The University Of Sheffield.

Part 9: Future high-speed devices

CMOS present and future

Beyond the lithography limit

Ballistic transport devices

Carbon and graphene

Single electron and spin devices

Optical switching and computation

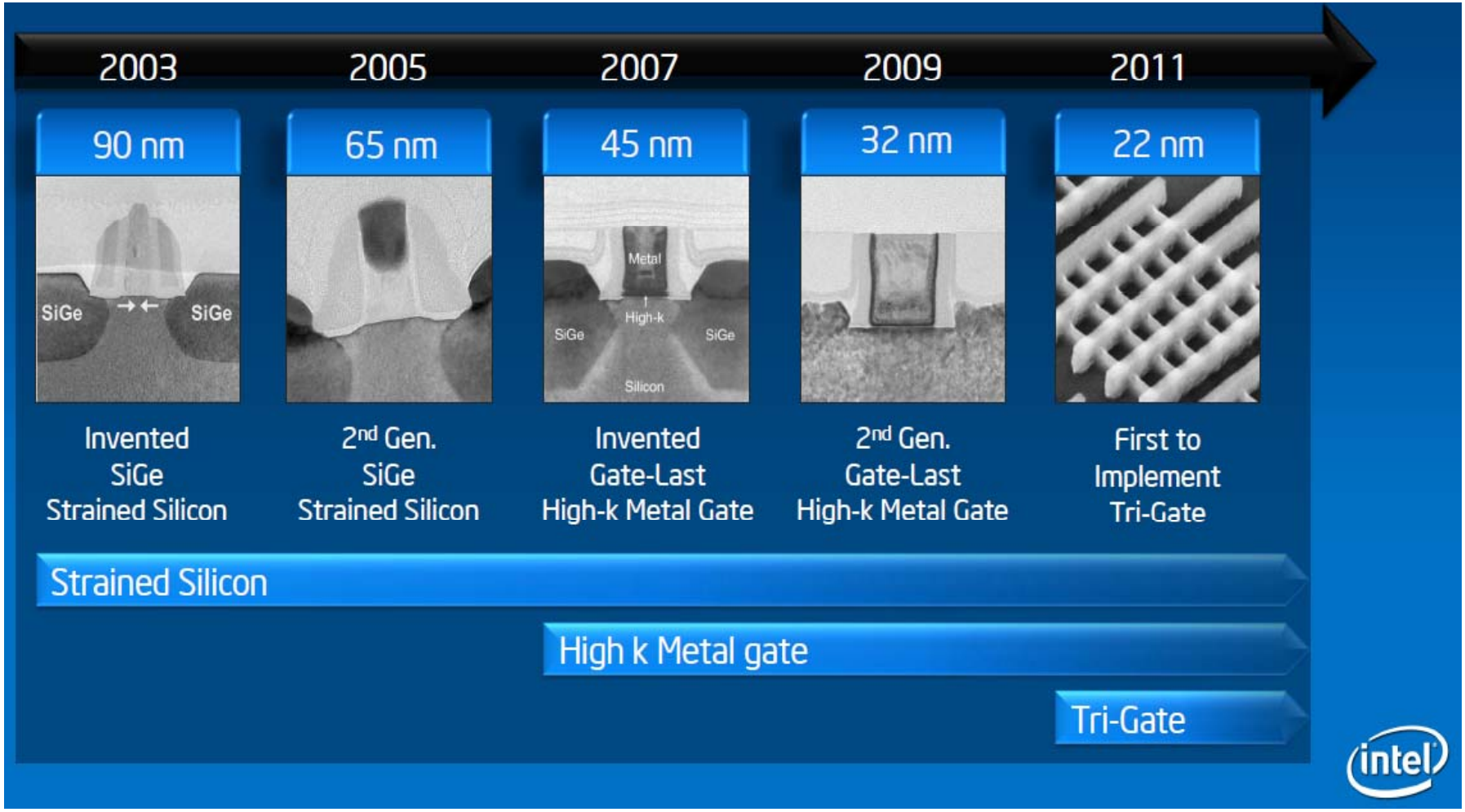
Future high-speed devices

The scaling of CMOS is becoming more and more difficult as we approach dimensions ~10nm. Also the cost is rising sharply per new technology mode (billions of dollars per fab)

