Review Approval

Prepare Request
Search Requests
Generate Reports
Approvals
Help
Wizard

Search Requests
New Search
Refine Search
Search Results

Clone Request
Edit Request
Cancel Request

Search Detail

Submittal Details

Document Info

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

Document Number : 5248345 SAND Number : 2006-7724 C

Review Type : Electronic Status : Approved

Sandia Contact : DEBENEDICTIS,ERIK P. Submittal Type : Conference Paper

Requestor : DEBENEDICTIS,ERIK P. Submit Date : 12/09/2006

Comments : Presented at the Exotic Technologies session at SC 06 (Supercomputing 2006) and stored in a time capsule.

Peer Reviewed? : N

Author(s)

DEBENEDICTIS,ERIK P.

Event (Conference/Journal/Book) Info

Name : Supercomputing 2006

City : Tampa State : FL Country : USA

Start Date : 11/13/2006 End Date : 11/16/2006

Partnership Info

Partnership Involved : No

Partner Approval : Agreement Number :

Patent Info

Scientific or Technical in Content : Yes

Technical Advance : No TA Form Filed : No

SD Number :

Classification and Sensitivity Info

Title : Unclassified-Unlimited Abstract : Document : Unclassified-Unlimited

Additional Limited Release Info : None.

DUSA : DIS-CS

Routing Details

Role	Routed To	Approved By	Approval Date
Manager Approver	PUNDIT,NEIL D.	PUNDIT,NEIL D.	12/11/2006
Conditions:			
Administrator Approver	LUCERO,ARLENE M.	FARRELLY,JEREMIAH	08/24/2007

Editorial note: Presented at the Exotic Technologies session at SC 06 (Supercomputing 2006). A hardcopy of this presentation has been placed in a “time capsule” along with presentations by Doc Bedard, Thomas Sterling, and a bottle of red wine. The plan is for the time capsule to be stored under the floor of a supercomputer machine room at Lawrence Berkeley National Laboratory. At SC 20 (14 years from now), the time capsule will be opened and the one of three entries closest to matching reality in 2020 will win. The winner will get the bottle of wine as a prize. The red wine is apparently of a type that will age effectively over a period of 14 years.

