Infrastructure for Petascale Biology

Dan Reed
reed@renci.org

Chancellor’s Eminent Professor
Vice-Chancellor for IT
University of North Carolina at Chapel Hill

Director, Renaissance Computing Institute (RENCI)
The most obvious differences between different animals are differences of size, but for some reason the zoologists have paid singularly little attention to them. … But yet it is easy to show that a hare could not be as large as a hippopotamus, or a whale as small as a herring. For every type of animal there is a most convenient size, and a large change in size inevitably carries with it a change of form.

J. B. S. Haldane
“In the last two decades advances in computing technology, from processing speed to network capacity and the Internet, have revolutionized the way scientists work.

From sequencing genomes to monitoring the Earth's climate, many recent scientific advances would not have been possible without a parallel increase in computing power - and with revolutionary technologies such as the quantum computer edging towards reality, what will the relationship between computing and science bring us over the next 15 years?”
Presentation Outline

- Petascale and beyond hardware
  - scalable, reliable and effective
- Complex, multilevel models
  - integration and *in silico* designs
- Scalable parallel applications
  - reliable, fault tolerance and efficient
- Easy to use portals for community codes
  - web-based tool and service interfaces
- Petascale data federation and storage
  - shared access and ontologies
- Ruminations on culture and scale
  - enabling petascale science
You Might Be A Big System Geek If …

• You think a $2M cluster
  – is a nice, single user development platform
• You need binoculars
  – to see the other end of your machine room
• You order storage systems
  – and analysts issue “buy” orders for disk stocks
• You measure system network connectivity
  – in hundreds of kilometers of cable/fiber
• You dream about cooling systems
  – and wonder when fluorinert will make a comeback
• You telephone the local nuclear power plant
  – before you boot your system
How Big Is Big?

- Every 10X brings new challenges
  - 64 processors was once considered large
    - it hasn’t been “large” for quite a while
  - 1024 processors is today’s “medium” size
  - 2048-16192 processors is today’s “large”
    - we’re struggling even here
- 100K+ processor systems
  - are being designed/deployed
  - we have fundamental challenges …
  - … and no integrated research program
- Petascale data archives
  - the personal petabyte is very near
  - Vannevar Bush’s Memex: still prescient
- See the 2005 PITAC report
  - www.nitrd.gov
The Coming of Consumer Parallelism

- Technology trends
  - slower rise in clock rates
  - multicore processors
    - IBM Power5/6 and SUN UltraSPARC
    - Intel Duo and AMD Opteron
    - quad core and beyond are coming
  - reduced power consumption
    - laptop and mobile market
  - greater I/O and memory integration
    - PCI Express, Infiniband, …
  - Justin Ratter (Intel)
    - “100’s of cores on a chip in 2015”

- Moore’s law isn’t a birthright
  - CMOS scaling issues are now a challenge
    - power, junction size, fabrication line costs, …
Influence: The Computing Food Chain

• Our influence has waned
  – overall market growth/technology maturation
  – incrementalism generally flourishes
• Only big events can kill computing dinosaurs
  – diversified, evolutionarily successful ecosystem
• We still have some influence
  – new foundational technologies
  – revolutionary approaches
  – disruptive product niches
but in much different ways …

• Only long term, coordinated strategies …
  – can make a fundamental difference
    • software, architecture, systems
    • algorithms, data models, ontologies
    • databases, web services, …
A Tool and A Market for Every Task

- 200K Honda units at 5 KW equals a 1 GW nuclear plant

- Each targets different applications
  - understand application needs

- Our missing architecture program …
We’re Still Trying to Get There …

February 1994

Sequential  Terascale  Petascale

PETAFLOPS II
2nd Conference on Enabling Technologies for Peta(flops) Computing

February 15-19, 1999
Doubletree Hotel  Santa Barbara, CA
Conference Chair  Program Chair  Steering Committee Chair
Paul Messina, Caltech  Thomas Sterling, Caltech/JPL  Paul H. Smith, DOE

CHOSE A NUMBER FROM 0 TO 10 THAT BEST DESCRIBES YOUR PAIN

CHOSE THE FACE THAT BEST DESCRIBES HOW YOU FEEL
..., But Peak Petascale Is In Sight

- LLNL BlueGene/L (June 22, 2006)
  - 200 teraflops on material science calculations
  - Qbox first-principles molecular dynamics (FPMD) code
    - metals under extreme temperature and pressure

