Present and Future Computing needs in Uncertainty Quantification for Large-scale Multi-physics Models

Charles Tong
Lawrence Livermore Lab

UQ Challenges for Multi-physics Applications

- **Difficult to prescribe parameter uncertainties (the priors)**
- **High-dimensionality of the uncertain parameters (10’s -100’s)**
- **High-dimensionality of the model outputs**
- **Models may be expensive to evaluate (many CPU-hours)**
- **Complex models show highly nonlinear (may be discontinuous) input-output relationships**
- **Data scarcity for the full system**
- **Models are often created by data far from operating conditions**
    - **extrapolation may be needed**
- **Structural uncertainties are difficult/expensive to quantify**
- **“Unknown unknowns” can greatly complicate the UQ process.**

In order to perform UQ, we need

- **A UQ methodology**
  - **A well-thought process/plan with a well-defined objective**
  - **Consisting of a number of steps**
  - **Each step may require expert judgment or suitable UQ algorithms**

- **Relevant UQ methods (forward propagation, SA, calibration)**
  - **Intrusive methods (one large simulation)**
  - **Non-intrusive methods (many smaller simulations)**
  - **Hybrid (intrusive+nonintrusive) methods (a number of medium simulations)**

- **Adequate hardware/software infrastructure to perform UQ**
  - **Job management: scheduling, monitoring**
  - **Data processing**
  - **Analysis and visualization of results**

An example UQ methodology

1. **Define the objective of the UQ study (e.g. quantify risk)**
2. **Problem specification (model, code version, assumptions, QOI)**
3. **Perform verification experiments (to assess numerical errors) (*)**
4. **Preliminary model/parameter identification and selection**
5. **Apply component/subsystem data fusion methodologies (***)**
6. **Apply parameter dimension reduction (**)**
7. **Response surface modeling (***)**
8. **Apply quantitative uncertainty/sensitivity analysis (*)**
9. **Validation, system level calibration, predictability assessment (*)**
10. **Expert reviews, documentation**

* Small ensemble runs (O(10))
** Moderate number of runs (a few hundred)
*** Large ensemble calculations (thousands)

Multi-physics UQ Project

- Demonstrate a predictive capability for a DOE NNSA application
- Project lasts until 2018 with first and second pegposts by 2012 and 2014
- Present focus: subsystem predictive capability
  - Use state-of-the-art physics models
  - Multiple events with common physics models
  - Physics-level data for calibration
  - Multi-event system level calibration
  - Assess uncertainties/sensitivities and prediction on a holdout event
  - Number of uncertain parameters in this case ~20
- Ensemble run modes (capacity computing?)
  - Grab a partition for n days
  - Use job submission system + daisy chain for automation

Current HPC Methods

- Algorithms used:
  - implicit/explicit hydrodynamics
  - material strengths and equation of states, etc.
- Codes: LLNL multi-physics codes
- Quantities that affect the problem size/scale of the simulations:
  - Spatial dimension, mesh sizes (each simulation)
  - Complexity of output of interest (sample size)
  - Number of events used (sample size)
- Parallelism expressed in
  - MPI within each simulation
  - Independent ensemble calculations
  - Multi-prob mode

Current HPC Requirements

- Facilities used: LLNL
- Architectures currently used:
  - Linux clusters (AMD/IBM PowerPC)
- Compute/memory load:
  - 1000's run each 128 processors/12 hours
- Data read/written: O(100) Mbytes per run (on LLNL storage)
- Necessary software, services or infrastructure
  - job submission system
- Known limitations/obstacles/bottlenecks:
  - resource availability when needed
- Hours requested/allocated/used in 2010 : ~10M CPU hours

HPC Usage and Methods for the Next 3-5 Years

- Upcoming changes to codes/methods/approaches
  - Same code (newer versions), different multi-physics models
  - UQ study of many models together (instead of one or a few)
  - Possibly look into 3D problems and/or 2D with higher resolution
- Changes to Compute/memory load
  - 2011 project: ~10000 runs (12 hours on 128 nodes)
  - 2011 project: 10000-100000 runs (12 hours on 128 nodes)
  - 2012 on: don’t know yet
- Changes to Data read/written: same for each run
- Changes to necessary software, services or infrastructure: none
- Anticipated limitations/obstacles/bottlenecks on 10K-1000K PE system.
  - Fault management, job submission
  - Parallel I/O (esp. I)
  - Programming model not a problem yet

Strategy for New Architectures

- How are you dealing with, or planning to deal with, many-core systems that have dozens or hundreds of computational cores per node?
  - So far, our codes has been modified to run multiple jobs in a single run (called multiprob mode) since the minimum processor count per job is 512
- How are you dealing with, or planning to deal with, systems that have a traditional processor augmented by some sort of accelerator such as a GPU or FPGA or similar?
  - Our code groups have taken steps to investigate how to take advantage of these new features
  - Not a problem directly related to UQ runs, for now

Summary

- Recommendations on NERSC architecture, system configuration and the associated service requirements needed for your science
  - Processor speed/number of compute nodes tradeoff
  - Parallel I/O (fast read, common libraries)
  - Memory per CPU (3D memory footprint for 3D explicit hydrodynamics)
  - More advanced job submission system for large ensemble (negotiated)
  - Fault detection, recovery, reporting
- NERSC generally acquires systems with roughly 10X performance every three years. What significant scientific progress could you achieve over the next 3 years with access to 50X NERSC resources?
  - Higher fidelity UQ on multiple-model predictive science studies
- What "expanded HPC resources" are important for your project?

- General discussion