, COLPERM, NSUP, NREL, nxbl, nybl], where nxbl and nybl are NIMROD parameters representing the tile size for matrix assembly. In addition, we used the number of time steps t to differentiate a high-fidelity simulation (i.e., large value of t) from a low-fidelity simulation (i.e., small value of t). In the table below, we allow running either 80 high-fidelity simulations (i.e., Single-task), or 20 high-fidelity ones plus 60 low-fidelity ones (i.e., Multitask). The table lists the minimum runtime and total tuning time in seconds. Running a multitask tuning can attain a similar optimal configuration to a single-task tuning while significantly reducing the total tuning time. Moreover, the optimal configuration is 20% faster than the default configuration. In FY21, we plan to further exploit the idea of the multi-armed bandit strategy to further improve the multi-fidelity tuning of M3D-C1 and NIMROD.

	M3D-C1		NIMROD	
	minimum	total	minimum	total
Single-task	11.19	12310	112.7	14710
Multitask	11.17	7797	112.8	9559

8.3 One-Sided Communication

Distributed memory SpTRSV exhibits irregular parallelism profuse with complex data dependencies expressed as fine-grained (small message) communication that can be interpreted in computer science terms as a directed acyclic graph (DAG). Traditional MPI send/rcv operations incur high overhead and thus, for such small messages, rarely attain high bandwidth. To rectify this, we implemented a synchronization-free task queue to manage the producer-consumer pairs within the DAG and leverage MPI’s one-sided communication capabilities. We leverage two forms of one-sided communication: one-sided puts [SIAMPP20] and one-sided gets. These two implementations can help us better understand the one-sided performance in different paradigms and provides a pathway to not only multi-GPU implementations but also one-sided communication on other architectures.

In one-sided, get-based communication, the sender first puts a counter to the receiver, the receiver keeps checking the counter until the value is valid. That procedure can be considered as *MPI_Isend*. The receiver then gets the data from sender’s buffer to its local buffer. Note that on Cori KNL and Cori HSW, *MPI_Get* is non-blocking. That is to say, when *MPI_Get* returns, there’s no guarantee the data is completely written into the local receiver buffer. *MPI_Win_flush_local* follows right behind *MPI_Get* is used to ensure the data completion. This procedure can be considered as emulating *MPI_Recv*. As a result, we need to perform three MPI operations, *MPI_Put*, *MPI_Get* and *MPI_Win_flush_local*, for every single point-to-point communication.

In the one-sided ‘put’ version, the sender puts the data directly to the receiver’s buffer with an extra payload representing a checksum of the data. The receiver calculates the checksum of its buffer until the newly calculated checksum equals the payload. As a result, users need to implement their own checksum (or other methods) to ensure data arrival completion. foMPI is an optimized one-sided MPI library for Cray Aries developed by ETH. It can be used to replace the standard Cray MPI library that was optimized for two-sided MPI. The following table shows our foMPI-based one-sided SpTRSV speedup compared to the two-sided (Cray MPI) version on Cori KNL for two different M3D matrices. Our one-side implementation always outperforms the two-sided baseline while our ‘get’ based variant outperforms our ‘put’ variant.

	nnz	Speedup using 256 processes (16x16)		Speedup using 1024 processes (32x32)	
		One-sided put	One-sided get	One-sided put	One-sided get
A30	13746850	1.26x	1.28x	1.13x	1.20x
s1	106015596	1.49x	1.58x	1.49x	1.61x

We constructed a performance model in order to understand SpTRSV performance and quantify the potential performance benefits without running the SpTRSV code. We first conduct a critical path analysis and refine the critical path based on the process decomposition. We then model the computation time and communication time separately based on the critical path. Figure 15 shows computation upper and lower bounds for matrix A30. The gap between the modeled lower bound and the measured data indicates the potential performance benefit from further optimizations that improve task placement. We plan in FY21 to expand the model’s applicability for the guidance for a smarter process decomposition.