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

Erik P. DeBenedictis

SAND2006-7724C
Approved for Unclassified Unlimited Release

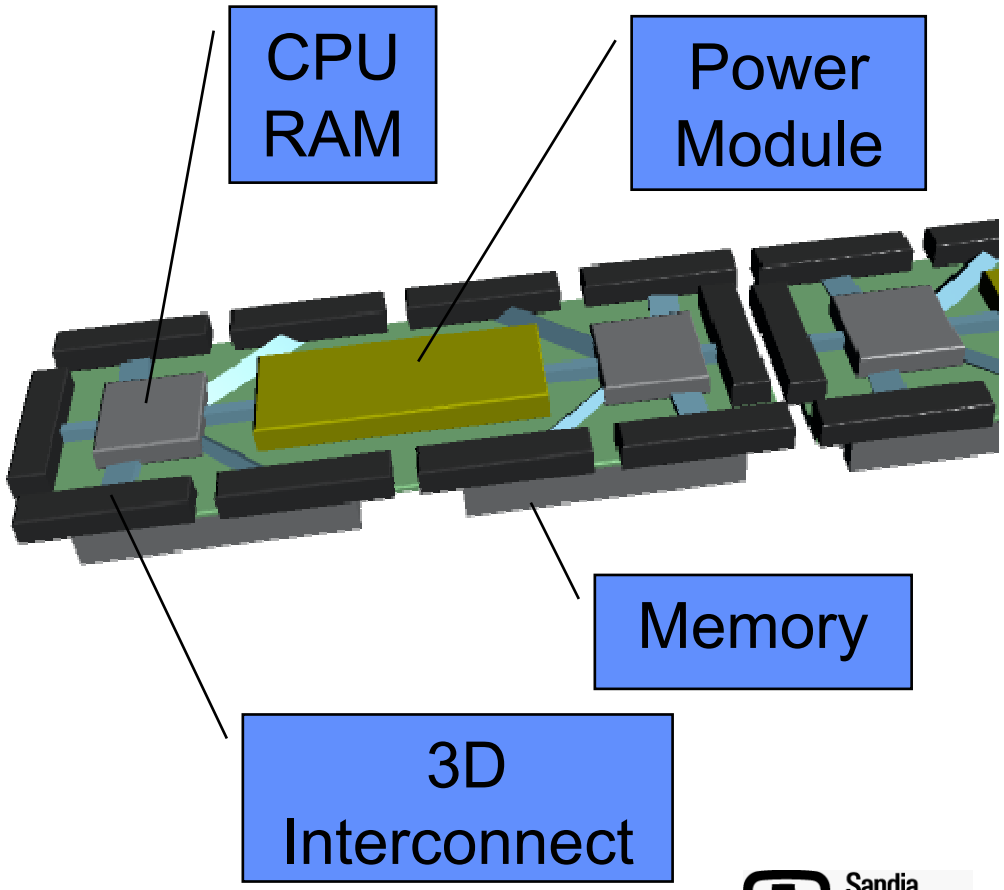
Sandia National Laboratories

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

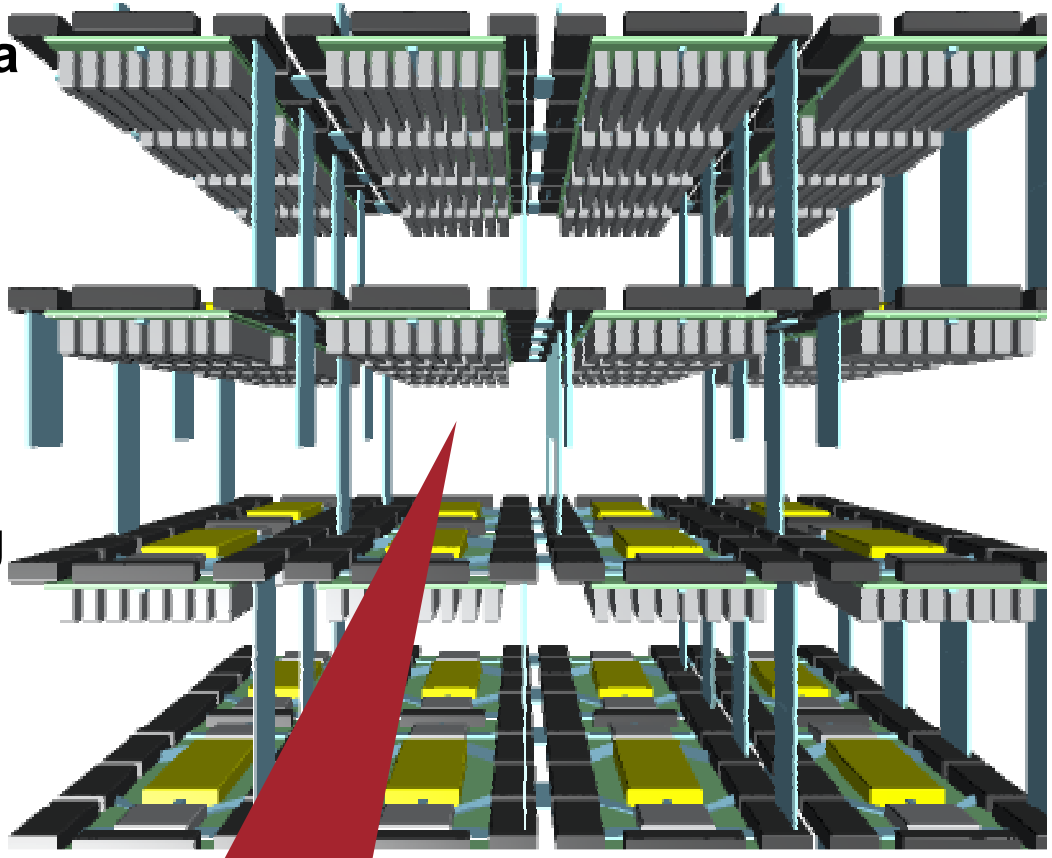
Packaging for a Spatial Locality

- **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**

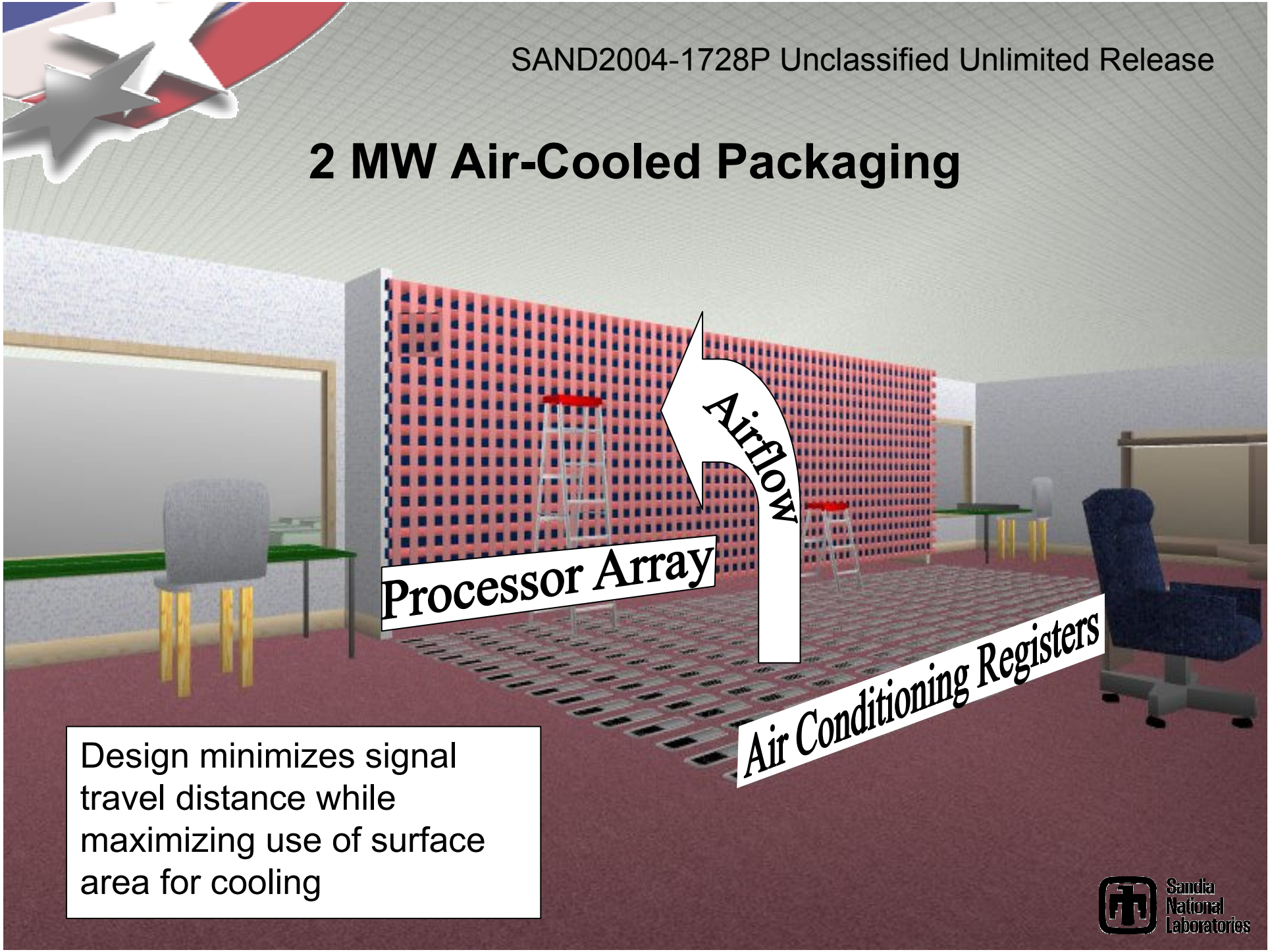


Packaging for a Spatial Locality

- **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**



2 MW Air-Cooled Packaging



Design minimizes signal travel distance while maximizing use of surface area for cooling

