The Godson-3 Multi-Core Processor and its Application in High Performance Computers

Weiwu Hu
Institute of Computing Technology, CAS

Elio Guidetti
STMicro Electronics

Contents

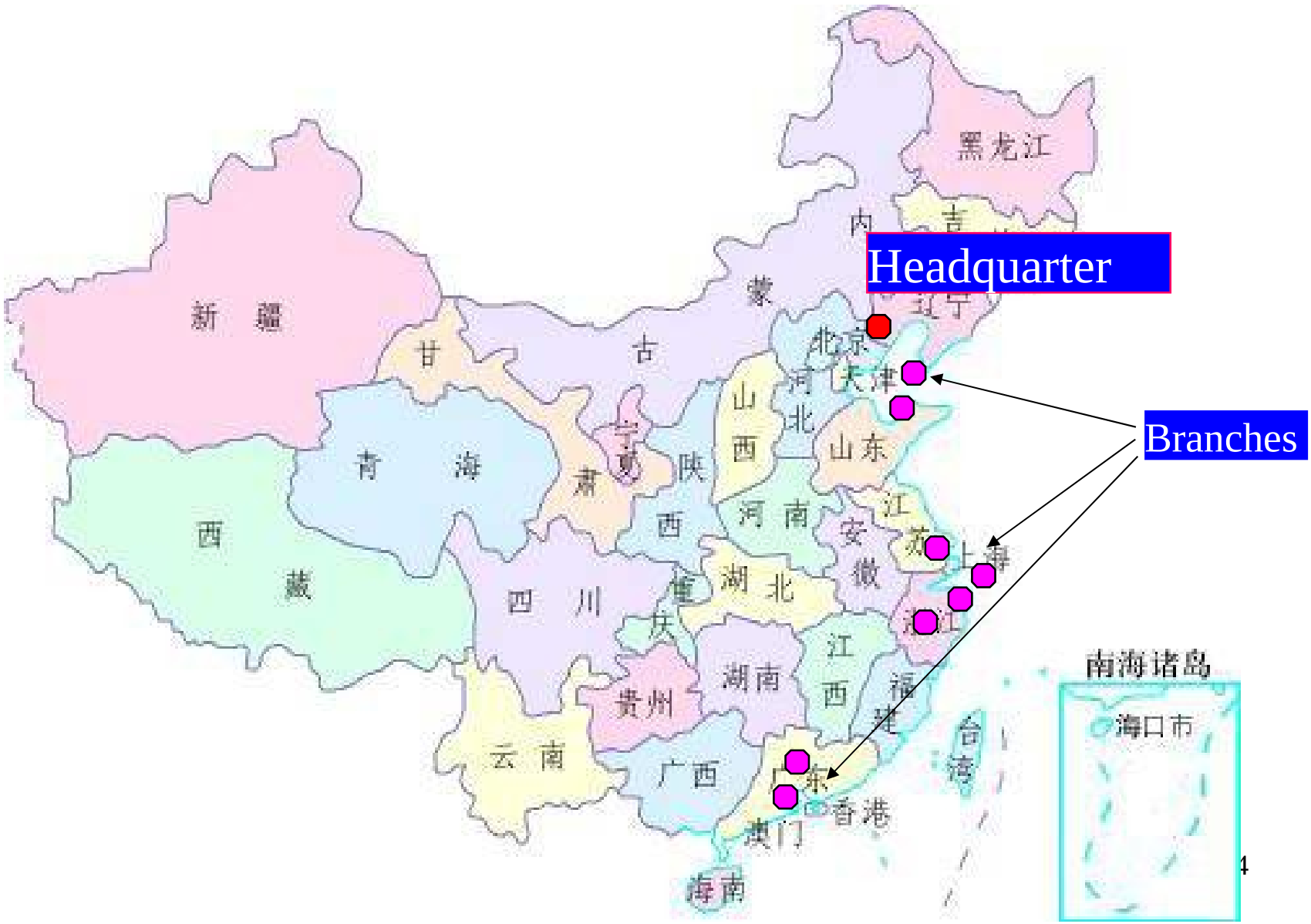
- **A brief introduction to ICT**
- **A brief introduction to Godson processors**
- **The Godson-3 multi-core processor**
- **PetaFLOPS and TeraFLOPS**

Godson is the academic name of Loongson™

ICT history contribution

- **Found in 1956, the first organization in China for computer science and technology research**
- **All original computer researchers in China are trained from ICT**
- **Built many computers for the country building before 1980**
- **Spin off big companies such as Lenovo and Dawning after 1980**
- **Spin off other institutes of CAS, such as institute of software and institute of microelectronics**

ICT Is a Networked Institute



ICT Organization

- R&D Divisions and Centers
  - Computer Systems (HPC, CPU, etc)
  - Network and Pervasive Computing
  - Intelligent Information Processing
  - Advanced Studies
- 8 regional branches
  - In cooperation with local government to promoting market
- Human Resource
  - 1500 people at headquarter ： including 1000 graduate students

Three Main Tasks of ICT

- **Solving the Nation's Big Problems**
- **Research and Development**
- **Graduate Education**

in computing

for the nation and the world

Solving the Nation's Big Problems

- **Improve Innovation Capability**
- **Pressing National Challenges**
  - **Energy, Healthcare, Environment, Education**
- **Benefiting the masses (1.3 billion)**

China Computer Market Trends

	GDP ($Trillion)	Computer Market ($Billion)	Internet Users (Million)	Client Devices (Million)
1995	0.69	7.4		
2000	1.08	25.9	22.5	8.9
2005	2.30	59.0	111	49.5
2010	*3.00*	*115.6*	*233*	*106*
2015	*4.75*	*217.3*	*411*	*191*
2020	*7.07*	*403.9*	*662*	*308*

China's IT market is still very shallow Source: IDC 2004

$Billion	PC	Server	Storage	PC : Server : Storage
China	11	1.8	0.7	1 : 0.16 : 0.06
Korea	2	1.2	1	1 : 0.60 : 0.50
Japan	10	8	9	1 : 0.80 : 0.90
North America	66	19	20	1 : 0.29 : 0.30
World	175	45	46	1 : 0.26 : 0.26

Expert from State e-Nation Office: Cannot copy the US route –
Cost > $10 trillion, Time > 30 years

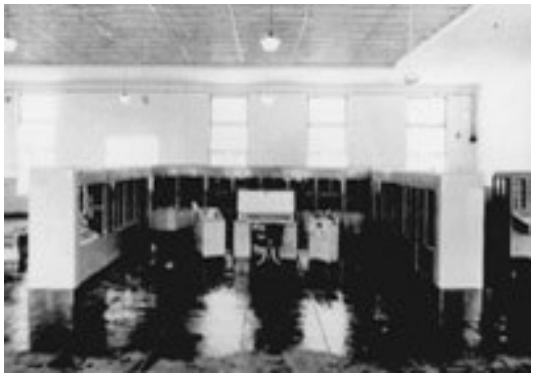
Computer Milestones

World Milestones		ICT Computers & Gaps		
1941	**1 flop/s**			
1945	100 flop/s	1958	Model 103	**13**
1949	**1 Kflop/s**			
1951	10 Kflop/s	1959	Model 104	**8**
1961	100 Kflop/s	1967	Model 109B	**6**
1964	**1 Mflop/s**	1976	Model 013	**12**
1968	10 Mflop/s	1983	Model 757	**15**
1975	100 Mflop/s			
1987	**1 Gflop/s**	1995	Dawning2000	**6**
1992	10 Gflop/s	2000	Dawning2000B	**7**
1993	100 Gflop/s	2003	Dawning4000L	**6**
1997	**1 Tflop/s**	2004	Dawning4000A	**4**
2000	10 Tflop/s	***2008***	***Dawning5000***	***3***
14	100 Tflop/s	***2010***	***Dawning5000A***	***2***
2008	**1 Pflop/s**			
2013	10 Pflop/s	High-Performance Computer		
2016	100 Pflop/s	Brand at ICT: **Dawning**		
2020	**1 Eflop/s**			

1941-2000 data barrowed from Jack Donggara, 2004

Computers designed by ICT

Model 103, 1958/8
First Computer in China



Model 104

Model 109C
First Large-scale Transistor Computer in China

Model 757, 1983/11
Vector Computer

Model **KJ8920, 1991/11**
Mainframe Computer

Dawning **Computers**

Evolution of HPC in China: Dawning HPCs

Dawning1000, 1995
intel i860
First MPP Computer in China
2.5Gflops Peak

Dawning2000, 1999,
Motorola PowerPC
First SMP Cluster in China
100Gflops Peak

Dawning3000, 2001
IBM Power3
SUMA Cluster in China
400Gflops Peak

Dawning4000-A, 2004
AMD Opteron
Grid-enabling Cluster
11.2 Tflops Peak

Dawning5000-A, 2008
6400 AMD 4-core Opteron
220 Tflops Peak

Dawning5000A Configuration

- **CPU** : 6400 AMD 4-core
- **Blade** : 1600 4-CPU-SMP
- **Node** : 160 10-blade
- **Cabinet** : 40 4-Node
- **Interconnection** : 10x12x24 DDR InfiniBand
- **System** : 200TFlops, 100TB Memory, 20Gbps
- **Storage** : 500TB, 50GB/s
- **Power** : 800KW
- **Cooling** : Air-cooling in Box + Water-cooling in Cab

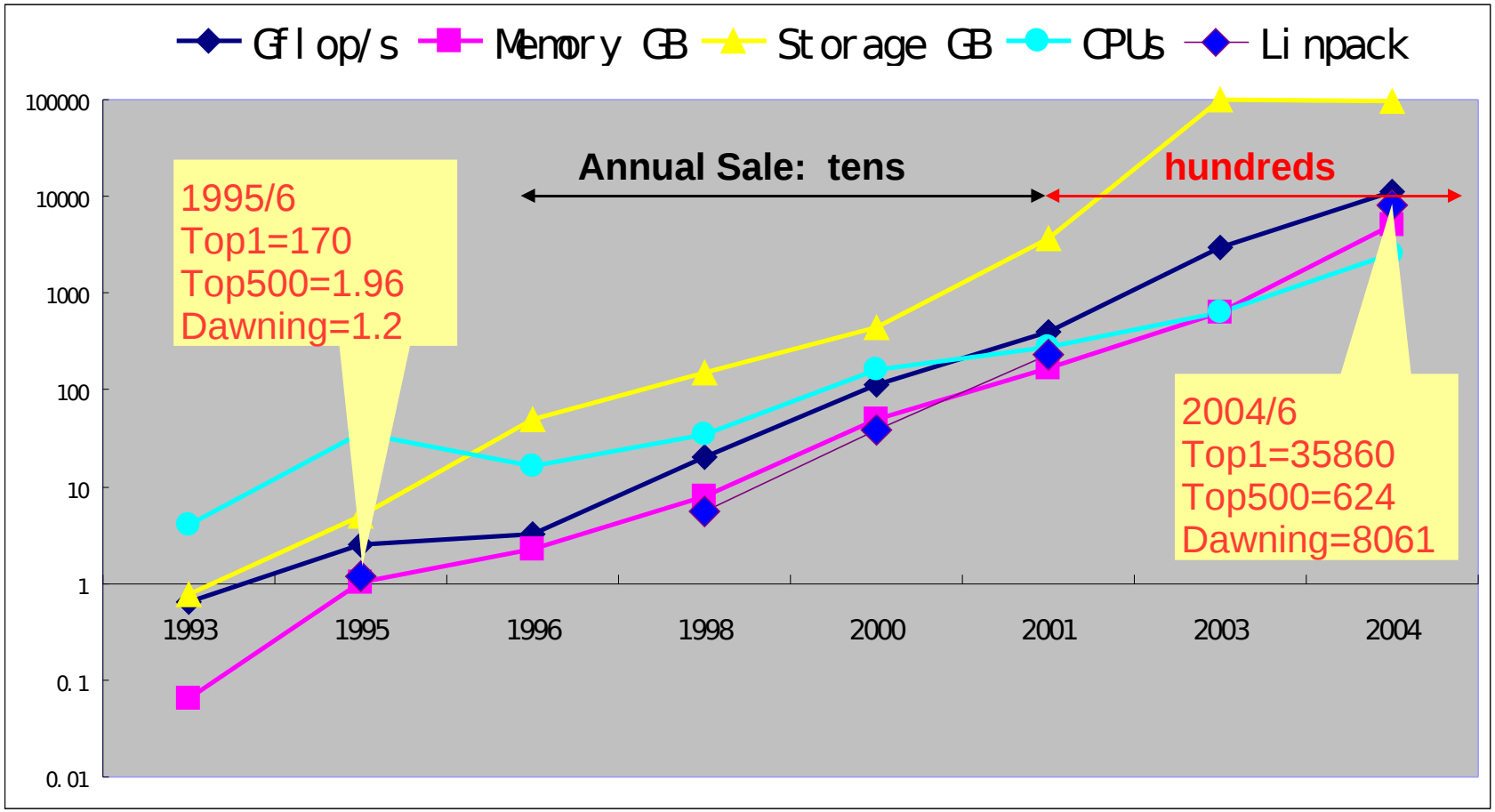
TOP500 SUPERCOMPUTER SITES

ISC 2004 — 19th International Supercomputer Conference

TOP500 List for June 2004

Rank	Site Country/Year	Computer / Processors Manufacturer	Computer Family Model	Inst. type Installation Area	R_{max} R_{peak}	N_{max} n_{half}
1	Earth Simulator Center Japan/2002	**Earth-Simulator** / 5120 NEC	**NEC Vector** SX6	**Research**	35860 40960	1.0752e+06 266240
2	Lawrence Livermore National Laboratory	*Thunder* **Intel Itanium2 Tiger4 1.4GHz**	**NOW - Intel Itanium** Itanium2 Tiger4	**Research**	19940 22938	975000 110000
3	Los Alamos National Laboratory United States/2002	**ASCI Q - AlphaServer SC45, 1.25 GHz** / 8192	**HP AlphaServer** Alpha-Server-Cluster	**Research**	13880 20480	633000 225000
4	IBM - Thomas Watson Research Center	**BlueGene/L DD1 Prototype (0.5GHz PowerPC 440**	**IBM BlueGene/L** BlueGene/L	**Research**	11680 16384	331775
5	NCSA United States/2003	*Tungsten* **PowerEdge 1750, P4 Xeon**	**Dell Cluster** PowerEdge 1750,	**Academic**	9819 15300	630000
6	ECMWF United Kingdom/2004	**eServer pSeries 690 (1.9 GHz Power4+)** / 2112	**IBM SP** SP Power4+,	**Research** Weather and	8955 16051	350000
7	Institute of Physical and Chemical Res. (RIKEN)	**RIKEN Super Combined Cluster** / 2048	**Fujitsu Cluster** Fujitsu Cluster	**Research**	8728 12534	474200 120000
8	IBM - Thomas Watson Research Center	**BlueGene/L DD2 Prototype (0.7 GHz PowerPC 440)** /	**IBM BlueGene/L** BlueGene/L	**Research**	8655 11469	294911
9	Pacific Northwest National Laboratory	*Mpp2* **Itanium2**	**HP Cluster** Integrity rx2600	**Research**	8633 11616	835000 140000
10	Shanghai Supercomputer Center China/2004	**Dawning 4000A, Opteron 2.2 GHz, Myrinet** / 2560 Dawning	**NOW - AMD** NOW Cluster - AMD - Myrinet	**Research**	8061 11264	728400 180000
11	Los Alamos National Laboratory United States/2003	*Lightning* **Opteron 2 GHz, Myrinet** /	**NOW - AMD** NOW Cluster - AMD -	**Research**	8051 11264	761160 109208
12	Lawrence Livermore National Laboratory United States/2002	**MCR Linux Cluster Xeon 2.4 GHz - Quadrics** / 2304 Linux Networx/Quadrics	**NOW - Intel Pentium** NOW Cluster - Intel Pentium - Quadrics	**Research**	7634 11060	350000 75000

Evolution of Dawning HPC Systems



Evolution of HPC in China: What Next?

