The Physical Limits of Computing Systems

Jim Garside

Department of Computer Science, The University of Manchester,

Oxford Road, Manchester, M13 9PL, U.K.

jgarside@cs.man.ac.uk

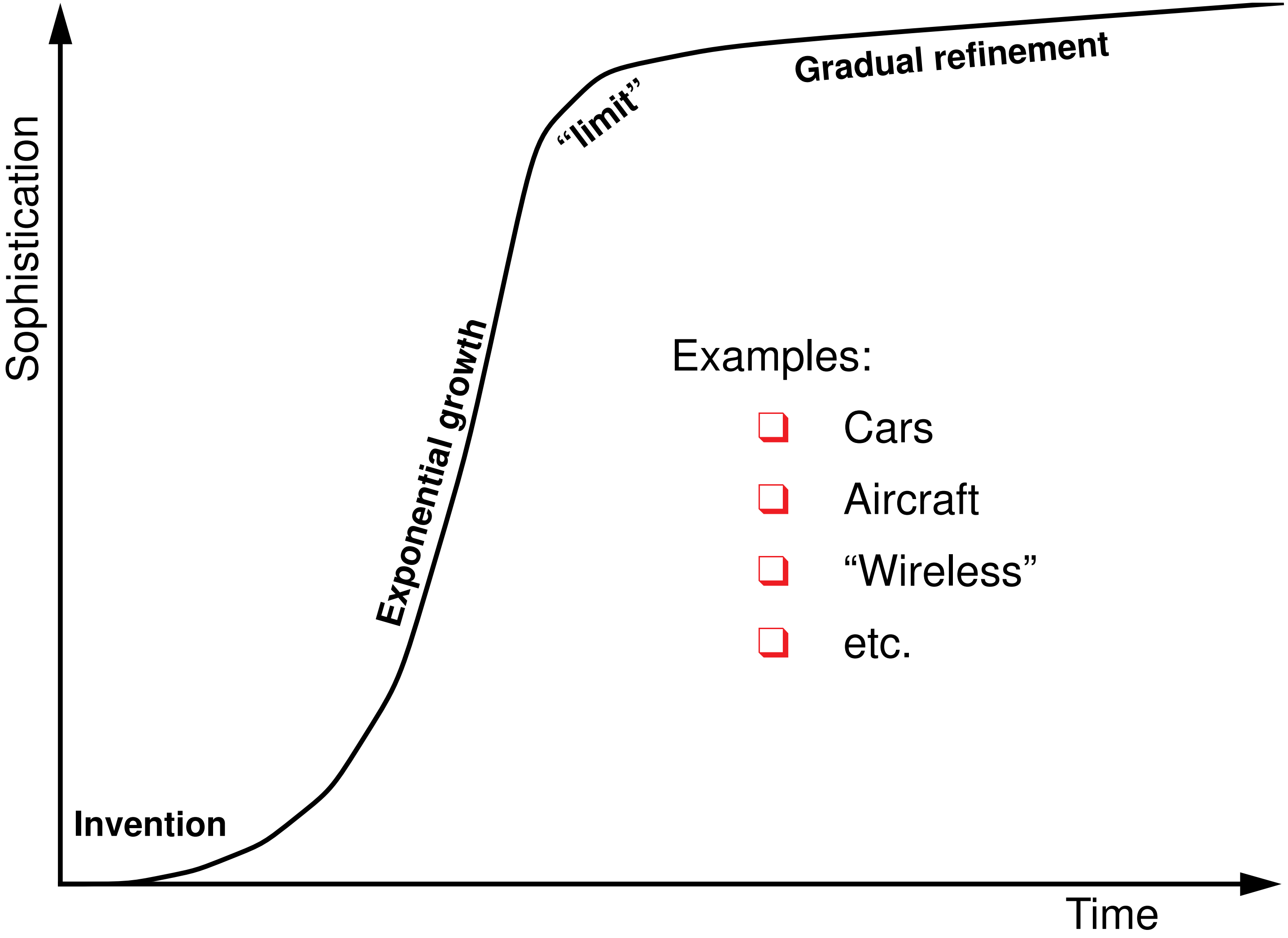
Contents

- Product development rate
- Extrapolations in microelectronics
- What may stop development
- Computers in 2010

THE UNIVERSITY of MANCHESTER

AMULET group

Typical Product Development Profile



Examples:

- Cars
- Aircraft
- "Wireless"
- etc.

Computing Systems Development

The "modern" computer has been around for about 50 years.

It is still in the period of exponential growth.

What is the "limit"?

