Computational Challenges and Needs for Academic and Industrial Applications Communities

IESP – Tsukuba

October, 2009

Methodology

Three ways to look at these issues

1. Preliminary (*i.e.* between the Paris and Tsukuba meetings): the (disciplinary) expert views
2. A view transversal to all application domains: 4 main items
3. Back to the disciplinary views: classification of issues with respect to expectation from the SW groups

Methodology

Three ways to look at these issues

1. **Preliminary (*i.e.* between the Paris and Tsukuba meetings): the (disciplinary) expert views**
2. A view transversal to all application domains: 4 main items
3. Back to the disciplinary views: classification of issues with respect to expectation from the SW groups

Engineering
12 contributions

Earth Sciences – Oil & Gas
Depth Imaging / Reservoir simulation

Expert name/affiliation - email: Henri CALAN DRA, TOTAL, Henri.CALANDRA@total.com

Scientific and computational challenges

- Sub Salt and Foothills Depth Imaging
- Fine scale reservoir simulation
- 4D monitoring
- Less approximation in the physics: non linear full waveform inverse problem
- Elastic, poro-elastic ground models…,

Software issues – short term (2009/2011)

- Mesh generation: scalability, load balancing
- Accurate and fast Wave Equation Solver
- Solvers (multi-grid, better pre-conditioner)
- Standard programming tools for addressing accelerating technology (e.g. GPGPU)

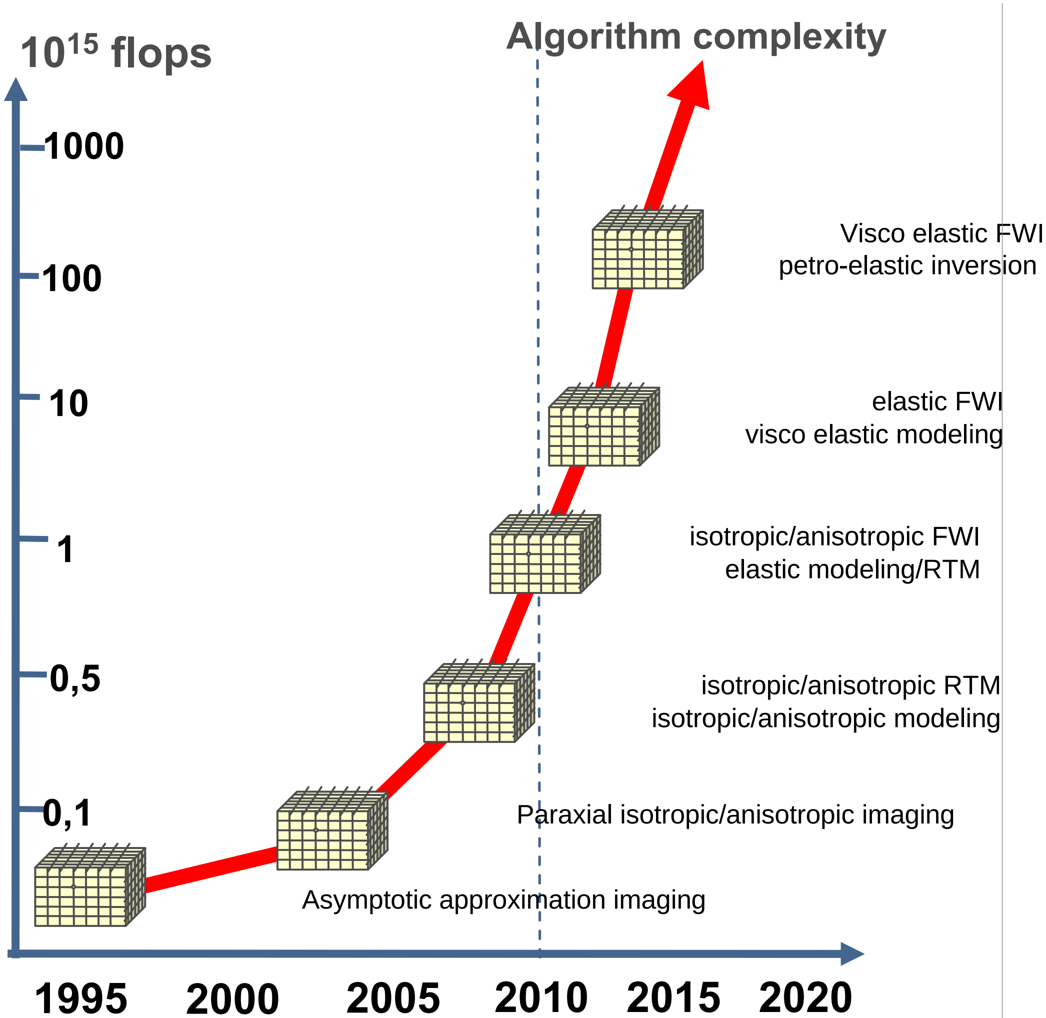
Software issues – long term 2015/2020

- New numerical methods for solving more complex Wave Equation formulation
- Scalable solvers for reservoir simulations
- Adaptive methods for heterogeneous platforms (hybrid e.g. CPU+GPU)
- New optimization methods (no gradient computations)
- Programming tools: PGAS language such as CAF ?

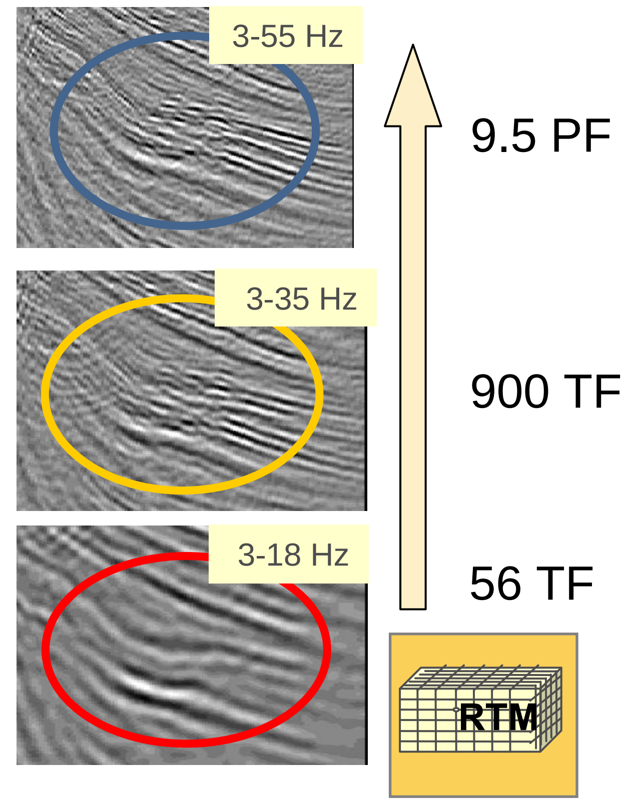
Impact of last machine changes (a few Tflops -> 100 Tflops)

- Last change (10=> 100 TFlops) was almost seamless, Depth Imaging codes were ready in OpenMP/MPI hybrid mode up to 4000 cores + scheduling of many jobs of different sizes to optimize the 100+ Tflops machine global workload – should scale up to 1+ Pflops/s 2010 NEXT: 10 PFlops 2012?
- Reinforcement of HPC expertise to harness petascale and beyond computers,
- Accelerating technology: load balancing on large systems with different kinds of compute units
- Impact of network technology: better, direct data migration, IO, initialisation; better SMP or distributed memory usage
- Impact of the many core technology on the design of the algorithm: will we have to revisit the physics?

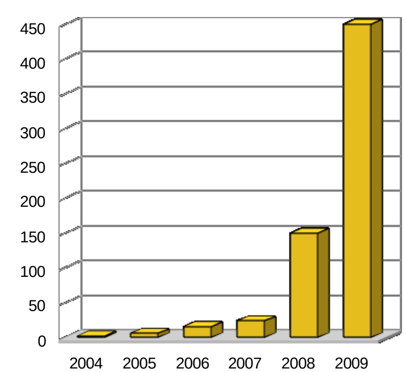
Industrial challenges in the Oil & Gas industry: Depth Imaging roadmap



Algorithmic complexity Vs. corresponding computing power



Substained performance for different frequency content over a 8 day processing duration



Computational Challenges and Needs for Academic and Industrial Applications Communities

AERONAUTICS – Eric CHAPUT / AIRBUS – eric.chaput@airbus.com

Scientific and computational challenges

Aero Optimisation & CFD-CSM coupling
Full multi-disciplinary optimization
CFD-based noise simulation
Real-time CFD-based in-flight simulation

Software issues – short term (2009/2011)

Parallel I/O, for CSM, for visualization
Multi-level parallelism
Load-balancing in industrial geometries, with adaptative meshing
Integrating and coupling (non-parallel) commercial codes
Data mining for constructing reduced models