Node	Demo.	Intro.	Processor	Technology
130nm		2000	Pentium III	248nm lithography
90nm		2002	Pentium 4	193nm lithography
65nm		2006	Intel Core	Proximity correction lithography, strained silicon
45nm	2004	2008	Core 2nd gen	High k dielectric
32nm	2006	2010	Core 3rd gen	Thinner high k dielectric, improved metal gate
22nm	2008	2012	Core 4th gen	3D Tri-gate
14nm	2009	2015-[1]	?	?

[1] In Feb 2011, Intel announced they would build 'Fab42' in Arizona at a cost of 4.5B$ to supply this technology. They delayed indefinitely these plans in Jan 2014.

Future high-speed devices



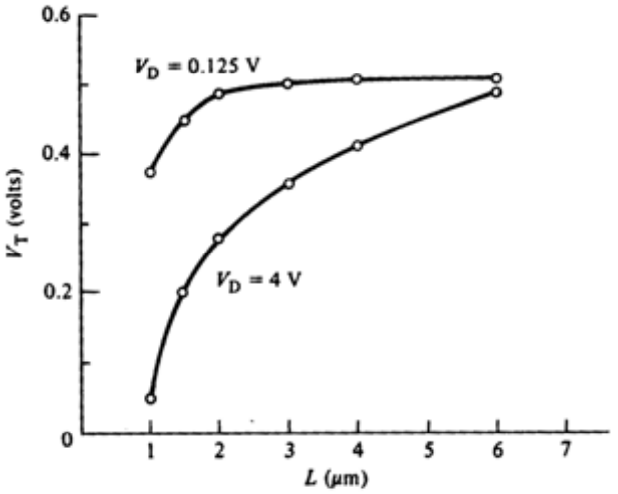
Moore's law is at work for now. But we should not underestimate the complexity and cost of the next steps.

Future high-speed devices

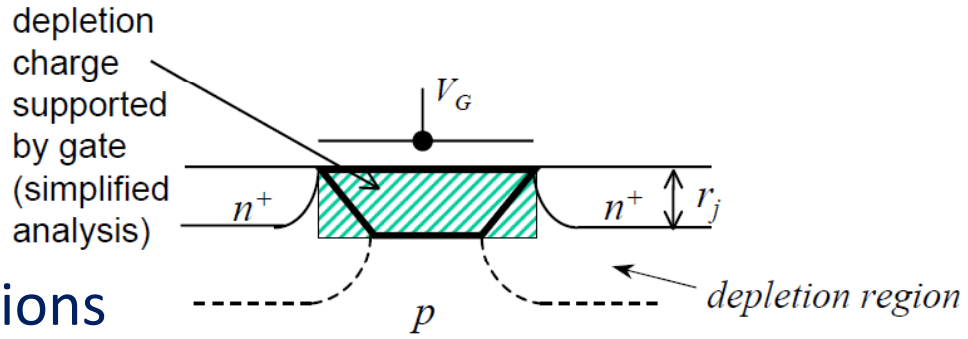
Present issues:

1) Short channel effect on V_T

There are a number of short channel effects present in MOS devices at small dimensions. Perhaps the most significant of these is the effect on V_T , which decreases with L.



This is good for low voltage devices. However at very small dimensions the threshold can collapse. This is undesirable: we need V_T to be finite and invariant with transistor dimensions



Future high-speed devices

and biasing.
The origin of this effect is as follows:

Before an inversion layer forms beneath the gate, the surface of the Si underneath the gate must be depleted.

This is done mainly by the gate voltage, but the source and drain p-n junctions help this process by lowering the potential barrier.