Sandia National Laboratories

Outline

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

Sandia National Laboratories

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

Sandia National Laboratories

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

Sandia National Laboratories

> 20

>16 - 18

>18 - 20

≤ 16

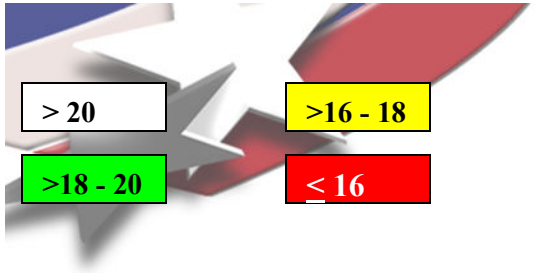
Critical Evaluation Memory

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

Sandia National Laboratories



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**

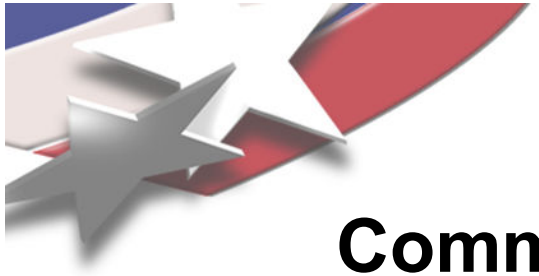
1 good option, and it is not a change for SC

