Case Study: FUSION SIMULATION PROGRAM

Planning for FSP Computing Infrastructure Needs:
Exploring Possible NERSC Collaborations

William Tang & Martin Greenwald
(on behalf of the FSP Planning Team)

Large Scale Computing & Storage Requirements for Fusion Energy

FES/ASCR/NERSC Workshop
Washington, DC

3-4 August 2010

FSP MISSION & VISION

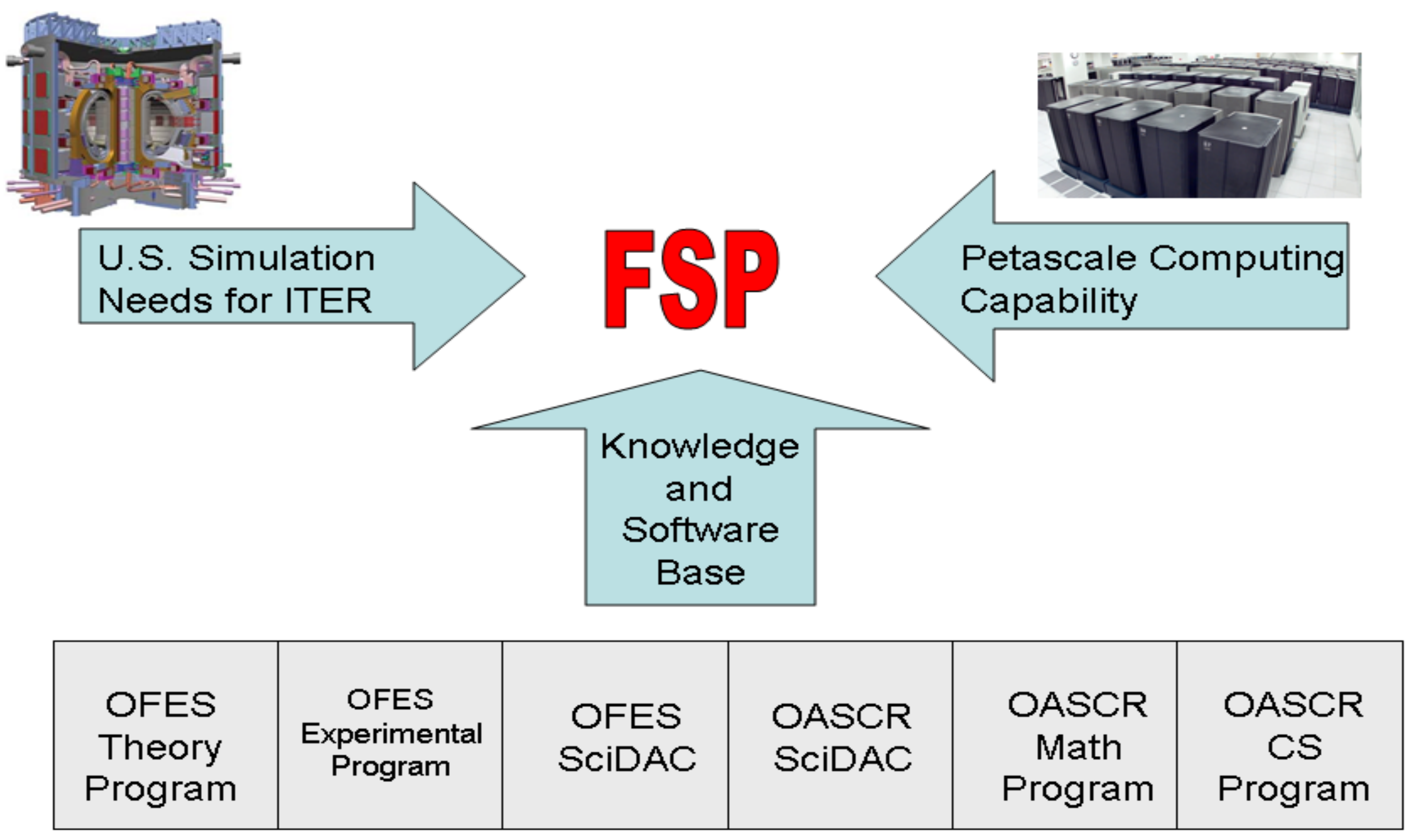
national FSP web-site [*http://www.pppl.gov/fsp/*]

