What goes to make a supercomputer?

James Davenport
Hebron & Medlock Professor of Information Technology

University of Bath

10 September 2013

Many thanks to Prof. Guest (Cardiff) and Jessica Jones

Computer Speed

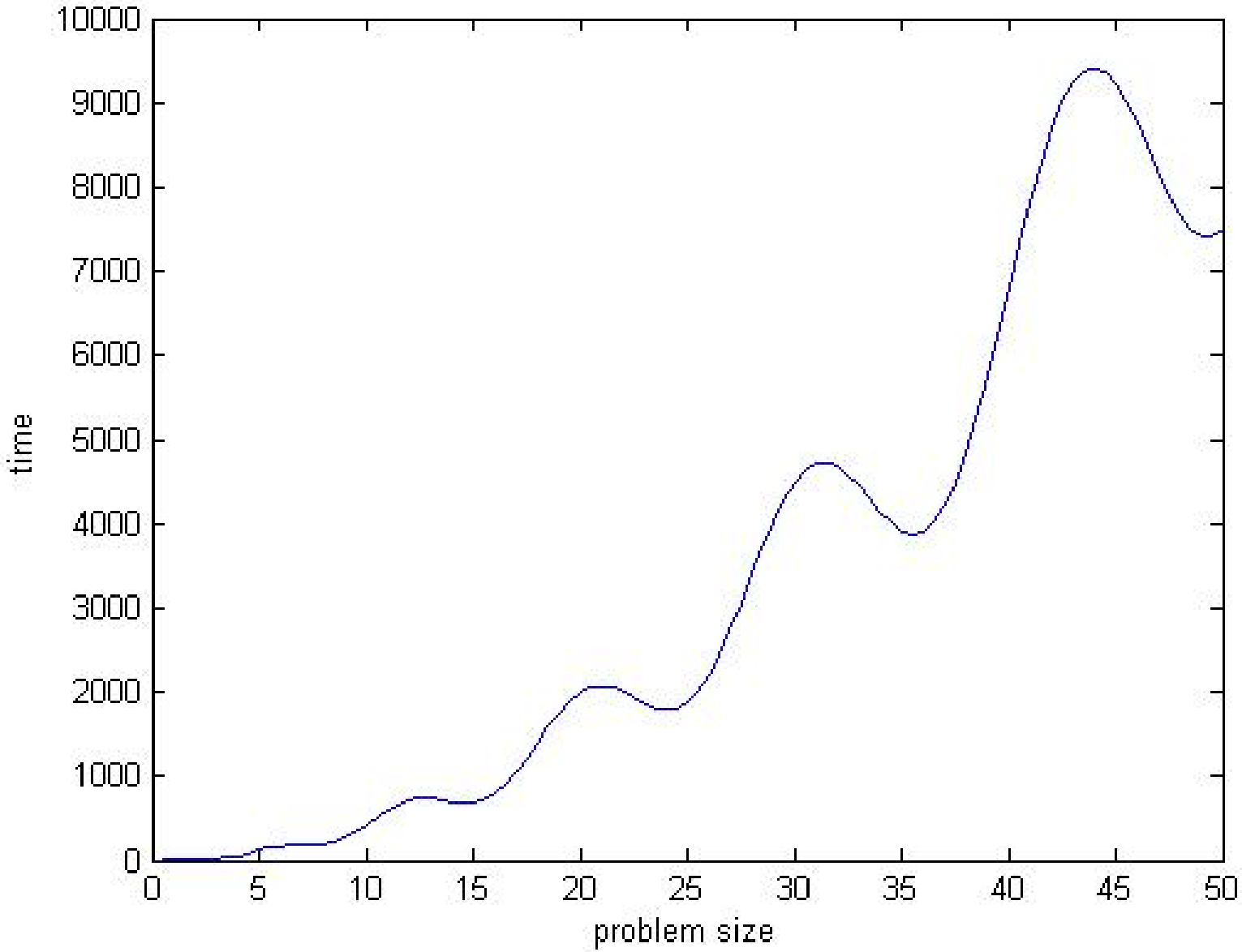
- More or less peaked at 3GHz in 2006

Why For a fixed feature size (current technology is 32nm or 22mm),

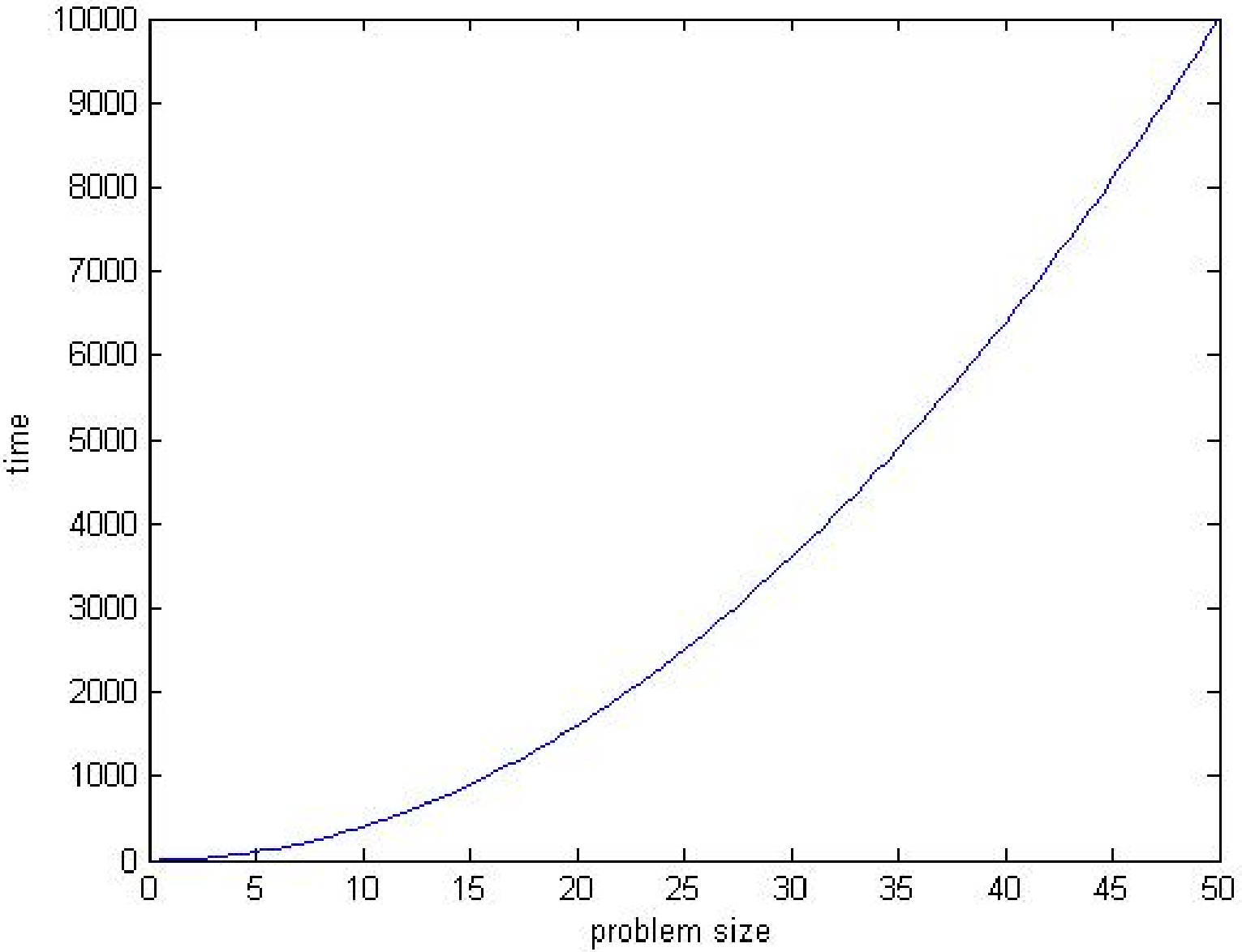
$$\text{Power} \propto \text{GHz}^3$$

- This explains why mobile 'phones run down unexpectedly!
- More seriously, cooling/power interaction is a major issue
- Especially for laptops

Measurements (details omitted to spare the author)



Rerun on a desktop



This is not a rare phenomenon

Statistics from Jessica's Laptop

Frequency	CPU0	CPU1
1.60 GHz	28.87%	26.27%
1.33 GHz	2.20%	0.45%
1.07 GHz	2.37%	0.48%
800MHz	66.55%	72.79%

Hence "half speed" is the most common.
"`current policy: frequency should be within 800 MHz and 1.60 GHz. The governor "ondemand" may decide which speed to use within this range.`"
In fact the "CPU"s are really threads within one CPU.

