Timeline

2005

R & D of technology for future supercomputer

2010

K-computer project

2012

Hardware Completion

Feasibility Study of next generation supercomputer

2013

Next Generation Supercomputer project (now planning)

2015

Target Day ?
K-computer Results

- Planned Goal: TOP500 No.1, HPCC awards
- TOP500 No.1 2 times (2011 and 2012)
- HPCC awards all classes 2012
- Gordon Bell prize 2 times (2011 and 2012)
Feasibility Study of next supercomputer

- Design and performance estimation of a system
- Target 2018

- Architecture
  - Scalar MPP (Fujitsu)
  - MPP with accelerator (Tsukuba)

- Target performance
- Maximum power $20 \sim 30$ MW
Our goal

- Highest processor chip performance
  - Ex. SPEC CPU rate
  - NAS parallel benchmarks

- Highest single core performance
  - SPEC CPU int,
  - SPEC CPU fp
  - Dhrystone

Single core performance is the starting point
x1000 in 11 years
x1.87 in a year
Single Core Performance

• Base for all the performance improvement

• Various speed-up methods
  – Faster clock frequency
    • New device --- Si, GaAs, HEMT, JJ device, Optical devices
    • Smaller device --- Semiconductor device width
      – Intel 4004 10,000 nm
      – Intel Corei7 3xxxx 22 nm

Clock speed is now saturating
  – Power consumption
  – Device density

東京大学
Power wall --- limitation of clock speed

• 100x faster clock from 1993 to 2003
  – Progress of CMOS technology

• No improvement from 2003 to Now
  – 150W/chip power limit
  – More transistor / area size
  – Faster clock requires higher threshold voltage
    • High-speed 1.2V
    • Low power 0.8V 40nm CMOS
Clock speed limit

Clock Freq of Top 10 machines of Top500

東京大学
Retrospective Study on Performance and Power Consumption of Computer Systems
Introduction

- A balance of performance and power consumption is needed for future HPC
- Predict the expected perf/power using the trend of existing computing systems
- Will extending current designs work for the future HPC design?
1. Pipeline design

- **Sequential execution**  
  ~10 cycles/instruction  
  - Old computers, Intel 8080

- **Basic pipelined execution**  
  2~4 cycles/instructions  
  - CDC6600, Manchester MU5, MIPS and many more

- **Out of order execution**  
  ~1~ cycle/instruction  
  - IBM 360/91, Alpha 21264, most of today’s processors

- **SuperScalar execution**  
  ~0.5 cycle/instruction  
  - Intel i960CA, Alpha21064, most of today’s processors

- **VLIW**  
  ~0.3 cycle/instruction  
  - Multiflow, Intel Itanium

- **Out of order, SuperScalar should be used with branch prediction**
2. Branch Prediction

- **Branch Target Buffer**  
  Short and local history
  - Manchester MU5, processors before 1995

- **Two level branch prediction**  
  History and pattern table
  - Intel Pentium III, and many more processors

- **Gshare and Gselect**  
  Use of global history
  - AMD, Pentium M, Core2,

- **Perceptron predictor**  
  Machine learning
  - Oracle T4, AMD

- **ITTAGE**  
  Cascaded history table

**Practical use of speculative execution**
High-speed features of a processor 3

3. Prefetch (hardware prefetch)
   – Memory address prediction for future accesses
   – Access throttling for optimal memory utilization

   – Sequential Prefetcher
   – Stride Prefetcher
   – Global History Buffer
   – **Address Map Matching**

Effective speculative execution

Practical use of global history
Other High-speed features

4. Cache memory, hierarchical cache memory
5. Cache replacement algorithm
   - Non-LRU algorithms to eliminate dead blocks
5. DRAM Memory access scheduling
6. Network on Chip (NoC) scheduling

7. Accelerator (floating point accelerator)
Problems on current performance measurements

- No consistent performance measurements
  - Dhrystone, SPEC CPU92, 95, 2000, 2006
  - Relationship between these benchmarks?
  - Old benchmarks on new machines?
- We need consistent measurement method to evaluate long-time performance trend.
Power consumption/power efficiency becomes important issues in recent years.

- Almost no records of power consumption of old computers
- Some computers had been measured, but by ad-hoc methods

Consistent power measurements are essential to design next generation supercomputers

- Measurements on 2-3 generation are insufficient
Measuring Setup

- Single method across all machines
- Use Power Meter and Current Probe to measure the current that flows into the chassis
  - **Power Meter**: Sanwa Supply TAP-TST7, Metaprotocol UbiWattMeter
  - **Current Probe**: Fluke 336
Computers

