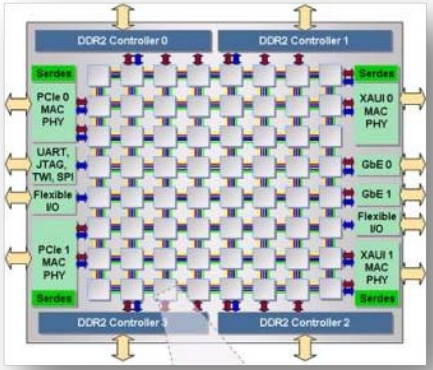
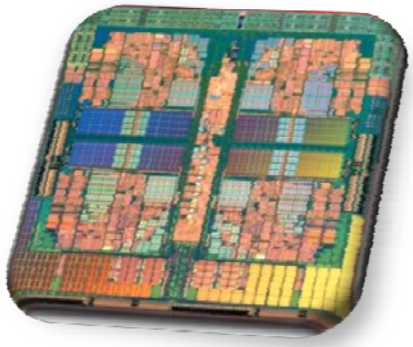
Multicore:
Let's Not Focus on the Present





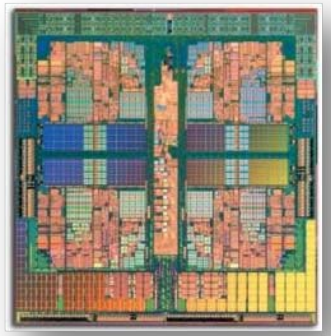
Dan Reed
reed@renci.org
www.renci.org/blog

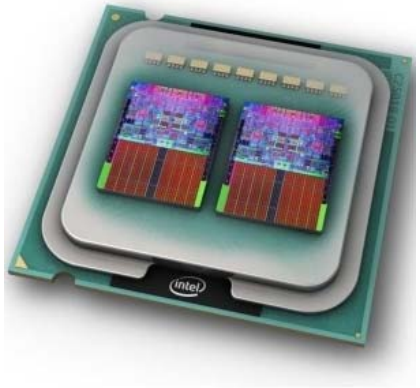
Chancellor's Eminent Professor
Senior Advisor for Strategy and Innovation
University of North Carolina at Chapel Hill
Director, Renaissance Computing Institute (RENCI)