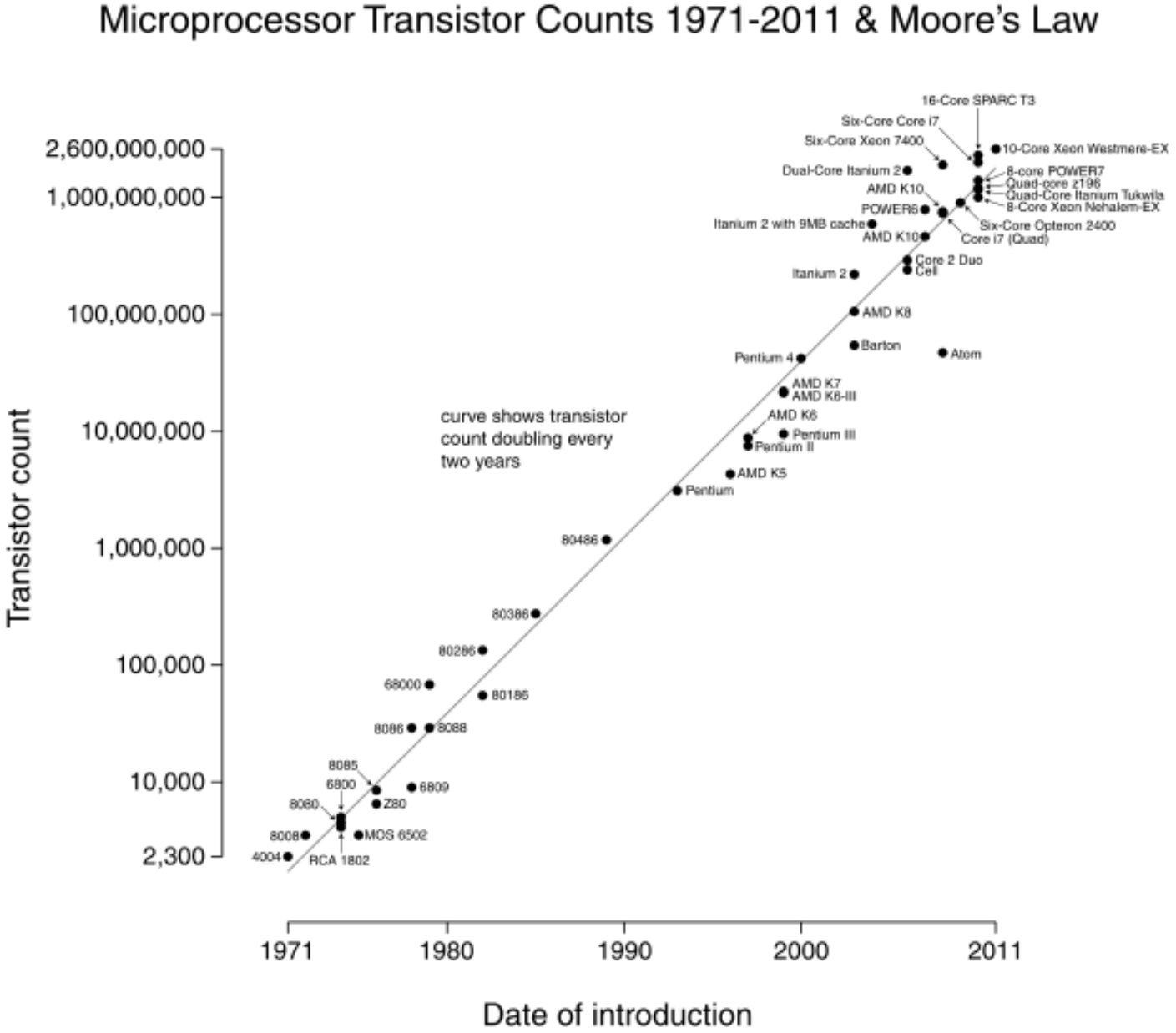
Moore's law

Moore was one of the founders of Intel:
INtegrated (chip) TEchnoLogy

He said	The number of transistors on a chip doubles every 2 years
Salesman	Computer speed doubles every 18 months