Large L:

Depletion charge supported by S/D

Small L:

Depletion charge supported by S/D

Future high-speed devices

This is not a significant issue for large gate lengths (gate length >> depletion width). However as the dimensions decrease the relative amount of charge supported by the source and drain increases. This makes it easier for the gate to deplete and so V_T falls and may even fall to zero.

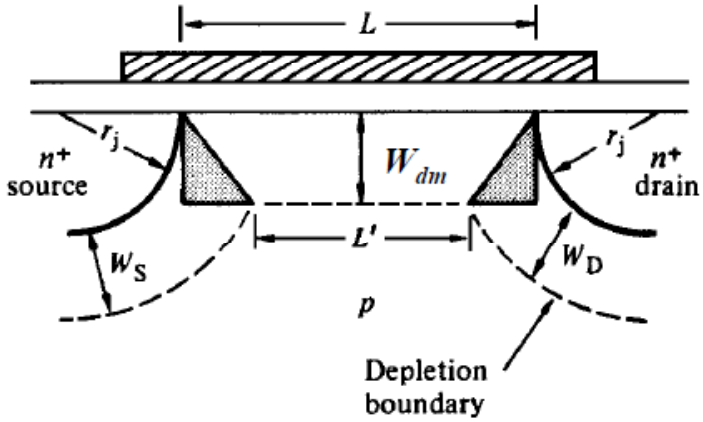
The gate now is only responsible for the reaming trapezoidal region (in white). This is less than the rectangular region by a factor

$$1-\frac{L+L'}{2L}$$

The effect on V_T is quite complex to derive. It turns out that:

Future high-speed devices

$$\Delta V_T = \frac{-qN_A W_{dm}}{C_{oxe}} \frac{r_j}{L}\left(\sqrt{1+\frac{2W_{dm}}{r_j}}-1\right)$$



So how to reduce ΔV_T (if we do not wish V_T to become infinitesimally small?)

- Increase C_{ox}- Reduce the oxide thickness
- Reduce R_J- shallower source drain implants
- Reduce W_{dm} (shallow devices) or reduce N_A

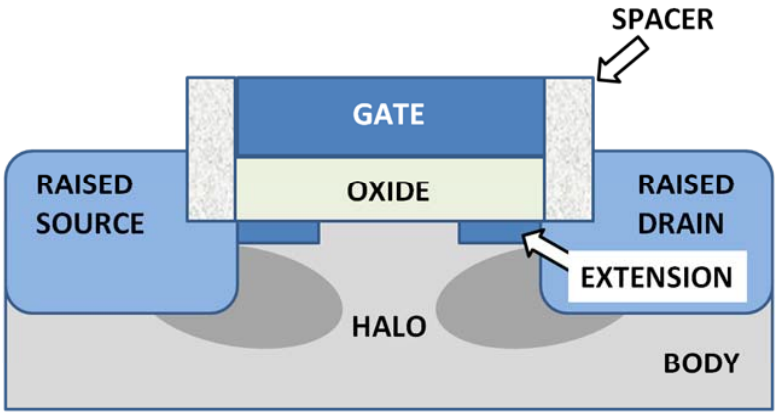
But their are problems for all of these options:

Future high-speed devices

- Increasing C_{ox} *increases gate leakage*
- Reducing r_J *increases the S and D resistance*
- W_{dm} and N_A are inversely related

$$W_{dm} \propto \frac{1}{\sqrt{N_A}}$$

One way forward is to use shallow source & drain extensions to effectively reduce r_j without increasing the S/D sheet resistance too much



Another is to vary the channel doping along its length using a 'halo' implant. This suppresses the effect of the source-drain depletion and also aids DIBL (see later)

Future high-speed devices

2) Gate Leakage

Major improvement in gate leakage was gained ~5 years ago by using the high-k dielectric Hafnium Oxide, HfO_2

This enabled the use of a thicker oxide (~3nm) whilst keeping the same capacitance value as before for SiO_2(~1.2nm) therefore reducing the gate leakage which exponentially reduces with thickness.