Dawning1000, 1995
intel i860
First MPP Computer in China
2.5Gflops Peak

Dawning2000, 1999,
Motorola PowerPC
First SMP Cluster in China
100Gflops Peak

Dawning3000, 2001
IBM Power3
SUMA Cluster in China
400Gflops Peak

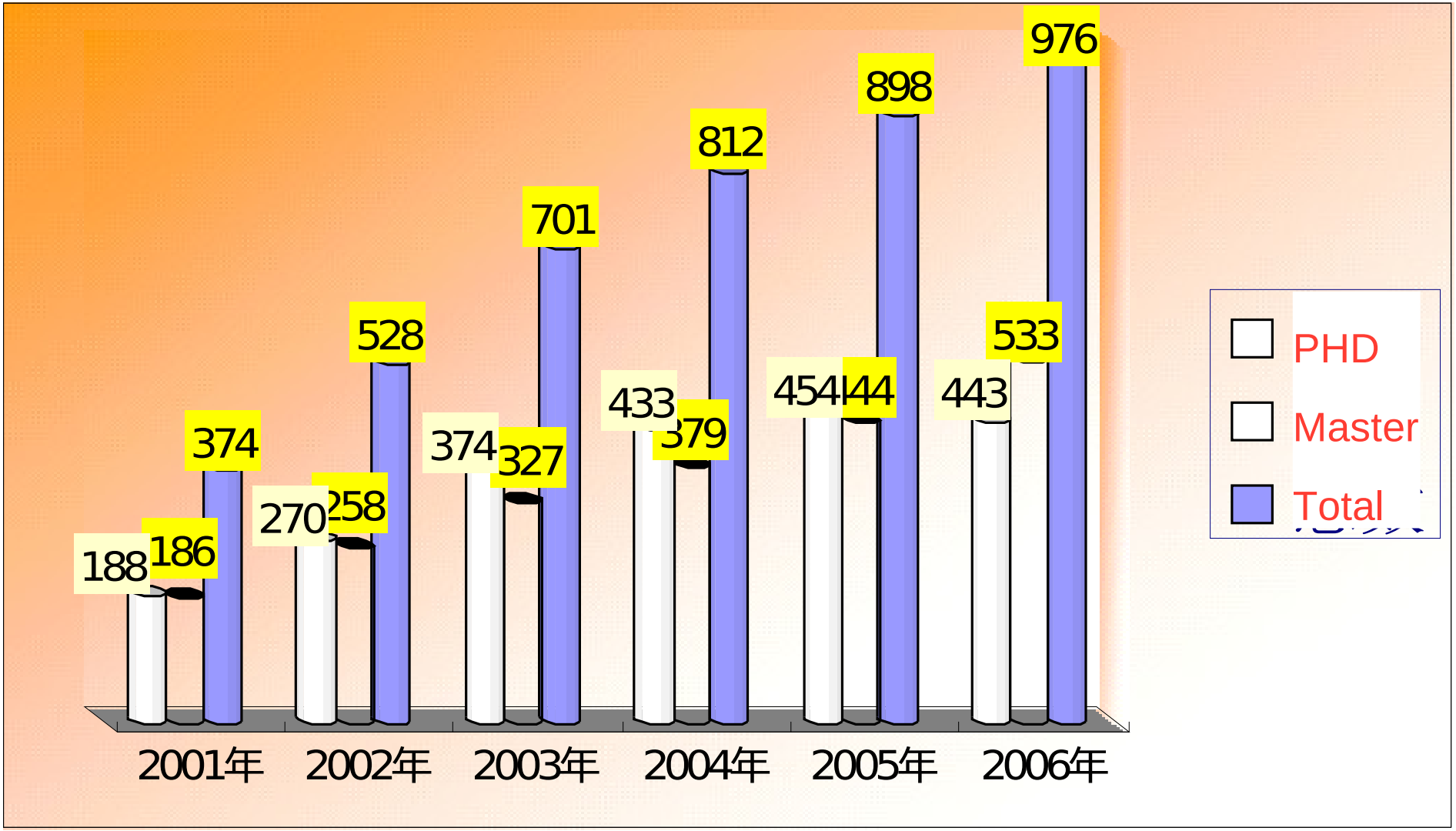
Dawning4000-A, 2004
AMD Opteron
Grid-enabling Cluster
11.2 Tflops Peak

Dawning5000-A, 2008
6400 AMD 4-core Opteron
220 Tflops Peak

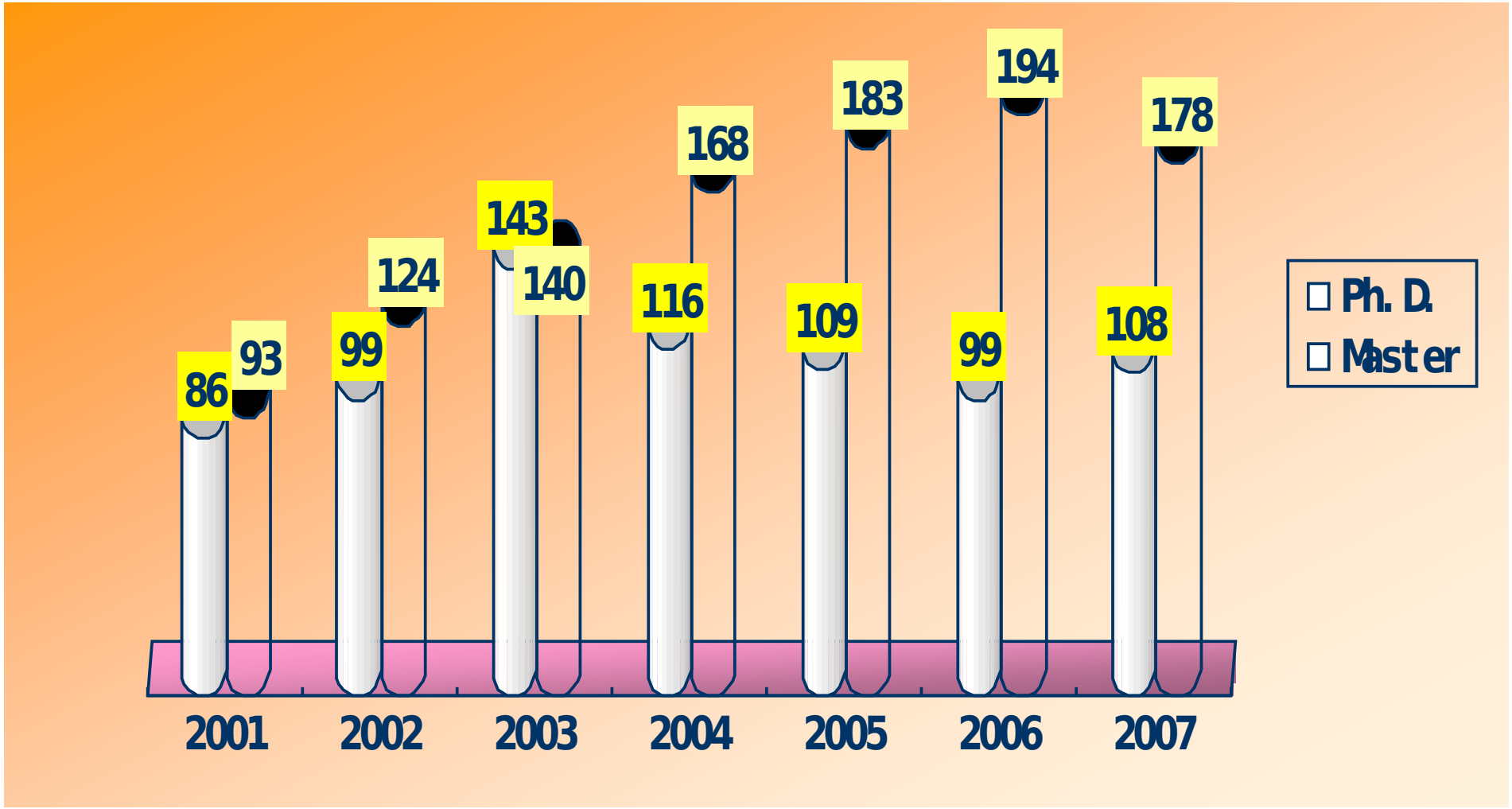
???

PetaFLOPS
Godson-3
2010

2001~2006 Graduate Student Enrolment



2000~2007 New Admissions



Academia and Professional

- International Journal Editorial Boards (>10)
  - IEEE Transactions on Computers
  - Parallel Computing
  - Journal of Systems and Software
  - Information and Management
  - ……
- IEEE Computer Society Beijing Center
- Journal of Computer Science & Technology
  - published by Springer
- International Conferences

Contents

- A brief introduction to ICT
- A brief introduction to Godson processors
- The Godson-3 multi-core processor
- PetaFLOPS and TeraFLOPS

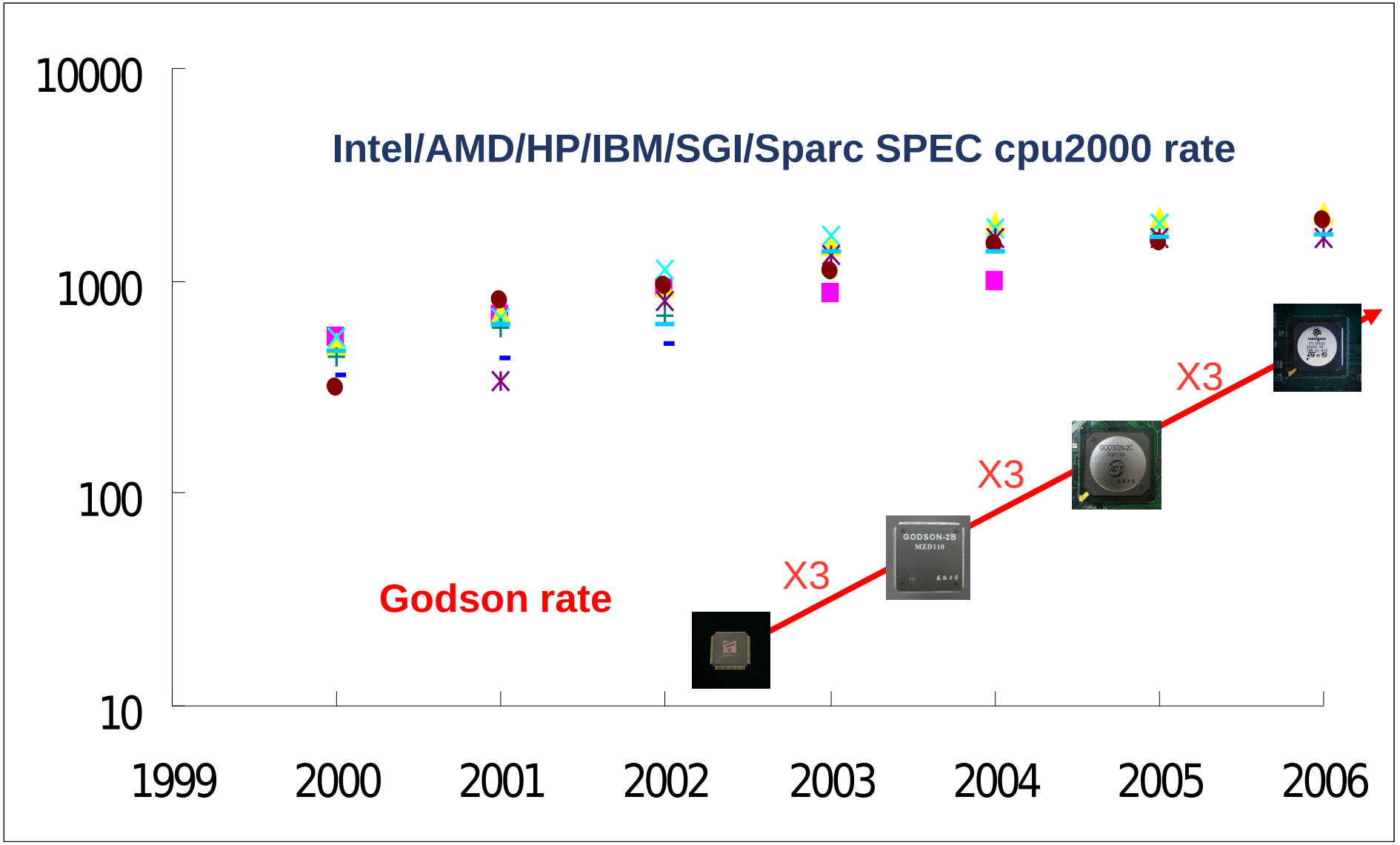
National Project

- High performance CPU is national strategic product
  - Chinese IT industry is big but not strong: 5.6 trillion RMB in 2007, only 22% by domestic companies, 3.75% profits
- Godson CPU is supported by
  - National 863 project
  - National 973 project
  - National Science Foundation of China
  - National key project
  - Key project of Chinese Academy of Sciences

Godson CPU Briefs

- ICT started Godson CPU design in 2001.
- The 32-bit Godson-1 CPU in 2002 is the first general purpose CPU in China.
- The 64-bit Godson-2B in 2003.10
- The 64-bit Godson-2C in 2004.12
- The 64-bit Godson-2E in 2006.03
- Each Triple the performance of its previous one.