The original statement still seems to be true, even though, for the last fifteen years, "experts" have said that Moore's Law stops in the next few years

Moore's Law


Microprocessor Transistor Counts 1971-2011 & Moore's Law
Transistor count
2,600,000,000
1,000,000,000
100,000,000
10,000,000
1,000,000
100,000
10,000
2,300
curve shows transistor count doubling every two years
16-Core SPARC T3
Six-Core Core i7
Six-Core Xeon 7400
10-Core Xeon Westmere-EX
Dual-Core Itanium 2
8-core POWER7
Quad-core z196
AMD K10
Quad-Core Itanium Tukwila
POWER6
8-Core Xeon Nehalem-EX
Itanium 2 with 9MB cache
Six-Core Opteron 2400
AMD K10
Core i7 (Quad)
Core 2 Duo
Cell
Itanium 2
AMD K8
Barton
Atom
Pentium 4
AMD K7
AMD K6-III
AMD K6
Pentium III
Pentium II
AMD K5
Pentium
80486
80386
80286
68000
80186
8086
8088
8085
6800
6809
Z80
8080
8008
MOS 6502
4004
RCA 1802
1971
1980
1990
2000
2011
Date of introduction

More and more transistors

What can I do with them?

- Add more and more instructions — the CISC trend that Mark Hayes mentioned.

So A modern Intel will do, in one instruction

```
c[1]=c[1]+a[1]*b[1]
c[2]=c[2]+a[2]*b[2]    /* useful for matrix multiply */
```

Or compute the reciprocals of the square roots of four single-precision numbers (someone must want this!)

- Put more and more memory on chip

* This is done (cache memory: see Hardware)

- Put more computers on a chip (multi-core)

CPUs, Cores, Processors etc.

Terminology is pretty confused.

Core	A part of a chip capable of processing its own instructions
Processor	The term schedulers use, typically $\approx$ "core", but:
Thread	A core processing a stream of instructions (some cores interleave multiple threads)
CPU	almost meaningless today
Chip	The smallest physical thing you can buy!
Die	Some chips actually contain two dies
Node	One or more chips with memory etc. ("a computer")
System	All the nodes and their interconnections, also "head nodes" (logging on,compiling etc.), disc etc

Typically (Intel machines)

ARM and Power PC machines have more, simpler cores.
Machines with GPU addons are more complicated still.

Cores/chip	2 on a laptop, 2–4 on a desktop, 4–16 on an HPC facility
Chips/node	1 on a laptop, 1–2 on a desktop, 2–4 on an HPC facility
Nodes/sys	10 for a Department, 100-1000 for a University, 10,000+ for a national resource (Tier 1), 100,000+ for Tier 0.
Record	Tianhe-2 (China) 3,120,000 cores, 1000GB main memory,
But	17.8MW power (plus cooling!)

Interconnect — types

How the various parts (core etc.) of a system talk to each other and memory

intrachip Built in (but non-trivial)

intranode Again built-in, and non-trivial. That's why standard hardware stops at 4 chips/node.

Dept. system typically 1Gb Ethernet — cost minimal

thirty men's heft of grasp in the gripe of his hand — Beowulf (8–11c), hence "Beowulf cluster"

Univ. system Possibly 10Gb Ethernet, more likely Infiniband — cost ≈25% of total

National Specialist technology, often the majority of the cost

Interconnect — features

Typically interconnect is node–node (as nodes have more cores, need better interconnect)