Sandia National Laboratories

Outline

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

Sandia National Laboratories



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…

Sandia National Laboratories

I do not have mystical clairvoyance, but I do have a VG set from an influential meeting that hasn't occurred yet…

Emerging Research Logic Technologies

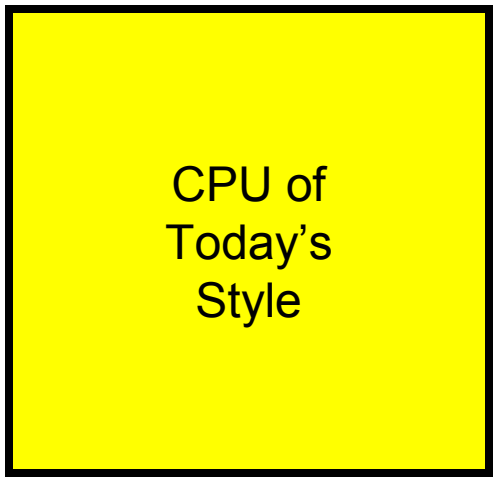
Traditional Goal

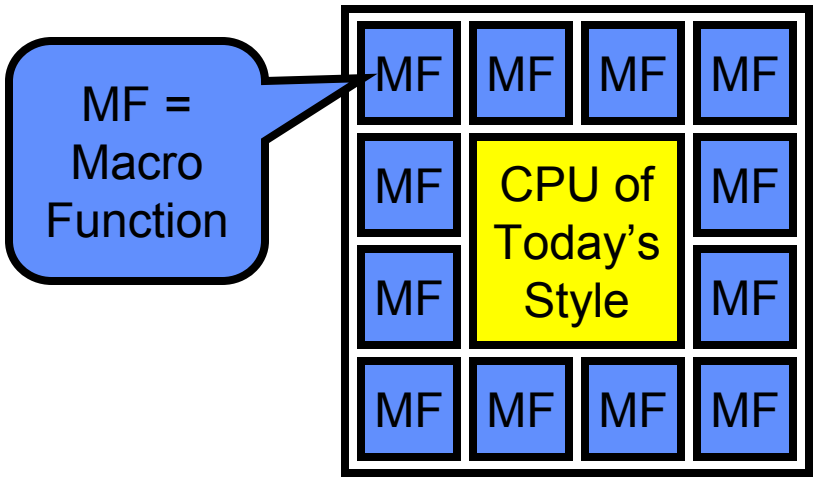
Logic technology that is scaleable beyond CMOS, high-speed, and low-power.

Macro Function Direction

- Current CPU style
- New direction proposed to industry will be to keep CPU but augment it with "macro functions."

- Macro functions may include non-CMOS logic devices specialized to nontraditional functions, such as speech recognition, etc.





Sandia National Laboratories

Outline

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

Sandia National Laboratories

Programmability Considerations

- Has code changed since the “late nCUBE” era?
  - MPI replaced proprietary message passing
  - We have a huge code base of math code (at Sandia)
  - We have frameworks (at Sandia)
- Conclusion
  - A lot of code written and put into reusable form, but little change in underlying programming method
- Implication
  - Further migration towards putting code into libraries, but the code will have the same basis

Programming

- Industry will integrate the following macro functions:
  - Graphics processors
  - Speech recognition
  - Visual recognition
- However, the hardware will be sufficiently general purpose to be used for supercomputing
- Still CMOS in this timeframe

- A small number of super-duper programming jocks will write supercomputing code for the macro functions
  - LAPACK
  - FEM meshing
  - Etc.
- Regular programmers will write C++/Fortran code interfacing like DirectX (Microsoft's GPU API)

Sandia National Laboratories

Programming Example

- I went by the PeakStream booth yesterday and see that they have a scientific programming library for graphics processors. I've never used it, but I think the approach might work with hardware up to 2020.

