Exotic Technologies Panel and Time Capsule Submission for Most Exciting Architecture at SC 20

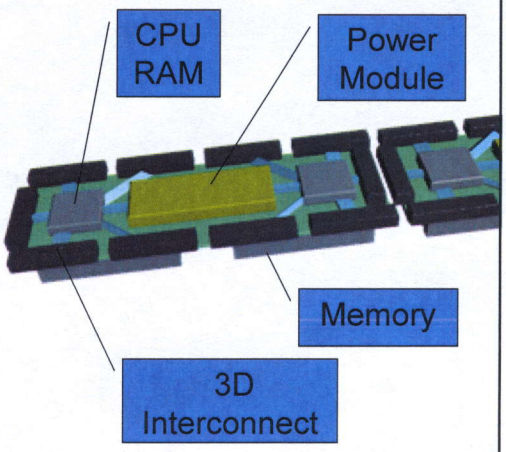
Erik P. DeBenedictis

SC 20 Supercomputer Projection

	Red Storm (Historical)	μP part only	My Entry
Total cores	13,000×2	50,000×4	50,000×40
Node Type	μP	μP	μP & macro function
Clock	2.5 GHz	20 GHz	20 GHz
Flops/chip	5×2 GF	50×4 GF	1.6 TF
Sys. Peak	125 TF	80 PF	800 PF
Maximum MPI Latency	10 μS	100 ns	100 ns
Power	2 MW	2 MW	2 MW

Packaging for a Spatial Locality

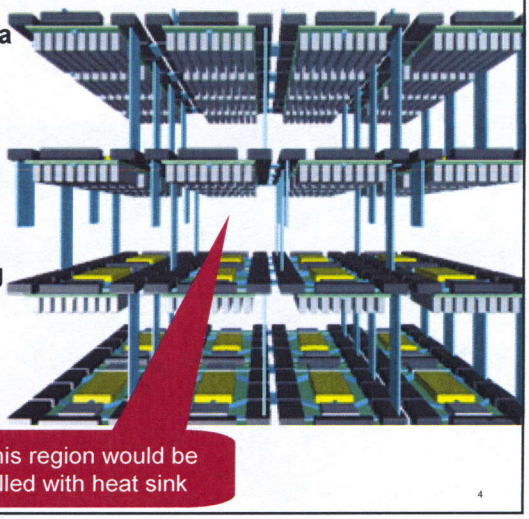
- **Basic Module**
  - **2 Chips**
  - **Each node 4 core conventional CPU plus**
  - **36 accelerator cores**
  - **1 GB+ on chip RAM**
  - **100 GB memory on bottom of module**
  - **Each module includes a power unit**
  - **Six optical interconnect channels, 3D mesh**

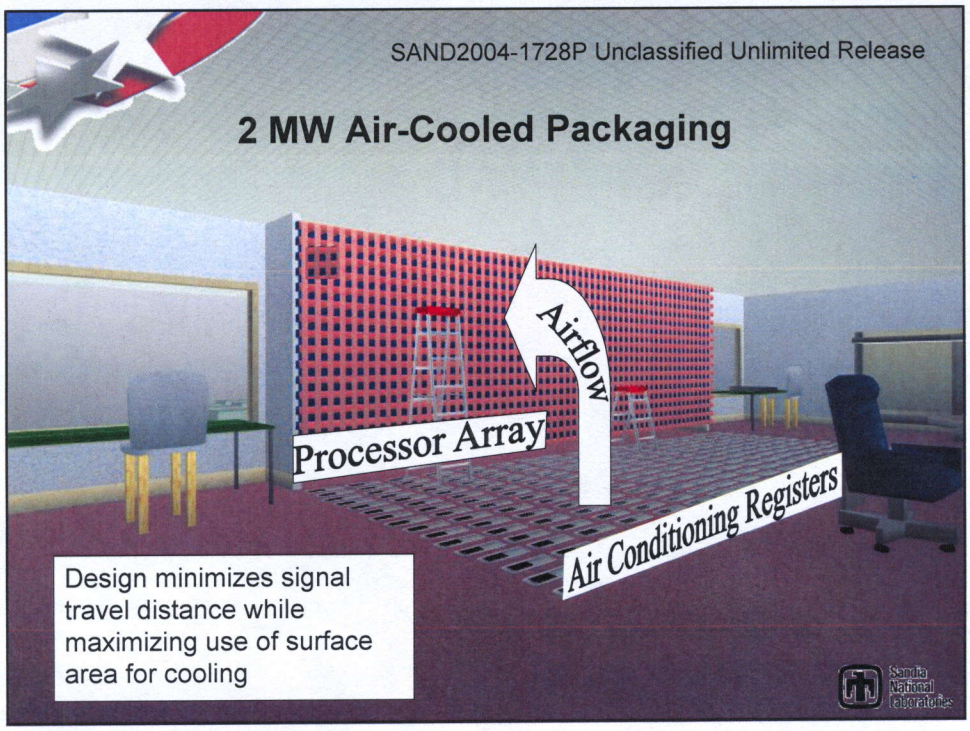


SAND2004-1728P Unclassified Unlimited Release

Packaging for a Spatial Locality

- **Entire supercomputer is a single structure**
- **All mesh network connections are of constant length (8" max)**
- **Air flows front to back**
  - **General approach will work for liquid cooling as well**