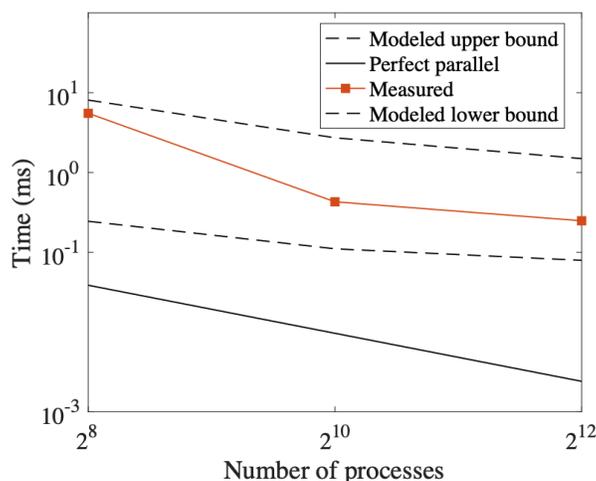


Figure 15 Comparison of modeled and actual performance of SpTRSV

8.4 GPU Acceleration

Although single-core CPU performance has stalled, aggregate multicore CPU performance continues to improve through the integration of more cores and more memory bandwidth. Nevertheless, GPUs can deliver as much as 5x more compute and bandwidth per socket making them an attractive choice for HPC. In FY20, we continued the development of the GPU implementation of SpTRSV and factorization in SuperLU_DIST. For both SpTRSV and factorization, we extended our CUDA-based implementation for NVIDIA GPUs to a HIP-based implementation for AMD GPUs thereby enabling portability to both NERSC’s forthcoming Perlmutter and OLCF’s forthcoming Frontier systems. For SpTRSV, we refactored the previous single-GPU implementation in preparation for the multi-GPU implementation from the following perspectives: 1) We removed the need for managed memory access in SpTRSV and used explicit device-host transfers. We ported the CPU communication data structures from C++ to C in order to access them from GPU. 2) For the GPU kernel, we kept the same parallelization strategy for the computation, i.e., we assign one column to one thread block and use the entire thread block to handle the GEMV inside one thread block. We also utilize GPU atomics to update the solution vector. In addition, we allocate extra thread blocks to handle row-wise reduction and column-wise broadcast communication operations. The following table shows the factorization and SpTRSV time and

flop performance using 5 CPU cores and 1 V100 GPU on a Cori GPU node for a Jacobi block $s1_mat_0_X$ in a typical 3D M3D-C1 simulation. Here X denotes the size of the Jacobi block, it is apparent that for both factorization and SpTRSV, the single GPU implementation performs better for larger blocks. In FY21, we plan to exploit the possibility of using larger Jacobi blocks for better utilization of the GPU compute power.

	Factorization				SpTRSV			
	5 Intel Gold 6148 cores		1 NVIDIA V100		5 Intel Gold 6148 cores		1 NVIDIA V100	
	Time	GFLOP/s	Time	GFLOP/s	Time	GFLOP/s	Time	GFLOP/s
$s1_mat_0_126936$	1.51E+00	61.31	1.97E+00	47.00	1.10E-02	9.62	7.30E-03	14.50
$s1_mat_0_253872$	7.26E+00	101.28	7.61E+00	96.62	3.62E-02	11.64	2.01E-02	20.96
$s1_mat_0_507744$	4.77E+01	120.54	4.79E+01	119.93	1.32E-01	12.47	7.79E-02	21.05

NVSHMEM is a one-sided communication library that implements the OpenSHMEM specification for communication across NVIDIA GPUs connected by different types of interconnects that include PCI-E, NVLink, and Infiniband. Our multi-GPU implementation of SpTRSV leverages NVSHMEM to perform GPU-GPU communications in one single CUDA kernel. The NVSHMEM (v1.1.3) version builds on our foMPI and single-GPU implementations. Unlike our efforts on CPU one-sided SpTRSV using foMPI (pure MPI communication), the one-sided GPU design leverages the multiple threads in thread blocks to handle the communication, and uses the NVSHMEM p2p synchronization API to manage the data completion. For each message, we use `nvshmem_quiet()` followed by a `nvshmem_signal()` to inform the receiver that the message is completed. The receiver uses a `nvshmem_wait_util()` to get the signal from the sender. The NVSHMEM p2p synchronization API requires a collective kernel launch using `nvshmemx_collective_launch()`. That collective launch function limits the number of thread blocks that can be concurrently scheduled on the GPU to 80. Therefore, in the current Px1 process decomposition NVSHMEM version, we increase the GEMV/TRSM size to reduce the number of block columns equal to or smaller than 80. In essence, hardware and algorithmic parallelism are tied together and capped at 80. We test matrix A30 on Summit Volta GPU, and we increase the GEMV/TRSM size to 2400 (default is 128). The single GPU version takes 36.1 ms to complete the solve. It takes 28.5 ms when using 6 GPUs on one Summit node. Unfortunately, with such a big super node size, the burden of performance becomes the computation of GEMV/TRSM. In addition, there is also a loss of concurrency, i.e., the number of GEMV/TRSM that can be solved concurrently. Therefore, in FY21 we plan to investigate a

proper multi-GPU design for SPTRSV that is cognizant and compatible with the limits imposed by the one-sided NVSHMEM library.