VISION: ***The Fusion Simulation Program (FSP) will enable scientific discovery of important new plasma phenomena with associated understanding that emerges only upon integration. It will provide a predictive integrated simulation capability for magnetically-confined fusion plasmas that are properly validated against experiments in regimes relevant for producing practical fusion energy.***

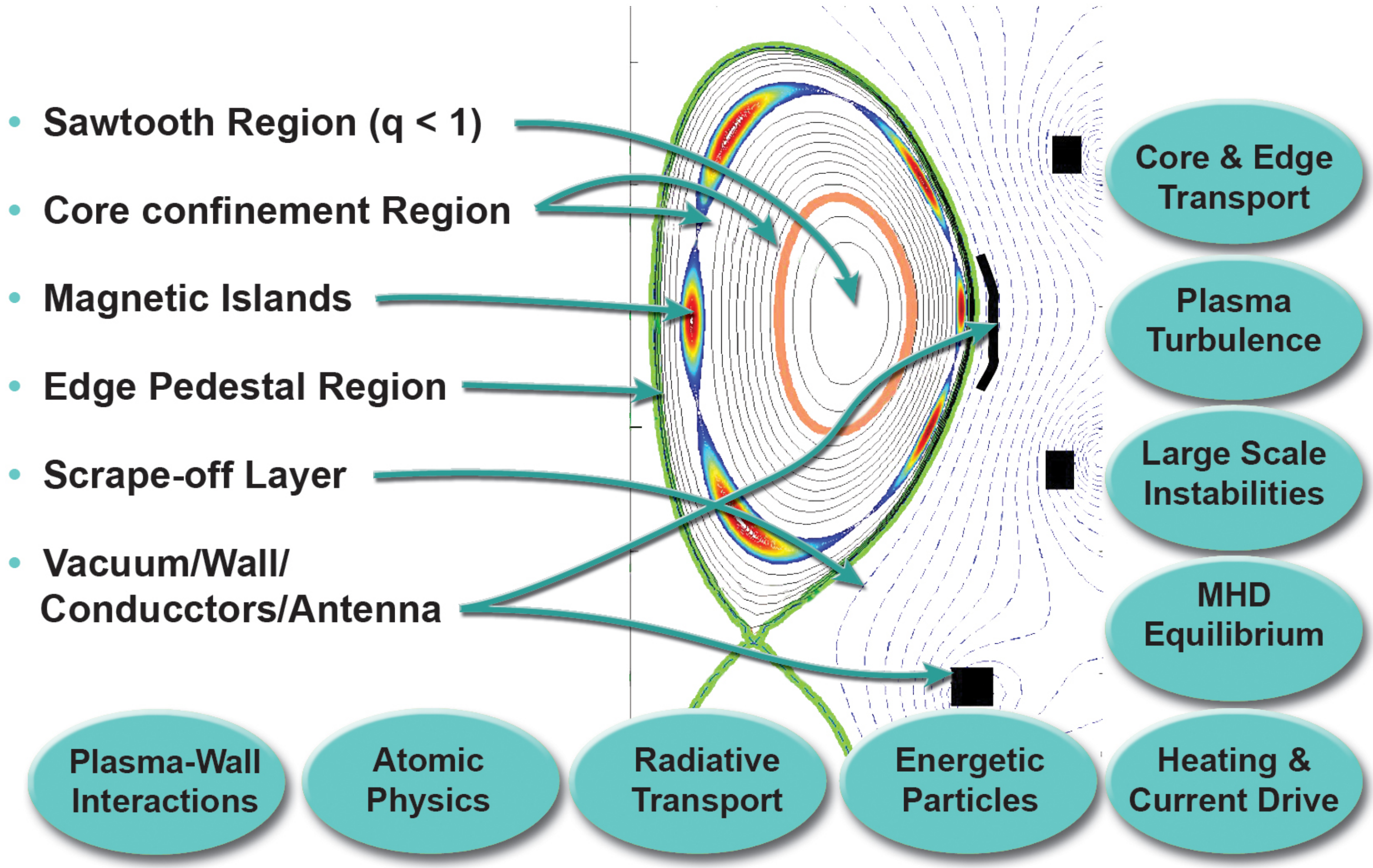
MISSION: ***The Fusion Simulation Program (FSP) will develop advanced HPC-enabled tools to help accelerate understanding of magnetized toroidal plasmas via efficient integration of multiple, coupled physical processes. This task will engage theory, experiment, and advanced HPC resources to deliver unprecedented capability for harvesting information from experiments and designing new devices with improved performance.***

FSP -- A Strategic *Opportunity* to Accelerate Scientific Progress in FES

- Need for reliable predictive simulation capability for *BP/ITER* *(especially in the US)*
- Powerful (“Leadership Class”) Computational Facilities moving rapidly toward petascale & beyond
- Interdisciplinary *collaborative experience*, knowledge, & software assembled over the course of nearly a decade under SciDAC plus OFES and OASCR base research programs in the US



Elements of an FSP Integrated Model



FSP Situation Analysis

- The FSP team* is currently funded to carry out a detailed "planning study" over two years (8/09-7/11) – with requirements as specified in the DoE RFP.