Godson Development



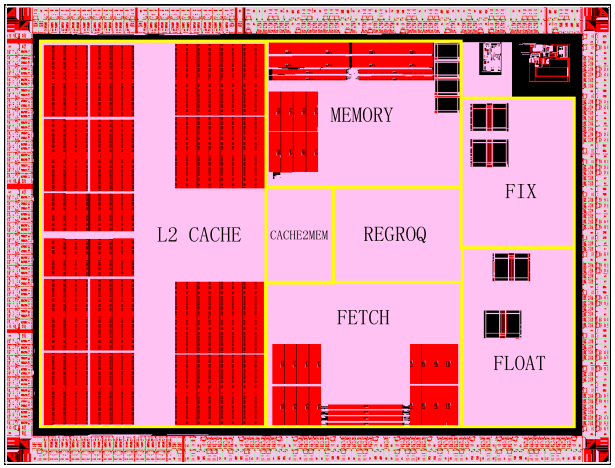
Godson-2E SPEC CPU2000 Rate

Programs	Reftime	Run time	Ratio
164.gzip	1400	403	347
175.vpr	1400	273	512
176.gcc	1100	221	497
181.mcf	1800	307	586
186.crafty	1000	167	598
197.parser	1800	472	382
252.eon	1300	188	690
253.perlbmk	1800	354	508
254.gap	1100	240	458
255.vortex	1900	263	722
256.bzip2	1500	365	411
300.twolf	3000	645	465
SPEC int2000			<503>

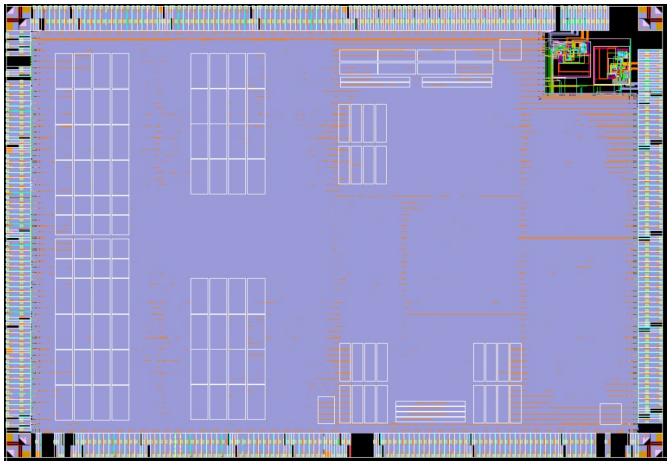
Programs	Ref time	Run time	Ratio
168.wupwise	1600	238	672
171.swim	3100	660	469
172.mgrid	1800	579	311
173.applu	2100	549	382
177.mesa	1400	221	634
178.galgel	2900	412	704
179.art	2600	416	624
183.equake	1300	208	624
187.facerec	1900	300	632
188.ammp	2200	432	509
189.lucas	2000	396	506
191.fma3d	2100	531	395
200.sixtrack	1100	345	319
301.apsi	2600	528	493
SPEC fp2000			<503>

Godson-2E and Godson-2F

- 1.0GHz@90nm CMOS, 5-7W
- 47M xtors, area 36mm^2
- Godson-2 CPU Core
  - 64-bit MIPS III Compatible
  - Four-issue, OOO
  - 64KB+64KB L1 (four-way)
  - 512KB L2 (four-way)
- On-chip DDR Controller
- SysAD Front-end bus



- 1.0GHz@90nm CMOS, 3-5W
- 51M xtors, area 43mm^2
- Godson-2 CPU Core
  - 64-bit MIPS III Compatible
  - Four-issue, OOO
  - 64KB+64KB L1 (four-way)
  - 512KB L2 (four-way)
- On-Chip DDR2 controller.
- PCI/PCIX, Local IO, GPIO, etc.
- Volume production



Some Applications

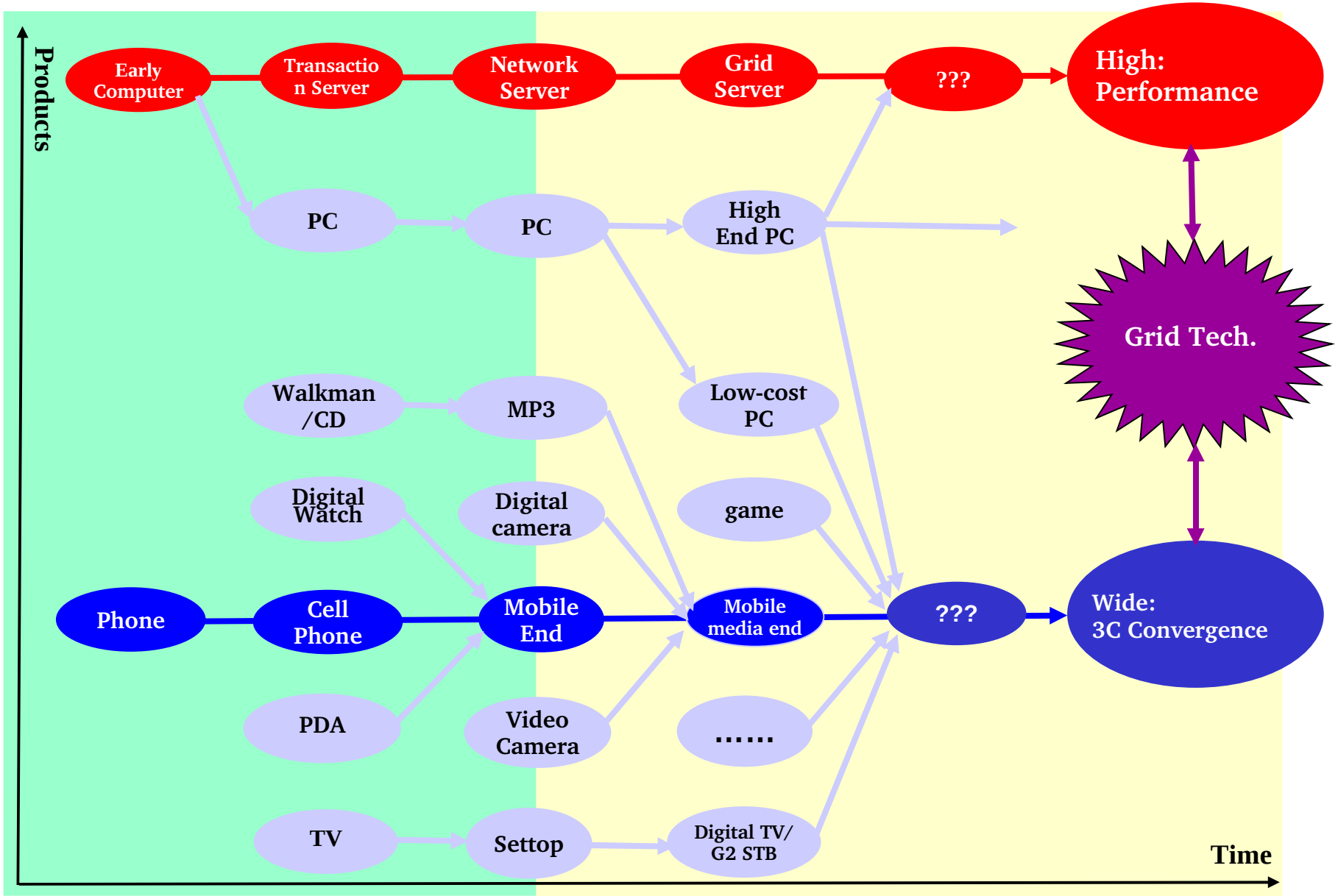
- With the high performance features, Loongson-2 CPU is welcome by many customers
  - Low-cost PC & notebook
  - Network applications
  - Low-end servers & HPC
  - High-end embedded applications.
- Million units order



Godson-2 Architecture Features

- 64-bit out-of-order execution pipeline
  - 9 stage pipeline, four issue
  - Dynamic scheduling: Group RS(16 fix+16 float), 64-entry ROB
  - Register renaming: 64-entry physical register file
  - Branch prediction: Gshare, BTB, RAS, 8-entry branch queue
  - Five Function units: tow fix, two float (SSE2-lke media), one memory
- Memory Hierarchy
  - 64KB instruction cache and 64KB data cache, 4-way set associated
  - TLB: 64-entry fully associated, two 4KB-4MB page each, separate 16 entry ITLB
  - 24 non-blocking accesses & on-the-fly memory disambiguation
  - Load speculation: return values on previous pending stores
  - 512KB-1MB L2 Cache
  - On-Chip memory controller
- Word-level CPU core
  - In-stat Report: The sophistication of the Godson-2 shows that the Chinese are poised to produce microprocessors as powerful as any in the world

Godson Roadmap: the Big Version



A View from Berkeley 2.0

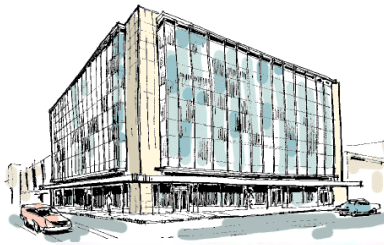
The Parallel Computing Landscape: A View from Berkeley 2.0

Krste Asanovic, Ras Bodik, Jim Demmel, Tony Keaveny, Kurt Keutzer, John Kubiatowicz, Edward Lee, Nelson Morgan, George Necula, Dave Patterson, Koushik Sen, John Wawrzynek, David Wessel, and Kathy Yelick

October, 2007

Re-inventing Client/Server

- "The Datacenter is the Computer"
  - Building sized computers: Google, MS, …



- "The Laptop/Handheld is the Computer"
  - '07: HP no. laptops > desktops
  - 1B Cell phones/yr, increasing in function
  - Otellini demoed "Universal Communicator"
    - Combination cell phone, PC and video device
  - Apple iPhone

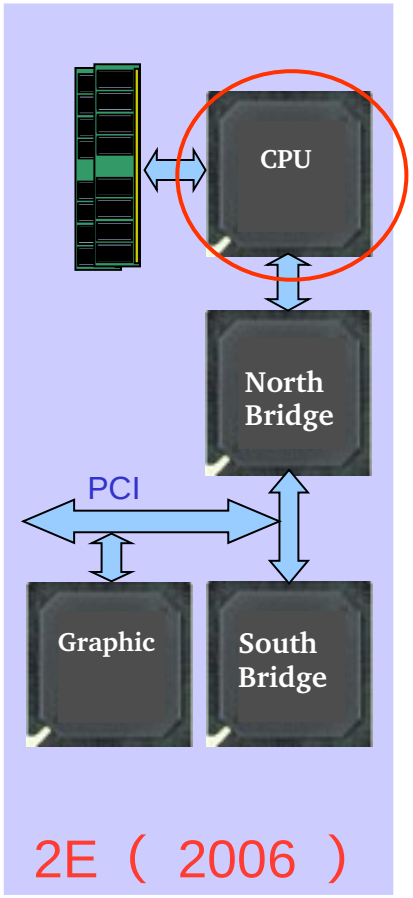
- Laptop/Handheld as future client, Datacenter as future server

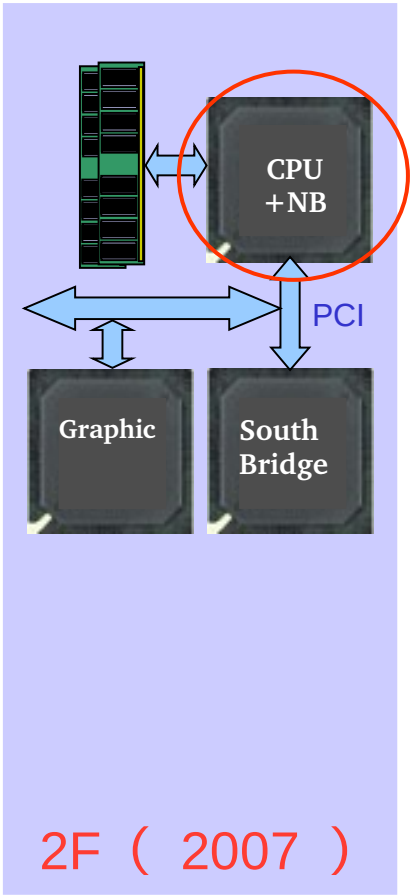
Laptop/Handheld Reality

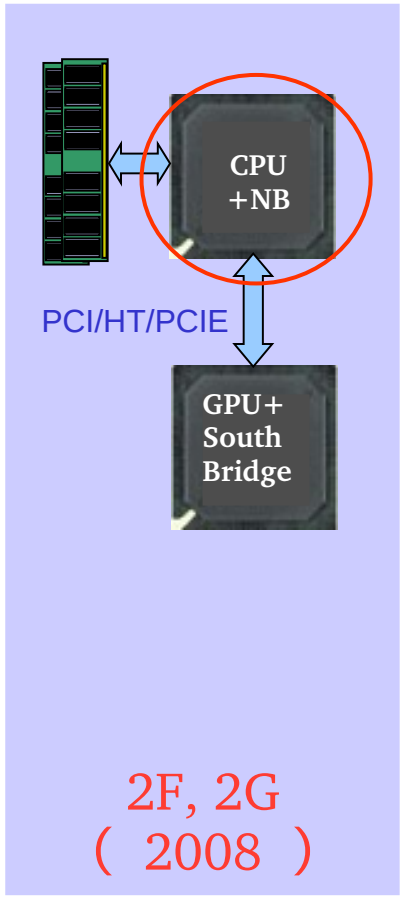
- Use with large displays at home or office
- What % time disconnected? 10%? 30% 50%?
  - Disconnectedness due to Economics
    - Cell towers and support system expensive to maintain
      ⇒ charge for use to recover costs ⇒ costs to communicate
    - Policy varies, but most countries allow wireless investment where make most money ⇒ Cities well covered
      ⇒ Rural areas will never be covered
  - Disconnectedness due to Technology
    - No coverage deep inside buildings
    - Satellite communication uses up batteries

