



Incorporating High Energy Physics Data Capabilities into Joint Forces Simulations

Julian J. Bunn and Thomas D. Gottschalk
Center for Advanced Computing Research, Caltech
julian@cacr.caltech.edu & tdg@cacr.caltech.edu



Basic Premise

- **The simulation and data management issues of “Big Science” and “Big Defense” have a number of similarities.**
- **These similarities can be exploited – to the benefit of each side.**
- **Promising areas of overlap include:**
 - **Globally distributed databases and data management.**
 - **Distributed computing and “User Toolkits”**
 - **Operational issues, such as resource scheduling.**

Game Plan

- **Briefly discuss status/activities for a particular large-scale High Energy Physics experiment (CMS/LHC)**
- **Note briefly issues already addressed within CMS/LHC**
 - Details/references in the submitted paper.
- **Most importantly: An invitation to exchange issues, experiences, failures, etc. between the two communities.**

HEP Primer

- **High Energy Physics (HEP) seeks to explore/understand the “fundamental interactions” of nature.**
- **Probing nature at small scales is a “big” problem, with huge data production systems and very large collaborations.**
 - **The CERN accelerator is about 18 miles in circumference.**
 - **Front-line collaborations have 2000+ scientists.**

The CERN site near Geneva.



The CMS Experiment at CERN



CMS Collaboration



36 Nations, 160 Institutions, 2008 Scientists and Engineers (November 2003)

TRIGGER & DATA ACQUISITION

Austria, CERN, Finland, France, Greece, Hungary, Italy, Korea, Poland, Portugal, Switzerland, UK, USA

TRACKER

Austria, Belgium, CERN, Finland, France, New Zealand, Germany, Italy, Japan*, Switzerland, UK, USA

CRYSTAL ECAL

Belarus, CERN, China, Croatia, Cyprus, France, Ireland, Italy, Japan*, Portugal, Russia, Serbia, Switzerland, UK, USA

PRESHOWER

Armenia, Belarus, CERN, Greece, India, Russia, Taipei, Uzbekistan

RETURN YOKE

Barrel: Czech Rep., Estonia, Germany, Greece, Russia, Endcap: Japan*, USA, Brazil

SUPERCONDUCTING MAGNET

All countries in CMS contribute to Magnet financing in particular: Finland, France, Italy, Japan*, Korea, Switzerland, USA

HCAL

Barrel: Bulgaria, India, Spain*, USA
Endcap: Belarus, Bulgaria, Russia, Ukraine
HO: India

FEET

Pakistan
China

FORWARD CALORIMETER

Hungary, Iran, Russia, Turkey, USA

MUON CHAMBERS

Barrel: Austria, Bulgaria, CERN, China, Germany, Hungary, Italy, Spain;
Endcap: Belarus, Bulgaria, China, Korea, Pakistan, Russia, USA

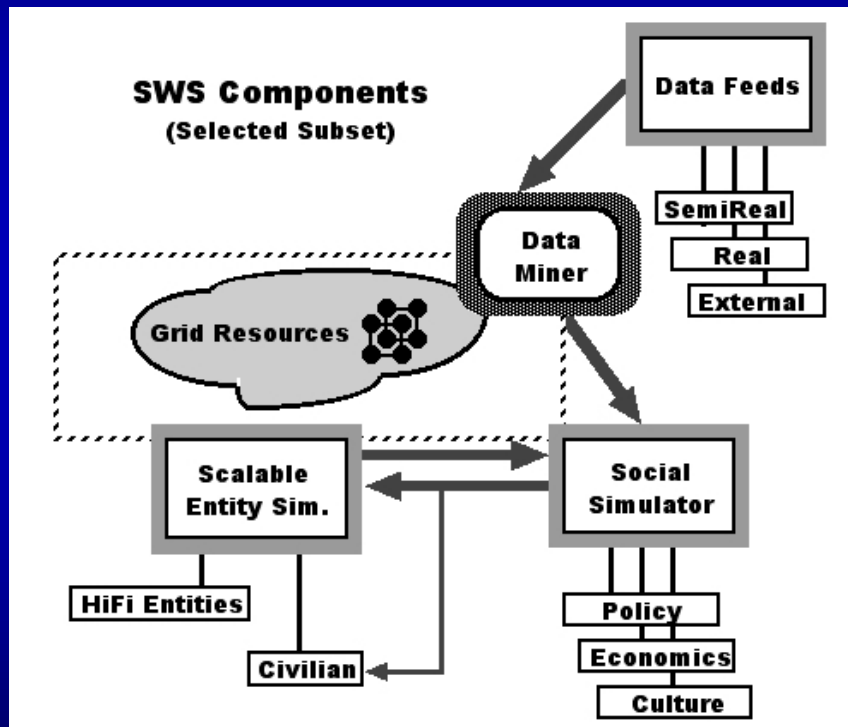
Total weight : 12500 T
Overall diameter : 15.0 m
Overall length : 21.5 m
Magnetic field : 4 Tesla