Renci
Renaissance Computing Institute

Microsoft®

Your potential. Our passion.™

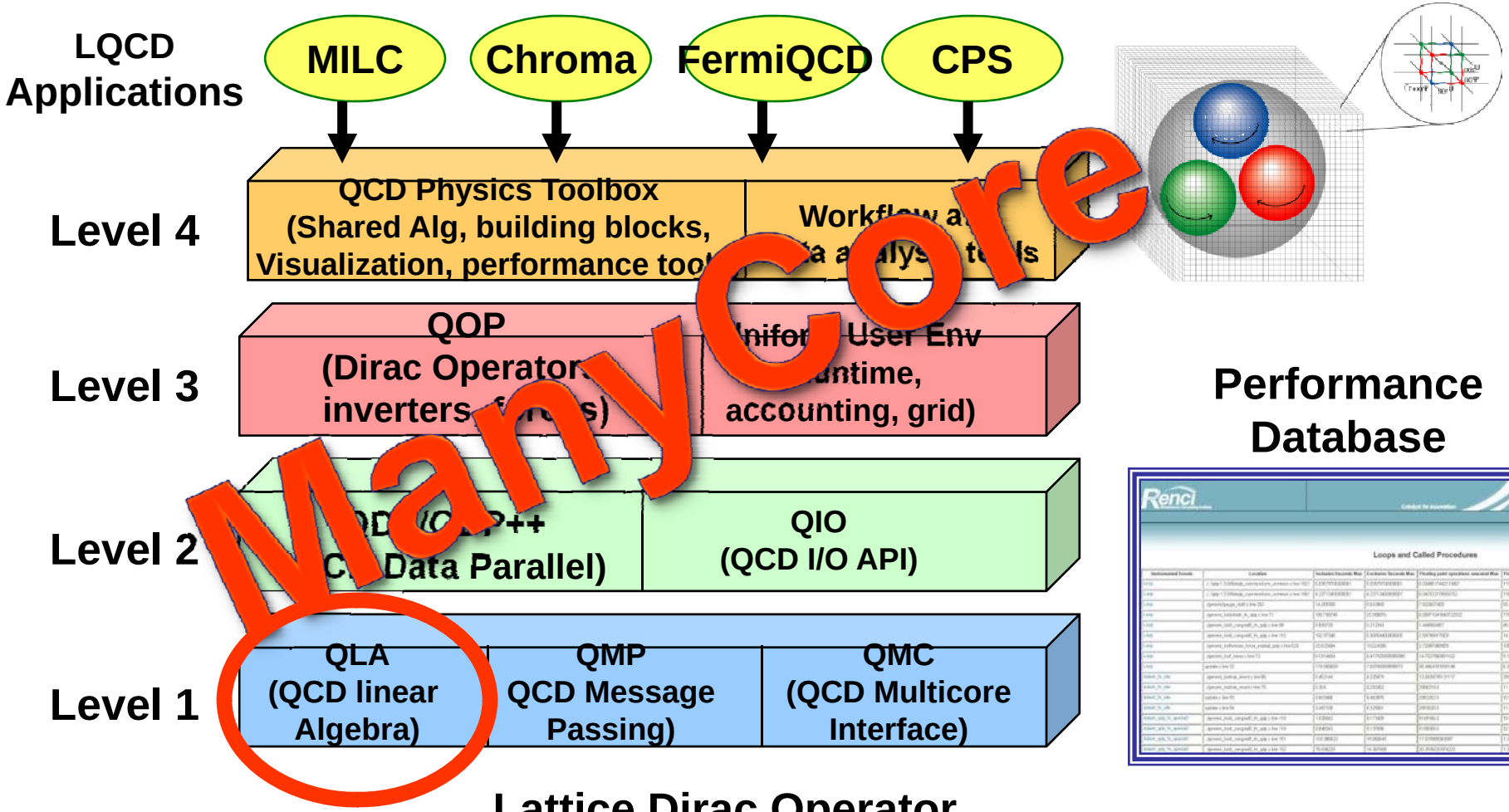




Renci

Renaissance Computing Institute

Lattice QCD Optimization



Lattice Dirac Operator

$$[D\psi]^{\alpha}(x) = \frac{1}{2a}\sum_{\beta,\mu}[U_{\mu}^{\alpha\beta}(x)\psi^{\beta}(x+\mu) - U_{\mu}^{*\beta\alpha}(x-\mu)\psi^{\beta}(x-\mu)] \quad \forall\alpha, x$$

Renci
Renaissance Computing In

Presentation Outline

"The future is here, it is just not evenly distributed." ***William Gibson***

- **Tools, culture and research**
- **Next generation applications**
- **Manycore heterogeneity**
- **Challenges and issues**

Renci Renaissance Computing Institute

Sapir–Whorf: Context and Research

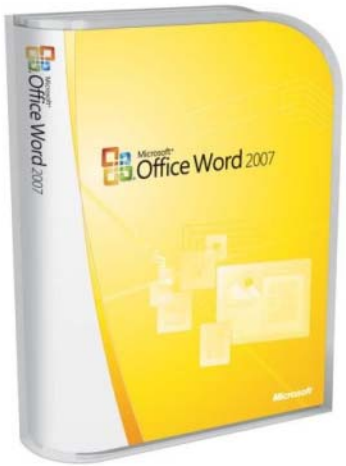
- **Sapir–Whorf Hypothesis (SWH)**
  - a particular language's nature influences the habitual thought of its speakers
- **Computing analog**
  - available systems shape research agendas
- **Consider some examples**
  - VAX 11/780 and UNIX
  - workstations and Ethernet
  - PCs and web
  - Linux clusters
  - clouds, multicore and social networks

Renci Renaissance Computing Institute

Post-WIMP Manycore Clouds

- **Mainframes**
  - business ADP
- **Minicomputers**
  - lab instrumentation
- **PCs**
  - office suites
- **Internet**
  - email, web …
- **The manycore killer app**
  - what's next?

- **It's *not* terascale Word™**



- **Exploiting**
  - hundreds of cores
  - heterogeneity
  - sensors

Renci Renaissance Computing Institute

Holistic Ecosystem Assessment

- **Applications**
  - WIMP and Linpack
- **Architecture**
- **Tools**
  - gcc, gdb
- **Services**
- **Systems**
  - Grids/clusters

- **Applications**
  - mobile services
- **Architectures**
  - heterogeneous manycore
- **Tools**
  - productivity frameworks
- **Services**
  - computing clouds
- **Systems**
  - massive data centers

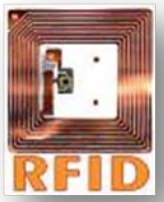
Renci
Renaissance Computing Institute

Convergence Device(s)

OR

Think About Mobility …

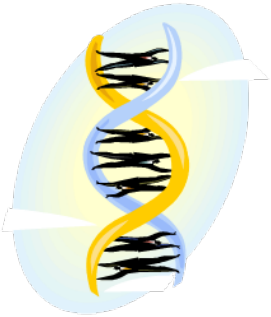
- **Technology drivers**
  - wireless communications
  - embedded processors
  - software services
- **Electronic tags and intelligent objects**
  - tags on everyday things (and individuals)
  - RFID, smart dust, …
- **Smart cars**
  - OBD II standard/Controller Area Network
  - navigation, active cruise control
  - road tracking, drowsy warning
- **Medical devices**
  - capsule endoscopy, ECG, pacemakers, …
- **Environmental sensors**
  - research and control



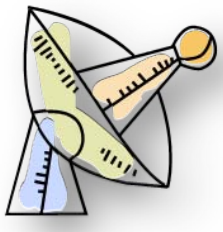
Renci
Renaissance Computing Institute

… and The Instrumented Life

- **Biological (static and dynamic)**
  - DNA sequence and polymorphisms (static)
  - gene expression levels (dynamic)
  - biomarkers (proteins, metabolites, physiological …)
- **Environmental**
  - air and pollutants, particulates
  - bacterial and viral distributions
  - food and liquids
  - mobility and exercise
- **Sociodynamic (physical and virtual)**
  - spatial dynamics
  - context and interactions
  - electronic infosphere

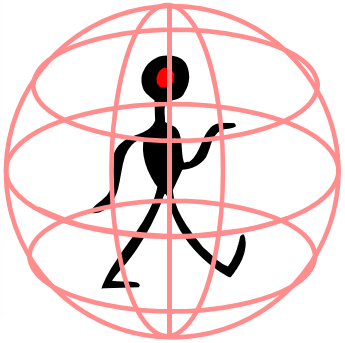


The Five Fold Way

- {Heterogeneous} manycore
  - on-chip parallelism
- Big, “really big” data centers
  - service hosting
- Web services
  - communities/capabilities
- Ubiquitous mobility
  - sensors, data and devices
- Bush’s Memex reborn
  - everywhere information
  - contextual, transductive