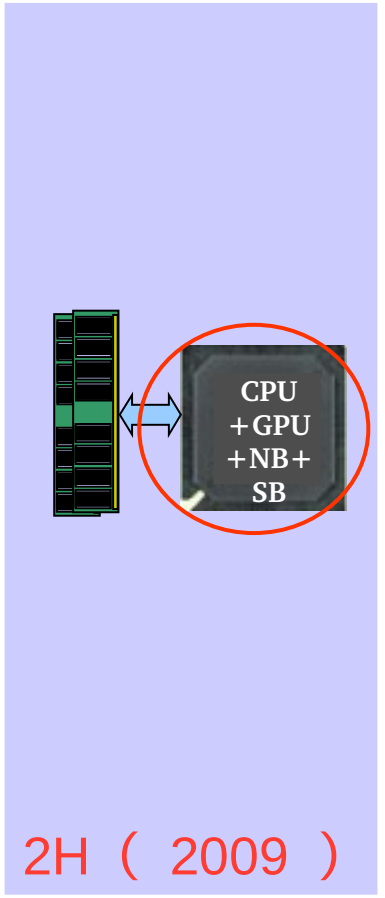
⇒ Need computation & storage in Laptop/Handheld

Low end roadmap: From CPU to SOC









2G vs. 2F

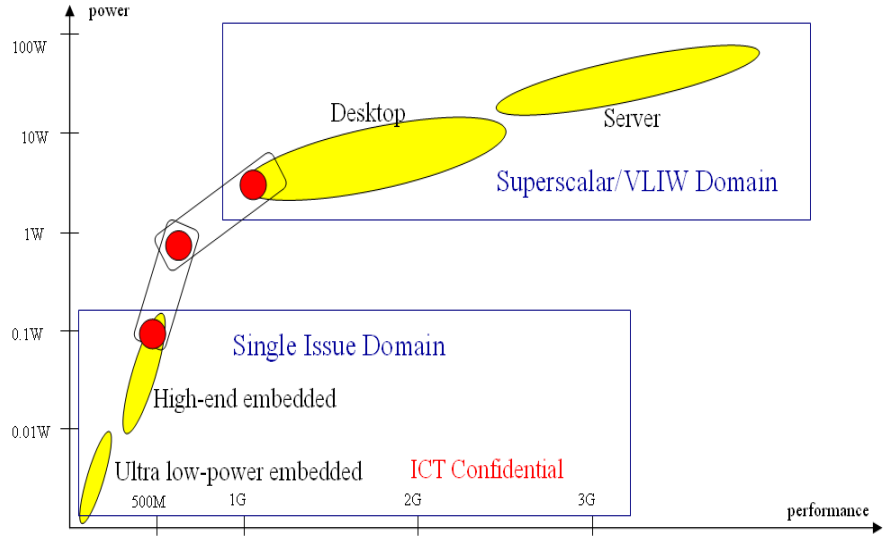
- 65nm vs. 90nm
- A little higher frequency: 5%
- Similar die size: 40-50 mm^2
- Half power consumption due to 65nm: 1w-3w
- Improved CPU core
  - MIPS64 compatible
  - X86 binary translation support
  - ECC for reliability, EJTAG for debug
  - 128-bit memory access
  - 1MB L2 vs. 512KB
- Much more flexible IOs
  - HT, PCI/PCIX, LPC, SPI, UART, GPIO

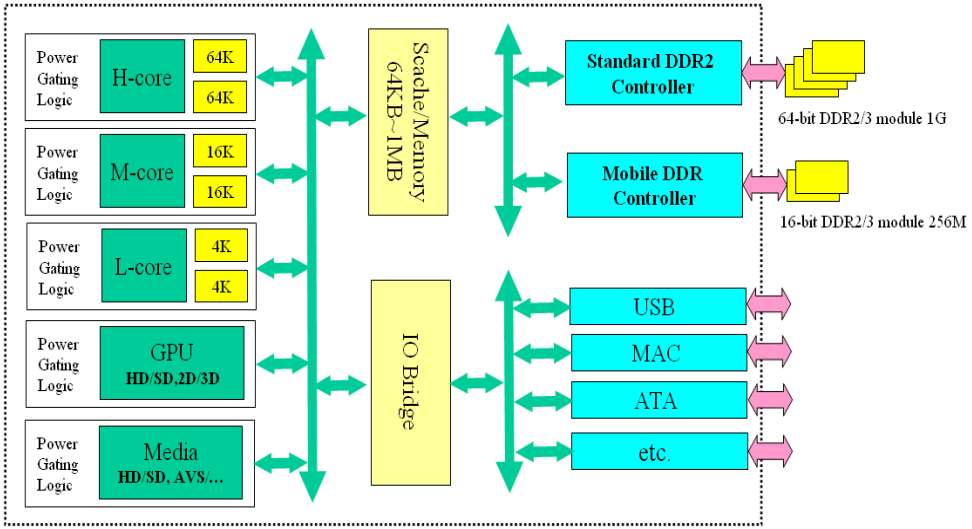
South Bridge of 2F and 2G

- For both 2F and 2G: PCI/PCIX interface
- Tolerate the current flaw of 2F PCI
  - Out of order instruction fetch
  - Low efficiency with SM502
- Only for low cost PC, as simple as possible
  - Simple GPU + 16/32 bit DDR2/3 for separate video RAM
  - USB, MAC, SATA, etc.
  - 2F+SB or 2G+SB: ~ $25
- If the user needs high end SB, can be connect to HT, PCIE, PCI/PCIX
- A low-end GS232 core integrated to be a stand alone SOC for some applications.

Loongson-2H: Single-chip Hetero multicore

- Closing the gap between
  - desktop performance and handheld low power consumption
- Single chip for mobile, desktop, and settop box
  - SP2: Scaling Performance by Scaling Power
- Hetero multicore
  - L-mode(0.1W):
  - M-mode(1.0W):
  - H-mode(5-10W):





Contents

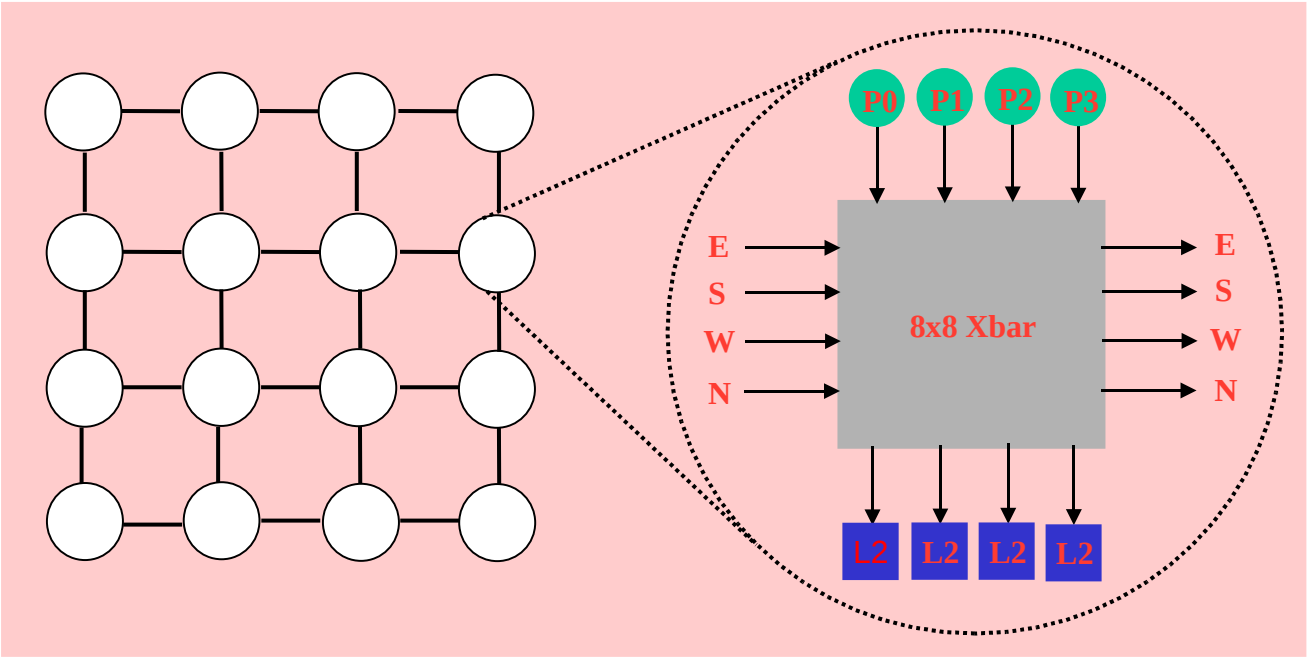
- A brief introduction to ICT
- A brief introduction to Godson processors
- The Godson-3 multi-core processor
- PetaFLOPS and TeraFLOPS

Godson-3 Briefs

- Distributed scalable architecture
- Reconfigurable architecture
- X86 binary translation speedup
- Low Power Consumption
- >1.0GHz@65nm

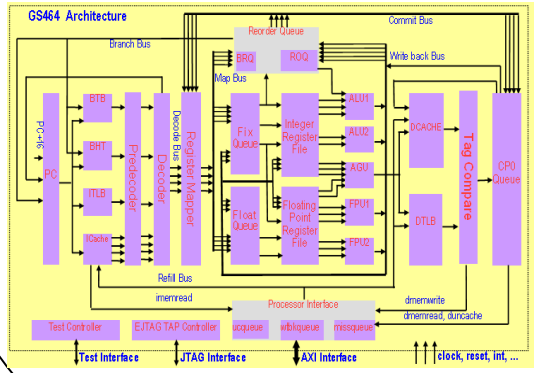
Scalable Architecture

- Scalable interconnection network
  - Crossbar + Mesh
  - Single crossbar connects cores, L2s, and four directions
- Directory-based cache coherence protocol
  - Distributed L2 caches are global addressed
  - Each cache block has a directory entry
  - Both data cache and instruction cache are recorded in directory

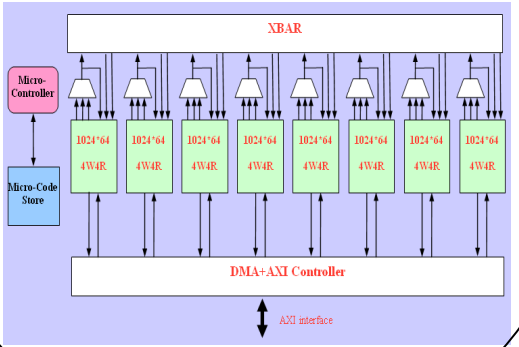


Reconfigurable Architecture

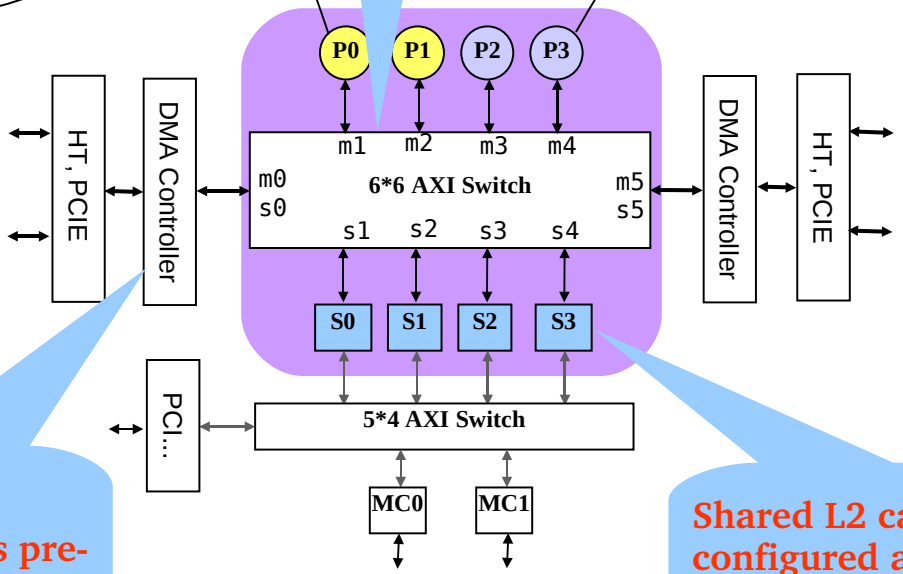
General Purpose Core, 64-bit, 4-issue, OOO, AXI interface



Multiple Purpose Core LINPACK, biology, signal processing, AXI interface



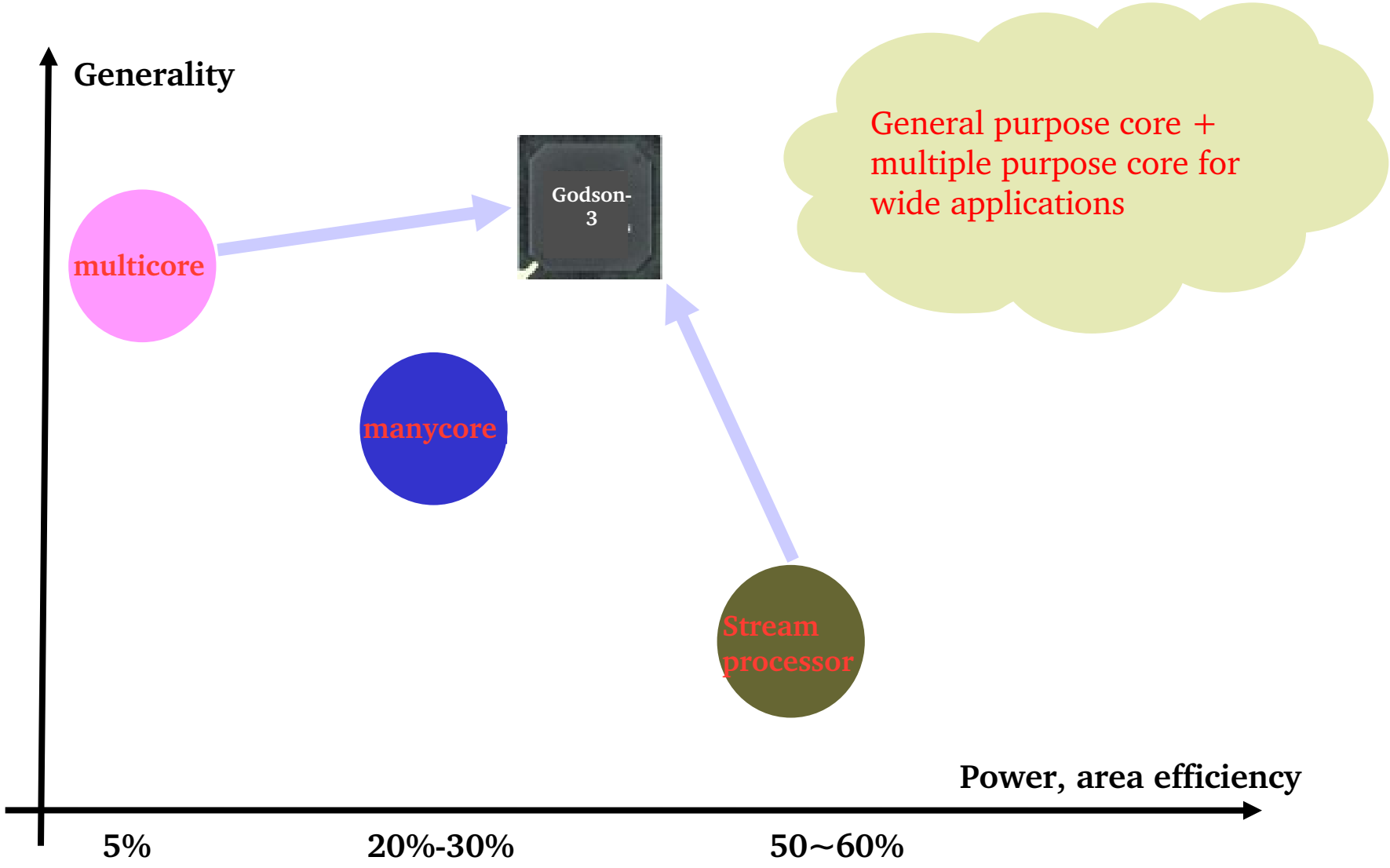
8 configurable address windows of each master port allow pages migration across L2 and memory