* Only through industrial contracts

Forces Modeling and Simulation

- **Large-Scale simulations for DoD are increasingly common – and important.**
- **Many issues in common with HEP applications:**
 - Large, distributed computation and data management.
 - Data exploration/exploitation issues.
 - Petabyte (and beyond) data sets.
- **Large problems lead to large, complex support systems**
 - JSAF/JUO, extending to SDG
 - Sentient World Simulation – an interesting pending possibility.

Idealized SWS Schematic

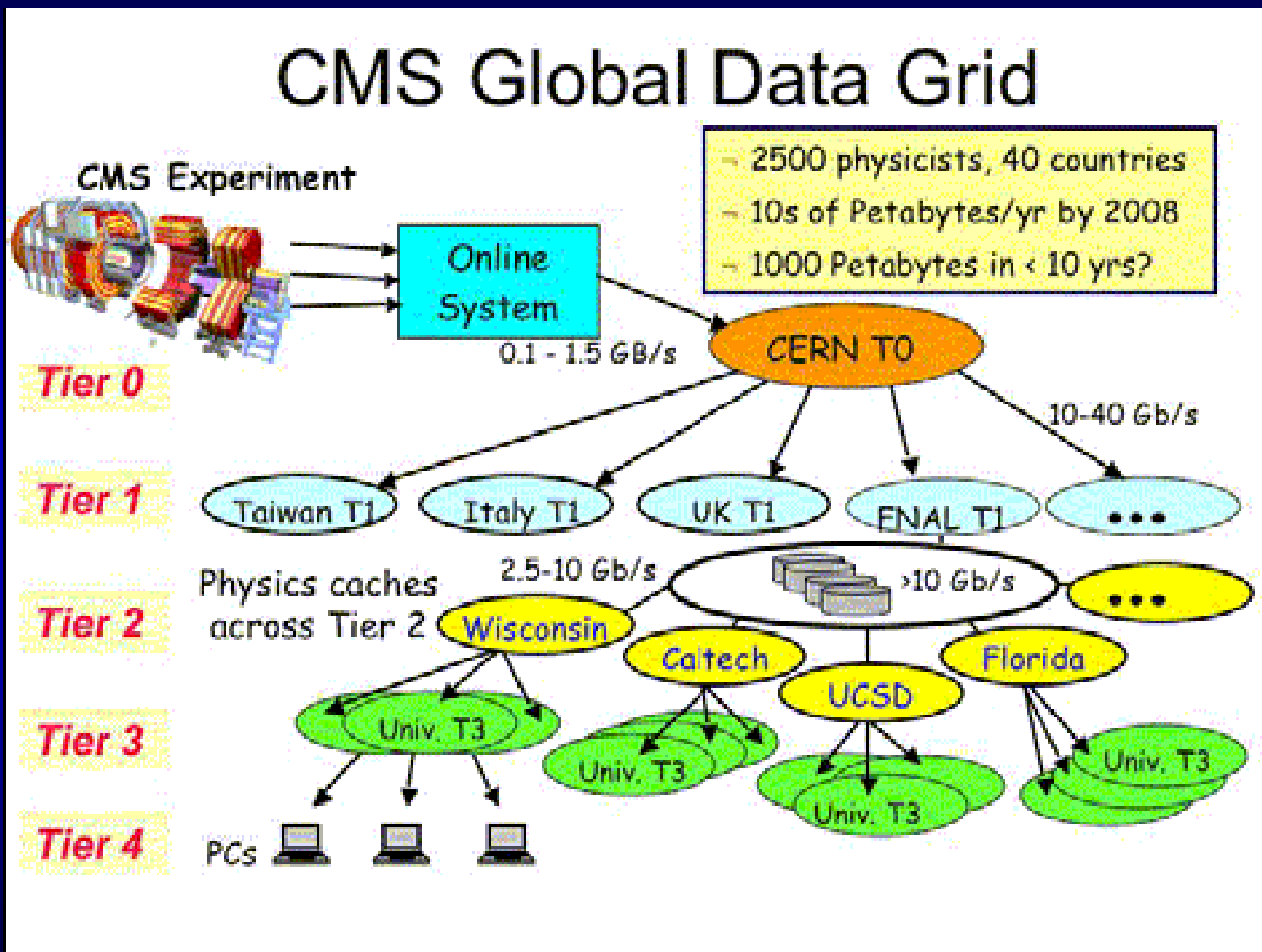


- Large-scale system
- Computational resources – nominal and “on demand”.