- ~200 Systems sold in years 1985 to 2012
- From embedded to large SMP systems:
  - Embedded: 68VZ328, ARM925, Cortex-A8, SH-4A
  - PC: 286, 68030, PPC601, ..., Ivy Bridge
  - WS: VAX, SPARC, PA-RISC, MIPS, Alpha
  - HPC: SX-9, POWER5, POWER7 Itanium 2, BGQ
<table>
<thead>
<tr>
<th>Year</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1989</td>
<td>SHARP X68000 PRO HD</td>
</tr>
<tr>
<td></td>
<td>SONY NWS-1460</td>
</tr>
<tr>
<td></td>
<td>Apple Macintosh IIci</td>
</tr>
<tr>
<td>1991</td>
<td>Sun SparcStation IPX</td>
</tr>
<tr>
<td></td>
<td>NEC PC-9801DA</td>
</tr>
<tr>
<td>1992</td>
<td>NEC PC-9801RA</td>
</tr>
<tr>
<td></td>
<td>Fujitsu FM TOWNS II HR</td>
</tr>
<tr>
<td></td>
<td>SGI IRIS Indigo R4000</td>
</tr>
<tr>
<td>1993</td>
<td>EPSON PRO-486</td>
</tr>
<tr>
<td></td>
<td>NEC PC-9821As2</td>
</tr>
<tr>
<td></td>
<td>NEC PC-9801BS2</td>
</tr>
<tr>
<td>1994</td>
<td>HP 9000 712/80</td>
</tr>
<tr>
<td></td>
<td>Sun SPARCstation 5/85</td>
</tr>
<tr>
<td></td>
<td>Sun SPARCstation 5/110</td>
</tr>
<tr>
<td>1995</td>
<td>Apple PowerMac 7100/80</td>
</tr>
<tr>
<td>1996</td>
<td>Advantech PCA-6144V</td>
</tr>
<tr>
<td></td>
<td>NEC PC-9821V13</td>
</tr>
<tr>
<td></td>
<td>SGI O2</td>
</tr>
<tr>
<td></td>
<td>Sun Ultra2 2200</td>
</tr>
<tr>
<td></td>
<td>DEC AlphaStation 255/300</td>
</tr>
<tr>
<td></td>
<td>DEC AlphaStation 500/400</td>
</tr>
<tr>
<td>1997</td>
<td>PalmPilot Professional</td>
</tr>
<tr>
<td>1998</td>
<td>Sun Ultra5</td>
</tr>
<tr>
<td></td>
<td>Sun Ultra60 2360</td>
</tr>
<tr>
<td></td>
<td>Symbol SPT 1500</td>
</tr>
<tr>
<td>1999</td>
<td>SGI VWS 320</td>
</tr>
<tr>
<td></td>
<td>Intergraph TDZ 2000 GX1</td>
</tr>
<tr>
<td></td>
<td>Sun Ultra60 1450</td>
</tr>
<tr>
<td></td>
<td>Compaq XP1000</td>
</tr>
<tr>
<td></td>
<td>API UP2000</td>
</tr>
<tr>
<td>2000</td>
<td>Apple PowerBook G3(Pismo)</td>
</tr>
<tr>
<td></td>
<td>SGI Octane2</td>
</tr>
<tr>
<td>2001</td>
<td>Shuttle FV25</td>
</tr>
<tr>
<td></td>
<td>Apple PowerMac G4 (Digital Audio)</td>
</tr>
<tr>
<td></td>
<td>Sun Fire 3800</td>
</tr>
<tr>
<td></td>
<td>Sun Blade 2000</td>
</tr>
<tr>
<td></td>
<td>Tyan Tiger MPX</td>
</tr>
<tr>
<td></td>
<td>Palm m130</td>
</tr>
<tr>
<td>2002</td>
<td>Cobalt Qube 3 Plus</td>
</tr>
<tr>
<td></td>
<td>Apple PowerMac G4 (FW800)</td>
</tr>
<tr>
<td></td>
<td>Apple PowerMac G5 (7,2)</td>
</tr>
<tr>
<td></td>
<td>Palm Zire 71</td>
</tr>
<tr>
<td></td>
<td>VIA EPIA-ML</td>
</tr>
<tr>
<td></td>
<td>IBM p5 570</td>
</tr>
<tr>
<td></td>
<td>Apple PowerBook G4</td>
</tr>
<tr>
<td></td>
<td>Intel SR870BH2</td>
</tr>
<tr>
<td></td>
<td>HP Integrity rx5670</td>
</tr>
<tr>
<td></td>
<td>Sun Fire V40z</td>
</tr>
<tr>
<td>2003</td>
<td>HP ProLiant DL145 G2</td>
</tr>
<tr>
<td></td>
<td>Leadtek Winfast K8N</td>
</tr>
<tr>
<td>2004</td>
<td>Sony Playstation 3</td>
</tr>
<tr>
<td></td>
<td>ASUS P5LD2 SE</td>
</tr>
<tr>
<td></td>
<td>Toshiba Dynabook CX/47E</td>
</tr>
<tr>
<td></td>
<td>XFX nForce 780i</td>
</tr>
<tr>
<td></td>
<td>SH-2007</td>
</tr>
<tr>
<td>2005</td>
<td>QNAP TS-409</td>
</tr>
<tr>
<td></td>
<td>DELL Inspiron 910</td>
</tr>
<tr>
<td></td>
<td>NEC SX-9 4P @CfCA</td>
</tr>
<tr>
<td></td>
<td>J&amp;W MINIX-780G-SP128M</td>
</tr>
<tr>
<td></td>
<td>Convey HC-1</td>
</tr>
<tr>
<td>2006</td>
<td>Buffalo Kuro-box/T4</td>
</tr>
<tr>
<td></td>
<td>SHARP PC-Z1</td>
</tr>
<tr>
<td></td>
<td>DELL PowerEdge R410</td>
</tr>
<tr>
<td></td>
<td>ASUS P7P55D LE</td>
</tr>
<tr>
<td>2007</td>
<td>Intel S5520HCR</td>
</tr>
<tr>
<td></td>
<td>Fujitsu Lifebook MH380/1A</td>
</tr>
<tr>
<td></td>
<td>Toshiba Dynabook AZ</td>
</tr>
<tr>
<td></td>
<td>ThinkPad X201s</td>
</tr>
<tr>
<td>2009</td>
<td>ASRock P67 Extreme6</td>
</tr>
</tbody>
</table>
Software Configuration