Software issues – long term 2015/2020

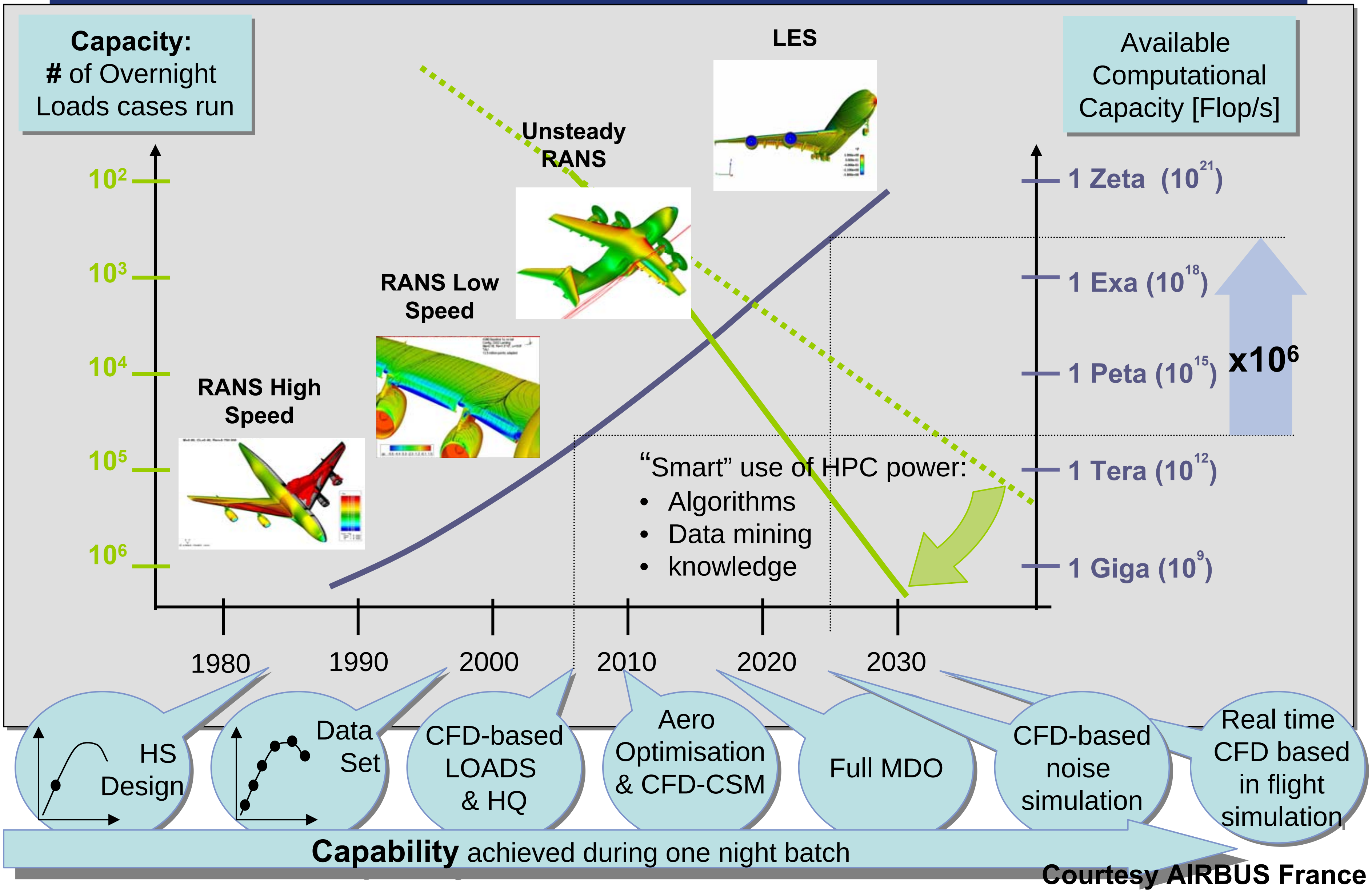
Increased efficiency (algorithms, compilers)
Compilers for hybrid architectures
Fault-tolerance, dynamic reconfiguration
Virtualization of matching between needs and resources

Impact of last machine changes (??flops -> ?? flops)

Better exploration of parameter space (embarrassingly parallel problem !)

Maintaining the scaling properties, maintaining the efficiency

High Performance Computing



Computational Challenges and Needs for Academic and Industrial Applications Communities

AERONAUTICS – Cord ROSSOW / DLR – h.mueller@dlr.de

Scientific and computational challenges

-Digital aircraft: complete design before starting industrial developments
- Preparation of certification before 1st flight

Software issues – long term 2015/2020

- Need for standards
- Operating systems taking care of task allocation, of load-balancing

Software issues – short term (2009/2011)

-Software and libraries should be applicable to all types of computers
- Need for solving the dilemma: implicit solvers not easy to parallelize, explicit solvers not very efficient
- Parallel IO for post-processing (to be done outside the mainframe), for allowing interaction with the simulation, for optimization using genetic algorithms, …

Impact of last machine changes (??flops -> ?? flops)

- Domain not using the newest architectures or machines (Top 25 rather than Top 5), for minimizing the impact of machine change

Computational Challenges and Needs for Academic and Industrial Applications Communities

Prof. Christian Allen, University of Bristol, CFD Rotor Aircraft

Scientific and computational challenges

Real-time simulation of full aircraft

Flight simulation using full Navier-Stokes equations

Software issues – long term 2015/2020

Effect of hardware changes.

Codes need to be rewritten to exploit GPUs

Software issues – short term (2009/2011)

Mainly efficiency issues due to scaling

Impact of machine changes

No problem so far with large distributed memory clusters

Need to write codes more in terms of memory and data management rather than message passing

Computational Challenges and Needs for Academic and Industrial Applications Communities

Prof David Emerson, STFC Daresbury Lab, CFD

Scientific and computational challenges

Fully utilise CFD in the engineering design cycle for

- Engine design
- Aerodynamics

Software issues – long term 2015/2020

Short-term problems (memory bandwidth, I/O etc) will become greatly exacerbated

Software issues – short term (2009/2011)

Multi-core with reduced memory bandwidth per core is seriously impacting most CFD codes

Input/output is becoming critical

Impact of machine changes

Machine change was OK until memory bandwidth per core began to drop dramatically

Investigating whether mixed MPI/OpenMP will help

Computational Challenges and Needs for Academic and Industrial Applications Communities

Dr Stewart Cant, University of Cambridge, CFD &Combustion

Scientific and computational challenges

Simulations at high Reynolds numbers given the strong scaling of memory and CPU time with Re

Software issues – long term 2015/2020

Exploitation of multi-core nodes

Input/output

Complete software chain from CAD to visualisation of solution

Software issues – short term (2009/2011)

Codes perform badly due to limited memory bandwidth on multi-core nodes

Impact of machine changes

So far painless to O(1000) cores

This is not expected to be the case for Exascale

CFD Simulation
Mechanical and vibratory behaviour of the fuel assemblies inside a nuclear core vessel – a developer point of view

Expert name/affiliation - email: Yvan Fournier/EDF – yvan.fournier@edf.fr

Scientific and computational challenges

Computations with smaller and smaller scales in larger and larger geometries for a better understanding of physical phenomena

⇒A better optimisation of the production (margin benefits)

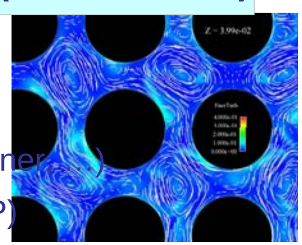
2007: 3D **RANS**, 5x5 rods, 100 millions cells, 2 M cpu.hours (4000 cores during 3 weeks)

2015: 3D LES Full vessel (17x17x196 rods) unsteady approach, >50 billion cells, 1000000 cores during few weeks

Software issues – long term 2015/2020

- New numerical methods (stochastic, SPH, FV)
- Scalability of linear solvers, hybrid solvers
- Code optimisation: wall of the collective communications, load balancing
- Adaptive methods (may benefit all of computation/visualisation/meshing)
- Data redistribution, IO (if flat MPI-IO model OK, good, otherwise require new “standard” data models)
- Fault tolerance
- Machine independent code optimisation & performance

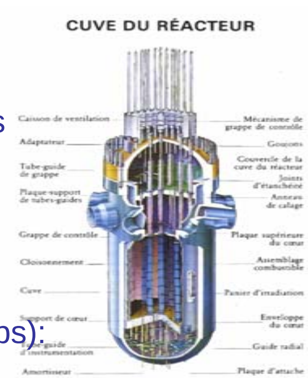
Software issues – short term (2009/2011)



- Mesh generation, visualization
- Scalability, load balancing
- Solvers (multi-grid, better&simpler pre-conditioner, ..)
- Mixing programming models (ex. MPI/OpenMP)
- Stability and robustness of the software stack (MPI, ..)
- API of scientific libraries (ex. BLAS!)
- Standardisation of compiler optimisation level pragmas
- Computing environment standardization (batch system, MPIExec,

Impact of last machine change (x10 Gflops -> 100 Tflops)

- Pre/post adaptation
- Reinforcement of the HPC expertise
- Few extra “simple” programming rules
- No rewriting, same solvers, same programming model, same software architecture thanks to technological evolution anticipation
- Expected impact (100 Tflops -> Xpflops): ie. 2015 software issues



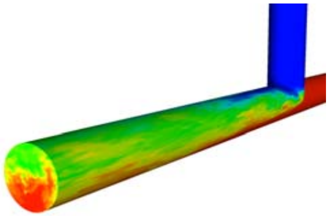
Computational Challenges and Needs for Academic and

2003	2006	2007	2010	2015

Consecutive thermal fatigue event

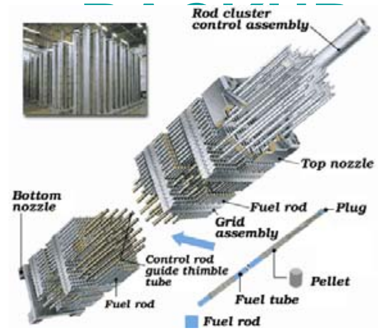
Computations enable to better understand the wall thermal loading in an injection.

Knowing the root causes of the event ⇒ define a new design to avoid this problem.



Computation with an L.E.S. approach for turbulent modelling

Refined mesh near the wall.



Part of a fuel assembly
3 grid assemblies

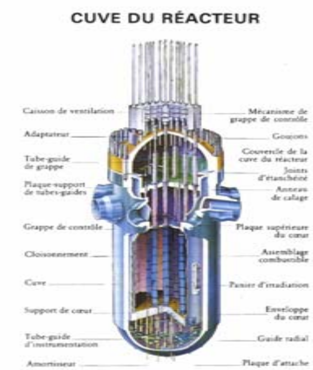
9 fuel assemblies

No experimental approach up to now

Will enable the study of side effects implied by the flow around neighbour fuel assemblies.