However in the recent 22nm process this HfO_x dielectric is back down to 0.9nm. This is only three atomic layers thick! Gate leakage is a major problem again.

One answer would be to find new dielectric materials with even higher k-values

Future high-speed devices

However several other conditions need to be met for these materials to be useful.

- A large height for electrons
- Thermal stability
- Mechanical stability
- A low bulk defect density (low number of traps)
- Low interface defect density (with Si)

	K	Gap (eV)	CB offset (eV)
Si		1.1	
SiO_2	3.9	9	3.2
Si_3N_4	7	5.3	2.4
Al_2O_3	9	8.8	2.8 (not ALD)
Ta_2O_5	22	4.4	0.35
TiO_2	80	3.5	0
$SrTiO_3$	2000	3.2	0
ZrO_2	25	5.8	1.5
HfO_2	25	5.8	1.4
$HfSiO_4$	11	6.5	1.8
La_2O_3	30	6	2.3
Y_2O_3	15	6	2.3
a-$LaAlO_3$	30	5.6	1.8

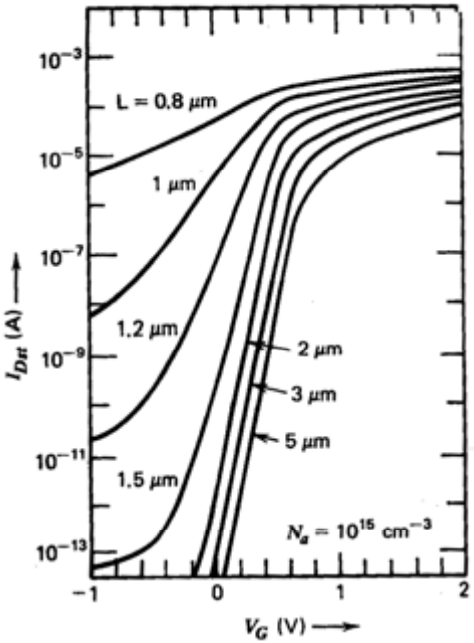
There are options for high-K dielectrics, but not all their properties are suitable

Future high-speed devices

3) Drain Induced Barrier Lowering (DIBL)

As the source & drain get very close together they become electrostatically coupled.

The drain bias can then affect the carrier flow at the source junction and cause an increase in sub-threshold current. In the extreme case devices never turn-off. Use of such devices would give a large static power dissipation

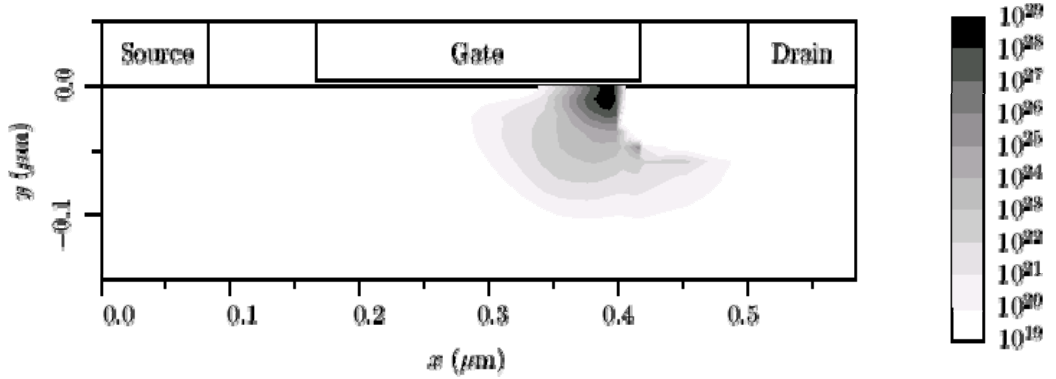


Future high-speed devices

4) Hot carrier effects.

As dimensions reduces, electric fields increase and impact ionisation can take place. The most likely region for this is the in the drain region close to the gate.