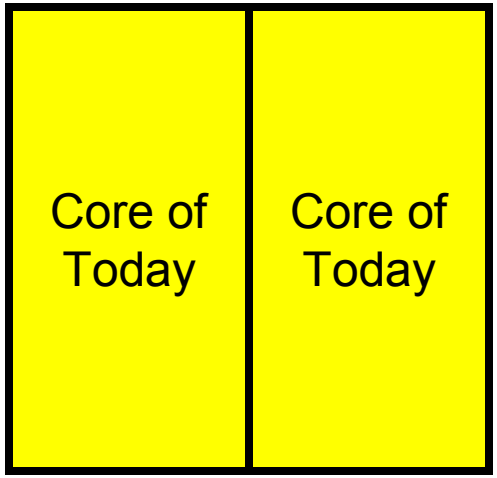
Outline

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

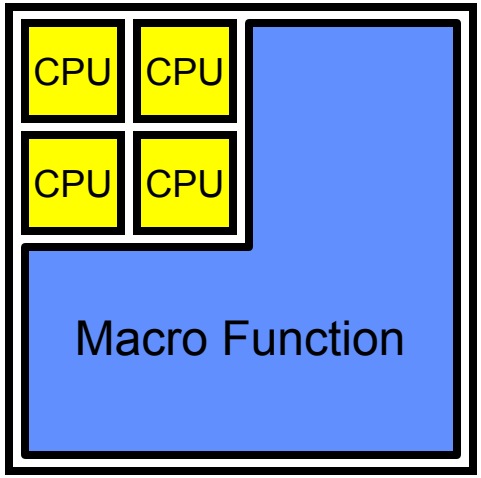
Sandia National Laboratories

Processor Chip Prediction

- ¼ of chip to be four CPUs each with 10× throughput of today's cores
- ¾ of chip to be a new Macro Function
- Layered nano memory



- Macro Function will be developed by industry and repurposed for supercomputing, originally
  - speech recognition
  - vision for robots



Sandia National Laboratories

CPU Detail

- Entry
  - Four cores at 50 GF Linpack-peak each, total 200 GF
  - 36 macro functions of 440 GF each, total 15.8 TF total
    - graphics, speech, vision, repurposed to scientific kernels
  - 16 TF per chip
  - Each chip to have 1 GB+ layered nano memory
  - As much external memory as you like (not a limit)
  - 50,000 chips in a 2 MW system → 800 Petaflops

Sandia National Laboratories

Memory Story – No Memory Wall

- I predict one of the 4+ nano memory options will succeed
- 1 GB+ memory will be integrated onto the CPU
  - I don't care if you call it cache, main memory, etc.
- Memory will be non-volatile
- This will boost CPU performance quite a bit over the 5× predicted by architecture study

Sandia National Laboratories

Interconnect

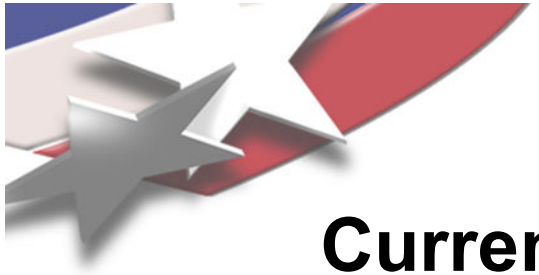
- Interconnect is likely to be optics, but not necessarily fiber
  - Free space
  - Waveguides
- Luxtera comes up often in discussions of optical interconnect. The Luxtera approach works with Si by having external lasers.

Sandia National Laboratories

Outline

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

Sandia National Laboratories



Current Activities to Watch and Why

- Cyclops – highly multicore architecture that could (with suitable systems software) blend legacy code compatibility with efficient use of multiple cores
  - Memory hierarchy is where the action is
  - I predict future will hold Cyclops + layered memory
- Layered memory (Nantero?)
- Optical interconnect (Luxtera?)
- Programming (PeakStream?)

Outline

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

Sandia National Laboratories

Conclusion I

- Industry is now putting additional resources created by Moore's Law into more cores and is talking about the same for graphics chips and Macro Functions
- Coders are getting further away from programming the bare hardware
- My solution has the following properties:

Sandia National Laboratories

Conclusion II

- The majority of users will program the conventional cores. They will see a fairly flat parallel Von Neumann computer. Of course, they are accustomed to using libraries for inner loops
- A small number of users will optimize low level code (libraries) for edge of the envelope hardware where the programmers need to be cognizant of data and operation placement
- I believe this is the most likely to happen, even if it does not make for the most exciting computer architecture research

Sandia National Laboratories