- Persistence
 - 24x7 Operations
- Knowledge Discovery
 - Finding “interesting” signals in an overwhelming background

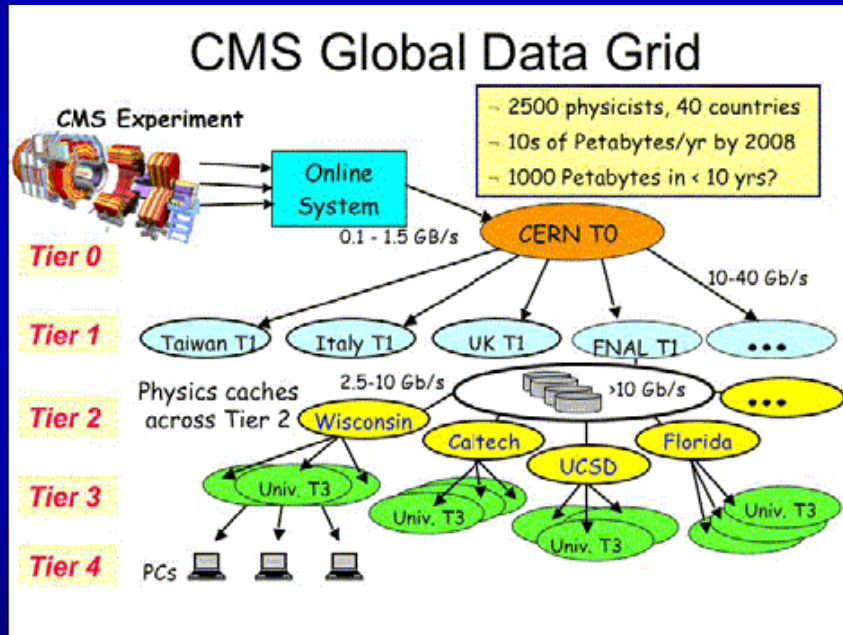
Large, Distributed Data Sets

- **Issues recognized for large-scale HEP**
 - The underlying database is distributed over a very large network.
 - Associated data set sizes huge (petabyte, exabyte)
 - Operations at the user level must be transparent.
 - Capabilities are required for significant, user-directed data manipulations.
 - Security and data integrity are critical.
- **Required system designed, implemented, tested**
 - The LHC (Large Hadron Collider) computational grid