amazon web services™

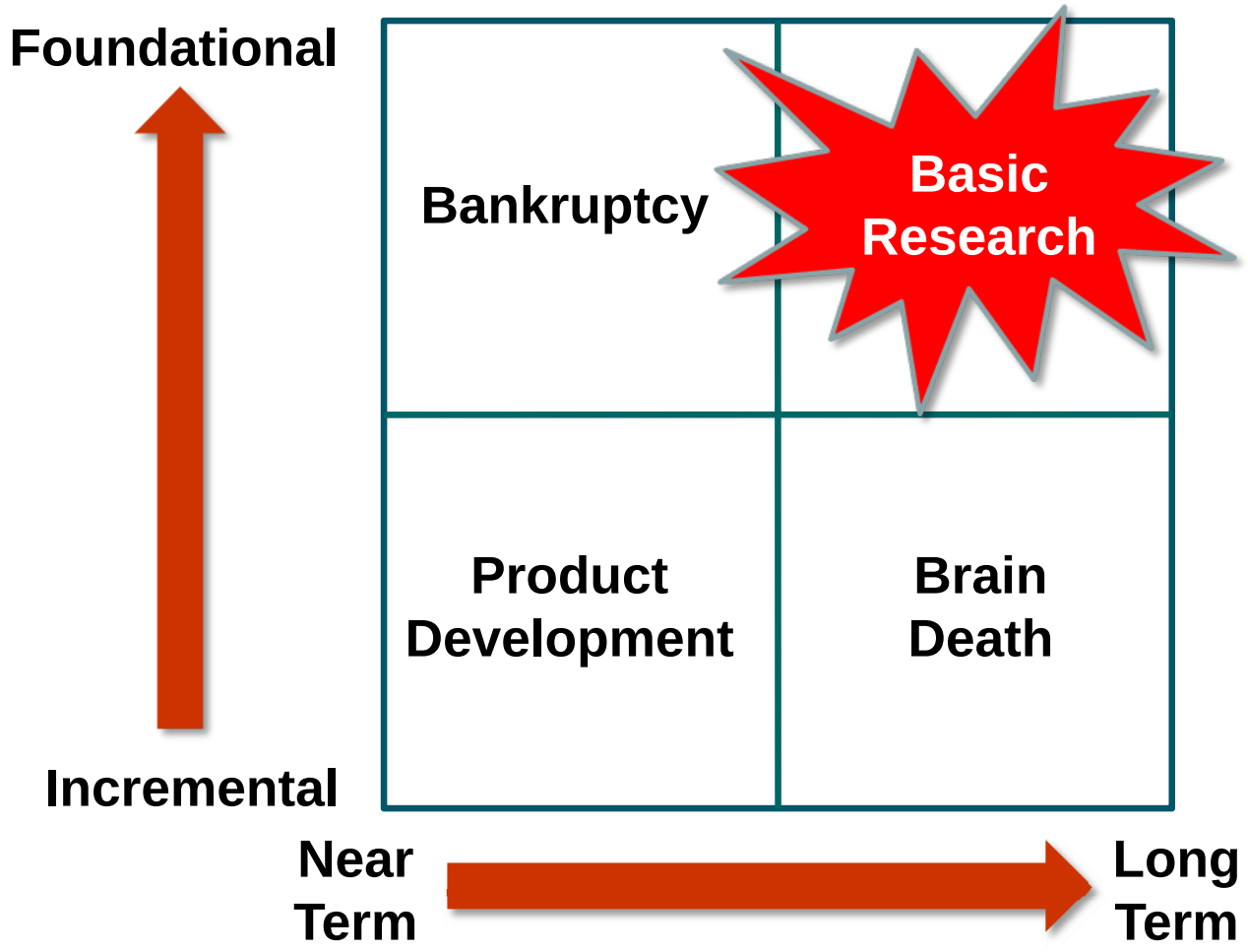


Renci Renaissance Computing Institute

Where'd The Big Visions Go?

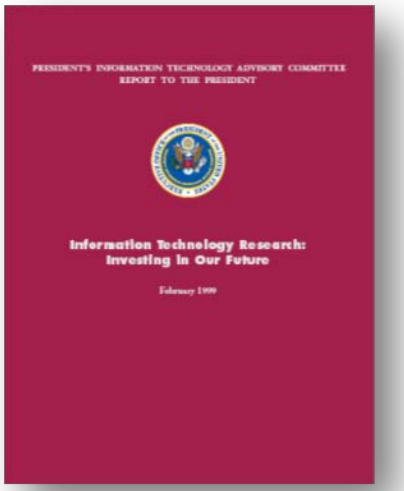
- **Remember …**
  - Project MAC
  - MULTICS
  - PLATO
  - ILLIAC IV
  - STRETCH
  - ARPANet
  - SketchPAD

Flavors of Innovation

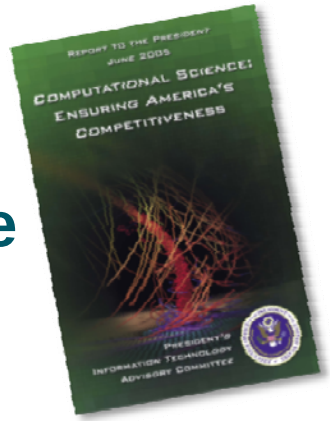


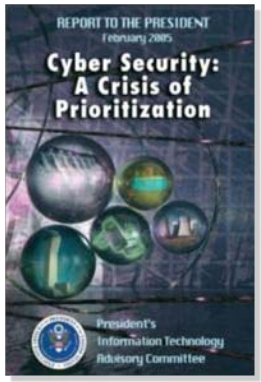
Prior NITRD Program Evaluations

- **PITAC's 1999 overall assessment**
  - ***Information Technology Research: Investing in Our Future***



- **During 2003-2005, focused PITAC assessments**
  - **health care and IT**
  - **cybersecurity**
  - **computational science**

