Bridging Gaps

The Role of Computer Science
In
Scientific Computing

David Greenberg

'A slow sort of country!' said the Queen. 'Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!'
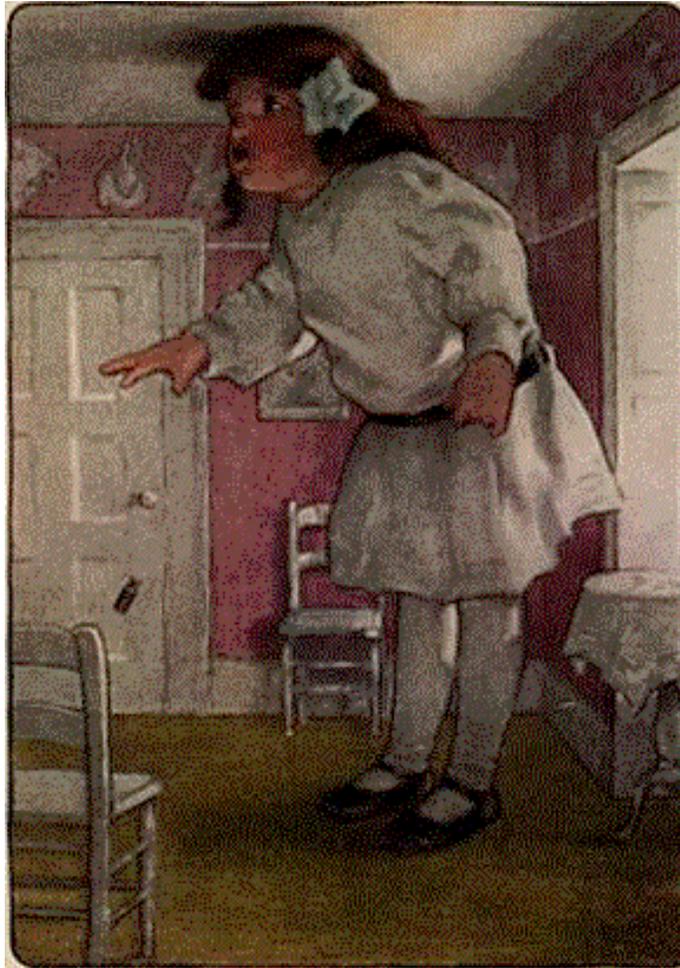
REDQUEEN

Run Away Architectures

- Ghz Processors --- Mhz memories
- Superscalar Processors --- old fashioned busses
- Multi-processor nodes --- PCI/ethernet interconnects
- Active/complex Processor mechanisms --- VAX era performance models and measurement tools
- Processor memory hierarchy increasingly finicky.

Higher Algorithmic Goals

- More precision
- Larger problems
- Larger parameter space

- Adaptive, irregular methods
- Resource aware
- Fault tolerant

Then it got down off the mushroom, and crawled away in the grass, merely remarking as it went, `One side will make you grow taller, and the other side will make you grow shorter.'

`One side of what? The other side of what?' thought Alice to herself.