DMA engine supports pre-fetch and matrix

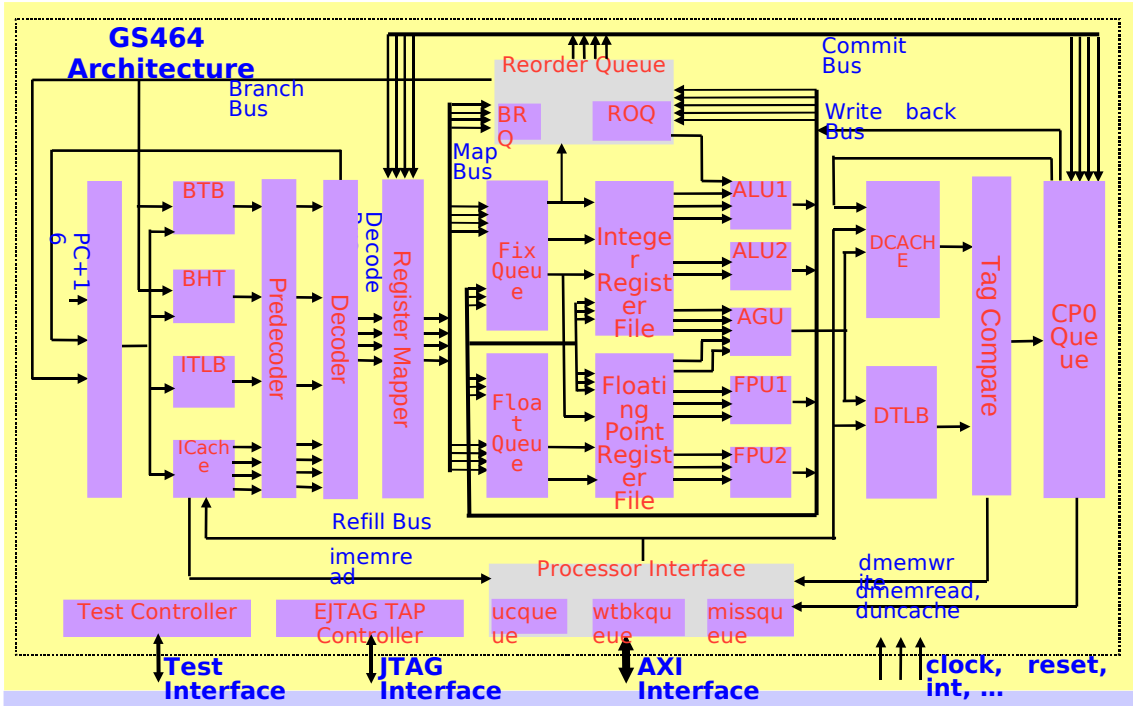
Shared L2 can be configured as internal RAM, DMA to internal RAM directly

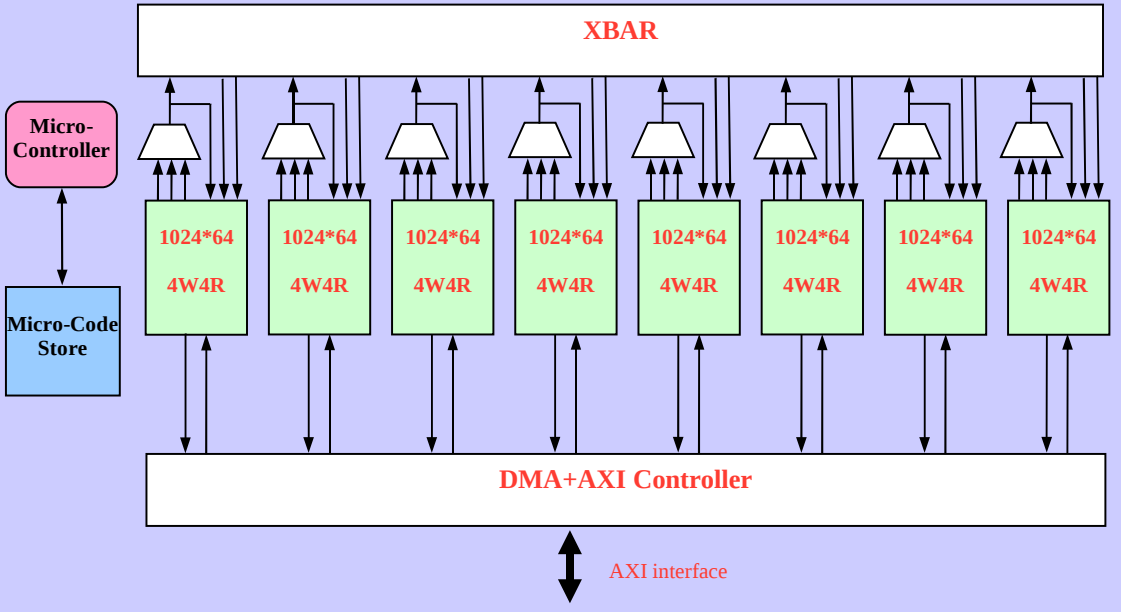
Generality vs. Efficiency



GS464 general purpose core

- MIPS64, 200+ more instructions for X86 binary translation and media acceleration
- Four-issue superscalar OOO pipeline
- Two fix, two FP, one memory units
- Two FP units each supports full pipelined double/paired-single MAC operation
- 48-bit VA and PA, 128-bit memory access
- 64KB icache and 64KB dcache, 4-way
- 64-entry fully associated TLB, 16-entry ITLB, variable page size
- Non-blocking accesses, load speculation
- Directory-based cache-coherence for CMP
- Parity check for icache, ECC for dcache
- EJTAG for debugging
- Standard 128-bit AXI interface



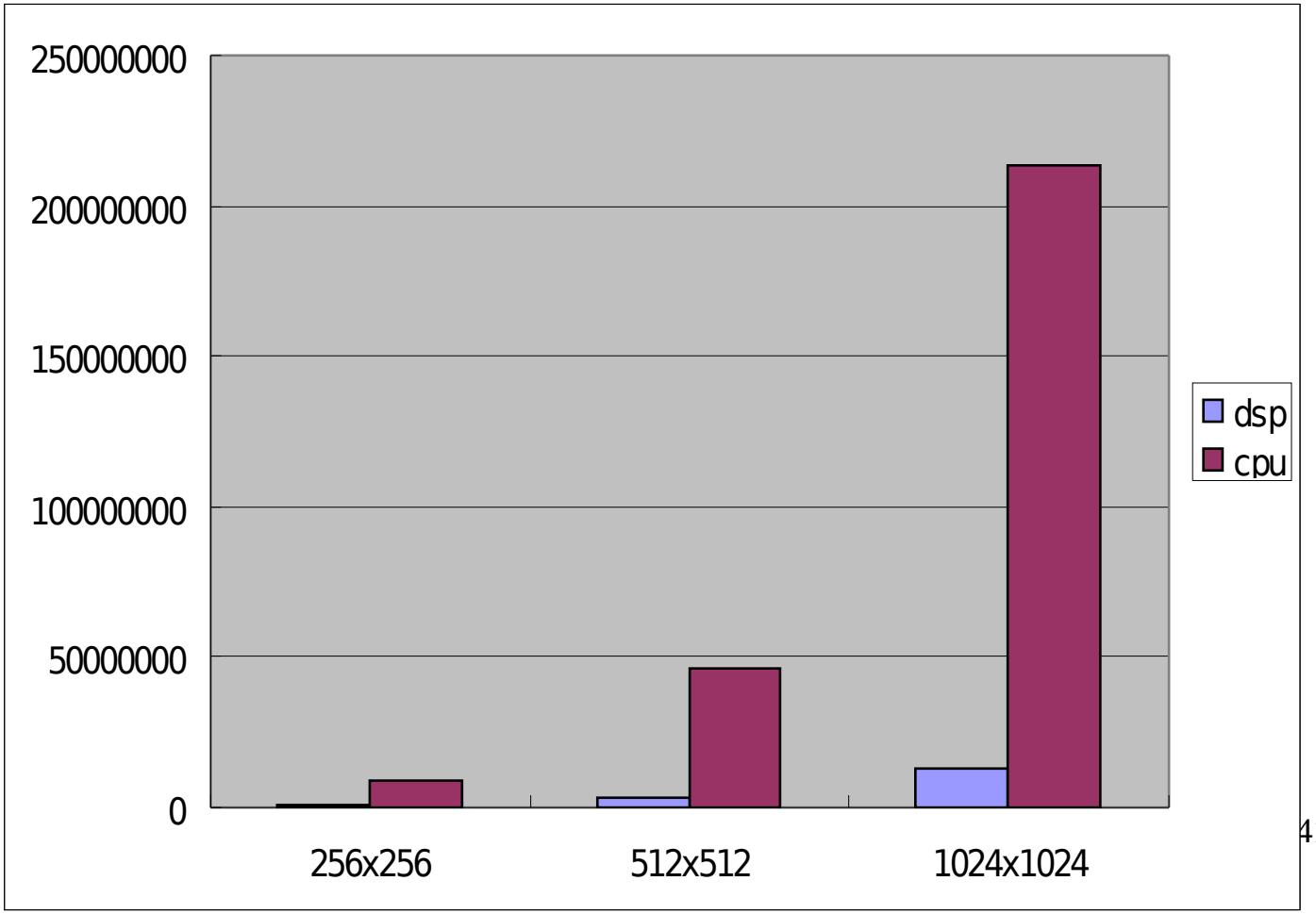


GStera multiple purpose core

- Target for LINPACK, biology computation, signal processing, etc
- 8-16 MACs per node
- Big multi-port register file
- Reconfigurable based on applications.
- Standard 128-bit AXI interface

Matrix Transposing Performance

- **15+times faster**



Hardware Support for X86 Binary Translation

- **Define new instructions**
  - X86 ISA function and MIPS ISA format
  - Binary translation mechanism supporting
  - >200 instructions are defined with 5% additional silicon cost
- **Speedup X86-to-MIPS binary translation by**
  - 10 times

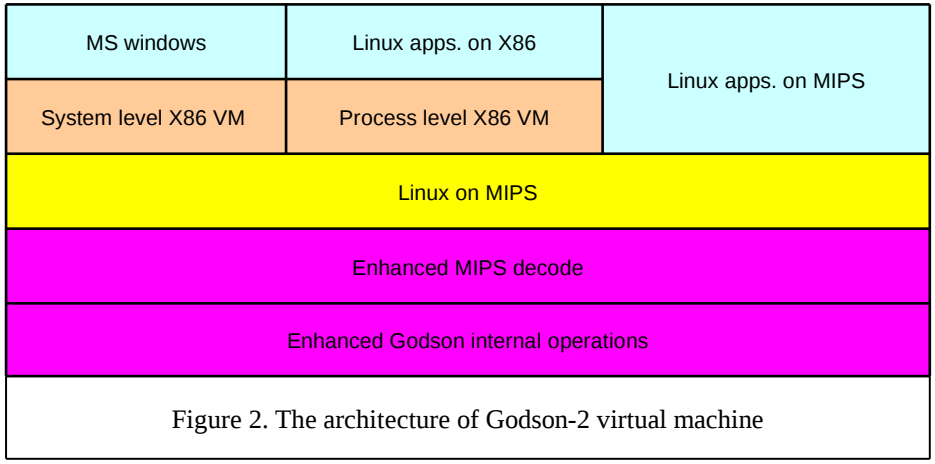
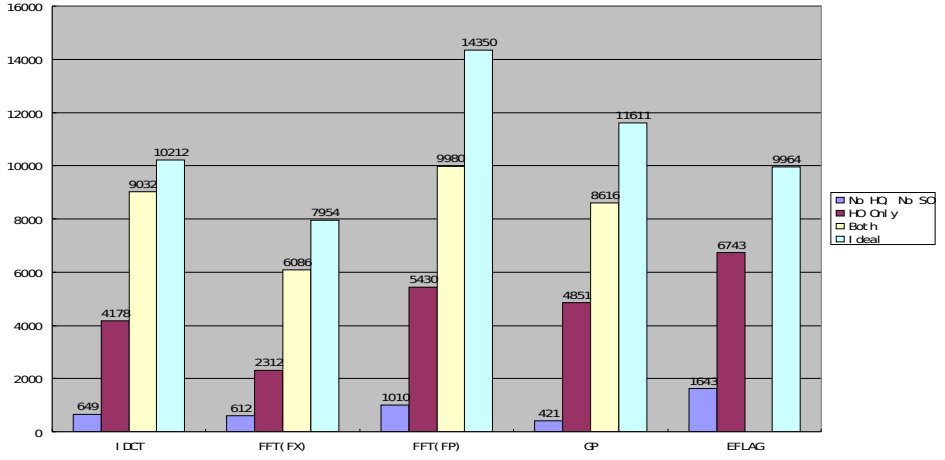
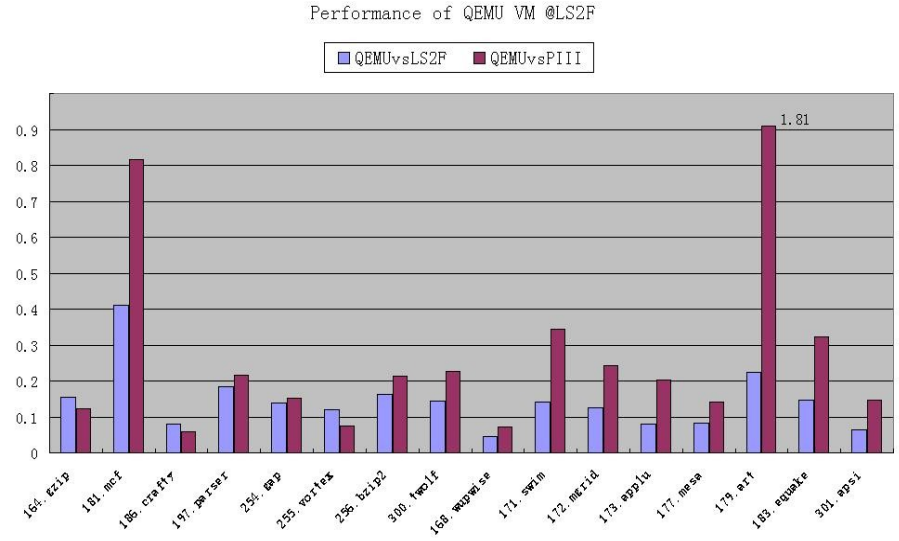