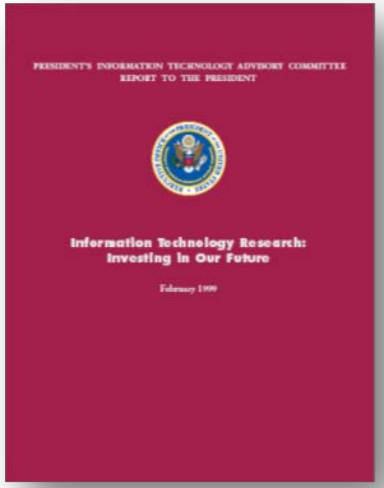




Renci
Renaissance Computing Institute

Kennedy Observations

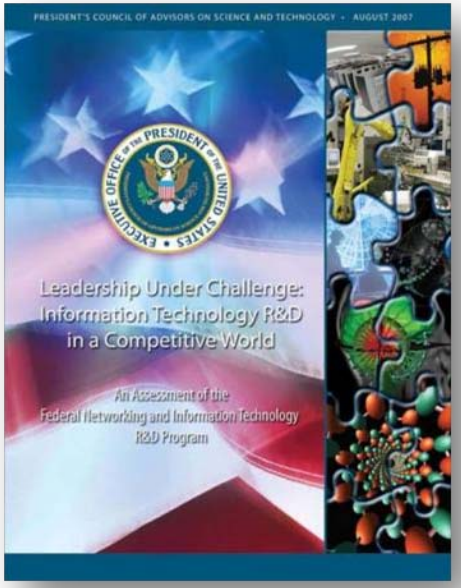
- **PITAC 1999 message: focus on long-term research**
  - **think big and make it possible for researchers to think big**
  - **increase the funding and the funding term**
    - unique responsibility of the Federal Government
- **Positive result: funding did increase**
  - **most of the measurable growth has gone to NSF**
  - **modes of funding diversified**
  - **new programs initiated**
- **Concerns**
  - **HPC software still not getting enough attention**
    - amounts and nature of funding
  - **Is the leadership and management adequate?**
  - **Are we returning to an era of short-term thinking?**



Renci Renaissance Computing Institute

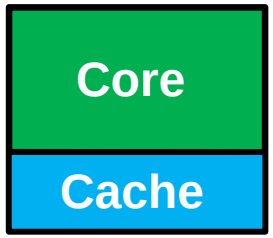
PCAST Recommendations

- **Revamp NIT education and training**
  - **new curricula and approaches to meet demands**
  - **increased fellowships/streamlined visa processes**
- **Rebalance the Federal NIT R&D portfolio**
  - ***more long-term, large-scale, multidisciplinary R&D***
  - ***more innovative, higher-risk R&D***
- **Reprioritize the Federal NIT R&D topics**
  - **increase**
    - systems connected with physical world
    - software, digital data and networking
  - **sustain**
    - high-end computing, security
    - HCI and social sciences
- **Improve planning/coordination of R&D programs**

