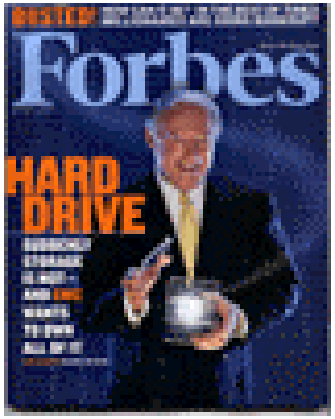
Patterson's Projects, People, Impact

- **Reduced Instruction Set Computer (RISC)**
  - What: simplified instructions to exploit VLSI: '80-'84
  - With: Sequin@UC, Hennessy@Stanford, Cocke@IBM
  - Direct Impact: Sun's SPARC, >90% embedded MPUs
- **Symbolic Processing Using RISCs (SPUR)**
  - What: desktop multiprocessor for AI: '84 - '89
  - With: Fateman, Hilfinger, Hodges, Katz, Ousterhout
  - Direct Impact: PPL => fast serial lines => Silicon Image
- **Redundant Arrays of Inexpensive Disks (RAID)**
  - What: many PC disks for speed, reliability: '88 - '93
  - With: Katz, Ousterhout, Stonebraker
  - Direct Impact:$19B/yr(EMC) 80% nonPC disks in RAID

Patterson's Projects, People, Impact

- **Networks of Workstations (NOW)**
  - What: big server via switched network of WS '94-'98
  - With: Anderson, Brewer, Culler
  - Direct Impact: Inktomi + many Internet companies
- **Tertiary Disk (TD: a NOW subset project)**
  - What: big, cheap, disk-intensive NOW (for SF Museum) '96-'99
  - Direct Impact: Scale8 (big, cheap, reliable Internet storage)



See 10/2/00 Forbes: EMC, RAID, Scale8

- **Intelligent RAM (IRAM)**
  - What: media processor inside DRAM chip: '97 - '01
  - With: Yelick
  - Direct Impact: nBand (easy-to-code DSP for wireless)

What's Next?

- **Berkeley View of PostPC Era**
  - Gadgets (mobile, wireless devices everywhere)
  - Services for Gadgets (big, fat, web servers)
- **For B.F.W.S. Challenge isn't what you think**
- **What its NOT: Performance, Cost**
- **What it is:**
  - Availability
    - » meet service goals despite HW/SW failures
  - Maintainability
    - » minimal human administration, regardless of scale
      Today, cost of maintenance = 10-100 cost of purchase
  - Evolutionary Growth
    - » systems evolve gracefully as upgraded/expanded

Principles for achieving AME

- **Introspection**
  - **reactive** techniques to detect and adapt to failures, workload variations, and system evolution
  - **proactive** techniques to anticipate and avert problems before they happen
- **Undo of any system administration event**
- **No single points of failure, redundancy everywhere**
- **Create AME benchmarks to measure progress**
- **Performance robustness, AME goals more important than peak performance, capital cost**
- **Brick building block to construct big servers**

Intelligent STORE (ISTORE) Brick

- **Webster's Dictionary: "brick: a handy-sized unit of building or paving material typically being rectangular: ~ 2 1/4 x 3 3/4 x 8 in."**
- **ISTORE-1 Brick: 2 x 4 x 11 inches (1.3x)**
  - Switched networks reduce need for CPU/bus model
  - Brick: Single physical form factor, fixed cooling required, compatible network interface to simplify physical maintenance, scaling over time
  - Contents should evolve over time: contains most cost effective MPU, DRAM, disk, compatible NI
  - Suggests network that will last, evolve: Ethernet
- **Bricks into big servers via Redundant Arrays of Inexpensive Network switches (RAIN)**

A glimpse into the future?

- **System-on-a-chip enables computer, memory, redundant network interfaces without significantly increasing size of disk**
- **ISTORE HW in 5-7 years:**
  - 2006 brick: System On a Chip integrated with MicroDrive
    - » 9GB disk, 50 MB/sec from disk
    - » connected via crossbar switch
    - » From brick to "domino"
  - If low power, 10,000 nodes fit into one 19" x 32" x 7' rack!
- **O(10,000) scale is our ultimate design point**