Better understanding of vibration phenomena and wear-out of the rods.

The whole vessel reactor



Computations with smaller and smaller scales in larger and larger geometries
⇒ a better understanding of physical phenomena ⇒ a more effective help for decision making
⇒ A better optimisation of the production (margin benefits)

10^6 cells	10^7 cells	10^8 cells	10^9 cells	10^{10} cells
3.10^{13} operations	**6.10^{14} operations**	**10^{16} operations**	**3.10^{17} operations**	**5.10^{18} operations**
Fujistu VPP 5000 1 of 4 vector processors 2 month length computation	Cluster, IBM Power5 400 processors 9 days	IBM Blue Gene/L 20 Tflops during 1 month	600 Tflops during 1 month	10 Pflops during 1 month
# 1 Gb of storage 2 Gb of memory	# 15 Gb of storage 25 Gb of memory	# 200 Gb of storage 250 Gb of memory	# 1 Tb of storage 2,5 Tb of memory	# 10 Tb of storage 25 Tb of memory
Power of the computer	Pre-processing not parallelized	Pre-processing not parallelized Mesh generation	… ibid. … … ibid. … Scalability / Solver	… ibid. … … ibid. … … ibid. … Visualisation

Computational Challenges and Needs for Academic and Industrial Applications Communities

Combustion – Parviz MOIN / Stanford Univ. – moin@stanford.edu

Scientific and computational challenges

- Uncertainty quantification: leading to a lot of additional computations, but critical for predictive science
-Jet noise using CFD methods
- Multiphysics and multiscale problems (turbulence, interfaces, combustion, multiphase flows, shocks)

Software issues – short term (2009/2011)

- Linear solvers running well on 10^3-10^5 cores
- Scalable Parallelization methods for complex and coupled systems and unstructured methods
- Parallel I/O and scalable management of large data sets

Software issues – long term 2015/2020

- Future of MPI-based codes is uncertain: New languages (domain-specific languages, DSL) for expressing parallelism will allow both performance and portability (10^3-10^5)

Impact of last machine changes (??flops -> ?? flops)

- Major rewrite of core infrastructure to support parallel I/O and parallel post-processing on multicore clusters

- For next major machine change, impact will be minimized somewhat due to long-standing interactions between computer scientists and applications scientists, and development of pde-specific DSL

CFD, Hydro-environmental Simulation
A developer and a user point of view

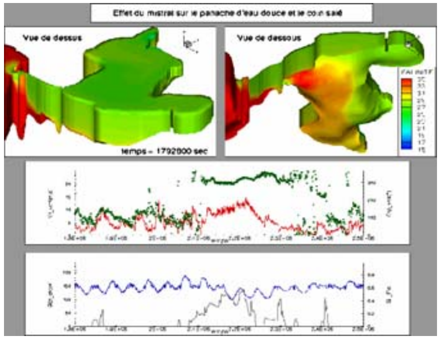
Expert name/affiliation - email: Jean-Daniel Mattei/EDF –jean-daniel.mattei@edf.fr

Scientific and computational challenges

3D free surface flow, sedimentology and ground water flow simulation

2009/10: 30 millions time step, 30 millions elements, 10 TB/run, 30 Tflops during several months/run, 10xrun/study

2015/20: model coupling (fluid/structure, sedimentology/wave/courant), LES, 3 Pflops during several months/run, 10xrun/study



Software issues – long term 2012/2020

- Coupling different scales, geometries, models, physics
- Inverse problem
- Uncertainty Quantification, Data Assimilation
- Numerical solvers

Software issues – short term (2009/2011)

- Porting the whole computation scheme (not only the computing kernel)
- Mesh generation, visualization
- Scalability, load balancing (characteristics method)
- Dealing with large number of time steps (30 millions), =>time parallelization?
- Mixing parallelism (MPI/OpenMP), use of GPU

Impact of last machine change (x10 Gflops -> 100 Tflops)

- Difficulty to “think” parallel
- Reinforcement of the HPC expertise and support
- Data management (data transfer, IO)
- Transparent access to computing power
- Portability/machine independent optimisation is still an issue

Computational Challenges and Needs for Academic and Industrial Applications Communities

Prof. Neil Sandham, University of Southampton – DNS & Turbulent Flows

Scientific and computational challenges

Improving the capability and reliability of CFD calculations for impact on industrial design, esp. for engine and airframe noise

Software issues – long term 2015/2020

New methods for spectral codes which currently use all-to-all communications

Data management

Software issues – short term (2009/2011)

Codes perform badly due to limited memory bandwidth on multi-core nodes

Exploring mixed MPI/OpenMP as a possible solution

Impact of machine changes

So far painless to dual-core with O(1000) cores

The dual-core to quad-core change has had a major impact on performance through limited memory bandwidth

Neutronic Simulation
3D PWR Core Calculation – a user point of view

Expert name/affiliation - email: Tanguy Courau/EDF – tanguy.courau@edf.fr

Scientific and computational challenges

Goal: optimize the nuclear fuel usage in power plants and in particular the reactor core reload pattern
Means : reference neutron transport calculations
2010: 3D full core calculation with homogenized fuel pins description 10 Tflops during 10 hours per study, x100 runs per studies

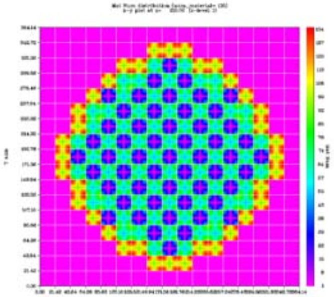
2020: 3D full core calculation with heterogeneous fuel pins description and thermal coupling, >1 Pflops during few days per study, x100 runs per studies

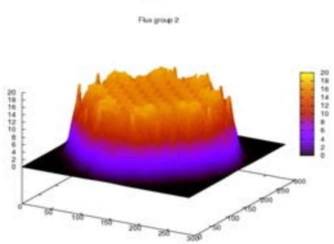
Software issues – long term 2020

- New numerical methods: advanced acceleration techniques, coupling stochastic with determinist methods
- How to deal with global operations (integral parameters evaluation, global spatial coupling of neutronic equations) with one million cores ?
- Using not only parallel solvers but parallel data management through all the calculation process including I/O
- Machine independent code optimisation & performance, hiding the hardware specificities

Software issues – short term (2009/2011)

- Pre & post processing: parallel visualizatio (capability to go from a global picture to the finest mesh level) and data management, IO management (dealing with 10-100 GB for each run)
- Dealing with more and more unknown induces algorithmic convergence issues, more iterations => more efficient acceleration techniques needed
- Solvers (multi-grid, better&simpler pre-conditioner, …)
- Computing environment standardization (batch system, MPIExec, …)
- Robustness of software stack





Impact of last machine change (x10 Gflops -> x10 Tflops)

- Revisiting parallel algorithm (PDEs)
- Higher machine dependence in code optimisation process
- Pre/post adaptation
- Strong reengineering is needed: few neutronic codes are natively well adapted to massively distributed memory architecture
- Reinforcement of the HPC expertise, support from dedicated high skilled HPC experts is needed

Stockastic Optimisation
Electricity production optimisation and risk simulation

Expert name/affiliation - email: Xavier Warin/EDF – xavier.warin@edf.fr

Scientific and computational challenges

Determine strategies for the electricity production that optimize specific economic criteria over varied time scales: maximum gain, minimum risk …: large scale stochastic optimization with millions of variables and constraints, in general continuous and integer values

Operational issue: energy stocks management (hydraulic reservoir, consumer contracts, nuclear fuel stocks) dealing with uncertainties of production, consumption, energy market, weather patterns

2010: taking into account 3 aggregated hydraulic stocks and 6 aggregated consumer contacts, 25 Tflops during few days

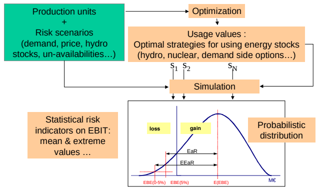
2015 : include fuel stocks management and power plants stops, dealing emission constraints, 20 Pflops during few weeks

Software issues – long term 2015/2020

Algorithmic is the key issue

Fault tolerance as transparent as possible for the programmer

Compiler issue: dealing with dynamic multi-level thread management



Software issues – short term (2009/2011)

Algorithmic:

- Used of well know algorithms (dynamic programming type, price decomposition methods) to deal x100 cores
- Investigating new algorithms to deal with binary constraints and adapted to x1000 cores

Programming model :

- Use of mixing programming models (MPI/OpenMP, MPI/Intel TBB)
- Investigating GPU programming model

Impact of last machine change (x Gflops -> 10 Tflops)

Algorithmic adaptation

Code rewriting

Increased of collaboration with academics, specially specialists in parallel programming model

Has popularised use of HPC in production optimisation domains, has opened new opportunities for energy management

Materials Science, Chemistry and Nanoscience
7 contributions

Computational Challenges and Needs for Academic and Industrial Applications Communities

Dr Matt Probert, University of York, Chemistry

Scientific and computational challenges

Electronic and structural properties of ever larger and more complex systems

New properties e.g. For nano-devices

Software issues – long term 2015/2020

Further scalability issues

Development of new algorithms

Mixed mode MPI/OpenMP

Software issues – short term (2009/2011)

Scaling of existing algorithms

Latency hiding

Impact of machine changes

So far painless to O(1000) cores but reaching limit of scalability with current algorithms

Use Exascale for ensembles simulations for parameter searches etc.