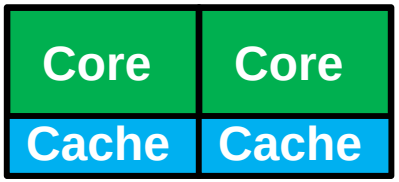


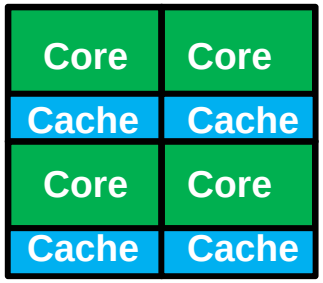
Renci
Renaissance Computing Institute

One, Two, Three, Many …

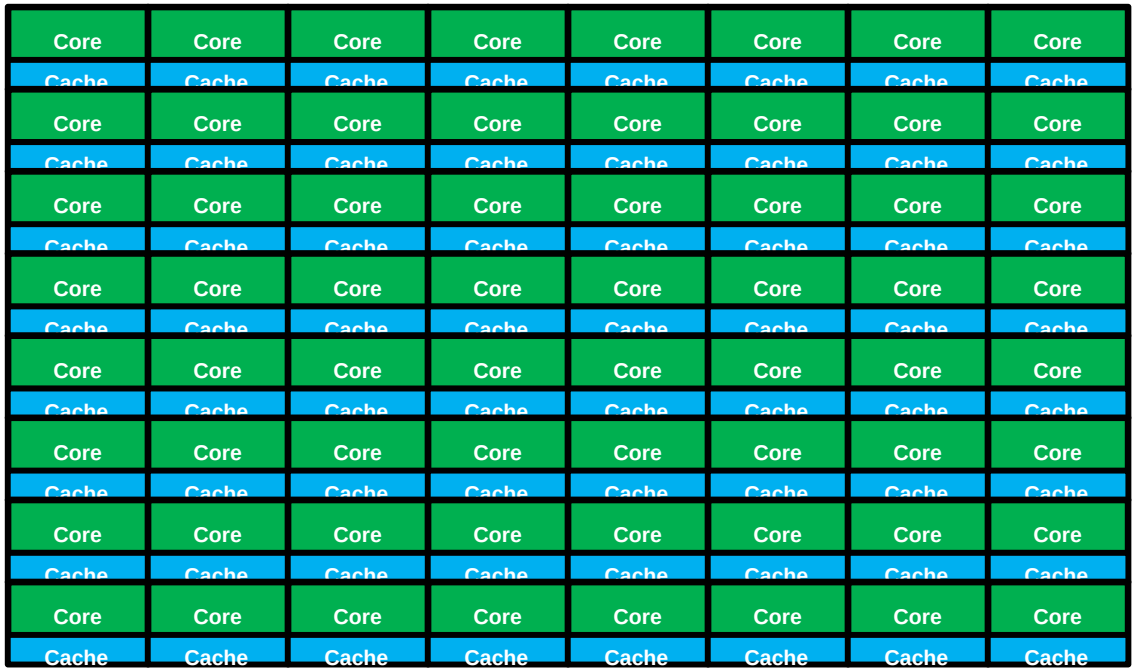


Single Thread





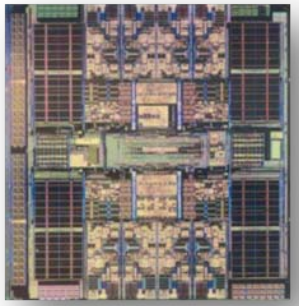
Single/Multiple Thread Balance



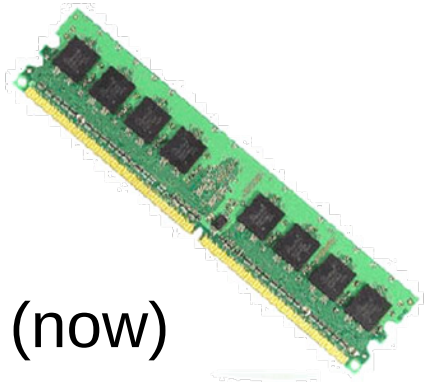
Serious Multithreading Optimization

Acknowledgment: Tim Mattson, Intel

Looking Forward …



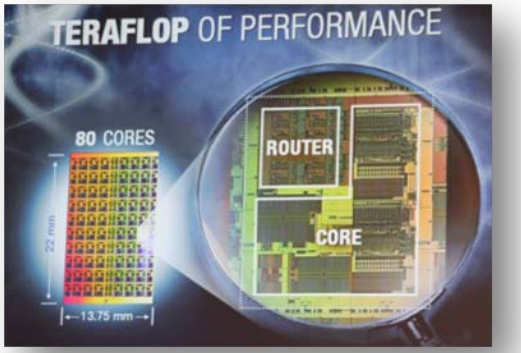
- **Cores**
  - **more, but simpler/smaller**
    - less out-of-order hardware, reduced power
  - **more heterogeneous**
    - multiple services
- **DRAM**
  - **getting bigger**
    - 64 Mb (1994) to Samsung 2 Gb DDR2 (now)
  - **but probably not enough faster**
    - 70 ns (1996) to Samsung DDR2 40-60ns (now)
  - **and banking has its limits (cost and pins)**

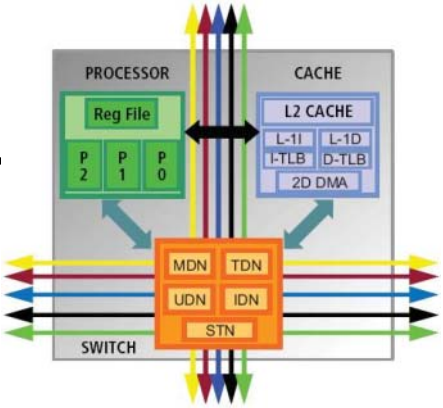


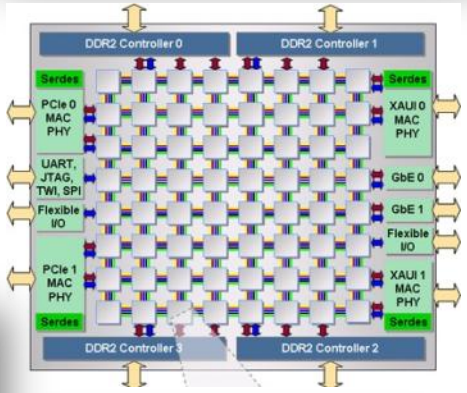
Renci Renaissance Computing Institute

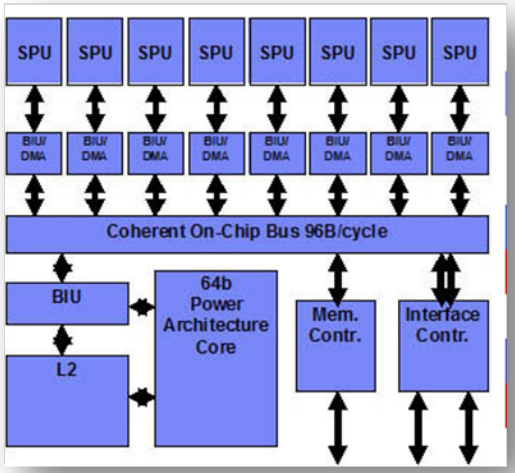
ManyCore Mashups

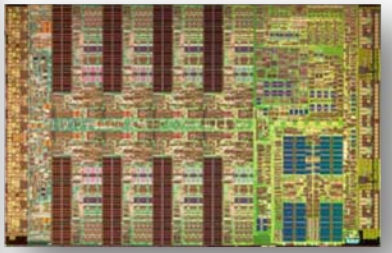
- **Intel's 80 core prototype**
  - 2-D mesh interconnect
  - 62 W power
- **Tilera 64 core system**
  - 8x8 grid of cores
  - 5 MB coherent cache
  - 4 DDR2 controllers
  - 2 10 GbE interfaces
- **IBM Cell**
  - PowerPC
  - and 8 cores











Architectural Futures

- **Replication of tweaked cores**
  - **interconnect (it really matters)**
  - **mix of core types**
    - heterogeneity and programmability
- **Or, more radical ideas …**
- **Other issues …**
  - **process variation and cores**
  - **performability**
    - performance and reliability
  - **dynamic power management**