The LHC Data Grid Hierarchy Model



The LHC Data Grid Hierarchy Model

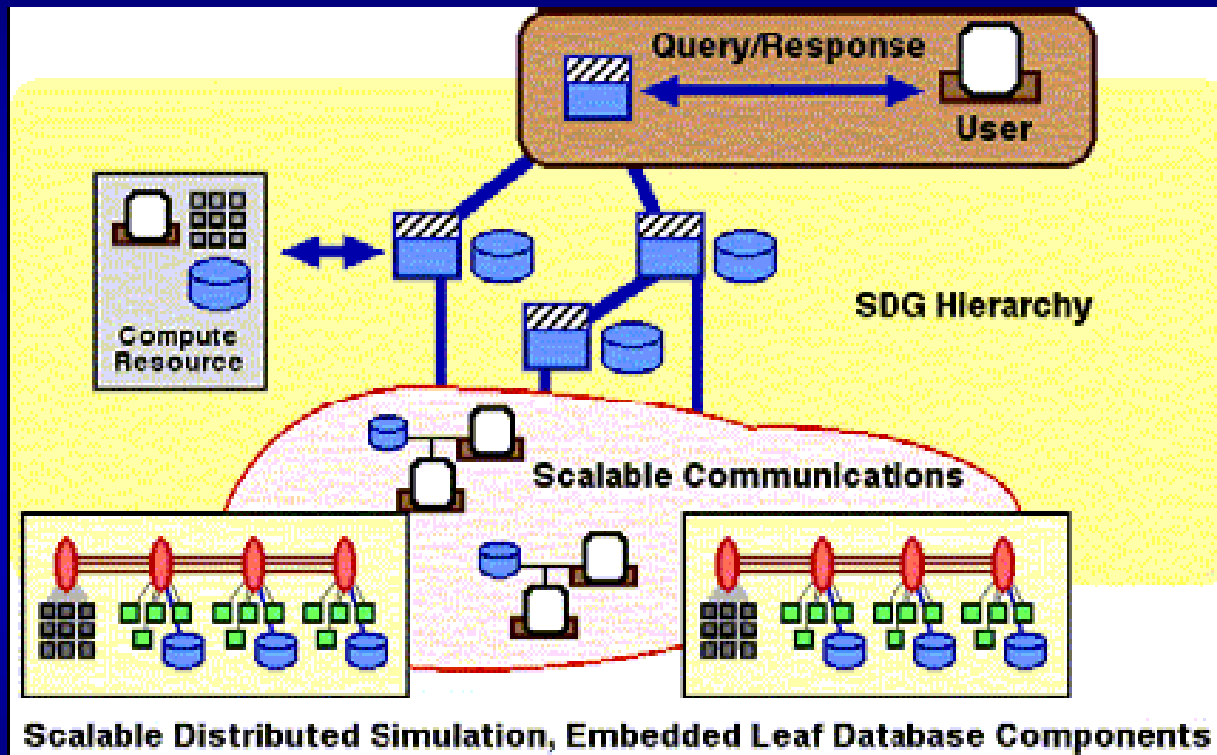


- Data “caught” at source, rate about 1 GByte/sec.
- Data replicated/distributed to “Tier 1” sites.
- Lower level Tiers copy data subsets into caches for analyses
- Lowest level Tier – single graduate student chained to a desk, but with access to all data, worldwide.

Distributed Computing and User Tool Kits

- **Not as simple as “Buy some disks, Linux clusters. Then negotiate for communications links”**
- **Front-line issues for LHC development:**
 - **Grid Computing**
 - **Real-Time Analysis Capabilities**
 - **User-Comfort Considerations**
- **All are important in overall operations**
- **Successes (and failures) from HEP can provide insights to similar developments for DoD/FMS**

Distributed Computing and User Tool Kits – Needed Now!



- Evolving JSAF/JUO/SDG – Transparent To End User
- Extensions: Data Mining, Real-Time Feeds, etc.

Real-time Considerations

- **No affordable system can support the suite of all possible user demands.**
- **Policies – and management/enforcement procedures**
 - **are needed to ensure best resource utilization.**
 - **Can't have a graduate student in Huntsville usurping the world-wide bandwidth with a poorly formed query!**
- **Security and data integrity are drivers, not “add ons”**

Conclusions

- **There are many similarities between “Big Science” and “Big Defense”**
- **Many issues for large FMS are likely already addressed within the HEP/academic community.**
- **Time has come for interactions and dialogue!**

Government Support

This material is based on research sponsored by the Air Force Research Laboratory under agreement number FA8750-05-2-0204. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Research Laboratory or the U.S. Government.