Bandwidth Speed of data transfer Gigabits/second: 1 or 10 for Ethernet,

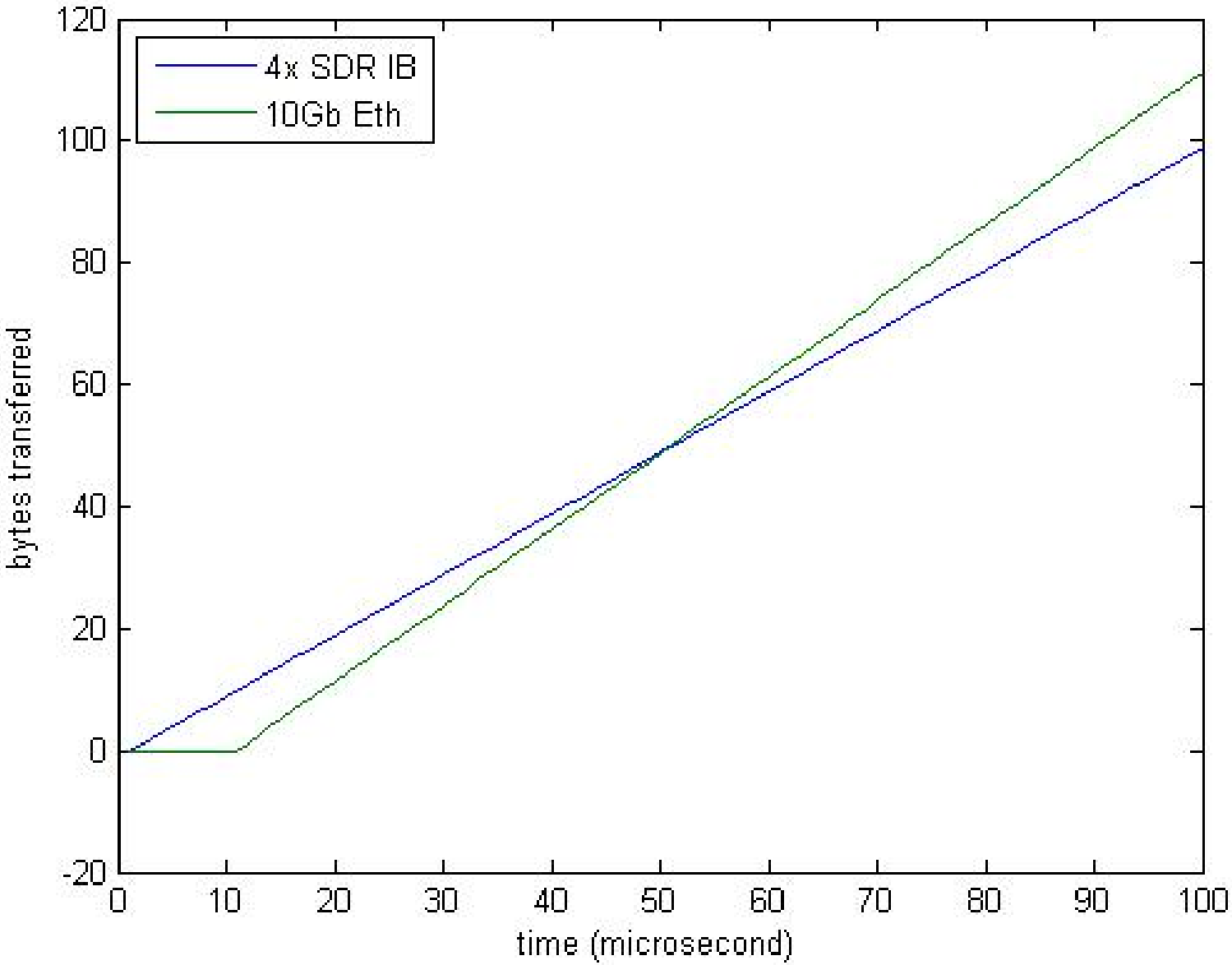
Inifinband 4×SDR (basic) is 8Gb/s, 4×QDR is 32Gb/sec

Custom Tianhe-2 has 51.2Gb/sec

Latency Time before data starts arriving at far end. custom sub-μsec, Infiniband is 1–2μsec; Ethernet is complicated — 11μsec minimum, also depends on message length (store and forward)

The best bandwidth is a jumbo jet full of CDROM, but the latency is terrible!

Bandwidth vs Latency (ideal)



Bandwidth vs Latency (closer to actual)