Variability Cause and Estimated Impact on Delay

Time domain (sec)	Mechanism	Delay impact (3σ)
1×10^{12}	Lithography node	20%
1×10^{9}	Electromigration	5%
1×10^{8}	Hot electron effect	5%
1×10^{6}	NBTI	15%
1×10^{4}	Chip electrical mean variation	15%
1×10^{1}	Across-chip L_{poly} variation	15%
1×10^{-4}	Self heating/temperature	12%
1×10^{-8}	SOI history effect	10%
1×10^{-10}	Supply voltage	17%
1×10^{-17}	Line-to-line coupling	10%
1×10^{-11}	Residual S/D charge	5%

Source: IBM

Source: Semiconductor International

Renci Renaissance Computing Institute

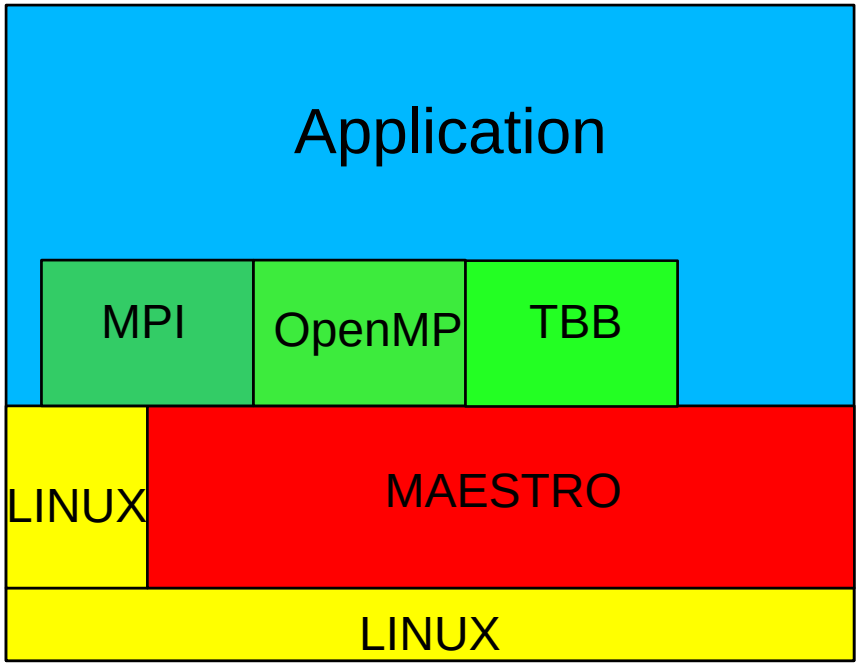
Maestro: Multicore Power Management

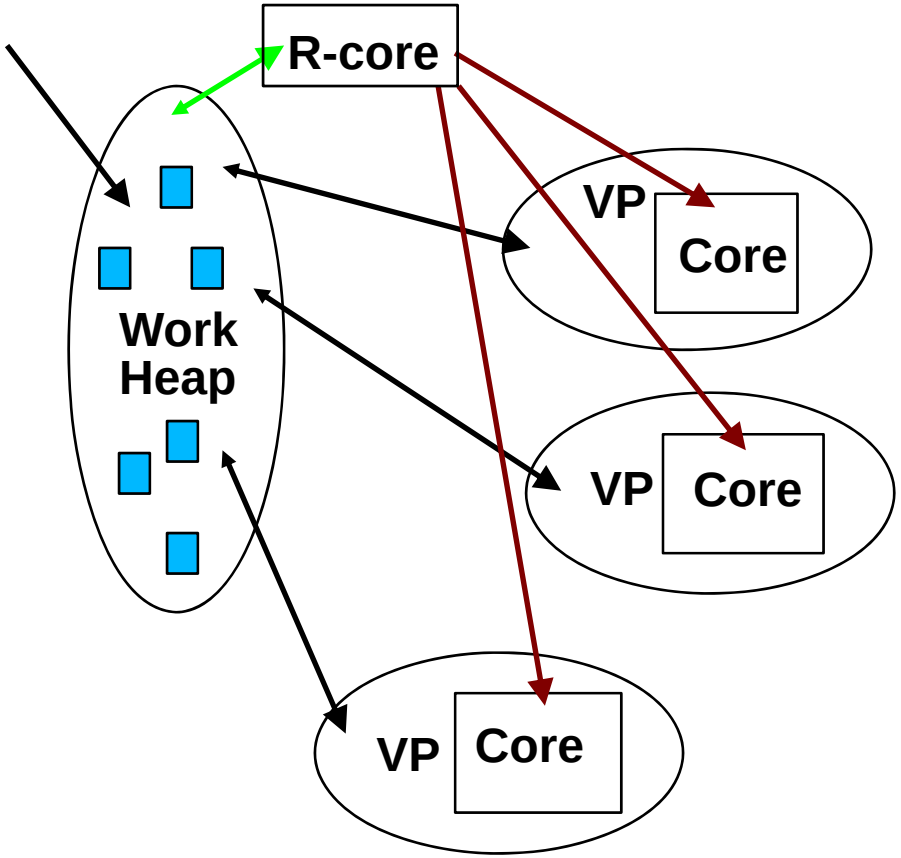
- **Approach**
  - **use “excess” computational power**
  - **Monitor/control application execution**
- **Concretely**
  - **manage power by turning cores down/off**
    - when performance limited
  - **manage parallelism to match available hardware**
    - over-virtualize threads for load balance
- **In the limit, memory performance constrains**
  - **monitor memory utilization and adjust frequency**