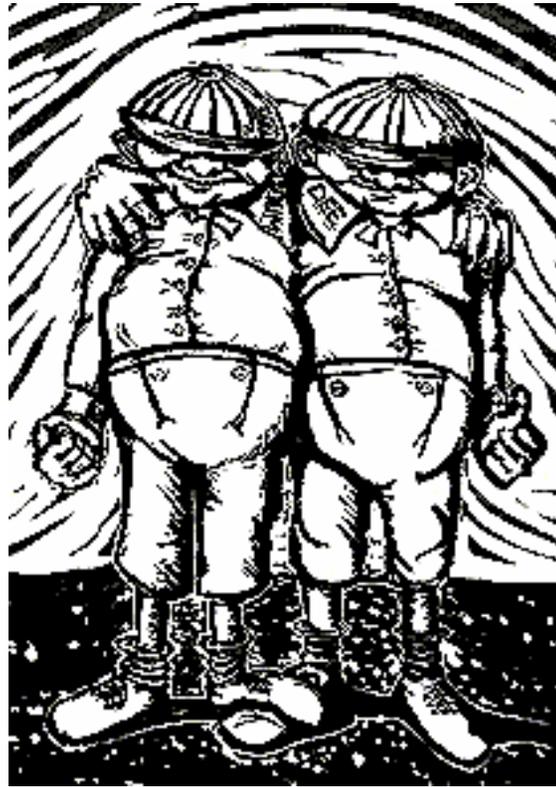Mushroom

The Parallelism Mushroom

- Parallelism
  - For typical science codes is easy to find, harder to express
- Balance the work
  - Sharing out work from a work queue
  - Partitioning for asymmetry in physical problem
  - Adjusting for local problem difficulty
- Don't move data
  - Cache line thrashing
  - Ghost regions

Expressing Task Parallelism

- Explicit threads, subroutines, loops
- How do you?
  - Choose next to do
  - Decide what processor to use
  - Encapsulate input and output
  - Monitor progress
- And
  - Keep work near its data
  - Do I/O
  - Scalably deal with fault tolerance

Expressing Data Locality

- Distributed Memory, Data layout pragmas
- How do you?
  - Adjust for multiple levels and sizes of memory
  - Move data from one local place to another
  - Dynamically allocate/reallocate
- And
  - Not use up all the processor cycles
  - Not trash the local memory hierarchy

'I know what you're thinking about,' said Tweedledum: 'but it isn't so,nohow.'

'Contrariwise,' continued Tweedledee, 'if it was so, it might be; and if it were so, it would be; but as it isn't, it ain't. That's logic.'

'I was thinking,' Alice said very politely, 'which is the best way out of this wood: it's getting so dark. Would you tell me, please?'

But the little men only looked at each other and grinned.   TWEEDLEDEEDUM

PVM,MPI

- Distributed Memory Model, Independent Processes, Explicit Data Movement, Library interface
- Successes
  - Standardized and codified, vendor acceptance, portability
  - Large messages bandwidth, Latency compared to TCP
  - Collective routines, communicators
- Challenges
  - Latency
  - Model Flexibility
  - Programmer overhead
- Continuing Work
  - MPI-2
  - Argonne, ORNL/Tennessee
  - MPI Software Tech

UPC - Unified Parallel C

- SALC Memory Model, Threads, Language Extension, Compiler/User share data movement
- Successes
  - Vendor's beginning to adopt
  - Very low latency/overhead
  - Simple programming model
- Challenges
  - Programmer Acceptance
  - Filling out tool set, combine with MPI?
  - Compiler optimization modules
- Continuing Work
  - LBL/NERSC, Compaq, IBM, CCS
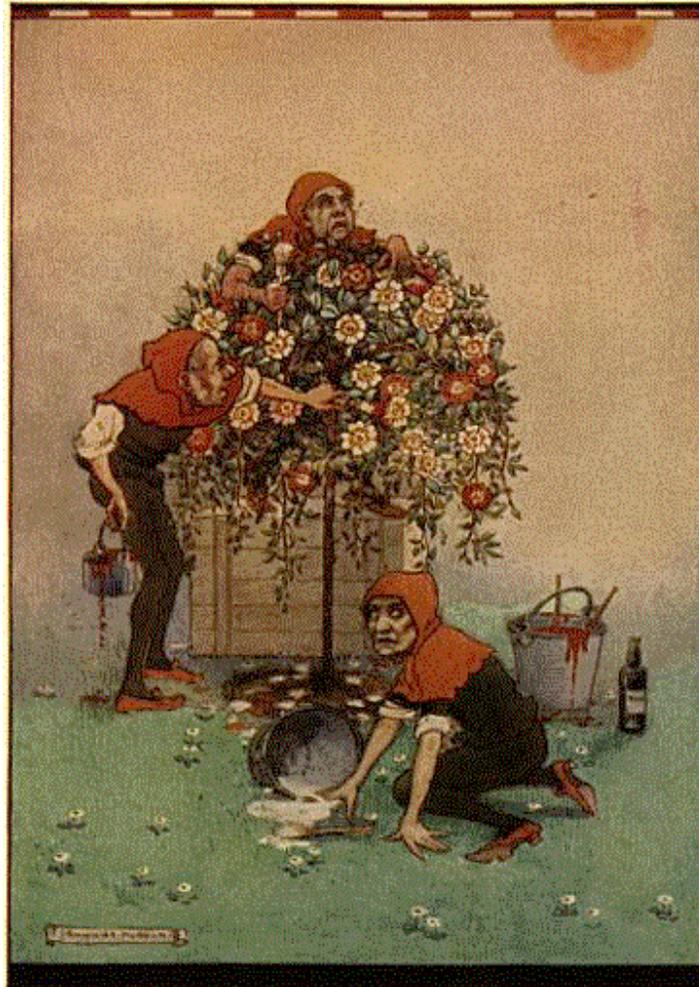
UPC - A few details