- ORNL “Baker” system (2008)
  - petascale peak performance
  - 24,000 socket, Opteron Linux system
    - Cray SeaStar 3-D interconnect

- NSF Petascale solicitation (2011)
  - sustained petascale performance
Petascale Hardware Challenges

- **Reliability**
  - rapidly growing system size
  - petascale systems
    - 50K-100K processor cores
    - ~1 petabyte of DRAM

- **Power management**
  - ~100 watts/processor (now)
    - multicore, lower power designs
  - petascale power
    - >20 MW plus cooling

- **Performance**
  - more than 100K parallelism
  - petascale application target
    - >10% of peak performance

- **Programmability**
  - 10 year software development timelines

David Barkai, Intel
Petascale and Beyond

- CMOS will support ~10 years of device progress
  - Moore’s law will likely hold at some level through 2016
- Current designs for petascale systems
  - IBM’s BG/P&Q and PERCS, Cray’s Cascade
  - SUN’s Hero and SGI’s Ultraviolet
  - large extrapolations of current parallel designs
- Architecture challenges are rising
  - “Internet in a Box” is the only model at present
- Petascale and beyond
  - power and cooling, packaging density
  - exposed concurrency in models
  - scalability of system software
  - application tailoring and optimization
  - programmability and ease of use
Petascale Biology Challenges

DNA sequence

Protein sequence and regulation

Protein structure

Protein/enzyme function

Sequence Annotation

Message

Promoter

T

A

T

A

C

A

G

T

Q

Y

C

A

G

T

Homology based protein structure prediction

Molecular simulations

Data integration

Organs, Organisms and Ecologies

Bacteria and cells

Pathway simulations

Network analysis

Metabolic pathways and regulatory networks

Multi-protein machines
Big Questions for Big Answers

Source: DOE Genomes to Life
Lessons From Other Domains

• Applications developed over a decade
  – incremental changes
    • solvers, science modules, tools
  – evolving development teams
    • lossy knowledge transfer

• Programmed to LCD
  – lowest common denominator (LCD)
    • tools and “fads” come and go
    • MPI – the assembly language of parallel programming
  – multiple execution platforms
    • codes live longer than systems

• Increasingly multidisciplinary
  – science and module interaction
    • local and global component optimization
  – diverse needs and demands
    • large memory, high I/O, compute intensive, …
Software Complexity and Growth

- 1971
  - Feature Extraction and Simulation
  - Physics Analysis and Results
  - Large Scale Data Management
  - Worldwide Collaboration (Grids)

- 2001
  - ~7 Million Lines of Code (BaBar)
  - ~100k LOC
  - ~500 people (BaBar)

Source: Richard Mount, SLAC
Petascale Applications Recipe

• Choose a world-class team
  – 5-10 researchers and staff
• Add $10M-$30M
  – software development and testing
• Wait 7-10 years (don’t second guess)
  – code maturity and capability
• Serve to broader community
  – fund research using the code
• Remember the old adage
  – too many cooks (consortia) spoil the broth (code)
• A few examples
  – MILC (QCD), NWChem, NAMD (MD)
  – Cactus (physics), ZEUS (cosmology)
Predicting Performance at Petascale

• No existing examples
  – a leap in parallelism
    • more processor sockets, more cores per socket
      – hundreds of cores/chip coming (e.g., IBM Cell)
    • deep memory hierarchies, huge interconnects
  – heterogeneous designs
    • shared and distributed memory
    • conventional and specialized cores very different system balance properties.
  – “petascale model problems” identified, but …
    • “petascale codes” are not ready
    • are numerical methods ready?

• Amdahl’s law writ large
  – If only .001 % of execution time is non-parallelizable
    speedup limited to <10,000
Software: We Need A New Approach

- It’s time for a radical change
  - complexity is rising dramatically
    - highly parallel and distributed systems
    - multidisciplinary applications
  - we need to get ahead of the curve
    - think about coming technologies
- *Deus ex machina* software model
  - it’s not working effectively
    - it will scale even less effectively to petaflops
  - profound implications for software design
- Two levels of software
  - *hero petascale codes*
  - complex workflows
Carolina Bioportal