Renci Renaissance Computing Institute

Source: Alan Porterfield/Rob Fowler

Maestro Structure





Source: Alan Porterfield/Rob Fowler

Programming Models/Styles

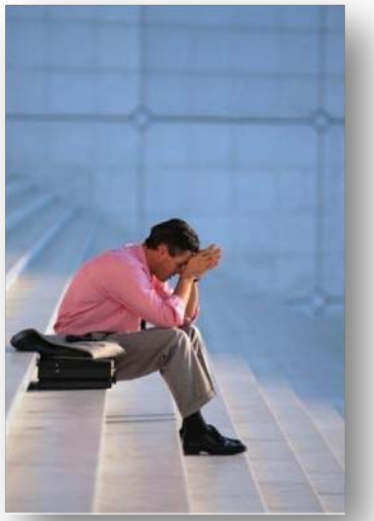
- Threads
  - several varieties
  - POSIX threads, May 1995
- Message passing
  - lots of vendor/research libraries (NX, PVM, …)
  - MPI, May 1994
- Data parallel
  - several dialects, including CM-Fortran
  - High-Performance FORTRAN (HPF), May 1993
- Partitioned global address space (PGAS)
  - UPC, CAF, Titanium …
- Functional languages
  - recently, F#
- Transactional memory
  - atomic/isolated code sections
  - lots of ferment; few, if any, standards

- Each creates resource definitions
  - input/output
  - communication
  - power/performance
  - scheduling
  - reliability

Renci Renaissance Computing Institute

Execution Models and Reliability

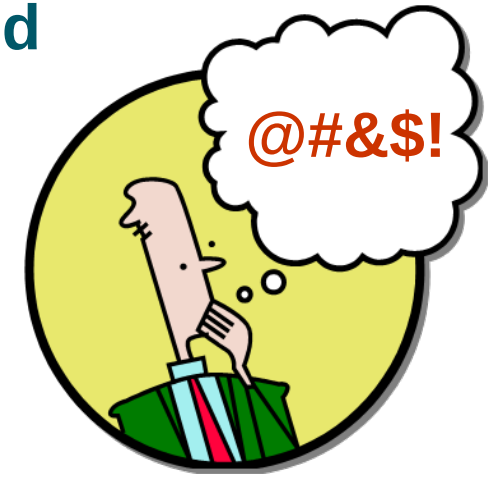
- **Accept failure as common**
  - ***integrated performability required***
- **Each model is amenable to different strategies**
  - **need-based resource selection**
  - **over-provisioning for duplicate execution**
  - **checkpoint/restart**
  - **algorithm-based fault tolerance**
  - **library-mediated over-provisioning**
  - **rollover and retry**



Renci Renaissance Computing Institute

A Gedanken Experiment

- **Select your ten favorite applications**
  - **measure the parallel execution time of each**
  - **rank the applications based on time**
- **Now, repeat for another system**
- **The rankings will be only semi-correlated**
  - **parallel systems are “ill conditioned”**
  - **wide variability and peak vs. sustained**
- **Why is this so?**
- **And, should/do we care?**



Renci
Renaissance Computing Institute

We're Speed Junkies

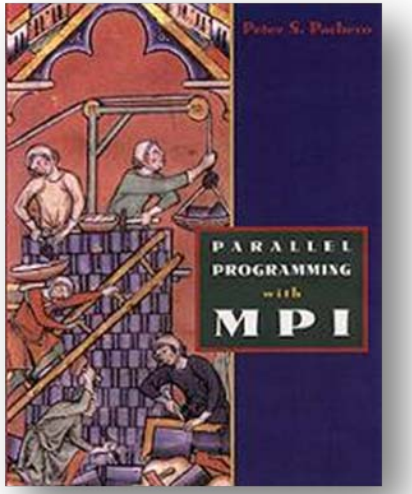
By sacrificing a factor of roughly three in circuit speed, it's possible that we could have built a more reliable multi-quadrant system in less time, for no more money, and with a comparable overall performance. The same concern for the last drop of performance hurt us as well in the secondary (parallel disk) and tertiary (laser) stores.

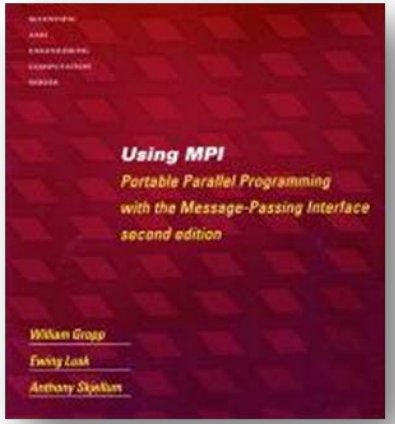
Dan Slotnick

Renci
Renaissance Computing Institute

MPI: It Hurts So Good

- **Observations**
  - **"assembly language" of parallel computing**
  - **lowest common denominator**
    - portable across architectures and systems
  - **upfront effort repaid by**
    - system portability
    - explicit locality management
- **Costs and implications**
  - **human productivity**
    - low-level programming model
  - **software innovation**
    - limited development of alternatives