Figure 2. The architecture of Godson-2 virtual machine



Preliminary Virtual Machine Performance

- QEMU@LS2F 800MHz vs. PIII 850MHz
  - Average efficiency 18% (except mcf and art)
  - INT 15%, FP 21%
- QEMU@LS2F 800MHz vs. LS2F 800MHz
  - Average efficiency 14%
  - INT 17%, FP 11%
- QEMU @LS3 vs. LS2F
  - Average efficiency 27%
  - INT 34%, FP %19



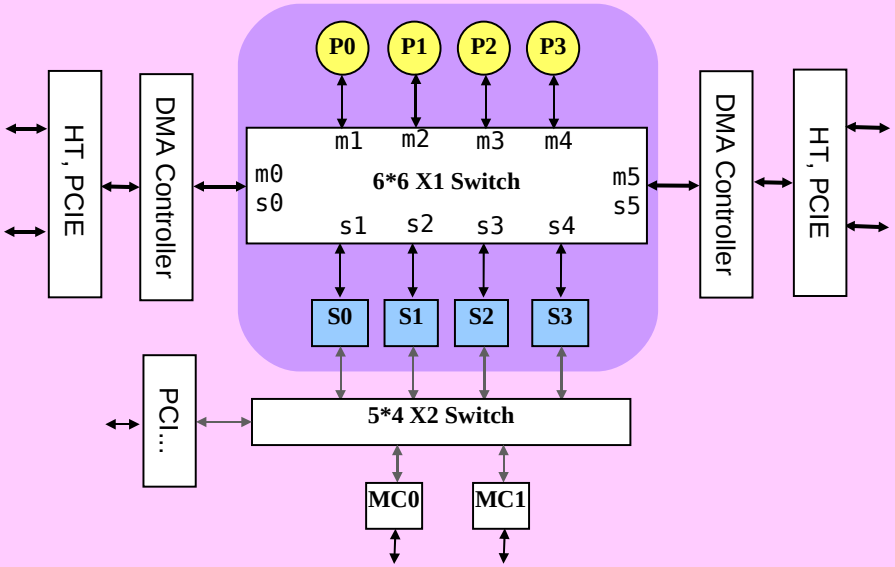
Low Power Design

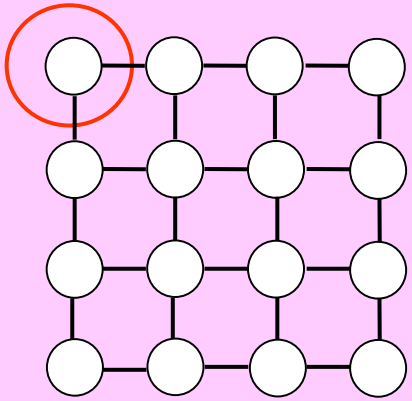
- Low leakage process is selected
  - LP/GP mixed
  - HVT/SVT mixed
- Manual clock gating regarding architecture
  - Much efficient in reducing power consumption compared to P&R tools
- Power management
  - Module level (CPU core, HT, PCIE, DDR2) clock gating
  - Frequency Scaling
  - Temperature Sensor

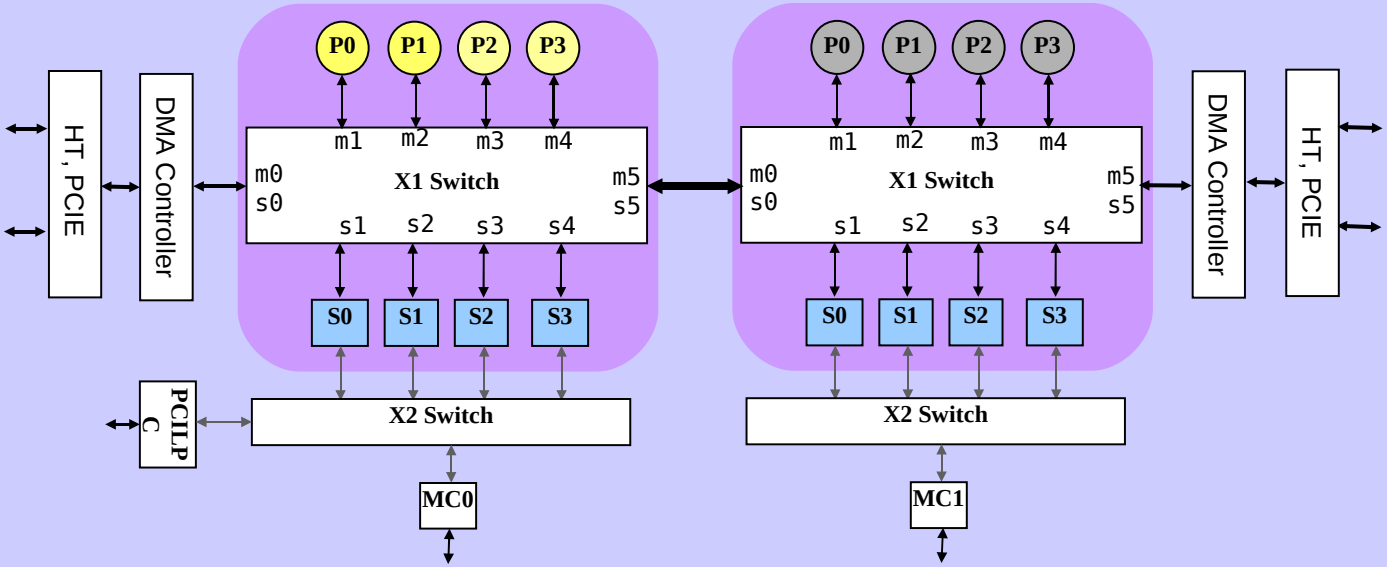
Physical implementation

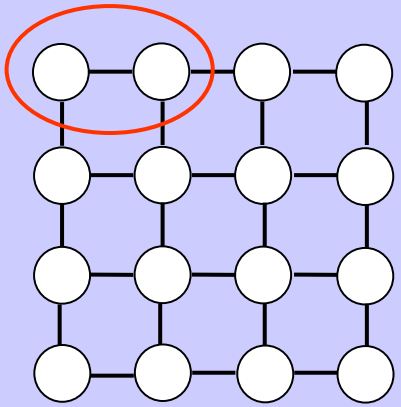
- 65nm CMOS LP/GP technology
- Cell-based design methodology
  - DC + ICC
  - Manual P&R for critical cells
- 2008: 4-core (4GP + 0MP) + 4MB L2
  - GP: General purpose core
  - MP: Multiple purpose core
  - 10w@1GHz
- 2009: 8-core (4GP + 4MP) + 4MB L2
  - 20w@1GHz

