Machines: Where Next?

Murray Cole

Technological Progress

Moore's "Law" suggests that the number of transistors embeddable per unit area doubles every 18 months (or so). Note that this is an exponential growth rate.

As transistors get smaller, they also get faster.

What should we do in our machine architectures to exploit this raw potential?

Spending the Transistor Budget

Three possibilities which have been, are being and/or will be investigated are

- overcoming the memory performance gap
- introducing (more) parallelism
- integrating more varied system components on-chip

The Memory Gap

The Moore speed up effect in processor technology is not reflected in the speed of bulk main memory and disc.

We can however trade-off speed for capacity: smaller and faster or larger and slower.

Locality of Reference

Extensive analysis and monitoring of real programs reveals that most memory accesses are far from "random". In fact they exhibit

1. temporal locality: an address which has just been accessed is quite likely to be accessed again soon

2. spatial locality: the next address to be accessed will quite often be close to one which has recently been accessed

We can exploit this by implementing a memory hierarchy in which recently (or hopefully soon to be) used locations are copied into faster temporary memory, for quick access. Such a copying memory is called a cache. Systems today have two, three or even four levels of cache, each increasingly larger and slower than the last.

The von Neumann Bottleneck

High-end modern processors can have several megabytes of cache memory on the processor chip, taking up 80-90% of the transistors.

This is "easy" technology (for designer and user), but is it sensible?

The resulting under-utilisation of transistors is a symptom of the von Neumann bottleneck: we have massive raw processing power, and massive storage capabilities, but we force our executions to proceed through a narrow conceptual bottleneck involving access to one word/one instruction at a time

To remove the bottleneck, we have to introduce and exploit the possibility of executing many instructions on many pieces of data concurrently.

In computer architecture, this is known as parallelism.

Parallelism in Computer Systems

Parallelism is already exploited on various scales within real computer systems, and will become increasingly attractive.

It is (and will be) visible to computer architects, and systems and applications programmers, rather than end users.

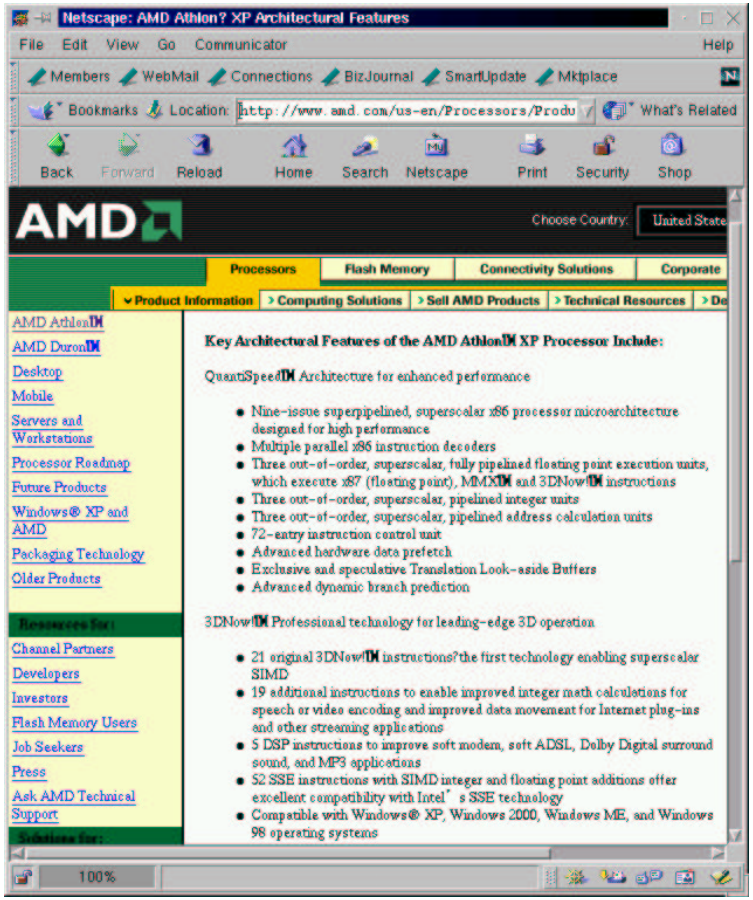
1. within the microarchitecture or instruction set architecture of a conventional von-Neumann processor (looking for parallelism between logically sequential instructions, or through single instructions working on multiple data items

2. between several processor cores on a single chip, executing concurrent threads within a single application process

3. between multiple processors (possible internally multithreaded) of "high performance workstation" or "supercomputer" architectures, located physically in a single box