8.5 Integrating 3D Algorithms

Recently we developed a new communication-avoiding algorithm framework for both sparse factorization and triangular solve in SuperLU_DIST [CA-SuperLU]. We showed that the theoretical communication complexity (both in latency and volume) is asymptotically reduced, the result of which is the ability to effectively use more than 10x more processes in strong scaling experiments. The key idea is to replicate multiple layers of 2D X-Y process grid along the third (Z) dimension, conceptually forming a 3D process grid. The Z-dimension is used to replicate some partial Schur-complement updates, hence reducing communication. We released a Beta version of the “Version-7” code on github for benchmarking purposes. Currently, we are finishing a preprocessing function, which redistributes matrix A and right-hand-side B from the 3D process grid (user input) into the Layer 0 of the 2D process grid that is required by the internal algorithm. The release of this v7.0.0 is imminent. Our next steps are: 1) integrate the GPU code into this new v7.0.0 released code base; 2) integrate the 3D triangular solve and the 2D one-sided triangular solve together into the v7.0.0 code base.

8.6 Software Releases

SuperLU_DIST releases since last report:

- Nov. 12, 2019 Version 6.2.0 [Version 6.2.0](#)
 - Use cmake's FortranCInterface module to automatically handle Fortran-C name mangling. Create a new file superlu_FCnames.h.
 - Several small fixes, see Change Log page.
- Feb. 27, 2020 Version 6.3.0 [Version 6.3.0](#)
 - A number of fixes for GPU code, Dec. 2019 - Jan. 2020
 - Redefine several structures as precision-dependent, so that both real and complex codes can be compiled at the same time.
 - Several small fixes, see Change Log page.
- April 2, 2020 Version 6.3.1 [Version 6.3.1](#)
 - update interface to CombBLAS, include both real and complex.
 - update FORTRAN/ wrapper to include CMake file.

9.0 Meshing and other M3D-C1 Support

RPI

9.1 Mesh Control for Pellet Simulations

The effective simulation of the pellet injections processes requires the construction of properly graded and aligned meshes. The current a-priori and adaptive mesh control procedures need improvement to provide the desired meshes. In the long run what is desired is an adapted mesh defined based on a posteriori mesh correction indicators determined during the analysis. In the short run, a procedure is being put in place where a knowledgeable user provide mesh size information that will drive mesh modification operations to produce a mesh that has a favorable mesh distribution. The manner in which

user will provide the needed mesh size information is by being able to satisfy an API routine that will be called at selected locations in a given initial, most likely uniform, mesh to provide the desired mesh sizes information at that location. This information will be used by a new version of the mesh adaptation procedures that use that API to modify the mesh to satisfy the requested mesh sizes. The API to do this is:

API: *m3dc1_set_adapt_size (r, z, h₁, v_{1r}, v_{1z}, h₂)*

Where r, z are the input coordinates of the point and h_1, v_{1r}, v_{1z}, h_2 represents the mesh size information needed to drive anisotropic mesh adaptation (Figure 16).

r, z : R and Z coordinates of a point that where the mesh size field is to be provided.

h_1 : Desired edge length along the direction v_1

v_{1r}, v_{1z} : Two components of the unit vector for direction corresponding to size h_1

h_2 : Desired edge length along the direction v_2 (perpendicular to v_1)

Note: h_1 and $v_1(v_{1r}, v_{1z})$ are the size and direction for one of the axis (It could be any of the major or minor axis). Followed by h_2 the desired size in the direction perpendicular to v_1 .