4-core and 8-core

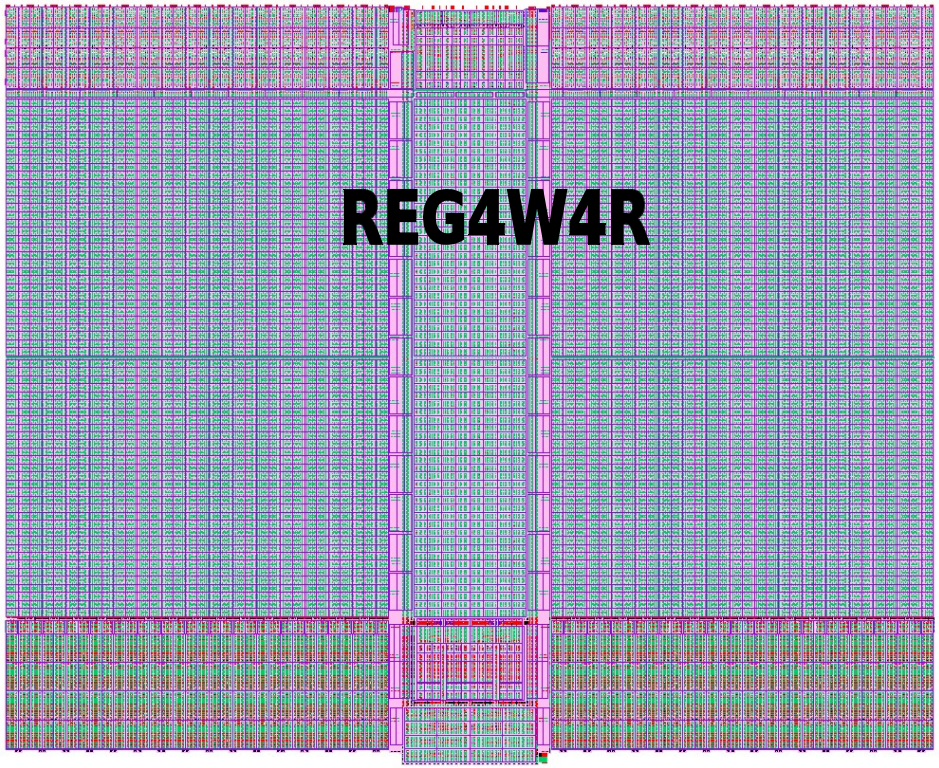








Full Customer Register file and CAM

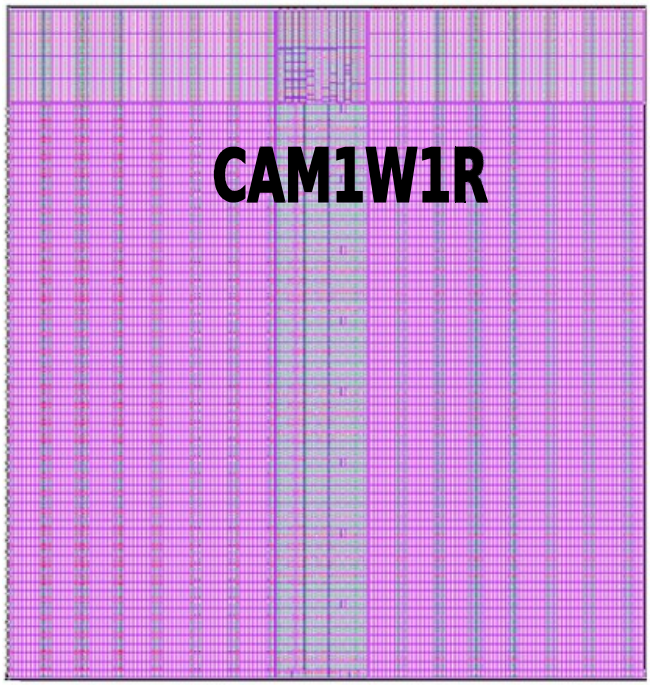


Physical register file

Size: 321um x 262um

Power: 50mW@1GHz

Delay: 470ps



TLB CAM

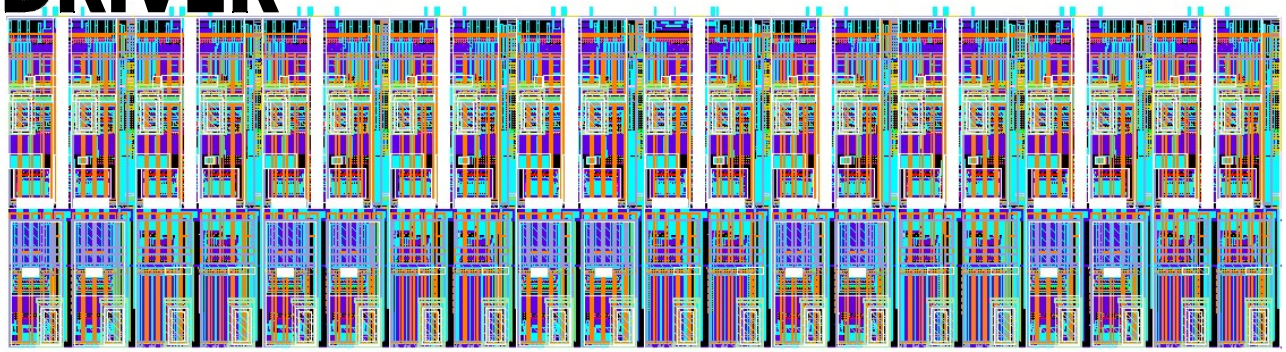
Size: 224 um x 235 um

Power: 55mw @ 1GHz

Delay: 550ps

HyperTransport PHY

DRIVER

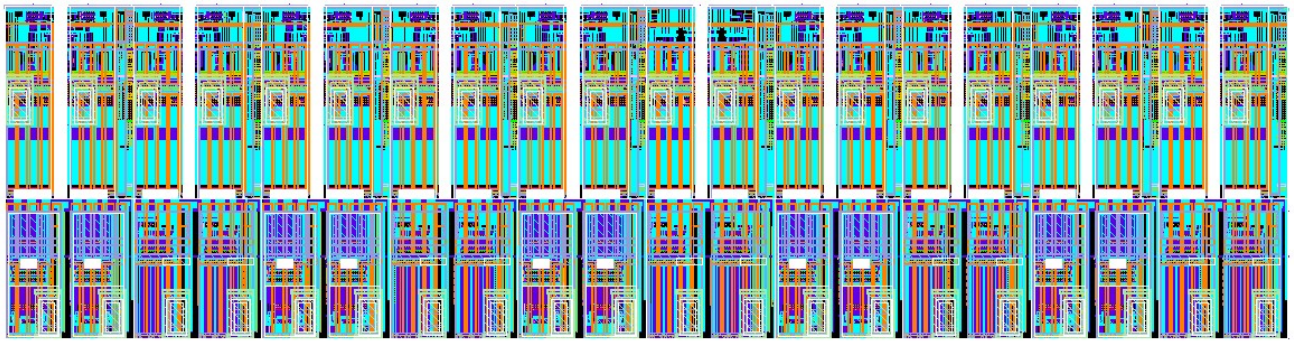


HT1.0 Driver & Receiver

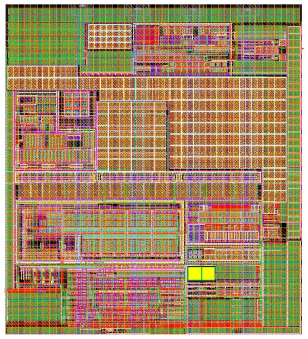
FlipChip Compatible 2Row design

800mw @ 1.6Gbps

RECEIVER



PLL



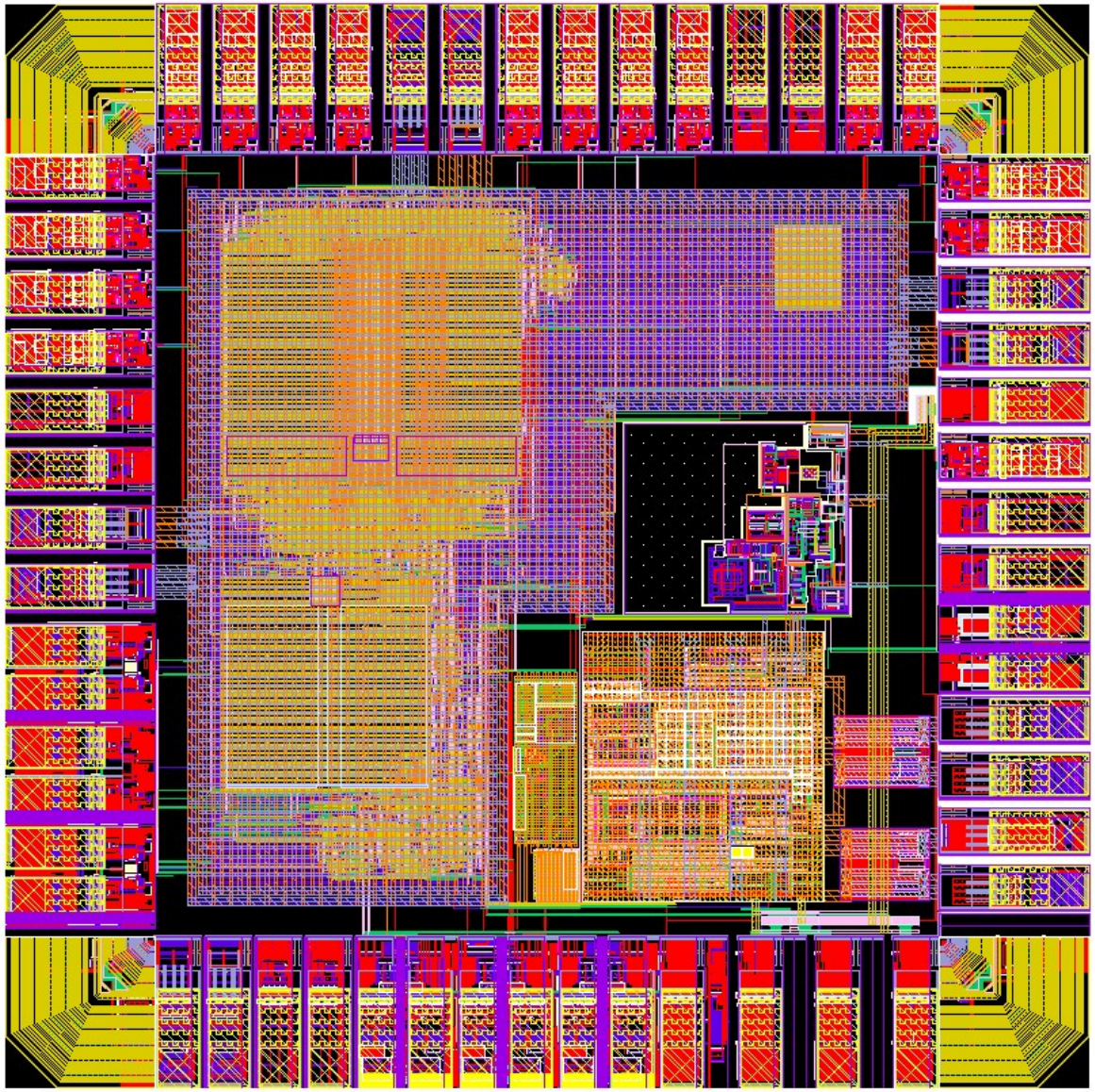
Size: 250um x 300um

Power: < 10mW

Freq: 1.6GHz

Jitter: 20 ps

Test Chip for Customer Blocks



TEST CHIP

ST 65nm

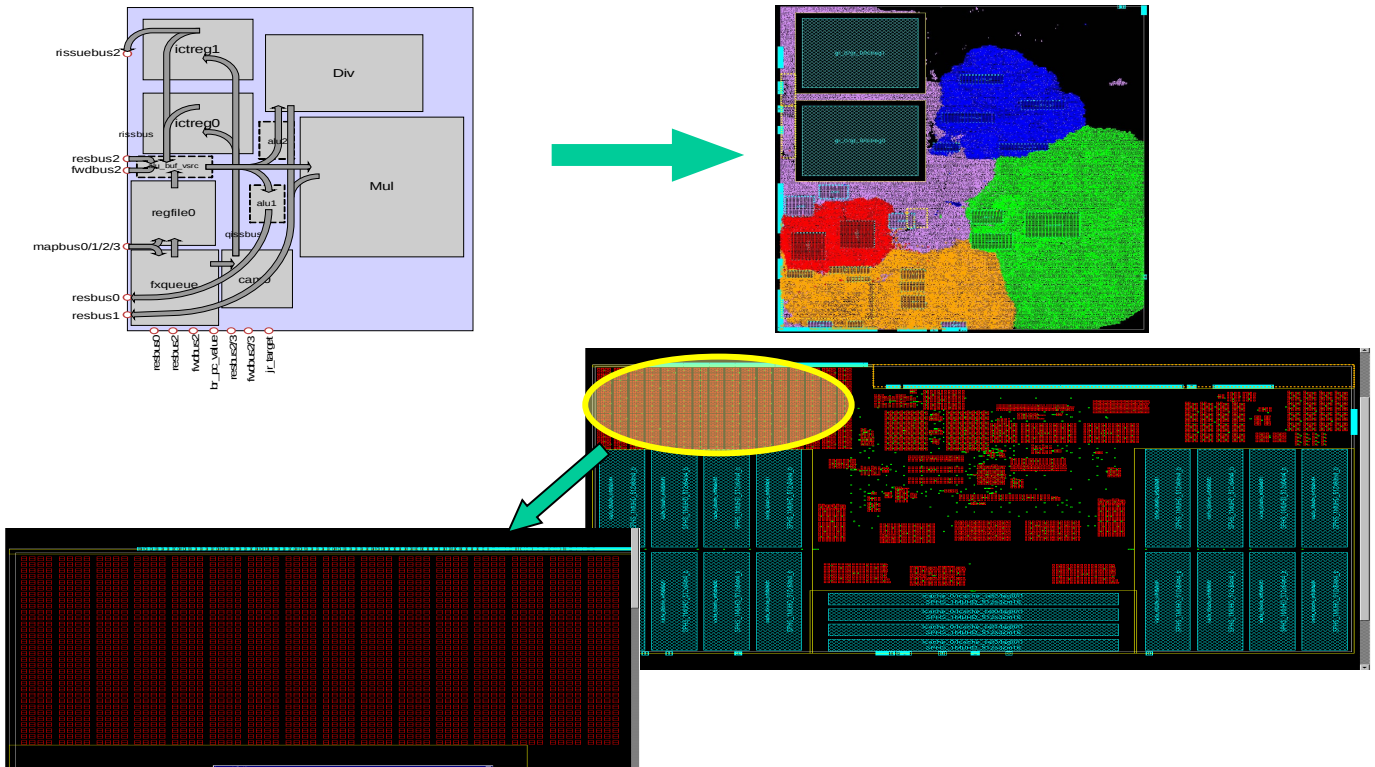
1206um x 1206um

Function:

- **CAM1W1R - BIST**
- **CAM1W1R - Scan**
- **RAM4W4R - BIST**
- **RAM4W4R - Scan**
- **ICT_PLL - Freq. test**
- **HT1.0 - BIST**
- **HT1.0 - Error rate test**

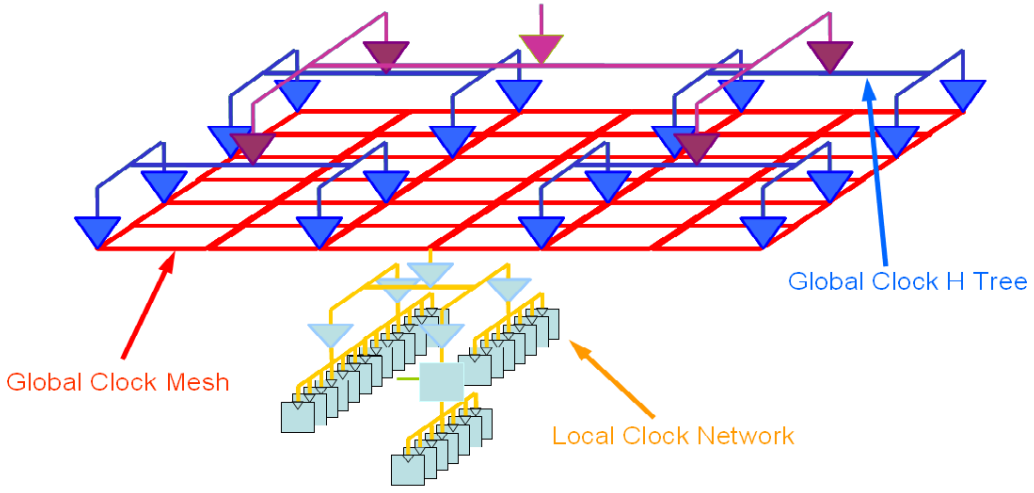
Cell-based high performance physical design

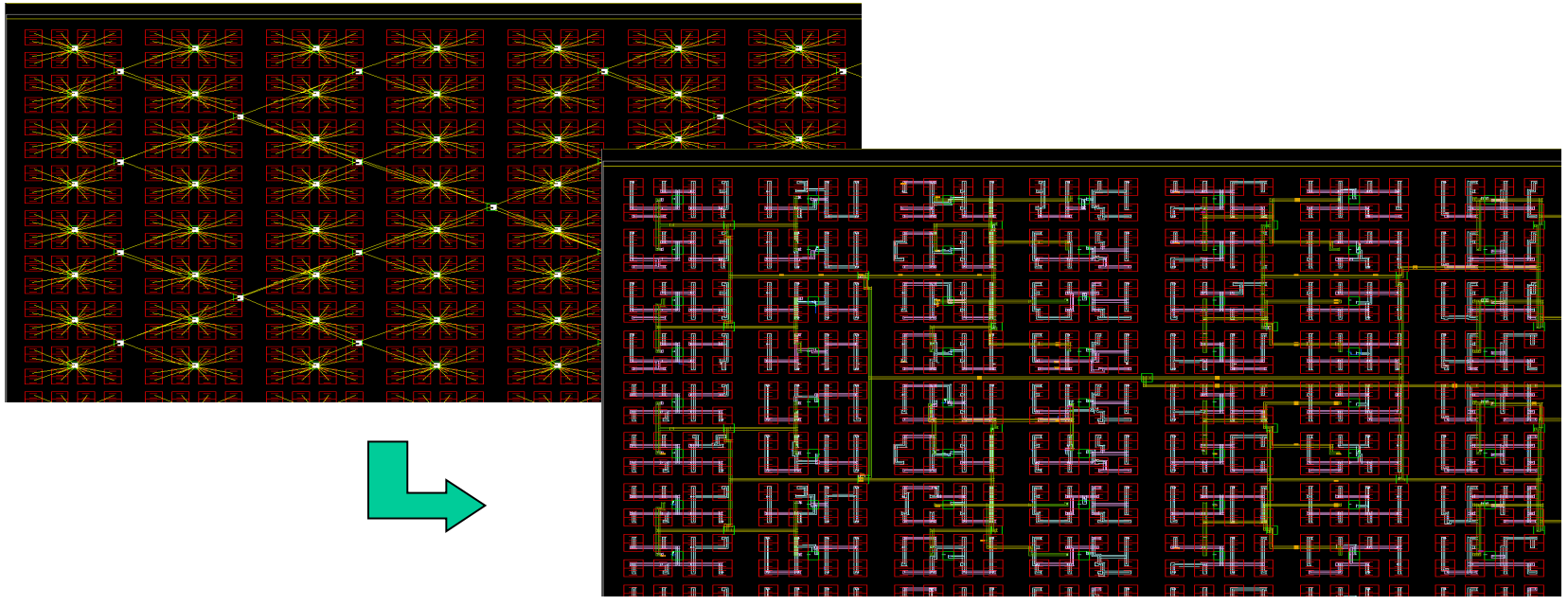
- The Full Hierarchical Design Methodology
- Manual placement & route for critical paths
- Manual placement of all FFs and clock buffers, manual clock gating
- Architecture optimization with physical feedback