  *Team of 6 national labs (PPPL, ORNL, LANL, LBNL, LLNL, ANL), 2 companies (GA, Tech-X), and 9 universities (MIT, Princeton, Columbia, NYU, UCSD, Chicago, Lehigh, Purdue, Texas)

  Current "project definition" phase *managed as a project*
  - Includes FSP program scope & deliverables and FSP planning scope & WBS
  - Targeted goals, schedules, milestones, responsible groups.
  - Build on "lessons learned" from other major scientific software development projects such as ASC [e.g. -- FY06 ASC Program Plan & more recent interactions @ LLNL and LANL]

- The FSP planning effort has an *active outreach* to the theory, modeling and experimental *national & international communities in FES* and the applied math and computer science *communities in ASCR* to help define scientific priorities and establish mechanisms for *productive collaborations* – e.g., visits to GA, MIT, Maryland, LLNL, ANL, LANL, ORNL, …..

- The FSP planning team has posted on its national web-site [*http://www.pppl.gov/fsp/*] an *FAQ section* and generally welcomes input, comments and suggestions from the FES and ASCR communities.

- A DOE-Office of Science review will be held at the end of the 2-year planning study (shortly after July 2011)

FSP Computing Infrastructure Needs

Current FSP Planning Task must estimate needs for computing requirements and growth:

- **Tuning of Systems for job mix – find most cost-effective platform for each job with flexibility (e.g., priority sometimes needed for small jobs!)**
- **Special requirements with respect to memory, storage, etc**
- **Availability of needed libraries and other supporting software**
- **Ability to respond to priorities set within the FSP domain**
- **Adequate CPU hours for software development (advanced components & frameworks), for V&V + UQ testing, and for production services**

FSP Computing Hardware Needs – 5 Year Horizon

- *Computational resource needs at this point are only notional*
- *However, we can provide a rough estimate by extrapolating from related computational programs in MFE (especially the proto-FSP's)*
- *FSP is envisioned to roughly double the scale and scope of the current MFE computational program*
    - *10s of large jobs using in aggregate >1M cores*
    - *100s of medium scale runs using 10,000s of cores*
    - *10,000s of small runs using 1000s of cores*
    - *Memory requirements from 0.1 GB/core for largest jobs to 2 GB/core for small and medium runs*

FSP Data Storage Infrastructure Needs

Current FSP Planning Task must estimate needs for data storage requirements and growth:

- Adequate, persistent bulk storage
  - Current TB/year range with, growth curve to PB range
- Very fast network transfer from data source
  - Possible Co-location - on same LAN as computing platforms
  - May require multiple U.S. data centers
- Fast access to community user base with modern connections via Esnet
  - Caching strategy?
- Support for FSP data management systems and software
  - Probably includes multiple file formats and relation databases
  - FSP control of access rules and mechanisms
  - FSP “management” of servers and software
- Provide long-term secure back-up/safe archive/disaster recovery

FSP Data Storage Needs – 5 Year Horizon

- *As with computation we can only make rough estimates at this point.*
- *Aggregate archival storage is likely to be in the multi-PB range in 1000s to 10,000s of files per year*
    - *Temporary storage needed by jobs during runs are also predicted to go into the PB range*
- *As noted in the previous slide, we are planning to catalog all FSP runs across all platforms regardless of physical location*
    - *UAL (universal access layer) planned for location independent data access*

FES SciDAC Collaboratory Experience

- Previous (SciDAC-I) FES SciDAC Collaboratory Project (GA, MIT, PPPL, + CS Partners) successfully implemented useful new collaborative technology
  - Addressed problems defined by fusion scientists
  - FusionGrid services used to benefit daily FES research

- Service oriented computation on FusionGrid proved successful
  - Simulation as a Service
  - Optimized the most expensive resource - people's time

- Future vision & work scope for Collaborative Control Room
  - Real-time support for experiments is critical
  - Encompasses most if not all FES collaborative needs
  - Software enhancements (dealing with larger data sets) clearly required for success