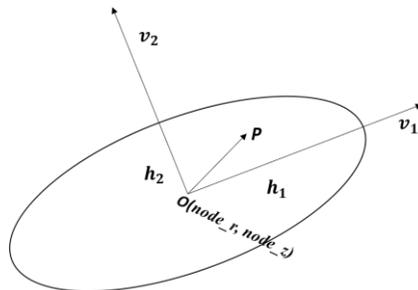


Figure 16. Representation of Mesh Size Field in terms of 2 sizes (h_1, h_2) and two unit vectors (v_1, v_2) at point O. OP is an edge in current configuration that needs to be adapted according to the size of ellipse

9.2 Supporting Mesh Adaptation of the 3D Extruded Meshes

The M3D-C1 3D meshes consist of wedge elements defined by the extrusion of a triangular mesh in the toroidal direction. To be able to better control the mesh discretization errors, it is desirable to provide procedures that can improve the 3D mesh when as an analysis proceeds. To maintain the wedge element only mesh any mesh improvement operation must separate mesh improvements into ones that alter the poloidal plane triangular meshes on all poloidal planes in the same way, and ones that change the number of poloidal planes. A procedure to support a uniform increase in the number of poloidal plane mesh was added and a procedure to adapt the poloidal plan meshes in the 3d mesh is being implemented.

The procedure to increase the number of poloidal planes in a 3D mesh can be applied at any selected restart. New planes are inserted equally as depicted in Figure 17, the 3D distributed mesh is constructed and solution fields are interpolated onto nodes on the newly introduced poloidal plane mesh nodes.

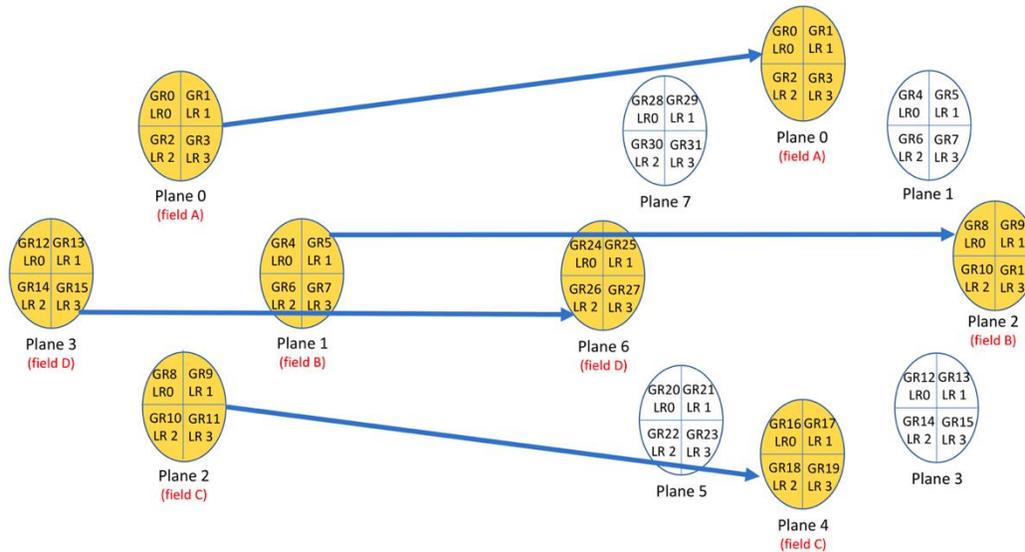


Figure 17. Multiplying # planes in 3D restart by factor 2 (new planes are equally inserted) GR: Global Rank, LR: Local Rank

The procedure to adapt the meshes on the poloidal planes within a 3D simulation is currently being implemented. The key steps in the process include:

- Determine the poloidal mesh size field.
- Adapt the poloidal plane mesh on all poloidal planes and up-date the fields
- Reconstruct the 3D wedge mesh accounting for the mesh adaption.

The procedure for defining the poloidal mesh size field first independently determines the desired mesh size field at the mesh nodes each poloidal plane based on the fields on that plane. Currently this is based on applying the current 2D error indicators. (Improved procedure will be considered in the future.) The single poloidal plane nodal mesh size value that will drive the poloidal plane mesh adaptation is the smallest associated with that node on any of the poloidal planes.

Given the resulting single poloidal plane mesh size field, the current 2D mesh adaptation procedure can be used with an extension that will do the incremental field transfer on each poloidal plane as each mesh modification operation is executed. Once the poloidal plane mesh is adapted and the fields set on each plane, the 3D wedge elements can be reconstructed. The procedures for these steps are currently being implemented. This approach works as long as the field information that needs to be transferred are associated with nodal quantities. If at some point integration point fields also need to be transferred, additional solution transfer procedures would need to be developed.