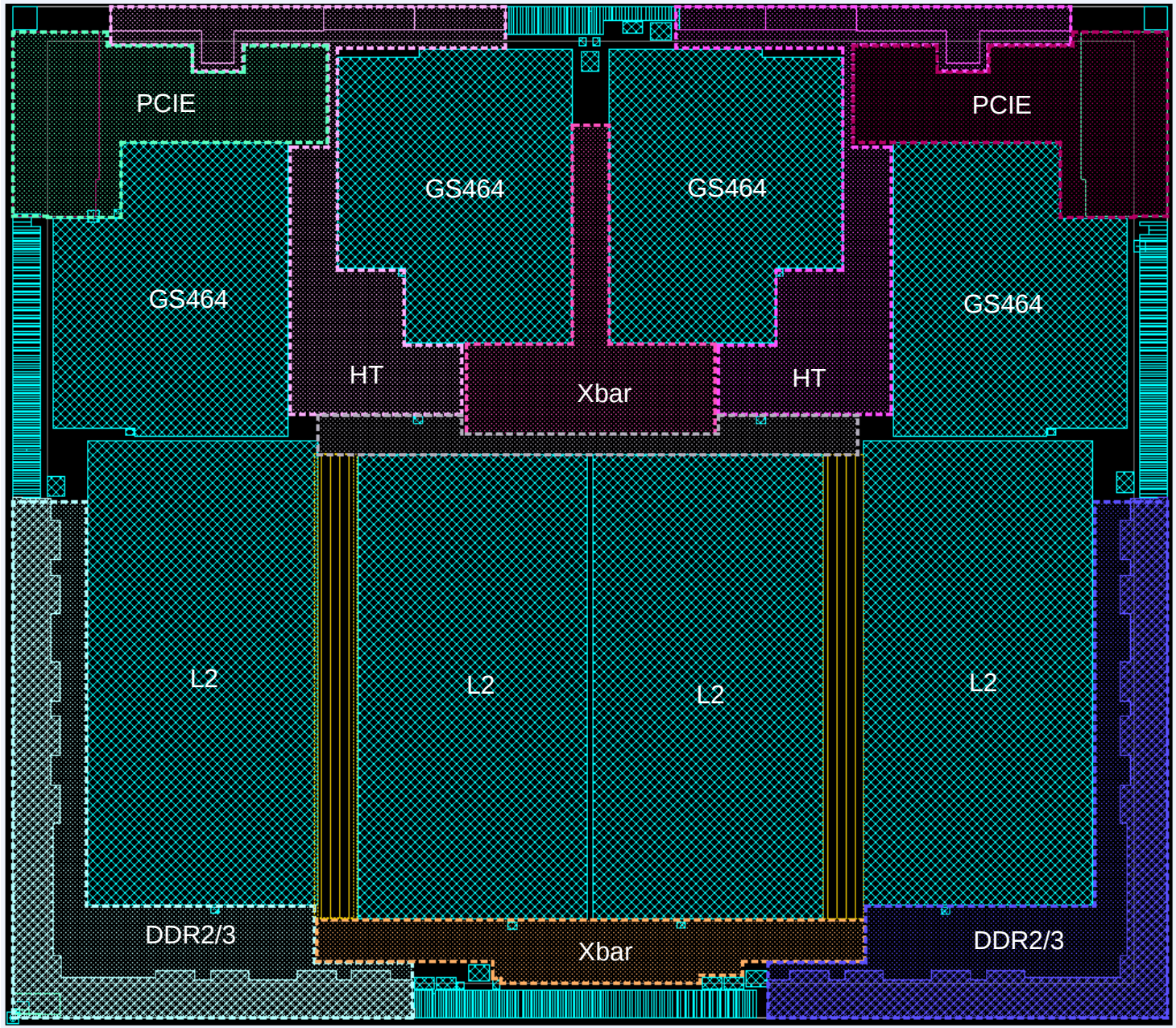
Clock Tree

- H-Tree + Mesh
- Manual placement of FFs
- Manual clock gate





Layout of 4-core Godson-3



Contents

- A brief introduction to ICT
- A brief introduction to Godson processors
- The Godson-3 multi-core processor
- PetaFLOPS and TeraFLOPS

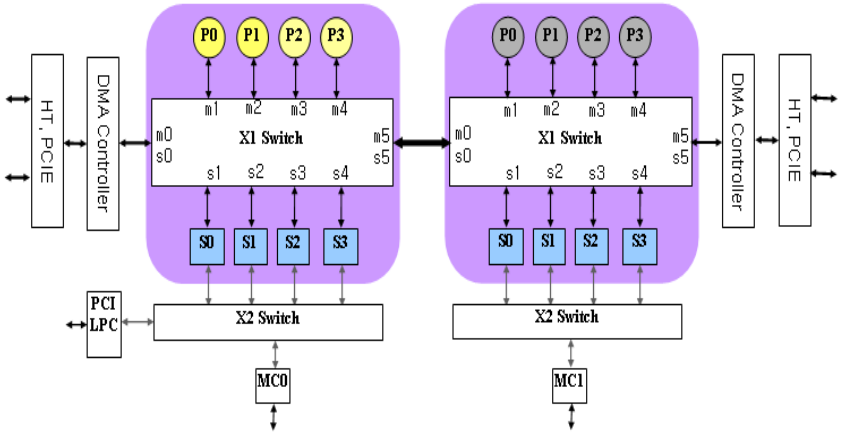
PetaFLOPS and TeraFLOPS

- **PetaFLOPS for National HPC**
  - **To build PetaFLOPS HPC with Godson-3 in 2010.**
- **TeraFLOPS for Personal HPC**
  - **Putting desktop to pockets**
  - **Putting TeraFLOPS to desktop: computing for the masses**

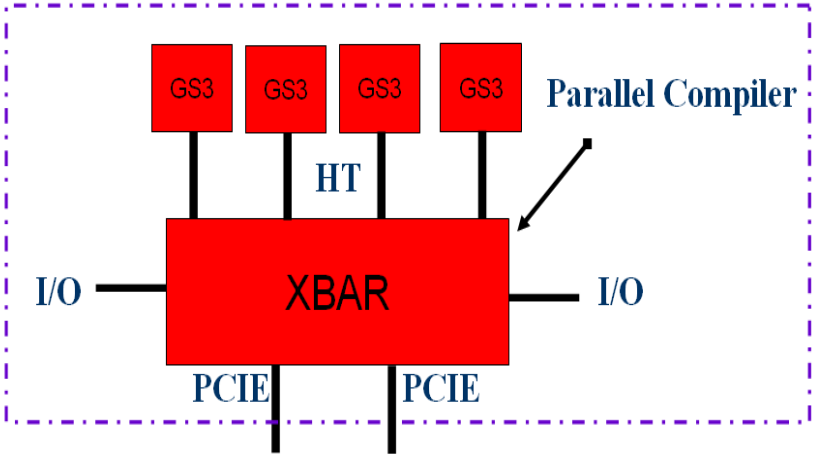
PetaFLOPS

- National 863 project
  - Based on Godson-3 CPU
- Hyper Parallel Processing
  - Good Scalability
  - Good Commodity
  - Low cost
  - Low power consumption < 1MW

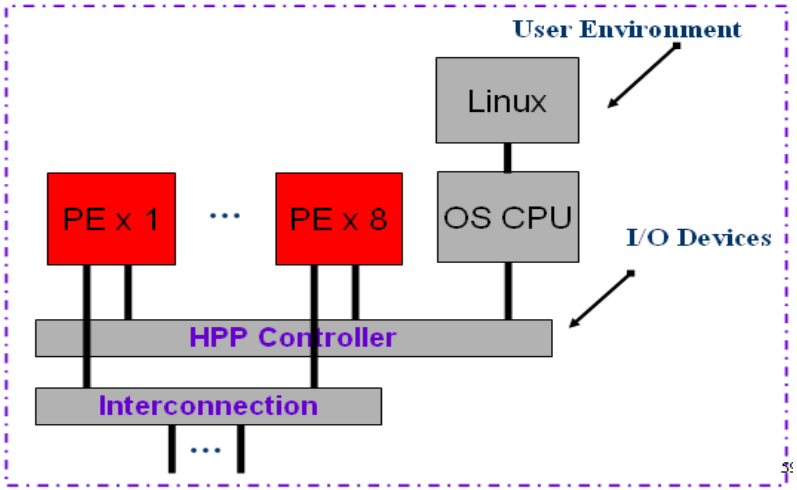
Dawning5000C Configuration



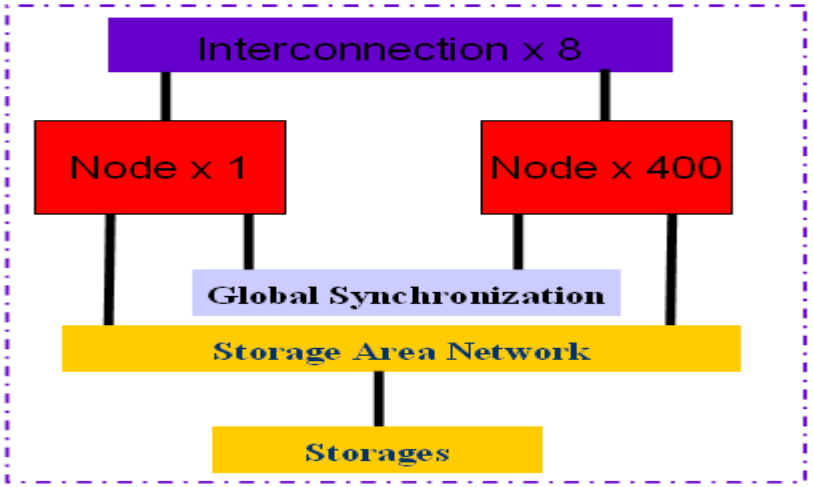
8 CORES (4GP+4MP) per CPU



4 chips per PE



8 PEs per node



400 nodes per system

Dawning5000C Configuration

- **CPU** ： 12800 80GF **8-core Loongson-3**
- **Blade** ： 3200 4-CPU PE
- **Node** ： 400 8-blade (+1-OS-module)
- **Cabinet** ： 60 7-Node
- **Interconnection** ： 16x18x36 InfiniBand, **Global Synchronization**
- **System** ： 1PF, 200TB Memory, 40Gbps, **1us Barrier**
- **Power** ： <**800KW**
- **Compatible** ： X86/Linux Application
- **Energy** ： **1000 MF/w** (Linpack/Power)

Personal High Performance Computers

- Computers are popular when they are personalized
  - IBM PC-XT
- HPC will be popular when they are personalized
  - Anti-Cloud?
- Personal HPC features:
  - High Performance: > 1TeraFLOPS
  - Low cost: < $10000/Teraflops
  - Low power: <1000W, connected to the wall of office
  - Ease to Use

Scaling down TeraFLOPS

TeraFLOPS in 1997

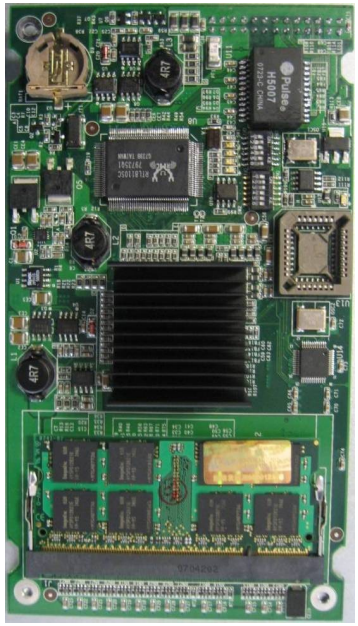
Scaling down

$100K/2007
"refrigerator"

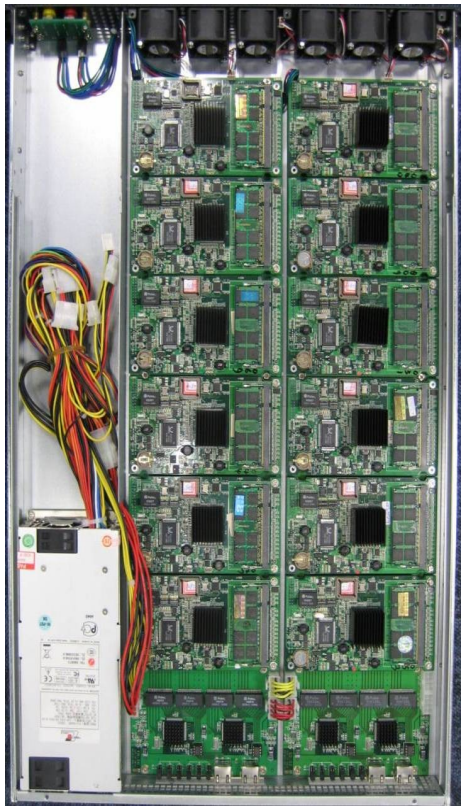
$50K/2008
"washing machine"

$10K/2009
"microwave oven"

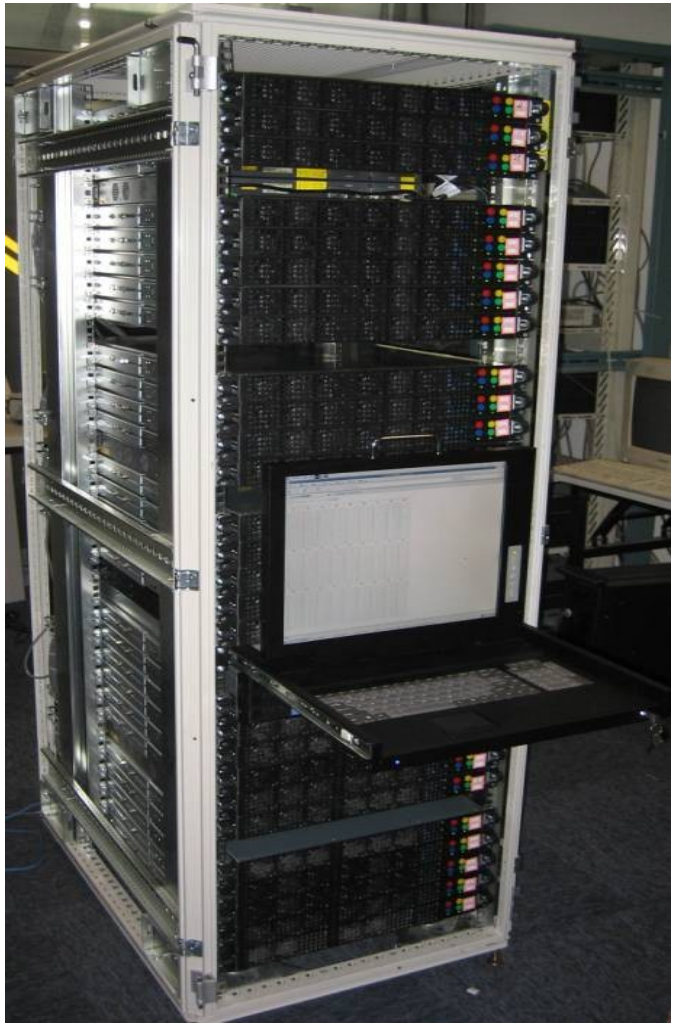
Refrigerator:
TeraFLOPS HPC based on Loongson-2F



2F node



1U12P



Proposals

- **Design the multiple purpose core according to the requirement oc CERN**
- **Integrated in the 8-core version of Godson-3**

Thanks