Computational Challenges and Needs for Academic and Industrial Applications Communities

Prof. Simon Hands, University of Swansea, QCD

Scientific and computational challenges

Extend current calculations with fully dynamical quarks into a regime where the quark masses are realistically light

Software issues – long term 2015/2020

Exploitation of multi-core nodes

Fault tolerance

Software issues – short term (2009/2011)

Relatively simple codes allow efficient exploitation of SIMD systems

Expert specialist help is absolutely crucial

Impact of machine changes

New algorithms required to parallelise in all four dimensions

Computational Challenges and Needs for Academic and Industrial Applications Communities

Prof. Richard Kenway, University of Edinburgh, QCD

Scientific and computational challenges

Simulations of the standard model of particle physics and theories beyond

Discover and understand new physics

Software issues – long term 2015/2020

There is no particular reason why sustained Exaflop/s should not be possible with modest adaptation of existing codes

Software issues – short term (2009/2011)

Main performance-limiting factor is memory latency/bandwidth

Support of community codes

Impact of machine changes

The step to Exascale should be smooth

A disruptive change may be required if checkpointing becomes highly inefficient

Computational Challenges and Needs for Academic and Industrial Applications Communities

Prof Nic Harrison, Imperial College & STFC Daresbury Lab, Materials Science

Scientific and computational challenges

Excited states

Thermodynamics

Multiple length and time scales

Software issues – long term 2015/2020

Completely new algorithms are required

Software issues – short term (2009/2011)

Current quantum codes do not scale beyond O(1000) cores

Exploitation of hierarchical parallelism

Impact of machine changes

Distributed memory codes have transitioned well across several generations up to O(1000) cores

Major re-code will be expected to exploit Exascale

Computational Challenges and Needs for Academic and Industrial Applications Communities

Prof Mike Payne, University of Cambridge, Chemistry

Scientific and computational challenges

No response

Software issues – long term 2015/2020

Biggest challenge is going multi-scale with multiple localised quantum regions

Need databases

Software issues – short term (2009/2011)

Currently adding functionality to linear scaling code ONETEP

Compute/communications imbalance will cause some problems

Impact of machine changes

Scrapping dusty deck codes and starting from scratch with new codes has been a big win

Ensemble of multiple instances can be used to exploit Petascale and Exascale

Materials Science, Chemistry and Nanoscience
Gilles Zerah - CEA

Scientific and computational challenges

The scientific challenge is mostly to develop tools to achieve predictive descriptions of response of materials, in conditions of usage as well as in their fabrication process.

Another challenge is "computational" synthesis of new materials. The two main computational challenge are: spatial scalability (more or less ok) and temporal scalability (difficult)

Software issues - 2009

Techniques for which communication is minimal efficiently address new architectures (eg GPU). This impose the development of "localized" techniques and basis sets. This is not really an issue, but points to the **necessity of standard libraries based on localized basis sets adapted to these new architectures.**

Software issues – 2012, 2015, 2020

One can envision a more and more tightly integration of materials simulations at many scales (**the multiscale paradigm**). This is probably the direction to go to achieve temporal scalability.

On an horizon of 10 years, one of the principal challenge will be to seamlessly integrate those scales which will rely on different description of matter (quantal, atomistic, mesoscopic etc..) which in turn must be adapted to the new hardware.

An efficient "communication" tool has yet to be developed to allow for scalable communication between the different scales.

This view is common to many engineering fields, but materials simulation naturally involve **discrete constituents (atoms, molecules, defects etc..) in very large quantities, which is somewhat favorable to the use of massively parallel machines.**

Material Science
Thierry Deutsch - CEA

Scientific and computational challenges

- Yield an accurate description of electronic interactions
- Simulate large size atomic systems
- Simulate the kinetics and dynamics of those large systems

Summary of research direction

- Better approximation of the N-Body effect
- O(N) algorithms instead of $O(N^3)$

Potential scientific impact

- Better understanding of the kinetics and the dynamics of materials
- Find new molecules by means of exploration of new atomic configurations
- Predict new physics

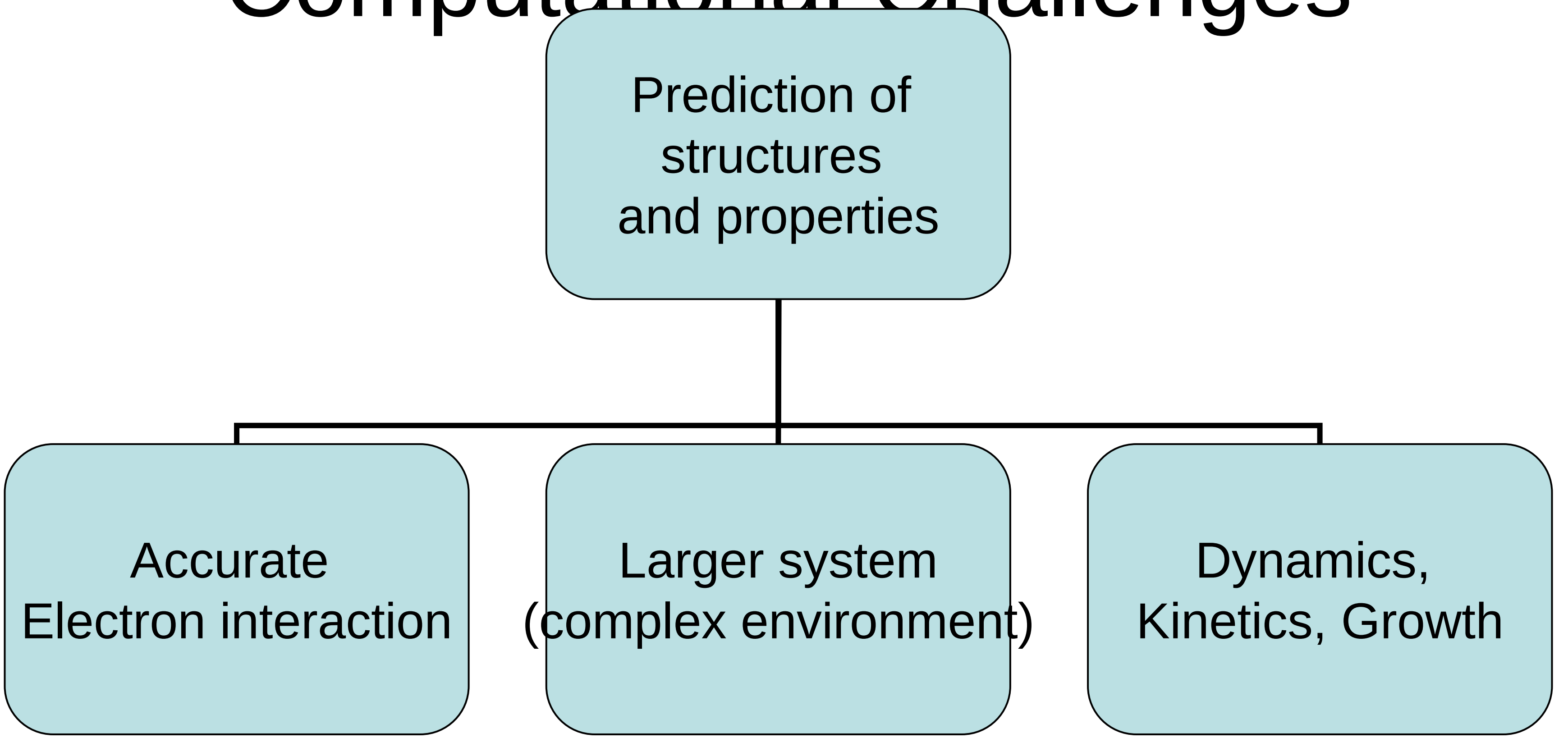
Potential impact on material science

- Better prediction of structures and properties of material for chemistry, materials ans nanosciences
- First results should be reached by the end of 2010.

Material Science

Thierry Deutsch - CEA

Computational Challenges



Astrophysics, HEP and Plasma Physics

2 contributions

Astrophysics: Bridging the many scale of the Universe

Expert name/affiliation - email: Edouard AUDIT, CEA/IRFU, edouard.audit@cea.fr

Scientific and computational challenges

Bridging the many scales of the Universe using simulations of increasing spatial and temporal resolution which include complex physical models ((magneto)hydrodynamics, gravity, radiative transfer, thermo-chemistry, nuclear burning,…)

- Physics of black hole and compact object
- Cosmology and large scale structures formation
- Dynamics of galaxies and of the interstellar medium
- Formation and evolution of star and planetary systems

Software issues – long term 2015/2020

- Scaling, especially for implicit solver
- Performances on special architecture (GPU, Cells,…)
- Manpower to follow the rapid change in programming paradigm
- IO, reliability (MTBF)
- Data handling, local vs. remote processing

Software issues – short term (2009/2011)

- Handling large data set (transfer, post-processing, visualisation)
- I/O on machines with over 10 000 core
- Scaling on a large number of cores (weak-scaling)
- Debbuging and optimisation on a large number of cores
- Shifting from memory to time limited runs
- NB: codes are mostly recent, some 10klines of source code + first hybrid CPU/GPU versions

Impact of last machine changes (several 10 Tflops -> 100+ Tflops)

- Design of a new I/O patterns
- Reduction of global communications
- Setup of a new local shared-memory system (256Gb) to post-process the data
- Hybrid (MPI/OpenMP) programming (not yet in production phase)

Computational Challenges and Needs for Academic and Industrial Applications Communities

Prof. S. Guenter   Max Planck Institute for Plasma Physics guenter@ipp.mpg.de

Scientific and computational challenges

- **Preparation and analysis of ITER discharges within days with resources between PF and EF.**
- **Advancement of plasma theory**

Software issues – long term 2015/2020

Evaluation of alternative, better scaling approaches e.g. multi grid, pure Monte Carlo methods

Software issues – short term (2009/2011)

- **Ensemble of various CFD solvers for 5 dim grid, FFTs**
- **Particle in cell approach, Monte Carlo codes in 5 dim phase space**

Technical Requirements

Extreme low latency for high communication requirements (high bandwidth less decisive)
Dedicated interconnect for synchronization and global operations required
Efficient and strong I/O system for handling of large input/output data in the PB range
In general weak scaling requirements
Multilevel of parallelism: Mixed mode possible to address core / node hierarchy
Pre- and post-processing: highly relevant