Simulated carrier density due to impact ionisation



These carriers can be collected by the drain, in which they contribute to the drain current, or can spill out of the channel to become trapped in the gate oxide or in the substrate where they become parasitic elements which impede the performance of the device.

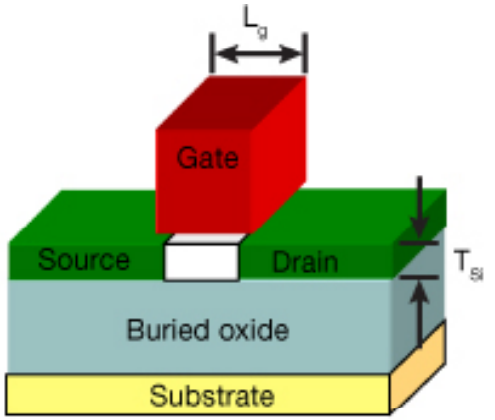
Future high-speed devices

3D Gates

3-dimensional gates are a response to some of the issues caused by short channel and leakage issues

Various groups have been investigating 3D FET geometries for the last 10 years. The basic 3D gate is usually called a Fin-FET.

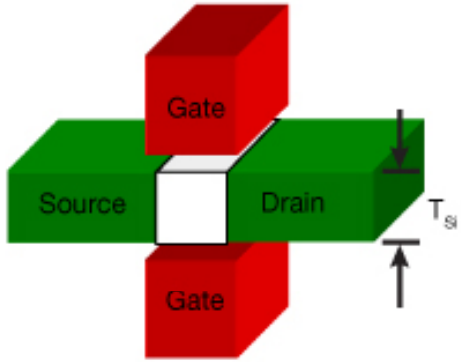
The concept developed from consideration of how to improve on the 'Ultra-thin body' MOSFET which attempts to solve short channel and leakage issues by keeping the gate as close as possible to all of the channel (shallow source & drain implants)



Ultra-Thin Body (UTB)

Future high-speed devices

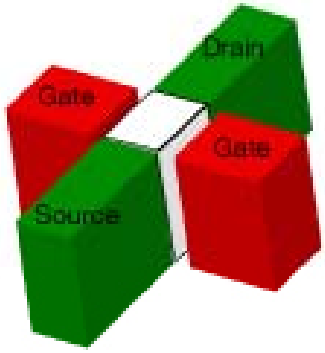
Hu at Berkeley suggested that a double-gate with depletion from both sides might be a better approach



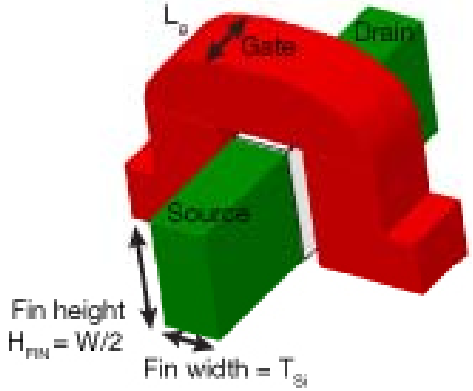
Double-Gate (DG)

The approach was highly successful and its natural extension is to put the gate on 3 sides- the FinFET

In the FinFET there is tight control of the conducting channel by a wrap-around gate. Very little gate field is spread and the influence of the source and drain regions is reduced



(90° rotation)