- All the expressive power of C (eg. structs, ptrs)
- Shared keyword tells compiler to distribute
- Barrier, strict variables control coherence
- Locks for synchronized activity

- Allocations and barriers are collective operations

- www.super.org/upc

Global Arrays

- SALC Memory Model, Data Parallel, Library based
- Successes
  - Match particular programmer needs
  - Multiple ports through use of layered model
- Challenges
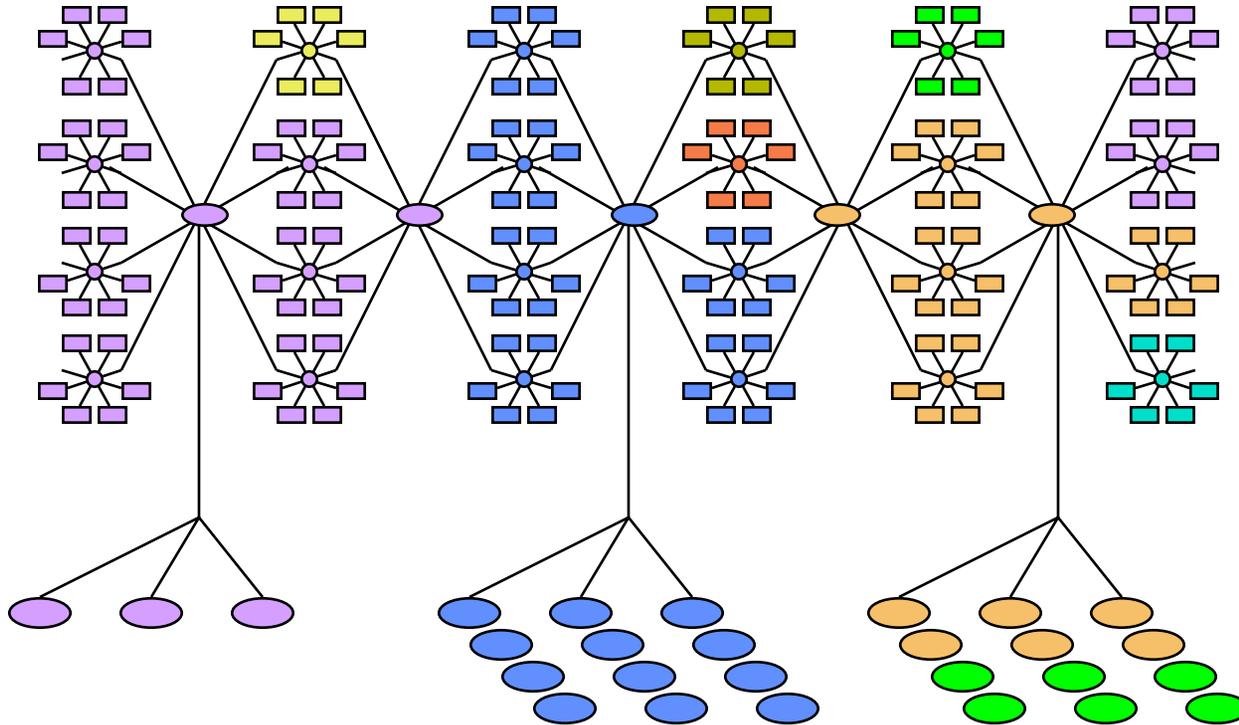 - Generality
  - User acceptance

`Would you tell me,' said Alice, a little timidly, `why you are painting those roses?'