9.3 Support of a Growing M3D-C1 User Community

RPI is in charge of maintaining the following libraries and programs to support M3DC1

- Mesh generation program with Simmetrix libraries
- Parallel mesh utility programs such as partitioning mesh, merging meshes etc.
- PUMI (Parallel Unstructured Mesh Infrastructure) to provide parallel mesh features to M3DC1
- m3dc1_scorec library to provide interaction between mesh, PETSc solver and M3DC1 Fortran code

As of now, RPI has been actively involved in debugging and maintaining the M3DC1 software on the following machines

- PPPL - Linux clusters
- Princeton - Linux clusters and GPU
- NERSC - Cori and GPU
- Brazilian National Lab of Scientific Computing - Supercomputer SDumont
- General Atomic - Linux clusters

10.0 Publications and Conference Proceedings

1. I. Krebs, F. J. Artola, C. R. Sovinec, S. C. Jardin, K. J. Bunkers, M. Hoelzl, N. M. Ferraro, "Axisymmetric simulations of vertical displacement events in tokamaks: A benchmark of nonlinear MHD Codes", *Phys Plasmas* **27** 022505(2020)
2. C. F. Clauser, S. C. Jardin, N. M. Ferraro, "Vertical Forces during VDEs in an ITER plasma and the Role of Halo Currents", *Nuclear Fusion* **59** 126037 (2019)
3. S. Jardin, I. Krebs, N. Ferraro, "A New Explanation of the Sawtooth Phenomena in Tokamaks, *Phys. Plasmas* **27**, 032509 (2020)
4. C. Zhao, C. Liu, S. Jardin, N. Ferraro, "Simulation of MHD instabilities with fluid runaway electron model in M3D-C1" *Nuclear Fusion* (2020)
5. C. Liu, C. Zhao, S. Jardin, A. Bhattacharjee, D. Brennan, N. Ferraro, "Structure and overstability of resistive modes with runaway electrons", *Phys. Plasmas* **27**, 092507 (2020)
6. S. Munaretto, B. E. Chapman, B. S. Cornille, A. M. DuBois, K. J. McCollam, C. R. Sovinec, A. F. Almagri, and J. A. Goetz, "Generation and suppression of runaway electrons in MST tokamak plasmas," *Nucl. Fusion* **60**, 046024 (2020),
7. X. Lin, D. Banerjee, P. Zhu, G. S. Xu, Y. Ye, Y. F. Wang, H. L. Li, Q. Zang, T. Zhang, and Y. J. Chen, "Stabilizing effect of enhanced resistivity on peeling-ballooning instabilities in EAST," *Plasma Phys. Control. Fusion* **62**, 035001 (2020), and
8. 4) Wenlong Huang and Ping Zhu, "Analytical model of plasma response to external magnetic perturbation in absence of no-slip condition," *Physics of Plasmas* **27**, 022514 (2020).
9. J. Ji, J. Spencer, and E. Held, "Exact irreducible moments of the Landau collision operator in the random velocity moment", *Plasma Research Express*, **2** (2020)
10. N. Bosviel, P. Parks, R. Samulyak, Near-field simulations of pellet ablation for disruptions mitigation in tokamaks, 2020, submitted to *Physics of Plasmas*, arXiv preprint arXiv:2009.08503

11. N. Bosviel, Near-field Models, Algorithms and Simulations of Pellet Ablation in Tokamaks, Ph.D. Thesis, defended in July 2020, Stony Brook University (advisor R. Samulyak)
12. FY2020 FES Theory Milestone Report "Modeling of Fully 3D Vertical Displacement Event Disruptions", <http://m3dc1.pppl.gov/2020TheoryMilestone.html>
13. K. J Bunders and C. Sovinec, "The influence of boundary and edge-plasma modeling in computations of axisymmetric vertical displacement events", to appear in Physics of Plasmas
14. Nan Ding, Samuel Williams, Yang Liu, and Xiaoye S. Li. "Leveraging One-Sided Communication for Sparse Triangular Solvers." In Proceedings of the 2020 SIAM Conference on Parallel Processing for Scientific Computing, pp. 93-105. Society for Industrial and Applied Mathematics, 2020.
15. Yang Liu, Wissam M. Sid-Lakhdar, Osni Marques, Xinran Zhu, James W. Demmel, Sherry Li, "GPTune: Multitask Learning for Autotuning Exascale Applications", (submitted to) PPOPP, 2020.
16. P. Sao, R. Vuduc, X. Li, "A communication-avoiding 3D algorithm for sparse LU factorization on heterogeneous systems", J. Parallel and Distributed Computing (JPDC), September 2019. doi: 10.1016/j.jpdc.2019.03.004