Life Sciences
5 contributions

Computational Challenges and Needs for Academic and Industrial Applications Communities

Prof. Sally Price, University College London, Biology

Scientific and computational challenges

Free energy simulations of thermodynamically feasible crystal structures of organic molecules

Ab initio methods for organic materials

Software issues – short term (2009/2011)

Software maintenance

Retention of key software developers

Software issues – long term 2015/2020

Funding of experienced and expert software developers in support of long-term software developments

Impact of machine changes

Limited by human resources

Stability, reliability etc of systems

Life Science - Simulations molecular ensembles

Prof. M. Scheffler - Fritz Haber Institut of the Max Planck Society
matthias.scheffler@fhi.mpg.de

Scientific and computational challenges

Simulations of 1-2 order of magnitude larger molecular ensembles (incl. solvent) over 2-3 orders of magnitude longer time scales are critical for:

• Structure prediction (e.g., protein conformation; combinatorial search for optimized structure of multinary materials)
• Direct simulation / evaluation of short-term dynamics,
• Meaningful parameterization of coarse-grained kinetic models (e.g., kinetic Monte Carlo, Markov models)
• lAdvancement beyond density functional theory in the local or generalized gradient approximation for large-scale problems

Software issues – short term (2009/2011)

• Efficient distribution of operations based on a real-space grid
• Fast robust eigenvalue solution for generalized, non-sparse eigenvalue problems

Software issues – long term 2015/2020

• Efficient parallel matrix algebra for "beyond-DFT" approaches
• Efficient parallel "distribution" of independent sub-processes with regular but infrequent data synchronization between runs
• Parallel post-processing of large amounts of output data (information collected during long molecular dynamics trajectories)

Technological requirements

• Extreme low latency for point-to-point communication operations Extremely fast global communication for synchronization of real-space grid based operations
• Dedicated interconnect for synchronisation and global operations required
• Large per-core memory for "beyond DFT" matrix algebra (large matrices, swapping to disk highly detrimental to performance)
• Efficient and strong I/O system for handling of large input/output data in the 10s of TB range

Protein Function Prediction : From sequences to structures
Michel Masella/CEA

Scientific and computational challenges

Regardless of the genome, 2/3 of its proteins belong to uncharacterized protein families.

Main goal : identifying the structure of these proteins and their biological partners => protein function prediction

- PLOS 2 (2004) e42 -

Software issues - 2009

Well established software for protein structure prediction : Modeller

⇒ Needs of high level of sequence similarity

Grand Challenge GENCI/CCRT 2009
CEA/DSV/IG-GNG

Software issues – 2011 and beyond

New bio-informatic algorithm => improving the proteinic structure prediction - SCOTCH software

- PNAS, 105 (2008) 7708 -

Refining protein structures and identification of protein partners using massive molecular dynamics simulations based on sophisticated force-fields - POLARIS(MD) code

- J Comput Chem 29 (2008) 1707 -

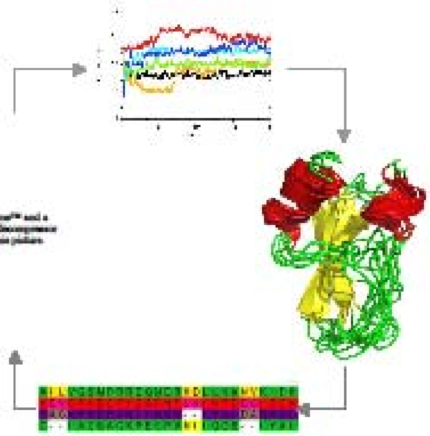
Coupling and scaling up both the approaches to propose a systematic functional annotation of new families

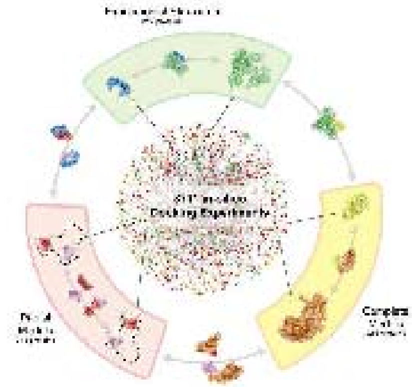
From sequences to structures : HPC Roadmap Michel Masella/CEA

2009	2011	2015 and beyond

Grand Challenge GENCI/CCRT

Proteins 69 (2007) 415





Identify all protein sequences using public resources and metagenomics data, and systematic modelling of proteins belonging to the family (Modeller software).

Improving the prediction of protein structure by coupling new bio-informatics algorithm and massive molecular dynamics simulation approaches.

Systematic identification of biological partners of proteins.

Computations using more and more sophisticated bio-informatical and physical modelling approaches ⇒ Identification of protein structure and function

2009	2011	2015 and beyond
1 family	**1 family**	**1 family**
5.10^3 cpu/~week	**5.10^4 cpu/~week**	**~ 10^{4*}KP cpu/~week**
		CSP : proteins structurally characterized ~ 10^4

# 25 Gb of storage	# 5 Tb of storage	# 5*CSP Tb of storage
500 Gb of memory	5 Tb of memory	5*CSP Tb of memory

Atomistic Simulations for Material Sciences and Biochemistry

Expert name/affiliation - email: Thomas SCHULTESS, CSCS, thomas.schulthess@cscs.ch

Scientific and computational challenges

- Strongly coupled electron systems
- More realistic free energy calculations

=> Application to material design, biochemistry

- Models are well know (quantum mechanics etc.), petascale codes are already running but numerical schemes that solve models in reasonable time are key (exponential complexity of models)
- Importance of strong scaling (time to solution) while being power efficient (CPU efficiency)

Software issues – short term (2009/2011)

- Codes are now ok for Petascale– parallelism that fits well on MPP machines
- Very high efficiencies in double or mixed precision were achieved on Jaguar/ORNL (up to 1.3 PF sustained w.r.t. 1.38 peak; i.e. > Linpack)

Software issues – long term 2015/2020

- Keep the ability to re-write or re-engineer codes with mixed teams (models, maths, s/w, h/w) and get suited funding for this

Since not every technology evolution is predictable, keep flexibility + capability of applications people to program

- Programming models or approaches able to harness heterogeneous cores/nodes, use both large memory nodes and address memory globally – how to further integrate partial promising approaches such as UPC, CUDA,OpenCL…
- Scalable and fault-tolerant communication (MPI or MPI-like)

Impact of last machine changes (1=Pflops ; 2=next/beyond)

1. major re-writing of codes; consolidation of "in situ" post-processing and data output filtering that lowered final I/O load
2. More code re-engineering, more in situ data processing co-located with computation

Computational biochemistry and molecular biology
T. Simonson – X/France

A few important scientific and computational challenges

Simulations of cellular nanostructures (10^6 particles: ribosome, spliceosome, molecular motors, etc) over microsecond timescales.

Prediction of the association modes of multiprotein assemblies from their component monomers, using extensive conformational searching and realistic energy functions and solvent models.

Ability to routinely simulate association/disassociation of libraries of biological complexes, such as protein:antibiotic or RNA:antibiotic librairies; requires microsecond turnaround for tens of ligands within a 24 hour timeframe (10^3 speedup from today)

Simulations of crowded, multicomponent cellular compartments: 10^7-10^8 particles, 100 microsecond timescales.

Software issues - 2009

Multilevel parallelism, integrating many nodes with many cores each

Better human interface; better integration of multiscale models

Weather, Climate, Earth Sciences

5 contributions

Computational Challenges and Needs for Academic and Industrial Applications Communities

Dr Adrian New, National Oceanography Centre Southampton, Ocean Science

Scientific and computational challenges

Global 1km model for internal waves

Global carbon cycle at hi-resolution

Large ensembles to address uncertainty in climate predictions

Software issues – long term 2015/2020

Mixed mode parallelism for maximum efficiency

Programming environments for accelerators

Software issues – short term (2009/2011)

Data handling, including input/output

Performance tuning tools

Impact of machine changes

New compilers for accelerator architectures

Data handling, including input/output

Mixed mode parallelism for efficient exploitation of multi-core nodes

Computational Challenges and Needs for Academic and Industrial Applications Communities

METEO-CLIMATOLOGY – Walter ZWIEFLHOFER / ECMWF – walter.zwieflhofer@ecmwf.int

Scientific and computational challenges

- High-resolution numerical weather prediction (NWP)
- Ensemble and high-resolution data assimilation

Software issues – long term 2015/2020

- Need for standard programming language's before giving-up with FORTRAN, MPI, …
- Need for new algorithmic approaches, allowing to look for the most adequate computer for solving the NWP problem

Software issues – short term (2009/2011)

- Next procurement (2013): going from 10^4+ to 10^5+ cores
- Parallel methods for minimization problems (data assimilation, i.e. strong scaling)
- Load-balancing methods at the lowest possible level, not at the programming level
- Effective performance analysis tools for 10^4-10^6 cores

Impact of last machine changes (37 Tflops -> 310 Tflops)

-No problem with I/O

-Still ok with parallelization paradigm (weak scaling for most parts)
- Incremental methods for data assimilation present the greatest challenge

Earth System Modeling

Mark Taylor, Sandia Nat. Labs., mataylo@sandia.gov

Scientific and computational challenges

Improved climate change predictions (decadal and long term) with reduced uncertainty, improved uncertainty quantification and better regional information.

Assess impacts of future climate change due to anthropogenic forcing and natural variability: global warming, sea level changes, extreme weather, distribution of precipitation, ice and clouds, etc…

Software issues – short term (2009/2011)

Short term issues dominated by scalability bottlenecks (i.e. strong scaling):

Largest bottleneck is existing atmospheric dynamical cores based on numerics, limited 1D domain decompoistion and insufficient scalability past t O(1K) cores. Ocean barotropic solver is stiff and limits scalability to O(10K) cores. Modern parallel I/O support needed in many legacy components. Scalability will now be required in every routine, impacting many previously computationally insignificant legacy procedures.