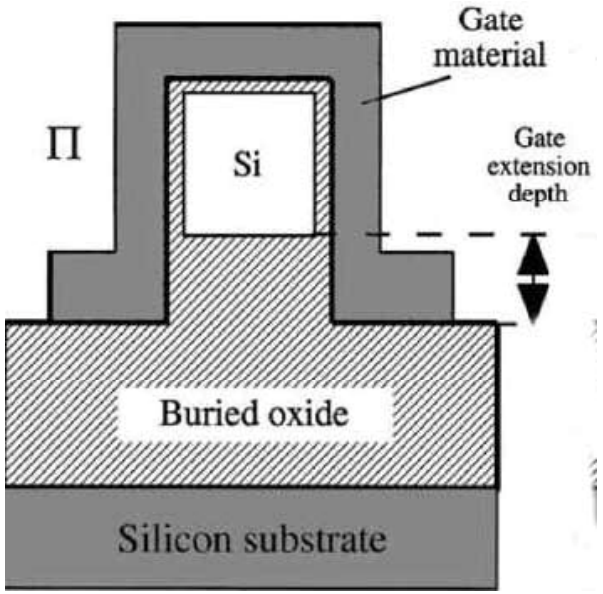


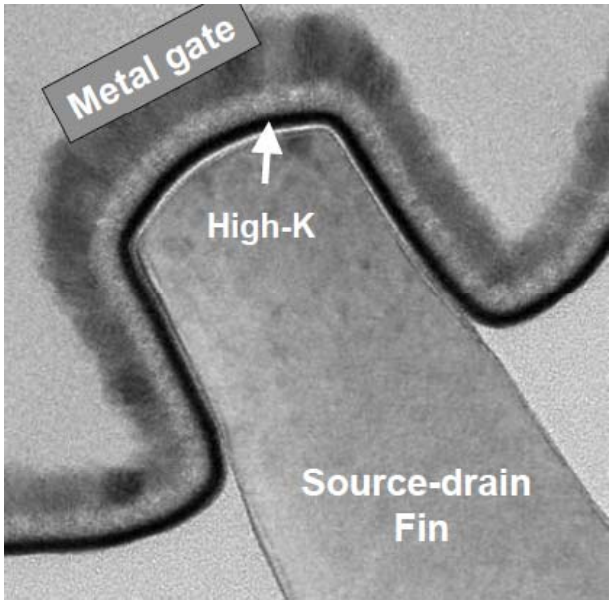
"FinFET"

Future high-speed devices

The construction of modern FinFET devices raises the level of the channel with respect to the source and drain. This allows the source and drain to be more highly doped by implantation.

The approach is based on silicon on insulator (SOI) technology in which a thin Si layer is re-grown on an insulating oxide. The quality of SOI growth has advanced considerably in the last decade





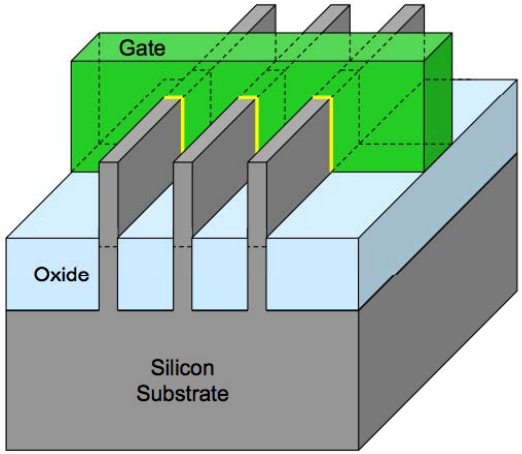
Future high-speed devices

When Intel introduced the Tri-Gate it took the FinFET concept and combined this with the idea of having multiple gates.

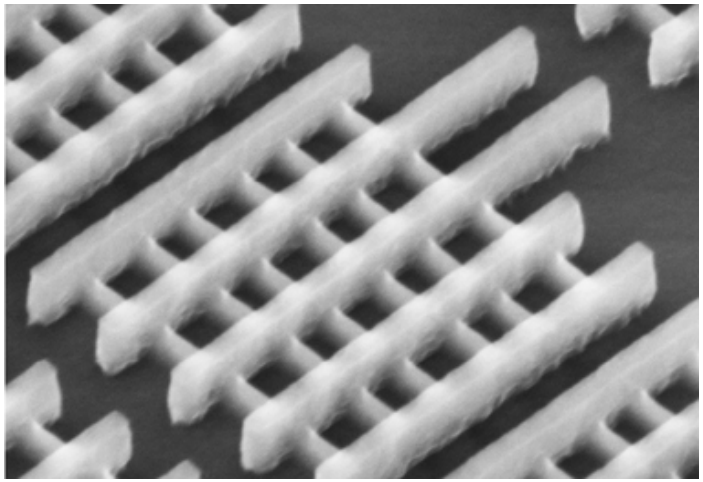
The approach has 3 gates and 6 source drain channels

Significant reductions in short channel effects and leakage current are seen.