Some fundamental limits are:

- Speed: the speed of light in free space
  186 000 miles/s or 3×10^8 ms^{-1} or 1 ft/ns
- Size: a single atom is about 1Å in radius
  (1Å $\equiv 10^{-10}$m $\equiv 10^{-4}\mu$m)

A current "state-of-the art" silicon chip:

- has a minimum feature size of about 0.25µm (i.e. ~1200 atoms)
- is about 2cm across, so electrical signals can cross it in about 0.3ns (somewhat slower than in free space)
- is clocked at speeds between 500MHz and 1GHz
  (i.e. a clock period of 1ns - 2ns).

THE UNIVERSITY of MANCHESTER

AMULET group

Going faster

To achieve "more" computing power a computer must improve by:

- doing more computations simultaneously ...
  parallelism has increased by 2-3 times over 50 years

- ... reducing processing inefficiency...
  a maximum of maybe 2x total

- ... and increasing the system frequency
  so far worth maybe 100 000 times improvement

The process technology has a dominant role in performance

Note:

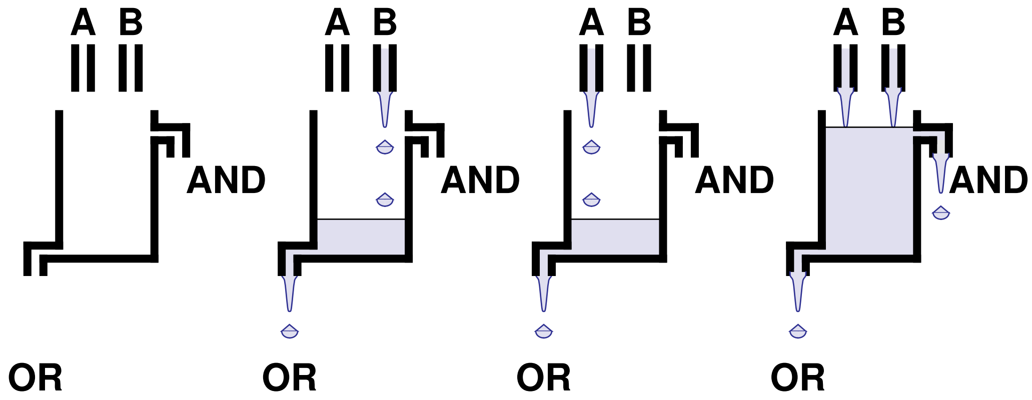
Computers are <u>not</u> electronic devices. A computer is a physical implementation of an algorithm. A computer could be made of anything!

THE UNIVERSITY of MANCHESTER

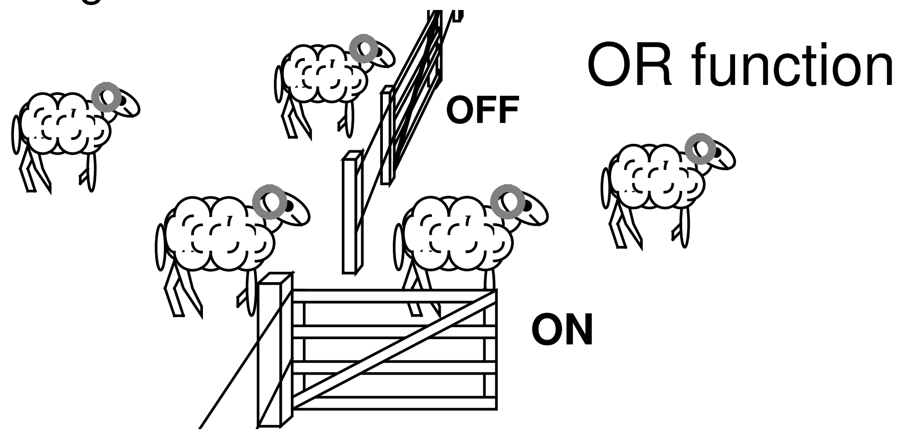
AMULET group

Logic gates are an abstract concept.

For example, a water powered gate:



or a sheep powered gate



A logic gate can be implemented in any technology.

AMULET group

THE UNIVERSITY of MANCHESTER

Practical systems

Past systems:

- Mechanics (e.g. Babbage's "difference engine")
- Electro-mechanical (relays)
- Vacuum tubes ("valves")
- Transistors
- Integrated circuits

- Fluidic/pneumatic logic (used in specialised applications)

Electronics has so far provided the only practical method of building large, complex computing systems.

Future *possible* technologies

Demonstrated

- Optical computers
  Interconnection – excellent, Switches – possible,
  Integration – difficult
- Superconducting technology (e.g. Josephson Junctions)
  Switching possible; low temperatures required
- Single electron switches & “quantum dots”
  Possibility demonstrated

Speculative

- Nanomechanical switches – molecular scale mechanics
- Biochemical switches
- Biological systems
- ...