Software issues – long term 2015/2020

Hybrid architectures require new programming models to expose all possible levels of parallism. Time-stepping bottleneck (perfect weak scalable models have linear reduction in simulation rate) becomes dominant. Exascale software needed for handling adaptive, multiscale and multiphysics approaches to simulation, data workflow and visualization.

Impact of last machine changes (100 Gflops -> 100 Tflops)

MPI/Fortran model still effective with some benefit from hybrid MPI/openMP model. Short term scalability bottlenecks identified (left panel) now become significant and have motivated much progress on these issues.

Limited scalability of existing models allows for increased focus on ensembles including multi-model ensemble, with dozens to hundreds of members.

Eflops machines with a petascale-ready Earth system model will allow for ensembles of regionally resolved century long simulations for improved uncertainty quantification and assessment of regional impacts of climate change.

Computational Challenges and Needs for Academic and Industrial Applications Communities: Weather and Climate

Christopher Kerr, NOAA, chris.kerr@noaa.gov

Scientific and computational challenges

Scientific goal: development of a global cloud resolving model for the study of climate change. The scientific goals for the project fall into two broad categories: improved estimates of cloud feedbacks and thereby improved estimates of the overall magnitude of climate sensitivity; and improved projections of the patterns of regional changes in precipitation and other aspects of hydrology, including extreme weather events.

Computational Challenges: The underlying requirement is the need for scalable algorithms for all components of the model infrastructure. This algorithmic development has been under development at the Geophysical Fluid Dynamics Laboratory, Princeton for the last several years. The software infrastructure needed to develop and support all phases of the climate experiments: pre-processing, post-processing, and model infrastructure does require significant institutional commitment.

Given the scientific and computational challenges of this multi-year project we expect this activity to evolve into a community project that will allow researches to utilize the tools under development to study climate change and related issues.

Software issues – past two years

Over the past two-years, the primary software issues addressed in the models have included implementation of:

•algorithmic schemes that scale with increasing numbers of cores

•hybrid programing model (MPI and OpenMP)

•scalable memory schemes

•scalable I/O schemes

Software issues – long term 2015/2020

Over the next couple of years, we expect to address the following software issues:

•continued development of the above schemes.

•study the implementation of different programming methodologies.

•develop methodologies for improving the single and multi core performance of the model.

•develop scalable pre-packages and post-processing packages for the models. The current packages are written for single-processor platforms.

•develop different strategies for performing post-processing on the model output. The current models write the model diagnostics and post-processing is done in a different job-step. We need to explore how the post-processing can be done as the model runs.

Coupled Climate Modeling

Robert Jacob/Argonne National Laboratory - jacob@mcs.anl.gov

Scientific and computational challenges

Make predictions of future climate statistics (average temperature, precipitation) on global and regional scales for the next several decades.

Models developed separately by sub-disciplines (atmosphere, ocean) and then coupled.

Approximately 1 million grid points in each model and 100's of variables. 512 cores. Bound by both memory and network speed.

Software issues – short term (2009/2011)

Conservative numerical methods that can scale to 100K nodes while still maintaining useable simulation speed (approx 5 simulated years/day)

Propagate mixed mode programming through entire climate model (only present in some components).

Visualization on irregular and unstructured grids.

Debugging at scale

Workflow and metadata for models with 100's of possible configurations.

Software issues – long term 2015/2020

Heterogeneous node programming.

Performance portability.

Possible loss of bit-for-bit reproduceability.

Revisit output strategy (all variables at all points at regular intervals may not scale)

Fault tolerance.

More comprehensive unit and system testing.

Inherent treatment of uncertainty.

Impact of last machine changes (??flops -> ?? flops)

Most disruptive recent change was from vector to MPP (only 10 years ago). Climate model's consume flops by more detailed non-fluid processes (e.g. radiation) or adding resolution.

Gflops -> Tflops: massive increase in storage requirement.

Tflops -> Pflops: current viz/analysis tools will break.

Pflops -> Exflops: Merging of weather and climate scales in model resolution.

Methodology

Three ways to look at these issues

1. Preliminary (*i.e.* between the Paris and Tsukuba meetings): the (disciplinary) expert views
2. **A view transversal to all application domains: 4 main items**
3. Back to the disciplinary views: classification of issues with respect to expectation from the SW groups

A view transversal to all application domains: 4 main items

A. Validation – verification - uncertainty quantification Bill Tang leader

- compare with experiment, evaluate how realistic is the simulation. How software tools can help that ?
- visualisation

B. Mathematical methods Fred Streitz leader

- algorithms
- solvers

C. Productivity and efficiency of code production Rob Harrison leader

- load-balancing, scalability
- tools for code development (debugging, performance analysis)
- programming model for actual and next computer generation
- use of scientific libraries

D. Integrated framework Giovanni Aloisio leader

-multi-code/model/scale

-CAE-computation-Viz

- Workflows

A. Validation – verification - uncertainty quantification

Bill Tang leader

Verification & Validation (V&V)

- Establishing the physics fidelity of modern simulation tools requires strong V&V -- *Reliable codes demand solid theoretical foundations and careful experimental validation at all appropriate scales*

  - Validation assesses degree to which a code (within its domain of applicability) "describes the real world."

    --- *improves fidelity of computational models by systematic, quantitative comparisons with experimental measurements*

  - Verification assesses degree to which a code correctly implements the chosen physical model

    --- *addressing accuracy of numerical approximations, mesh/space and temporal discretization, statistical sampling errors, etc.*

    --- *code verification approaches also include:*

    (1) comparisons with analytic theoretical predictions

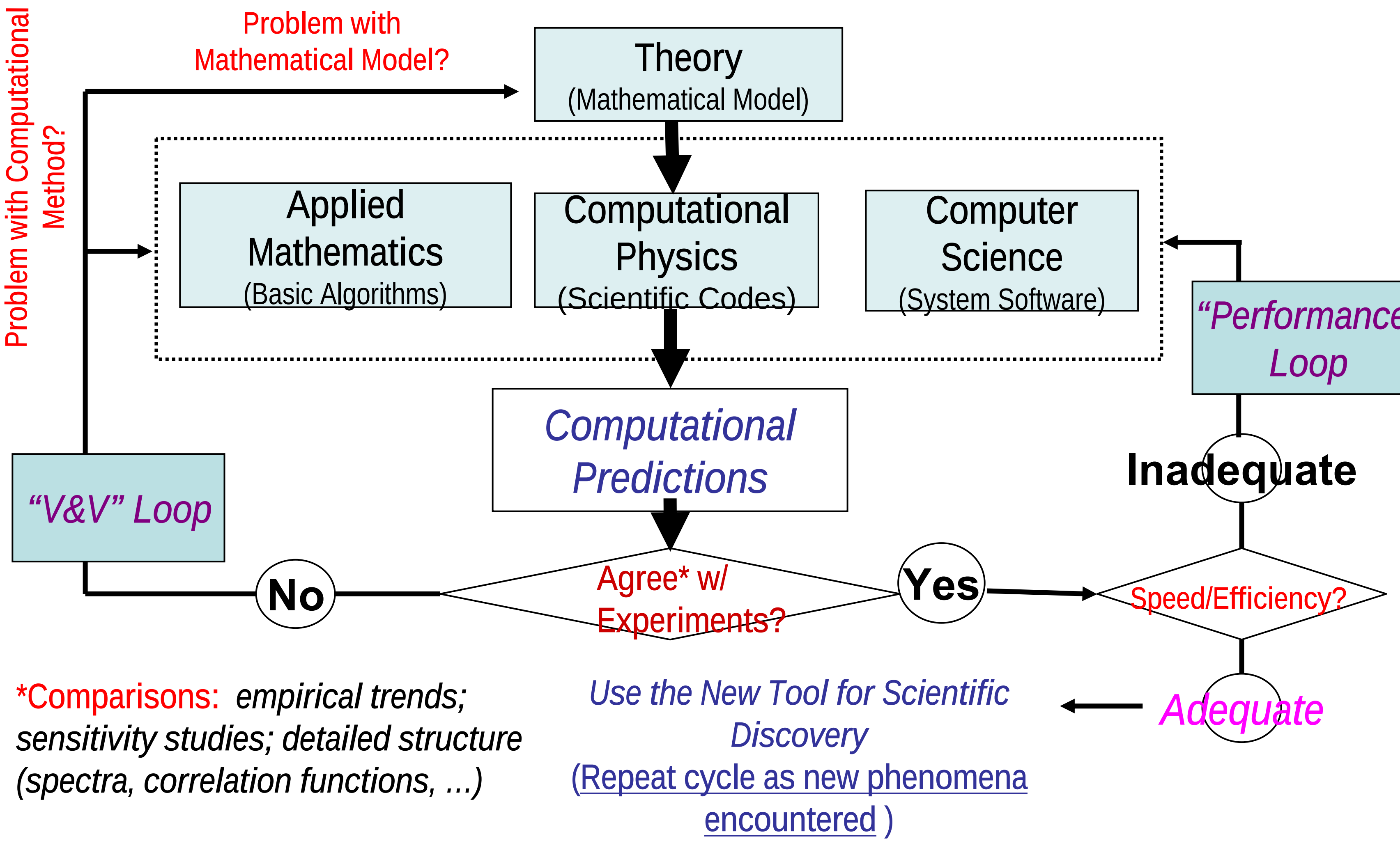
    e.g. – threshold/onset conditions for instabilities; etc.

    (2) cross-code benchmarking – involving codes based on different mathematical formulations/algorithms but *targeting the same generic physics*

    e.g. -- finite difference, finite elements, spectral methods, implicit schemes, etc. and/or models such as Kinetic [Particle-in-Cell, Vlasov/Continuum], Fluid [Hydrodynamic], Hybrid Kinetic-Fluid, etc.