- Helps to position US for optimizing its major investment ($1B) in ITER

Experimental FES Demands Rapid Data Analysis in Near-Real-Time

- Pulsed Experiments
  - currently about 10s duration plasma every 20 minutes; much longer duration with far more data for ITER in future
- 20-40 people in control room + more from remote locations currently
  - Much greater scale for ITER in future
    $\sim 10^6$ named data items are anticipated
  - kHz to MHz sample rates today
  - strong focus on between pulse analysis

- Not batch analysis and not a "needle in a haystack" problem
  - Rapid near-real-time analysis of many measurements in current FES experiments are successfully carried out; BUT future large data challenges are formidable

- More informed decisions result in better experimental planning
  - Efficient collaborative control room essential

FUTURE ISSUES BEYOND 5-YEARS:
FES DATA ANALYSIS CHALLENGES FOR ITER

- **DATA TRANSFER FROM ITER TO US**
  - Current estimates of data size is roughly 40 TB per shot for long-pulse shots of 400 seconds duration

    -- would demand 100 GB/sec bandwidth

    -- likely need to be able to parallelize at least a significant fraction of this data for streaming
  - Current estimates of time between shots is roughly 1600 seconds -- a rather limited period of time

    -- I/O will be very stressed for:

    (i) reading even a fraction of this amount of data from memory into CPUS & then writing back to disk

    (ii) displaying of the information

    realistic development of such capabilities is a major challenge

  - Current MDS+ capabilities unable to deal with future parallelism and streaming issues

NOTE: *The ITER data system and modes of operation, remote participation, remote data caching are all still being determined.*

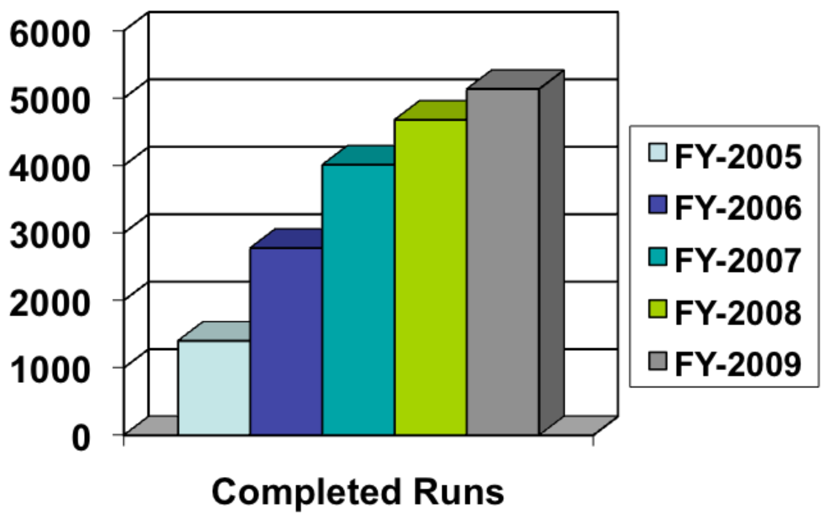
FES DATA ANALYSIS CHALLENGES FOR ITER (continued)

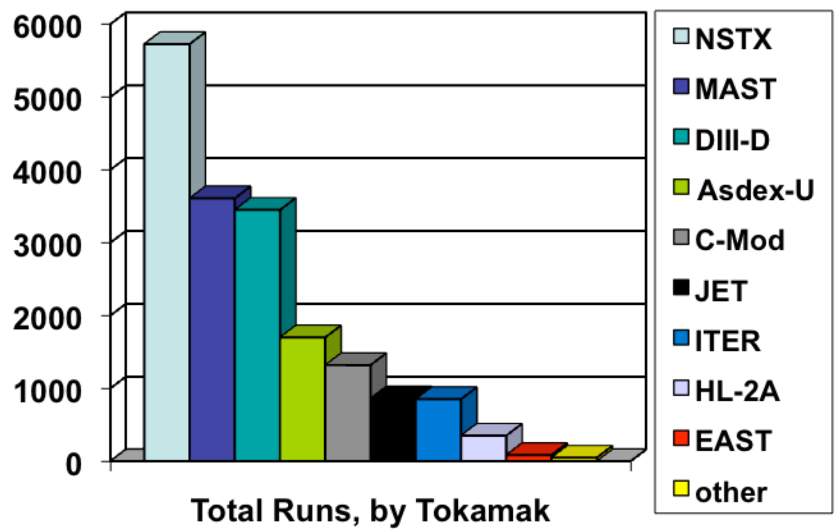
- **POSSIBLE CHANGE IN PARADIGM:** possible movement from current "data file paradigm" to "data streaming paradigm" to accommodate much larger data sets
  - analogous to looking at various frames of a movie while the movie is still being generated
  - advance image processing capabilities could enable end-users/physicists to examine/analyze information while shot in progress
- **ASSOCIATED HARDWARE CHALLENGES**
  - Most present-day computer systems do not have the memory (50 TB or so) needed to deal with large data collection
    -- might lead to approach of examining one stream at a time or possibly processing one stream on one machine while simultaneously moving another stream
- **ASSOCIATED SECURITY CHALLENGES**
  - Users can access parts of data per shot but not allowed access to other associated information
  - Users need to add information/annotate shots & query off their own and other collaborators annotations
  - Important to keep connections "alive" for long periods & keeping the security channels open

