The role of HPC & first-principles simulation in whole-device-modeling.

D L Green[*1,] J M Park[1]
[1] *Oak Ridge National Laboratory*
* Corresponding Author: greendl1@ornl.gov

Oral Presentation Requested: No
Sub-Panel : (C) Whole Device Modeling

Summary

- **Motivation :** The SciDAC program has proven successful at enabling valuable high-risk, compute-intensive investigations. However, transitioning the knowledge and capability these discovery efforts produce to a persistent, production-level whole-device-model for fusion will require modifications to the present program.
- **Approach & Impact :** (**1**) Emphasize the delivery of reduced models derived from HPC discovery efforts; (**2**) Emphasize HPC code availability for benchmarking and validation via inclusion in an integration framework; and (**3**) Focus the utilization of ASCR tools and expertise in any future call for proposals targeting an eventual whole-device-model capability beyond that of just HPC performance and massive-scale computing. Such measures will facilitate progress towards a useful whole-device-model, as well as HPC code validation, and community wide benchmarking.

Some clarification; For the purposes of this paper we refer to all simulations that require stand-alone leadership class computing resources, e.g., "first-principles" or "high-fidelity" simulations, as high-performance-computing (HPC) codes. The point is that some codes require massive compute resources, and here we recommend a possible role for these codes going forward in establishing a community wide integration and whole-device-modeling program.

Motivation : Fusion is a complex problem, where predictive capability in a single niche area can take many years to develop. It is also true that the multi-scale nature of the plasma can be addressed by either a single first-principles simulation that resolves everything from the smallest scales up, or by breaking the problem into manageable pieces and coupling those approaches together. The latter "integrated" approach is best suited to understanding a single part of the problem, then taking that understanding or predictive capability and integrating it within a larger collection of such physics models. The full first-principles (at least to the Debye length and plasma frequency scales), extreme-scale simulation approach is used for example in cases where separation of spatial and / or time scales is not possible. And while useful for discovery, the is not practically useful for most of the use cases of a whole-device-model, e.g., design point identification or scenario design. And really, the first-principles approach, or HPC codes that do one piece of the problem at high fidelity, will always need further integration as the scope extends beyond that of the code itself (e.g., to include materials, neutrons, engineering, cost, risk, etc). The question we address here is how the use of HPC codes may usefully contribute to a WDM or larger integration effort.

We see WDM as a subset of a larger integration effort. Traditionally the idea of a WDM has been an integrated simulation involving components that are computational "lightweight". However, by starting with an approach to integration that is agnostic to the computational weight of its components (e.g., frameworks that facilitate integration at any scale, specifically the Integrated Plasma Simulator [IPS]), we need not limit ourselves to such small scales. Certainly a computationally lightweight WDM should be one of the major elements of a larger integration effort, but there are both other applications of integrated simulation, and applications of a WDM beyond the lightweight components to consider.

In expanding the lightweight WDM idea to what we will call "medium-fidelity" (beyond the 0-D WDMs we have now, where each component requires a varying number of compute cores ranging from 10's of cores to 100's of cores, or to even a few 1000's of cores), the metrics that define usefulness are likely turn-around time (queue wait time included) and component code robustness. We envisage such medium-fidelity time dependent simulations with reduced models for core transport, the pedestal, and actuators are within reach now, with the limiting factor being the availability of robust, medium-fidelity models for all components of a WDM, e.g., models for the edge, engineering components, neutrons, etc. In addition to turn-around time, what's probably important is to match the level of physics fidelity among components, rather than their computational resource requirements; especially given the load balancing and resource management capabilities of frameworks like the IPS. Such load balancing and resource management, whereby physics components might even be able to be run concurrently, is capability we have now, meaning minimum wall clock time. This type of WDM calculation is by no means computationally lightweight, but could certainly be run in less than 24 hours, i.e., in order to inform experimental planning for example. We expect steady progress to a medium-fidelity WDM in this way and propose the following role for HPC to complement such an effort.

Approach : We describe minor modifications to the SciDAC program where a portfolio of high-risk, **discovery** focused efforts proceed in parallel to a persistent community-wide, **product development / useful predictive capability** focused integration effort - one part of which is a WDM. Based on the experience of the authors in the day-to-day issues of trying to achieve a practical integration of multiple physics components of all compute scales from across the community, we expect the following steps will help facilitate progress in a useful WDM.

1. SciDACs deliver reduced models : Increase the emphasis the discovery SciDAC's place on producing reduced models based on their HPC / first-principles studies, i.e., focus on discovering, developing, and delivering robust parameterizations of the physics of interest as a reduced model. An excellent example here is the running of the HPC GYRO turbulence code to produce the TGLF transport model for use in modular core transport solvers.

2. Make HPC codes available for benchmarking / validation within the integration framework : The use of lower compute requirement components within a WDM capability often means the physics fidelity is less than ideal for a given use case. One use for first-principle simulations would be to validate the accuracy of the reduced model. The most practical approach to this, as well as for the use cases of code-to-code benchmarking, single code validation by external analysts, and any general code-to-code coupling (HPC or otherwise), is to emphasize the availability of HPC codes within a common HPC capable framework - while not their direct inclusion in a WDM.

3. Utilize ASCR tools & expertise beyond HPC performance : Clearly, a whole-device-model will not make complete and optimal use of an ExaScale computing system, whereas a single first-principles simulation might. This is not to say that an integration approach wouldn't require large scale computing, it certainly would, just not in a way that can produce a single strong scaling curve showing optimal utilization. What an integration based whole-device-model would require from ASCR expertise is advanced coupling methods, uncertainty quantification, data management, workflow, resource and scheduling management for the collection of disparate models that will ultimately compose an integrated whole-device-model. This is in contrast to the first-principles approach where the dominant requirement from ASCR is computational performance (speedup) expertise. We assert here a strong role for ASCR in both approaches, but it needs to be recognized that these roles are different, especially when issuing calls for proposals for integrated efforts.

Impact : By implementing a program with closer ties between the discovery focused SciDACs / base-theory programs with a whole-device-modeling & integration focused effort, the previous HPC onus on integrated efforts can be left to the discovery aspects of the program, while ensuring places for the products of these efforts within the larger integration.