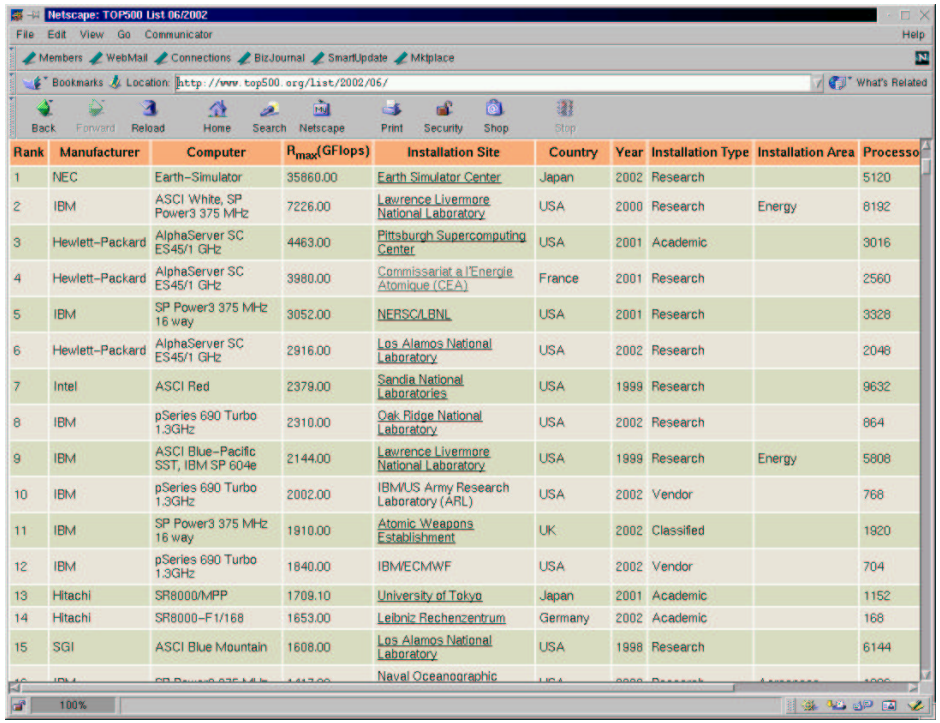
4. between physically distributed computers (of the various types above) collaborating across buildings or continents on the solution of very computationally and/or data intensive problems

As well as the immediate problems of how to physically and architecturally organise such systems, there are many open questions concerning how best to design efficient algorithms which can exploit them, and usable programming constructs, languages and libraries with which to express these algorithms.



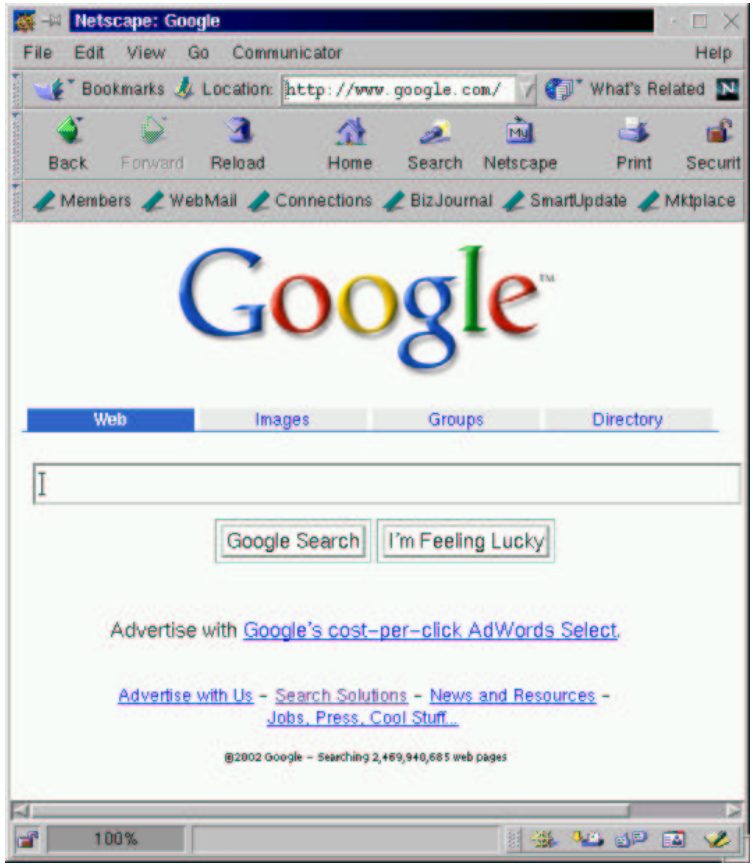
- a modern processor with extensive internal parallelism

- how many processors in five years?