Five and Seven said nothing, but looked at Two. Two began in a low voice, `Why the fact is, you see, Miss, this here ought to have been a redrose-tree, and we put a white one in by mistake; and if the Queen was to find it out, we should all have our heads cut off, you know. So you see, Miss, we're doing our best, afore she comes, to--' At this moment Five, who had been anxiously looking across the garden, called out `The Queen! The Queen!' and the three gardeners instantly threw themselves flat upon their faces. There was a sound of many footsteps, and Alice looked round, eager to see the Queen.

Roses

Computational Needs

- Tflop scale computers without single vendor constraints.

- Very-large scale simulations.
- Optimization and parameterization studies.
- Data analysis.

- Growth as need and budget develop.
- Flexibility in use.
- Ability to prototype and experiment.
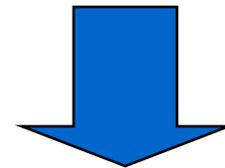
Computational Plant Concept

- **Grow with need:**
  Always at leading edge
- **Power plant:**
  Fuel simulation revolution
- **Specialize:**
  Best of everything
- **Spread:**
  From SAN to WAN
- **Colonize:**
  Every desk and machine room

Cplant, Chiba City, HPTi, Turbo

- Recognition that HPC industry is precarious
- Leverage commodity parts, build on the Beowulf trend
- But need much more scalability, managability,
- And expect many more varied users per machine

Scalability, scalability, scalability

- Tens of thousands of threads.
- Thousands of interconnect links
- Multiple terabytes of RAM
- Tens of terabytes of disk
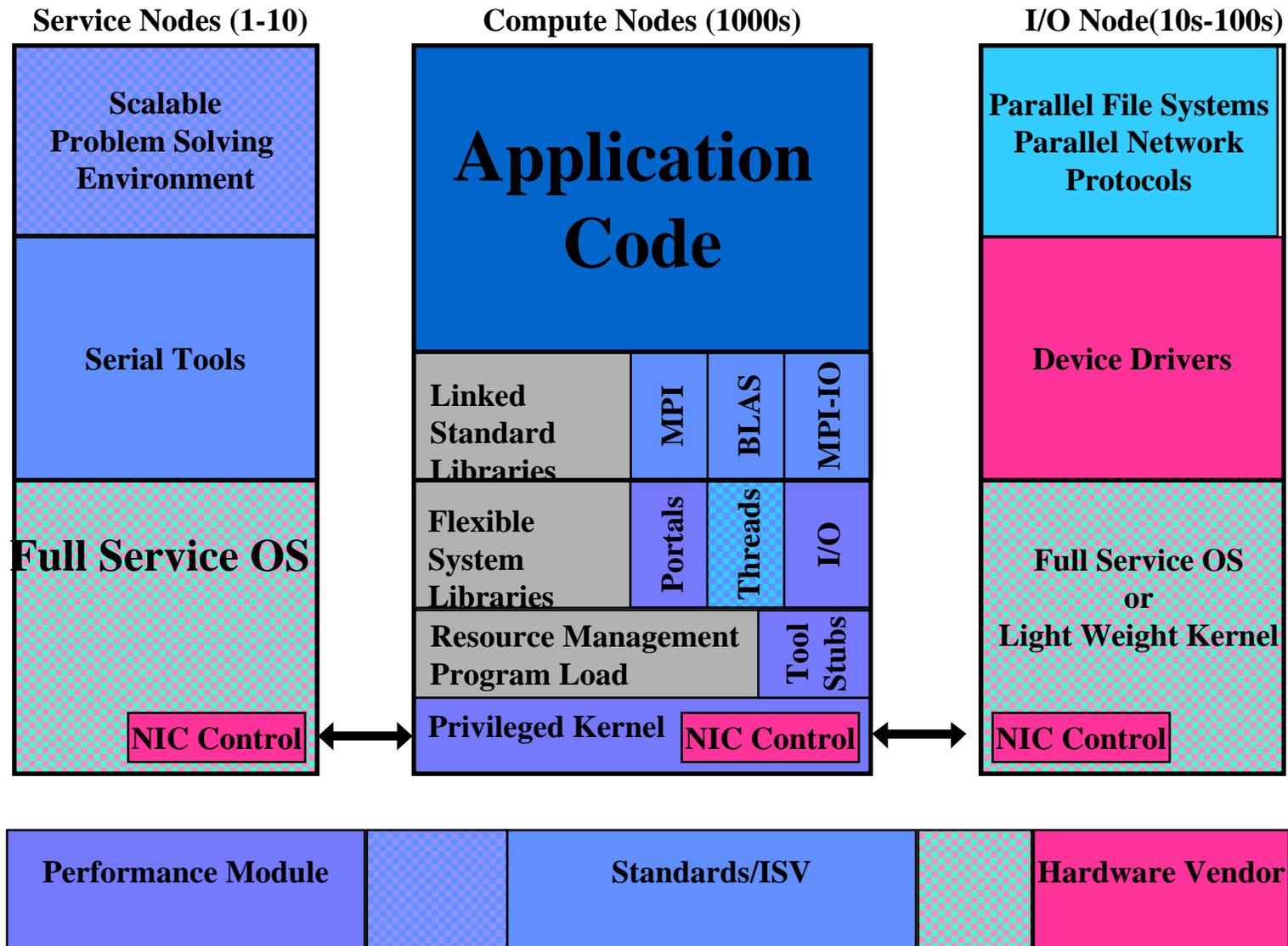- Petabytes of tertiary storage