Renci
Renaissance Computing Institute

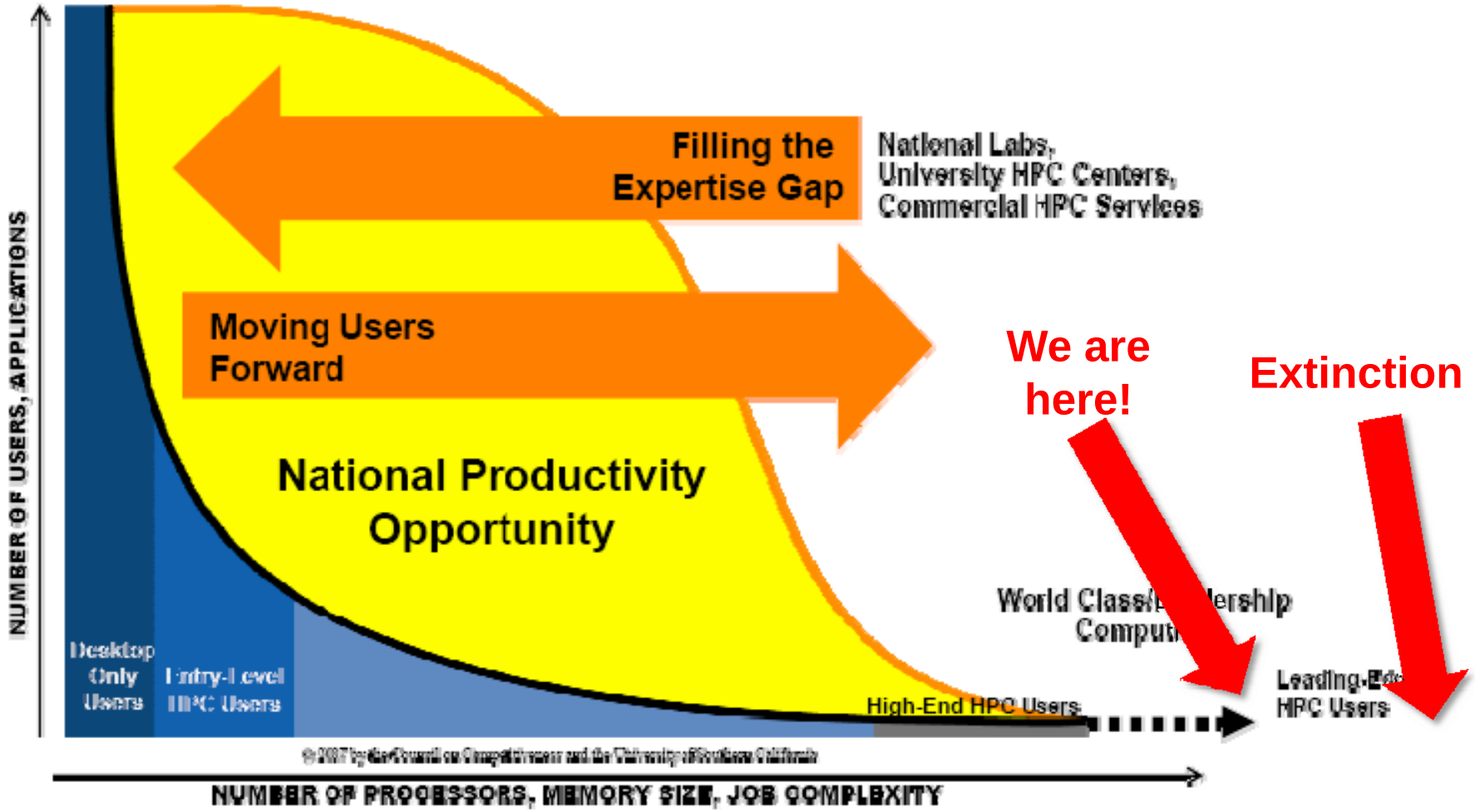
Choices, Choices, …

- **High performance**
  - **exploiting system specific features**
    - cache footprint, latency/bandwidth ratios, …
  - ***militates against portable code***
- **Portability**
  - **targeting the lowest common denominator**
    - standard hardware and software attributes
  - ***militates against ultra high-performance code***
- **Low development cost**
  - **cost shifting to hide human investment**
    - people are the really expensive part
  - **specialization to problem solution**
  - ***militates against portable, high-performance code***

Performance

Portability

Council on Competitiveness



Source: Council on Competitiveness

Virtualization and Programmability

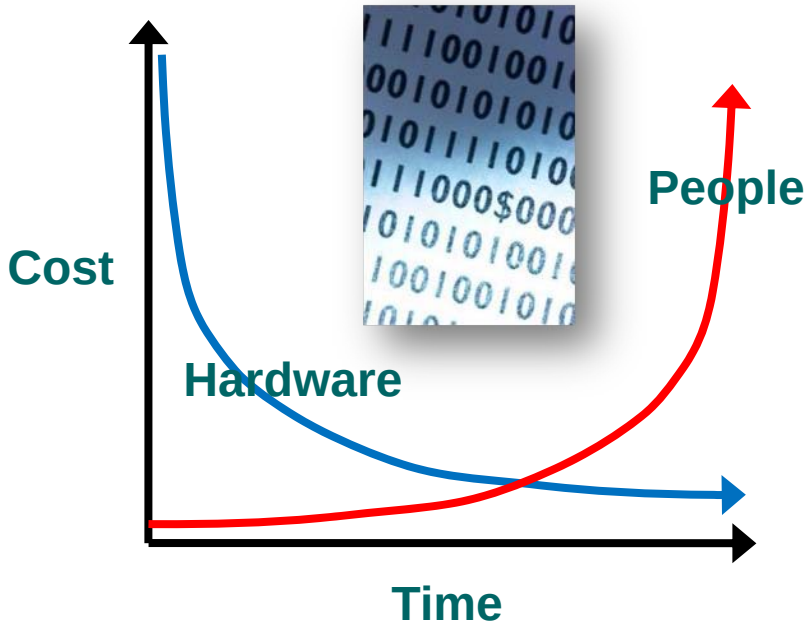
- **Simple quality of service (QoS)**
  - performance
  - reliability
  - power

- **Virtualization and complexity hiding**
  - user assertions/specifications
  - implementation/mediation
- **The great mashup**
  - cloud computing/clusters
  - multicore/ManyCore
  - software complexity

Renci
Renaissance Computing Institute

Economic Divergence/Optimization

- $/teraflop-year
  - declining rapidly
- $/developer-year
  - rising rapidly
- Applications outlive systems
  - by many years



- Machine-synthesized and managed software
  - getting cheaper and more feasible …
- Feedback directed optimization
  - an older, based on run-time data
  - increasingly blurred compilation/execution boundaries
  - deep optimization (hours, days, weeks …)

Renci Renaissance Computing Institute