- **Bioportal features**
  - **web service access**
    - common bioinformatics tools
  - **extensible toolkit and infrastructure**
  - **national reach and community**
    - NSF TeraGrid deployment
      - www.tgbioportal.org
    - NIH Center for Exploratory Genetics
- **HapMap bakeoff**
  - **resample from HapMap haplotypes**
    - create individuals with properties of data
  - **model disease**
    - create large populations with families/select individuals
- **Complex workflows**
  - **Genes2Life (phylogenetic trees)**
  - **complex traits (obesity)**
Carolina Bioportal Infrastructure

- **Application Framework**
  - HTML files
  - Application XML description

- **User Workspace**
  - Remote File Access
  - User Profile

- **Bioportal**
  - Job History Database
  - Application Processing
  - Workflow Processing
  - Command files

- **Application Services**
  - GridFTP
  - Gatekeeper

- **Grid Framework**
  - MyProxy

- **Security and Account Management**
  - OGCE User Databases
  - Authentication, Grid credential

- **Application Processing**
  - Velocity files

- **Interface Generator**
  - PISE
  - Application XML description
Data Heterogeneity and Complexity

Genomic, proteomic, transcriptomic, metabalomic, protein-protein interactions, regulatory bio-networks, alignments, disease, patterns and motifs, protein structure, protein classifications, specialist proteins (enzymes, receptors), ...

Source: Carole Goble (Manchester)
Sensor Data Overload

- High resolution brain imaging
  - 4.5 petabytes (PB) per brain

Source: Robert Morris, IBM
PITAC Findings and Recommendations

• Traditional disciplinary boundaries within academia and Federal R&D agencies severely inhibit the development of effective research and education in computational science. There is a paucity of incentives for longer-term multidisciplinary, multi-agency, or multi-sector efforts stifles structural innovation.

• The National Science and Technology Council (NSTC) must commission the National Academies to convene, on a fast track, one or more task forces to develop and maintain a multi-decade roadmap for computational science.
The Path to Petascale

• An integrated roadmap
  – algorithms and applications
  – programming systems and tools
  – adaptive system software
  – scalable, efficient architectures

• We’re still far from that …
  – limited applications
  – brittle software and tools
  – Inefficient architectures
Some Truisms

• **Petascale science is BIG science**
  – **expensive, not cheap**
    • instruments, software, databases, people
    • long-term (decadal) commitments
  – **deep, not broad**
    • driven by strategic priorities and investments

• **NSF culture**
  – **small, rather than big science**
    • a thousand flowers not ten redwoods
  – **conservative and risk averse**
    • peer review is not without problems
Big Means What?

- Big projects are getting smaller!  
  - remember the effects of inflation 
  \[ P = C (1 + R)^N \]
- We need to think bigger!  
  - what is a >$100M project?
Taking The Long View

• **Go long**
  – sustained, long-term exploration
  – long-term, strategic plans and support
  – twenty years is about the right time scale
    • think about major defense procurements

• **Go deep**
  – multidisciplinary examination of problems
    • understand applications and needs
  – resource scale matched to problem scale
    – $50M-$100M/year is about the right monetary scale
      • for each substantive project

• **Implications**
  – ask strategic questions
  – make long-term commitments
Science: The Top Questions

- What Genetic Changes Made Us Uniquely Human?
- How Are Memories Stored and Retrieved?
- How Did Cooperative Behavior Evolve?
- How Will Big Pictures Emerge from a Sea of Biological Data?
- How Far Can We Push Chemical Self-Assembly?
- What Are the Limits of Conventional Computing?
- Can We Selectively Shut Off Immune Responses?
- Do Deeper Principles Underlie Quantum Uncertainty and Nonlocality?
- Is an Effective HIV Vaccine Feasible?
- How Hot Will the Greenhouse World Be?
- What Can Replace Cheap Oil and When?
- Will Malthus Continue to Be Wrong?
- What Is the Universe Made Of?
- What is the Biological Basis of Consciousness?
- Why Do Humans Have So Few Genes?
- To What Extent Are Genetic Variation and Personal Health Linked?
- How Much Can Human Life Span Be Extended?
- How Much can Organ Regeneration be improved?
- How Can a Skin Cell Become a Nerve Cell?
- How Does a Single Somatic Cell Become a Whole Plant?
- How Does Earth's Interior Work?
- Are We Alone in the Universe?
- How and Where Did Life on Earth Arise?
- What Determines Species Diversity?