The ~~Physical~~ Practical Limits of Computing Systems

At present **integrated circuits** using transistor switches still appear to be the only practical way forwards.

- High gain
  transistors not only switch signals, they amplify them in the process
- Solid state
  the absence of moving parts means that they are fast, reliable and do not wear out easily
- Easily integrated
  many millions of components can be manufactured on a single device, keeping both size and cost down
- Process tolerant
  although irregularities in manufacturing cause differences in “identical” devices, parts will tolerate significant differences whilst still functioning

Moore’s Law

The number of transistors on a chip doubles every year. (1964)

In the late ’70s this ceased to be true; since then it has required 18 months for each doubling.

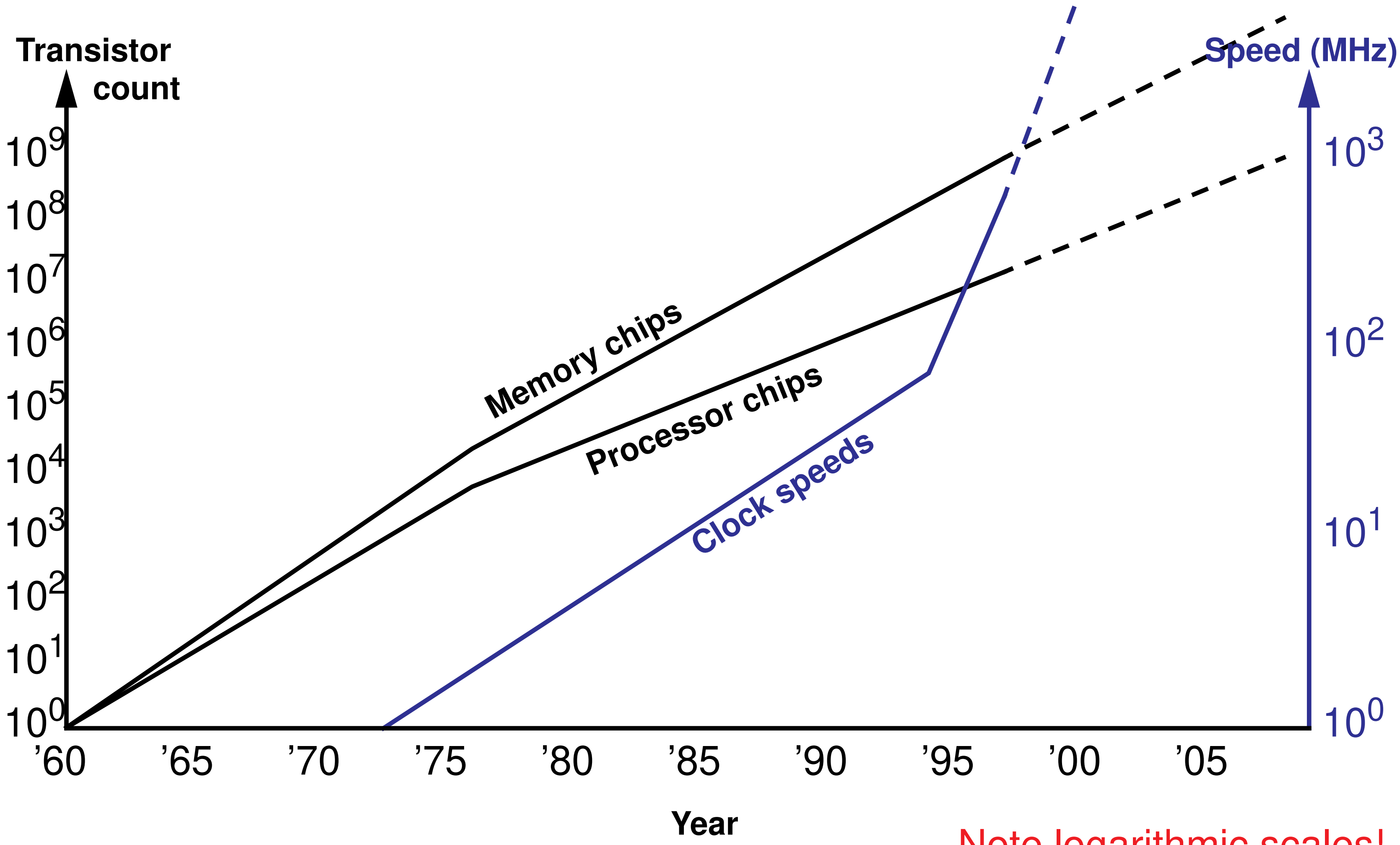
This reduced rate of progress means that:

- Transistor sizes shrink at about 25%/year
- Speeds increase by about 60%/year

Meanwhile

- Price/chip is roughly constant
  (e.g. price/byte of memory has dropped by over 10 000x in 25 years)

Rate of Process change



Note logarithmic scales!