Fusion Grid Production Services

Example: PPPL (prepared by D. McCune)

5 Years, Oct. 2004 – Sep. 2009





- TRANSP/PTRANSP – possible early FSP predecessor
- Serial and small parallel (Np~16) jobs
- Approximately *100 users from world wide fusion program*
- Wide use by tokamak experimental projects
  - Data Analysis
  - Experimental Proposals
  - Design of new experiments, upgrades, diagnostics…

Near-Term Expansion of Capacity Computing Services
Example: PPPL (prepared by D. McCune)

- Tokamak Auxiliary Heating: ICRF Simulations – 96 processors per toroidal mode number per antenna using TORIC
  - Each mode/antenna combination an independent calculation
  - Repeated calls in a time dependent simulation
  - Estimated Time: *24-48 hours per simulation*
  - Estimated Number of Simulations: *100's of simulations*
- Associated Capacity Computing Services Needed:
  - Fairly strong interconnect (within each 96 processor block) such as Infiniband– 24x speedup demonstrated on PPPL IB cluster.
  - User turnaround reduced from 6 weeks to 1.5 days.
  - Typical Capacity Computing Load Estimate: 100-2000 processors
  - Well below level of LCF computing but well beyond current capability of local clusters at FES facilities

Current Fusion Grid System

Example: PPPL (prepared by D. McCune)

- Globus/SSL Certificate proxy-based authentication
  - Globus gatekeeper and gridftp daemons
  - Original development: *SciDAC Collaboratory, 2000-2004*
- Distributed Authorization System (ROAM – Resource Oriented Authorization Mechanism)
- Systems (hardware) made available to support service
- Service software installation in service provider account
- Service execution in user service accounts
  - Service provides configuration without requiring end user involvement
  - Process remote user requests with tasks including:

    Job queuing, monitoring, data access, job cancellation
  - No interactive access by end users
  - Interactive access by maintainers for trouble shooting

Associated Questions for NERSC

- Can systems be made available?
- Can GLOBUS be used?
- Can Service Provider Accounts be set up?
- Can end-user service accounts be established?
- Can expert access be allowed for trouble shooting?
- Will storage/systems on FSP-administered devices reside on the NERSC LAN?
- How could more flexible computing environments be implemented?

Possible Proto-type Project with NERSC:

Explore using the TRANSP/PTRANSP Auxiliary Heating (ICRF) Service as a near-term example for delivery of FSP production services to user community

Concluding Comments

• The current *motivation for the FSP* is compelling

-- next major milestone in MFE research is a *burning plasma experiment* --- the multi-billion dollar *ITER* facility located in France & involving the collaboration of 7 governments representing over half of world's population

-- ITER targets 500 MW for 400 seconds with gain > 10 to demonstrate *technical feasibility of fusion energy* & *DEMO* (*demonstration power plant*) will target 2500 MW with gain of 25

-- recognized need for using *advanced computation to harvest knowledge from ITER and for designing DEMO*

• Future Integrated Modeling Tools will target realistic simulations of fusion and energy systems with unprecedented physics fidelity

-- involves delivering shorter-term opportunistic HPC software tools (built largely from modestly improved existing tools); &

-- parallel longer-term development emphasizing new, more rigorous, more engineered performance capabilities

• *In general, the FSP (Fusion Simulation Program) is expected to accelerate progress in delivering reliable predictive capabilities in Fusion Energy Science -- benefiting significantly from access to HPC resources -- from terascale to petascale & beyond -- together with a vigorous verification, validation, and UQ program.*