Outline

- **Degree of Innovation**
- **Non-Architecture Projections**
- **Architecture Projections**
- **Programming**
- **Architecture Summary**
- **Current Activities to Watch and Why**
- **Conclusions**

Perspective on Innovation

- 1992 + 14 = 2006; 2006 + 14 = 2020
- If rate of innovation stays the same, we should see as big an advance to 2020 as we saw from "late nCUBE" through now
- However, I think SC is maturing. I think the community will only accept innovations backwards compatible with what we have now. If there is major innovation, I think it will be best represented in a new conference, say "I Robot 2020."

Outline

- Degree of Innovation
- Non-Architecture Projections
- Architecture Projections
- Programming
- Architecture Summary
- Current Activities to Watch and Why
- Conclusions

Scaling Implications for CPUs

- 5× performance increase for a single core
  - Burger and Keckler study, slide follows
  - NOTE: Integrated RAM will increase this another 2×
- 64 cores of today's complexity
  - 90 nm → 18 nm is 5×. Dual core × 5^2 → 50 ≈ 64
- I think we'll see a hybrid – to be discussed later

UT Austin Study (2000)

- The Study
  - Clock Rate versus IPC: The End of the Road for Conventional Microarchitectures, Vikas Agarwal, M.S. Hrishikesh, Stephen W. Keckler, Doug Burger. 27th Annual International Symposium on Computer Architecture
- Conclusions (to be Explained)
  - Modified ITRS roadmap predictions to be more friendly to architectures
  - Concluded there would be a 12%/year growth…
  - However, recent growth has been ~30%, with industry's maneuver to cheat the analysis instructive

> 20

>18 - 20

>16 - 18

≤ 16

Critical Evaluation Memory

For each Technology Entry (e.g. 1D Structures, sum horizontally over the 8 Criteria
Max Sum = 24
Min Sum = 8

Memory Device Technologies (Potential)	*Scalability [A]*	*Performance [B]*	*Energy Efficiency [C]*	*OFF/ON "1"/"0" Ratio [D1]*	*Operational Reliability [E]*	*Operate Temp [F] ****	*CMOS Technological Compatibility [G]***	*CMOS Architectural Compatibility [H]**
Nano Floating Gate Memory	2.5	2.5	2.5	2.5	2.2	2.7	2.7	3.0
Engineered Tunnel Barrier Memory	2.2	2.3	2.3	2.3	2.4	2.8	2.8	3.0
Ferroelectric FET Memory	1.9	2.3	2.5	2.2	2.0	3.0	2.6	3.0
Insulator Resistance Change Memory	2.5	2.5	2.0	2.2	1.9	2.8	2.6	2.8
Polymer Memory	2.1	1.5	2.3	2.2	1.6	2.9	2.3	2.5
Molecular Memory	2.3	1.5	2.4	1.6	1.4	2.6	1.9	2.3

4 good options

> 20

>18 - 20

>16 - 18

≤ 16

Critical Evaluation Logic

For each Technology Entry (e.g. 1D Structures, sum horizontally over the 8 Criteria
Max Sum = 24
Min Sum = 8

Logic Device Technologies (Potential)	*Scalability [A]*	*Performance [B]*	*Energy Efficiency [C]*	*Gain [D2]*	*Operational Reliability [E]*	*Room Temp Operation [F] ****	*CMOS Technological Compatibility [G]***	*CMOS Architectural Compatibility [H]**
1D Structures (CNTs & NWs)	2.4	2.5	2.3	2.3	2.1	2.8	2.3	2.8
Resonant Tunneling Devices	1.5	2.2	2.1	1.7	1.7	2.5	2.0	2.0
SETs	1.9	1.5	2.6	1.4	1.2	1.9	2.1	2.1
Molecular Devices	1.6	1.8	2.2	1.5	1.6	2.3	1.7	1.8
Ferromagnetic Devices	1.4	1.3	1.9	1.5	2.0	2.5	1.7	1.7
Spin Transistor	2.2	1.3	2.4	1.2	1.2	2.4	1.5	1.7

1 good option, and it is not a change for SC

Outline

- Degree of Innovation
- Non-Architecture Projections
- Architecture Projections
- Programming
- Architecture Summary
- Current Activities to Watch and Why
- Conclusions

Commodity μP Architecture in 2020

- Industry is now ramping the number of cores per die
- Intel and AMD are making serious noises about integrating graphics processors into CPU die
- I have special information that upcoming ITRS direction will advocate "macro functions" (to be explained later)
- These are self-confirming data points that answer the commodity μP architecture question
  - Note: this answer could be wrong…