  - Uncertainty Quantification (UQ) is a key element of the V&V process

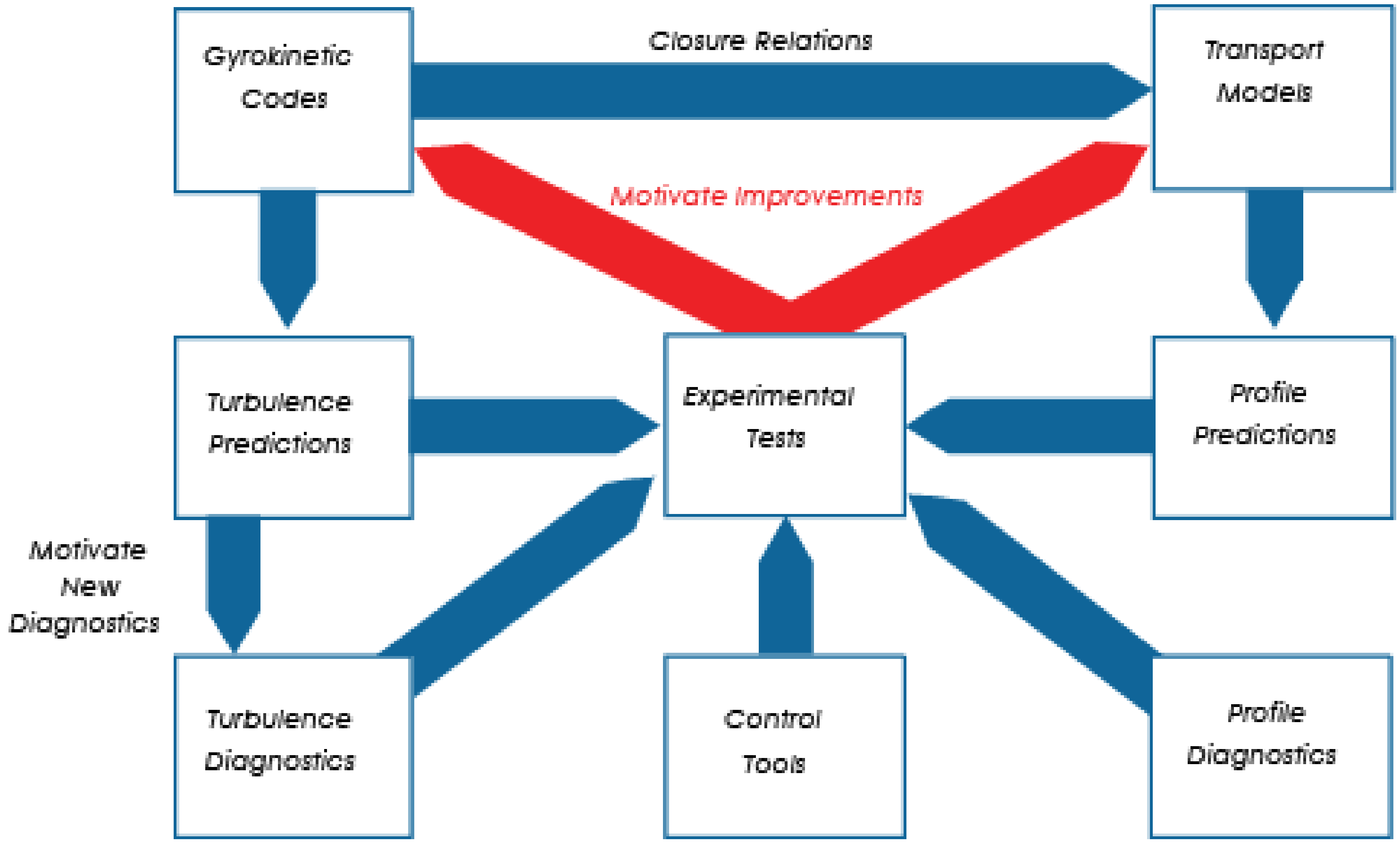
1. V & V within Advanced Scientific Code Development



V & V efforts require efficient Workflow environments with the capability to analyze and manage large amounts of data from experimental observations and from advanced simulations at the petascale and beyond.

Example of V&V from Fusion Energy Science

- *Combined Efforts from Theory/Modeling/Experiment for Development of Realistic Simulation Capability of Turbulent Transport in the Core Region of a Fusion System*



- *V & V efforts require efficient Workflow environments with the capability to analyze and manage large amounts of data from experimental observations and from advanced simulations at the petascale and beyond.*

UQ Defined

Uncertainty Quantification is the end-to-end study of the reliability of scientific inferences.

- Ideally, UQ results in

  (i) a quantitative assessment of that reliability,

  (ii) an inventory of possible sources of error and uncertainty in the inferences and predictions,

  (iii) an inventory of the sources of error and uncertainty accounted for in the assessment, and

  (iv) an inventory of assumptions on which the assessment is based.

- UQ studies all sources of error and uncertainty, including: systematic and stochastic measurement error; ignorance; limitations of theoretical models; limitations of numerical representations of those models; limitations on the accuracy and reliability of computations, approximations, and algorithms; and human error.

UQ with Extreme Computer Architecture

Scientific and computational challenges

Petascale models require Exascale UQ
Extreme data management
Usage model continuum from Exa-capacity to Exa-Capability

Summary of research direction

Develop new UQ methodologies
Change requirements for extreme scale HW/SW to reflect usage model
Couple development of UQ Pipeline, applications and scientific data mgmt & storage
Improve system IO balance

Expected Scientific and Computational Outcomes

New UQ methods with broad impact on every area of simulation science
Adjoint enable forward methods
Gaussian process models
Local approximations, response surface, filtering

Potential impact on Uncertainty Quantification and Error Analysis Problems that arise in various apps?

Enables use of extreme computing in a variety of usage models

Curse of Dimensionality

Scientific and computational challenges

Sampling of topological complexity in high dimensions (>100)

Maximizing information content/sample

Summary of research direction

- **Adaptive sample refinement**
- **Dimension reduction**
- **Variable selection**
- **Advanced response surface methodology**
- **Topological characterization techniques**
- **Embedded UQ, e.g., adjoint methods**

Expected Scientific and Computational Outcomes

- **Self-adapting, self-guiding UQ pipeline**
- **UQ-enabled application codes**

Potential impact on Uncertainty Quantification and Error Analysis Problems that arise in various apps?

Consistent uncertainty estimates in global climate sensitivity

- **Predicting regional climate impacts (hydrology) and extreme events**
-

B. Mathematical methods

Fred Streitz leader

Bulk of algorithm design work will be done *internally*

- development of innovative algorithms to solve both new and familiar problems at the exascale requires research in (and utilization of) applied mathematics,applied statistics,numerical methods, …

Certain desirable design elements can exploit X-stack (*external*)

- optimize data flow: tools to map cache use, to inform of cache hits/misses (with cost), need for software stack to hide latency, for user- accessible tools to manage memory hierarchy
- exploit coarse/fine grain parallelism: parallelization parameters resulting from hardware expressed in way that can be incorporated into algorithms, option of hand/auto tuning
- load-balance aware: tools/hooks to that provide tuning information (user managed load-balance), "Automagic" load balancing (OS managed load-balance) design for load balance first
- utilize mixed/variable precision: user specifies precision requirements, at a minimum: information available to users about int/double/single resources available, at best: stack automatically uses correct hardware
- manifestly fault tolerant: failure information available to users, fault tolerant OS, MTBF info available to users, allow tuning of restart strategies, inimize need for full restart files?

C. Productivity and efficiency of code production

Rob Harrison leader

Scientific application user productivity

Key challenges	Summary of research direction
Remote interaction with HPC resources (data volume) Automating work flow Automating data analysis Non-expert use of complex codes	Data reduction methods and hierarchical representations Automation and expert systems including VV & UQ Evolution/sampling methods for rare-events Data analysis and mining methods

Potential impact on software component	Potential impact on usability, capability, and breadth of community
Tools for capturing and employing expert knowledge Exascale work flow framework (differs from petascale in 1000x volume and much broader deployment)	Exascale simulation moves beyond basic science discovery (knowledge creation, informing decisions)

Scientific application developer productivity

Key challenges	Summary of research direction
HPC entry barrier already too high Life-cycle cost of exascale codes Correctness and code quality Enabling rapid science innovation Breadth of science at exascale	Standard, transparent programming model for hybrid systems Resilient programming paradigms Scalable distributed-shared-memory environments (beyond local node) X-PACK: efficient & robust math libs

Potential impact on software component	Potential impact on usability, capability, and breadth of community
Reduced cost to develop & deploy exascale applications Rapid deployment of new exascale applications Inter-operable science components	Many more disciplines at exascale Deep capability for critical sciences Capacity science enabled on tera and petascale subsystems

D. Integrated framework

Giovanni Aloisio leader

Integrated framework

- ➢ Support for multi-scale and multi-physics S/W
- ➢ Interoperability between scientific components (codes), between scientific components and transversal services (meshing, Visualization, Uncertainties Quantification, Data Assimilation, …)
- ➢ Ability to instantiate the framework for dedicated usage/community
  - Component programming model and standard/portable implementation of the execution model
  - Tools for defining and supervising workflows (coupling scheme)
  - Common data model and associated libraries for data exchange
  - Transparent access to computing power (massive and distributed)

Methodology

Three ways to look at these issues

1. Preliminary (*i.e.* between the Paris and Tsukuba meetings): the (disciplinary) expert views
2. A view transversal to all application domains: 4 main items
3. **Back to the disciplinary views: classification of issues with respect to expectation from the SW groups**

High-Energy Physics,
Astrophysics and Plasma Physics

High Energy Physics

Key challenges

- To achieve the highest possible sustained applications performance
- Exploiting architectures with imbalanced node performance and inter-node communications
- To develop multi-layered algorithms and implementations to fully exploit on-chip (heterogeneous) capabilities and massive system parallelism
- Tolerance to and recovery from system