AMULET group

THE UNIVERSITY of MANCHESTER

Physical Limitations in Microelectronics

Thermal energy (kT)

- many orders of magnitude available before cooling essential

Electric field intensity

- already at practical limits (~500kV/cm)

Quantum tunnelling

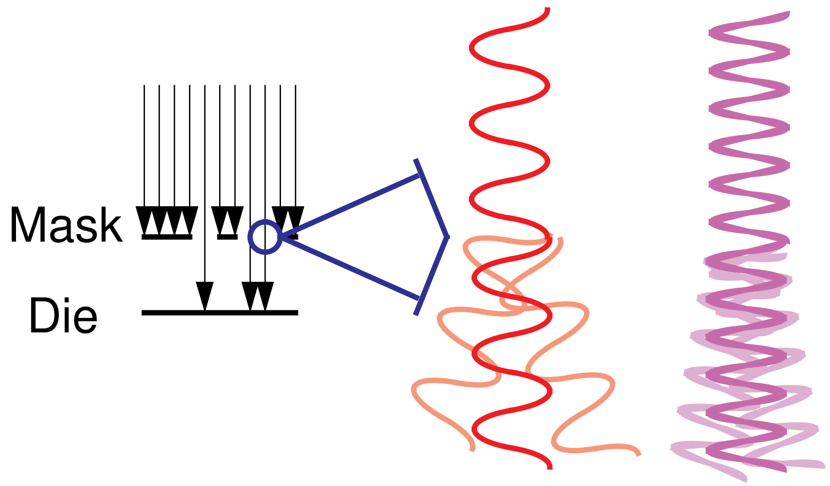
- significant for features with <100Å spacing; approached by prototypes

Lightspeed

- increasingly difficult to ensure “simultaneous” switching across a chip; even harder $\Rightarrow$ impossible off chip

Practical Limitations (1)

Lithography



- Diffraction causes “fuzziness” of image
- Use short wavelength light to reduce diffraction
- Already in ultra-violet – next step is to X-rays
- X-rays are hard to focus and mask

Alternative is to use a beam of electrons to write onto the die; this is comparatively easy.

However this is analogous to handwriting a book, rather than printing it; i.e. rather slow!

Practical Limitations (2)

Interconnection density

- transistors are single switches which need to be connected together.
- As geometries shrink & transistor count increases the interconnection problem gets worse
- Chip area is beginning to be dominated by interconnections rather than processing components.

- Optical interconnection (if feasible) could circumvent this limit

Note: interconnection is one thing that biological systems (i.e. brains) are good at.

Practical Limitations (3)

Power Dissipation

All the energy used to power a chip ends up as heat.

This was a relatively insignificant consideration until recently.

Power dissipation per transistor decreases as the devices shrink, but this is offset by the increase in the number of transistors on the chip. Furthermore the power dissipated increases with the clock speed of the device.

Modern microprocessors burn ~50W; this is like compressing a domestic light bulb to the size of a fingernail, or about 10x the heat flux of a frying pan. It will continue to increase.

However

- Cooling technology can keep pace for some time yet (at least 10x or 5 years worth)
- New design processes are capable of making chips much more power efficient

Practical Limitations (4)

Cost

- The cost of making a chip has been inversely proportional to feature size.
- Feature size halves every 3 years ...
- ... therefore cost doubles every 3 years.
- A current state-of-the-art silicon foundry costs c. $2 000 000 000
- It will remain useful for “leading edge” devices for just a few years.
- Many large companies now cannot afford foundries by themselves

THE UNIVERSITY of MANCHESTER

AMULET group

Where will it all end?

- The end is nigh!
  The end has been nigh for some time and is no nigher!
- The theoretical limits are still some way off
- The engineering limits are a lot closer
- The economic limits are closer still

What is the ‘limit’?

- The current rate of increase in available computing power will continue for at least 10 years.
- Within 5 years it will start to become apparent if the increase will continue beyond that point.
- Sometime <u>beyond 2010</u> the economically available computing performance will reach an effective plateau.

AMULET group

THE UNIVERSITY of MANCHESTER

Computers in 2010

Desktop computers

- Basically the same technology currently in use
- Comprise a single chip, maybe $1cm^2$
- Somewhat cheaper than they are now
- About 100-1000 times the computing power available today
- Hot enough to glow in the dark!

Embedded computers

- Most of the market
- Small, powerful and very cheap
- Ubiquitous