- OS: Linux, Solaris, IRIX, MS-DOS, NetBSD, ...
- Compiler: GCC 4.5 where available
- Optimization: High
  - Equivalent to -O3 in GCC
Benchmarks

- SPEC CPU2006
  - Popular in architectural research papers
  - Requires 1GBytes of memory
  - Takes around 1 week on Pentium III
  - Some compilers have difficulty handling large and complicated code
Benchmarks (2)

- A single benchmark that run on all machines:
  - Dhrystone - Mainly function calls and string op.
- NAS Parallel Benchmarks
  - Floating-point oriented benchmarks
  - Much cleaner and simpler compared to SPEC CPU

- We'll compare these benchmarks to SPEC CPU later.
Performance remarks

- Trend changes after 1995
  - UltraSPARC, Pentium, EV56, ...
- Improvement of embedded processor faster than other segments
  - ARM
- Higher-performance processors are still increasing their performance nevertheless
Performance/power remarks

- Performance/power improvement: driven mainly by performance
- Recent ARM processors (Cortex-A8/9) are on the same range as Intel counterparts
SPEC CPU2006 vs Dhry or NPB

- SPEC CPU2006 is **HUGE**
  - Simply does not work on older hardware
  - Unable to compare old and new hardware using the same benchmark
- Dhrystone and SPEC CINT2006
  - Dhrystone sufficient for measuring perf/power?
- NPB and SPEC CFP2006
  - NPB mainly written in FORTRAN and is much smaller
CINT2006/Dhrystone

- Very good correlation observed
- Dhrystone runs completely in-cache
  - WSS in CINT2006 is around 1GB in 32b systems
- Too much string operations
  - Level of optimization in string functions in standard C library
CFP2006/NPB

Cor. Coe. 0.979
(w/o SX-9)
SPEC CPU2006 vs NPB and Dhry

- Dhrystone reflects overall system performance as well as CINT2006
  - CINT2006 still useful for analyzing it
- NPB reflects overall floating-point application performance as well as CFP2006
  - Yet much simpler to compile for specific system configuration
Extrapolate performance and power consumption using the data we obtained.

In year 2018 (we want an exaflops machine then):

- 2,963 VAX MIPS/W
  - translates to 3.3e-10 J/Instruction
- For 20MW operation at 1 Exaflops
  - 2E-13 J/Floating operation
- Integer instruction 1000x heavier than floating point ones! (in terms of power consumption)
Actions needed

- Int-to-Float instruction mix in Strong-scaling application will stay the same
- Improve the integer instruction perf/power by increasing the performance
- Otherwise applications that we can run on exaflops machine will be limited
Related work

- SPECpower [Lange 09]
  - Runs on Java VM, performance impact from this is unknown

- Power consumption of Google servers [Barroso 05]
  - Only three generations of their specific servers
Supercomputers in 2018

- Conventional multi-core processor
  - Up to 300 Pflops with 30MW power
  - Improvement of microarchitecture will realize E flops
  - Accelerator is also effective for some applications

- GPGPU will be dead around 2018
- Naïve many-core will have single-core performance difficulties (Xeon Phi: Currently 926 VAX MIPS/ 1.2GHz clock)
Conclusion

- Performance have driven perf/power improvement
- We need to act now for the exaflops machine design
  - Integer performance is the key
- Dhrystone and NPB can be used in place of SPEC CPU benchmarks in many cases
Donation please!

- We are collecting old computers
- It’s computer zoo. Not computer museum.

All the computer in the Zoo is running

What we want to have

- Any Robotron computers
- ISA Graphics cards
- IBM 5110 APL computer parts
- Transputers