Summary of research direction

Applications community will be involved in developing:

- Multi-layer, multi-scale algorithms and implementations
- Optimised single-core/single-chip routines for complex linear algebra
- Support for mixed precision arithmetic
- Tolerance to numerical errors to exploit eg GPU/accelerators
- Data management and standardization for shared use

Potential impact on software component

Generic software components required by the application:

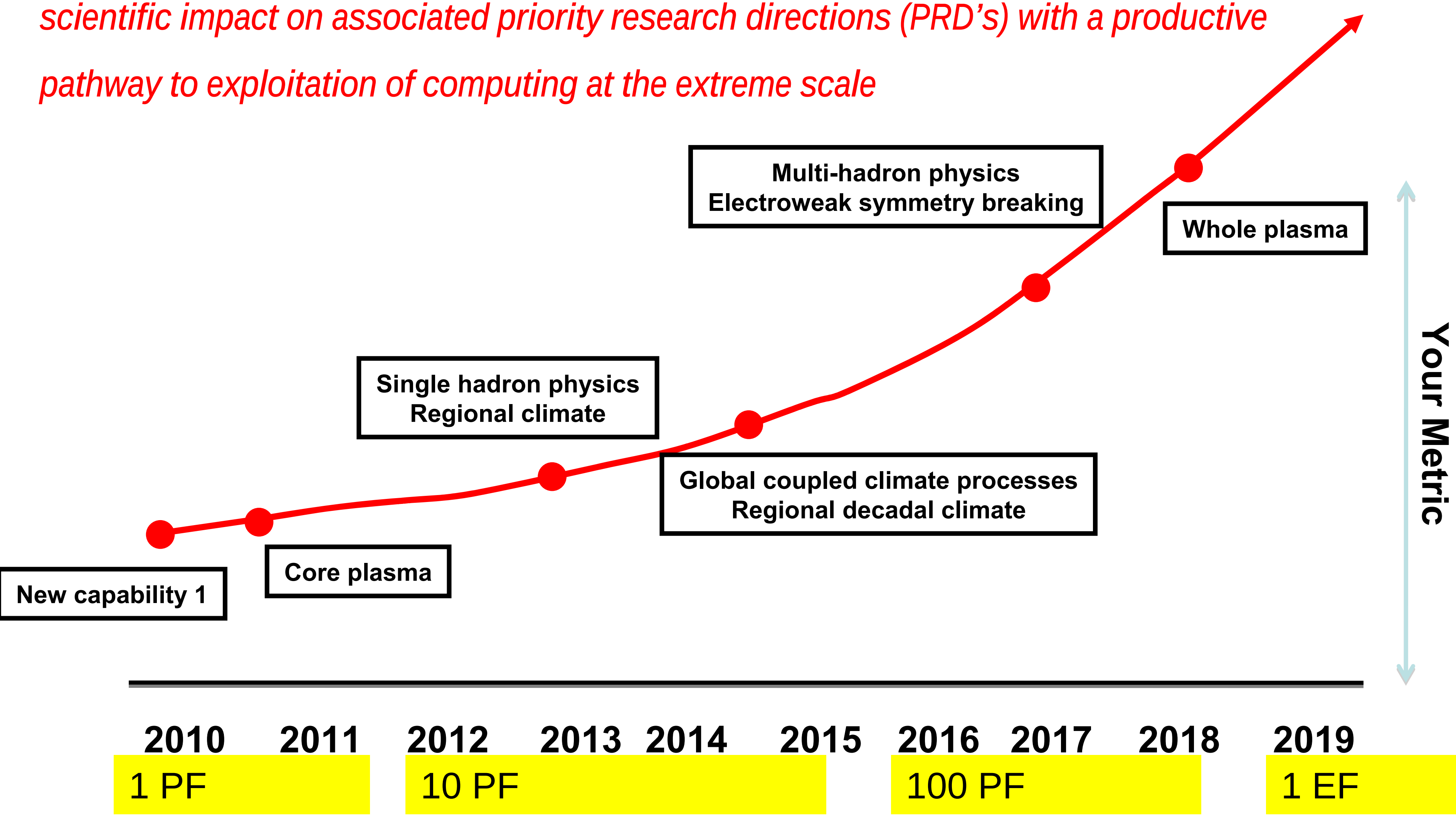
- Highly parallel, high bandwidth I/O
- Efficient compilers for multi-layered parallel algorithms
- Automatic recovery from hardware and system errors
- Robust, global file system

Potential impact on usability, capability, and breadth of community

- Stress testing and verification of exascale hardware and system software
- Development of new algorithms
- Reliable systems
- Global data sharing and interoperability

Pioneering Applications

Pioneering Applications with demonstrated need for Exascale to have significant scientific impact on associated priority research directions (PRD's) with a productive pathway to exploitation of computing at the extreme scale



Materials Science, Chemistry and Nanoscience

Challenges for materials, chemistry and nano community

- Transition codes from replicated, dense data structures to distributed, sparse data structures
  - *Runtime, programming models, libraries*
  - Reduce algorithmic complexity to increase system size to nanoscale
- Transition from data-focused algorithms to compute-focused algorithms
  - *I/O, runtime, libraries*
  - Identification of characteristic motion and rare events in molecular dynamics
- Transition to less tightly coupled algorithms to increase strong scaling (at expense of computing)
  - *Programming models, libraries, runtime*
  - Stochastic sampling of multiple coupled trajectories
  - Extends effective time scale of simulation

Challenges for materials, chemistry and nano community

- Transition to hybrid/heterogeneous parallelism to expose scalability in algorithms
  - *OS, Runtime, programming models, languages*
  - Overlapping execution of multiphysics codes
  - Expressing and managing fine-grained concurrency
  - Gain factor of 1000 in parallelism?
- Develop new data handling paradigms
  - *I/O, runtime, programming models, frameworks, libraries*
  - can't save everything – need to carefully design the simulation
  - Data reduction must occur prior to post-analysis
  - need embedded analysis/visualization
- Transition to multiphysics codes
  - *Frameworks, libraries, I/O, programming models*
  - Mission-driven science demands greater interoperability between disciplines
  - Device level simulations couple physics/chemistry/engineering co

Engineering

Computational Engineering Issues

Preliminary remark: different concerns between code developers, simulation environment developers, end users

Productivity.

- Programming model: *Exaflop machines will first run Petaflop grade apps (x1000 runs)*

⇒ *dealing with* ***hierarchical and heterogeneous*** *architectures* addressing **portability (functional & efficiency), maintainability** …. but using actual standards Fortran/C/C++, Python, MPI/OpenMP

- Debugging/perf. tools

- Fault Tolerance: strong fault tolerance for production (result within the night, non human interaction), weak fault tolerance for "reference" computations (run during several weeks/months, possible human interaction)

Computational Engineering Issues

X-Algorithms. Libraries, solvers, numerical method, algorithms: portable, efficient on cross architectures, unified interfaces

- multi-grid, better and simpler pre-conditioner
- new numerical methods for CFD: stochastic, SPH, FV
- Adaptive methods for heterogeneous platforms
- Advanced acceleration techniques,
- Coupling stochastic with determinist methods (Neutronic)

Verification and validation, UQ. *i.e.* dedicated slides

Rmqk: UQ type simulation needs management of very large data set and large number of data set:

Computational Engineering Issues

Integrated framework

- Framework: support for multi-scale and multi-physics S/W, interoperability between scientific components (codes), between scientific components and transversal services (meshing, Vis, UQ, DA, …), ability to instantiate the framework for dedicated usage/community
  - Component programming model and standard/portable implementation of the execution model
  - Tools for defining and supervising workflows (coupling scheme)
  - Common data model and associated libraries for data exchange
  - Transparent access to computing power (massive and distributed)
  - Meshing and visualization (pre and post)

Example: producing/adapting visualizing 50 billions of cells mesh for CFD simulation, impact on scalability, load balancing

Computational Engineering Issues

Other concerns:

- Need (more) dedicated high skilled HPC experts in application teams
- Keep the ability to re-write or re-engineer codes with mixed teams (models, maths, s/w, h/w)
- Strong links to be established/reinforced between high end computing facilities design and engineering communities in order to anticipate (at least 5 to 10 years) application breakthrough (through pioneers apps?)

Climate,
Weather, and Earth Sciences

Computational Climate Change Issues

From the application people (*internal*)

Model Development at exascale : Adopt a system view of climate modelling, improving model resolution, model physics, data analysis and visualization

Expectations from the software groups (*external*)

Productivity: All Climate models have to be rewritten for exascale =>Climate scientists would have to be parallel-computing experts unless the community can define software engineering guidelines encoded in community frameworks (software library in Physics and Numerics, new programming infrastructures to enable sustained extreme scale performance

How climate scientists can efficiently interact with the climate code (e.g. Exascale SDK and/or through advanced workflow tools)

Reliability: fault detection and resilience strategies in order to reduce the likelihood of undetectable errors, hardware checkpoint restart, Improved debugging tools

Performance: programming models and auto-tuning technologies for performance portability, fault resilience and a greater understanding of causality to understand performance

Load Balancing: efficient strategies

I/O: advanced parallel I/O support for many legacy components.

Scalability: scalable memory schemes

Programming models: Clarity in the programming model for exascale

Data management climate change issues

Data Storage: caching algorithms to move in/out data from dynamic storages providing high level of performance

Parallel File System: improvements in parallel I/O libraries (concurrency, scalability, bandwidth usage)

Parallel file systems are vendor specific => Integration issues in heterogeneous solutions! Open solutions…

Data movement : improvements in replication strategies, caching/replication schemes, optical connectivity

Metadata/Knowledge management*:* Efficient search algorithms (keyword based, full text, etc.)

Data analysis and visualization: mathematical & algorithms approaches and related parallel implementations able to scale with the high number of available processors

Active storage processing studies, software libraries to embed functions within storage, data analysis techniques (clustering, statistical analysis, etc.)