- *Inexpensive parts*
- *Hot swappable parts*
- *Interoperable parts*
- *Everything plugs into the network*

Hardware/commodity enablers

- Cheap commodity nodes.
- High-speed network interfaces.
- Low cost, cascadable switches.
- Network attachable RAIDs.
- Robust racking systems.

- Remote management functions.
- Large-scale system management software.

Tinker, tailor, prototype, buy

Service Nodes (1-10)

Scalable Problem Solving Environment
Serial Tools
Full Service OS
NIC Control

Compute Nodes (1000s)

Application Code

Linked Standard Libraries	MPI	BLAS	MPI-IO
Flexible System Libraries	Portals	Threads	I/O
Resource Management Program Load		Tool Stubs
Privileged Kernel	NIC Control

I/O Node(10s-100s)

Parallel File Systems Parallel Network Protocols
Device Drivers
Full Service OS or Light Weight Kernel
NIC Control

Performance Module		Standards/ISV		Hardware Vendor

Ongoing efforts

- Extreme Linux Forum
  - Conference in association with USENIX, developer meetings
- Some major efforts
  - Surmounting BIOS/startup issues. LOBOS, Bproc
  - Scalable system management
  - Reasonable IO
- Vendors entering market
  - IBM, SGI, Compaq, HPTi, TurboLinux, Skyld, Alta

Linux issues

- Customizing for compute vs. IO vs. login partitions
- Customizing for high-performance scientific computing
  - Big pages, page coloring, myrinet drivers, quadrics drivers
- Microkernel work

Open system progress

- PVFS
- Skyld's new distribution
- Need debugger and performance tools

- 'Who are you?' said the Caterpillar.

- This was not an encouraging opening for a conversation. Alice replied, rather shyly, 'I-I hardly know, sir, just at present- at least I know who I was when I got up this morning, but I think I must have been changed several times since then.'

Caterpillar

Co-evolution

- Architecture
- Programming Model
- Applications

The Network is the Battleground