Rank	Manufacturer	Computer	R_{max}(GFlops)	Installation Site	Country	Year	Installation Type	Installation Area	Processo
1	NEC	Earth-Simulator	35860.00	Earth Simulator Center	Japan	2002	Research		5120
2	IBM	ASCI White, SP Power3 375 MHz	7226.00	Lawrence Livermore National Laboratory	USA	2000	Research	Energy	8192
3	Hewlett-Packard	AlphaServer SC ES45/1 GHz	4463.00	Pittsburgh Supercomputing Center	USA	2001	Academic		3016
4	Hewlett-Packard	AlphaServer SC ES45/1 GHz	3980.00	Commissariat a l'Energie Atomique (CEA)	France	2001	Research		2560
5	IBM	SP Power3 375 MHz 16 way	3052.00	NERSC/LBNL	USA	2001	Research		3328
6	Hewlett-Packard	AlphaServer SC ES45/1 GHz	2916.00	Los Alamos National Laboratory	USA	2002	Research		2048
7	Intel	ASCI Red	2379.00	Sandia National Laboratories	USA	1999	Research		9632
8	IBM	pSeries 690 Turbo 1.3GHz	2310.00	Oak Ridge National Laboratory	USA	2002	Research		864
9	IBM	ASCI Blue-Pacific SST, IBM SP 604e	2144.00	Lawrence Livermore National Laboratory	USA	1999	Research	Energy	5808
10	IBM	pSeries 690 Turbo 1.3GHz	2002.00	IBM/US Army Research Laboratory (ARL)	USA	2002	Vendor		768
11	IBM	SP Power3 375 MHz 16 way	1910.00	Atomic Weapons Establishment	UK	2002	Classified		1920
12	IBM	pSeries 690 Turbo 1.3GHz	1840.00	IBM/ECMWF	USA	2002	Vendor		704
13	Hitachi	SR8000/MPP	1709.10	University of Tokyo	Japan	2001	Academic		1152
14	Hitachi	SR8000-F1/168	1653.00	Leibniz Rechenzentrum	Germany	2002	Academic		168
15	SGI	ASCI Blue Mountain	1608.00	Los Alamos National Laboratory	USA	1998	Research		6144
				Naval Oceanographic					

- rated on LINPACK (number-crunching) benchmark



- >10,000 processors, petabytes of data (1,000,000 Gigabytes)

- >3,900,000 processors (not all at the same time)
- >1,000,000 years of CPU time

Systems on Chip

Even more adventurously, we can now begin to integrate processors, memory solar/vibration charged batteries, sensors and transmitters all one the same unit.

This is known as SLI (Systems Level Integration) or SOC (System on Chip) technology.

One ambitious project at UC Berkeley proposes Smart Dust: hundreds or thousands of such devices are dispersed over an area of interest, find each other, and build up information on what's happening which can be used for various purposes.

Project target is to do all this with a one cubic millimetre device.

Many may break, or fall upside down, or whatever... It doesn't matter, the rest will collaborate to get the information!

The Future with Smart Dust?

(From **http://robotics.eecs.berkeley.edu/~pister/SmartDust/**)

In 2010 everything you own that is worth more than a few dollars will know that it's yours, and you'll be able to find it whenever you want it. Stealing cars, furniture, stereos, or other valuables will be unusual, because any of your valuables that leave your house will check in on their way out the door, and scream like a troll's magic purse if removed without permission (they may scream at 2.4 GHz rather than in audio).

In 2010 your house and office will be aware of your presence, and even orientation, in a given room. Lighting, heating, and other comforts will be adjusted accordingly.

In 2010 a speck of dust on each of your fingernails will continuously transmit fingertip motion to your computer. Your computer will understand when you type, point, click, gesture, sculpt, or play air guitar.

In 2010 your car will know the freeway conditions on your favorite route home, not at the level of some pathetic traffic announcer telling you that it's slow on I5, but with detail of the instantaneous speed and history of every vehicle between you and your destination, as well as he ones that are likely to get on the freeway, should you choose to look at that detail. Most likely your software will just tell you which route to take, and how many minutes it will take. Your spouse will know too, if you so choose.

In 2010 you won't have to hunt for a parking space. You'll call ahead and find (and maybe reserve) the most convenient open space in the lots that you use.

In 2010 everything of any value that you own will have it's own set of sensors, letting you know when your tyre pressure is low, the bridge ahead is out (or unsafe), your milk is going bad, or your water heater is about to die.

In 2010 MEMS sensors will be everywhere, and sensing virtually everything. Scavenging power from sunlight, vibration, thermal gradients, and background RF, sensors motes will be immortal, completely self contained, single chip computers with sensing, communication, and power supply built in. Entirely solid state, and with no natural decay processes, they may well survive the human race. Descendants of dolphins may mine them from arctic ice and marvel at the extinct technology.