ISTORE as Storage System of the Future

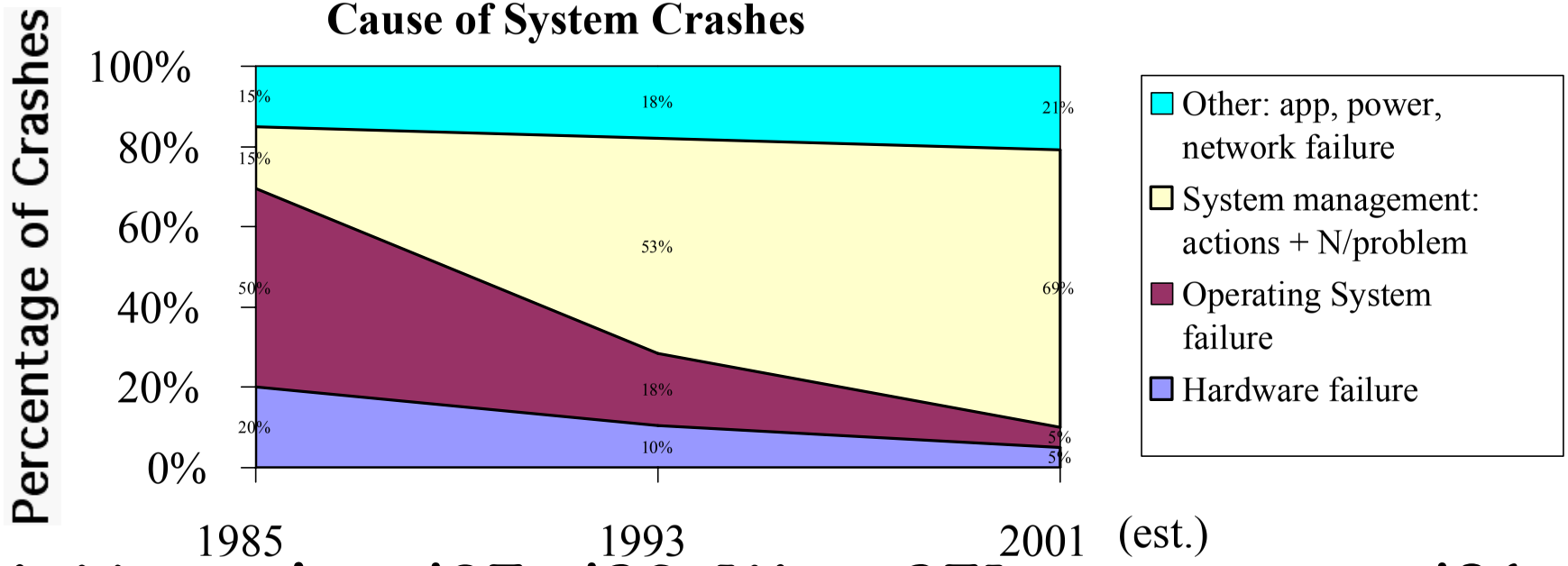
- **Availability, Maintainability, and Evolutionary growth key challenges for storage systems**
  - Maintenance Cost ~ >10X Purchase Cost per year,
  - Even 2X purchase cost for 1/2 maintenance cost wins
  - AME improvement enables even larger systems
- **ISTORE has cost-performance advantages**
  - Better space, power/cooling costs ($@colocation site)
  - More MIPS, cheaper MIPS, no bus bottlenecks
  - Compression reduces network $, encryption protects
  - Single interconnect, supports evolution of technology
- **Match to future software storage services**
  - Future storage service software target clusters

Questions?

Contact us if you're interested:
email: **patterson@cs.berkeley.edu**
http://iram.cs.berkeley.edu/istore
phone: (510) 642-6587
FAX: (510) 643-7352

Is Maintenance the Key?

- **Rule of Thumb: Maintenance 10X to 100X HW**
  - so over 5 year product life, ~ 95% of cost is maintenance



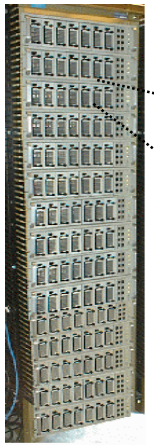
- **VAX crashes '85, '93 [Murp95]; extrap. to '01**
- **Sys. Man.:** N crashes/problem, SysAdmin action
  - **Actions: set params bad, bad config, bad app install**
- **HW/OS 70% in '85 to 28% in '93. In '01, 10%?**

ISTORE-1 hardware platform

- **80-node x86-based cluster, 1.4TB storage**
  - **cluster nodes are plug-and-play, intelligent, network-attached storage "bricks"**
    - » a single field-replaceable unit to simplify maintenance
  - **each node is a full x86 PC w/256MB DRAM, 18GB disk**
  - **more CPU than NAS; fewer disks/node than cluster**

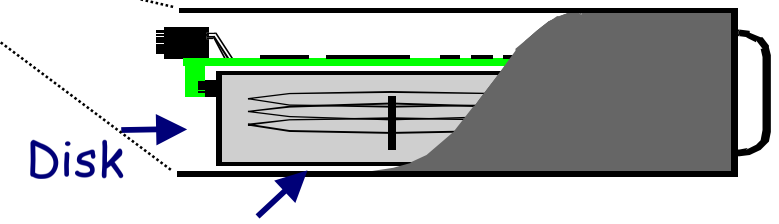
ISTORE Chassis

80 nodes, 8 per tray
2 levels of switches
•20 100 Mbit/s
•2 1 Gbit/s
Environment Monitoring:
UPS, redundant PS,
fans, heat and vibration
sensors...



Intelligent Disk "Brick"

Portable PC CPU: Pentium II/266 + DRAM
Redundant NICs (4 100 Mb/s links)
Diagnostic Processor



Cost of Space, Power, Bandwidth

- **Co-location sites (e.g., Exodus) offer space, expandable bandwidth, stable power**
- **Charge ~$1000/month per rack (~ 10 sq. ft.)**
  - Includes 1 20-amp circuit/rack; charges ~$100/month per extra 20-amp circuit/rack
- **Bandwidth cost: ~$500 per Mbit/sec/Month**
- **ISTORE-1: 2X savings in space vs. Sun 10K**
  - ISTORE-1: 1 rack (big) switches, 1 rack (old) UPSs, 1 rack for 80 CPUs/disks (3/8 VME rack unit/brick)
- **ISTORE-2: 8X-16X space?**
- **Space, power cost/year for 1000 disks: Sun $924k, ISTORE-1 $484k, ISTORE2 $50k**

Cost of Bandwidth, Safety

- **Network bandwidth cost is significant**
  - 1000 Mbit/sec/month => $6,000,000/year
- **Security will increase in importance for storage service providers**
- **XML => server format conversion for gadgets**

=> Storage systems of future need greater computing ability

- Compress to reduce cost of network bandwidth 3X; save $4M/year?
- Encrypt to protect information in transit for B2B

=> Increasing processing/disk for future storage apps