22 nm Tri-Gate Transistor

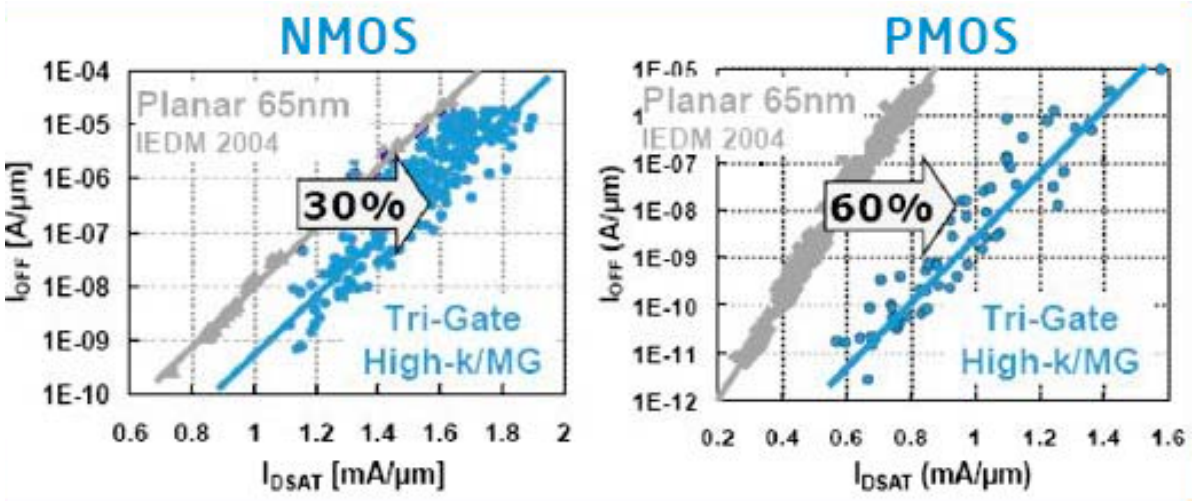


Tri-Gate transistors can have multiple fins connected together to increase total drive strength for higher performance



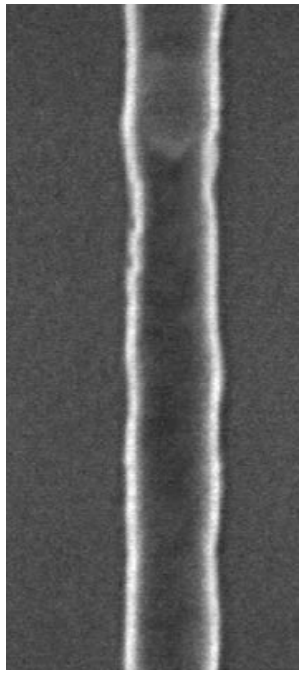
Future high-speed devices

Intel has published results which show major reductions in the off-state current, particularly for the PMOS



The FinFETs approach has alleviated problematic performance versus power trade-off issues. Designers can run the high-speed devices faster and use the same amount of power, compared to the planar equivalent, or run them at the same performance using less power. This enables design teams to balance throughput, performance and power to match the needs of each application. It has been a significant factor in extending Moore's law over the last few years

Future high-speed devices



But can Moore's law of scaling go on forever?

1) Limits of photolithography

Significant line-edge roughness at <20nm due to statistical variations in resist concentration and photon shot noise

Photolithography is a so-called *top down* nanotechnology approach. One starts with a piece of semiconductor and then pattern and etch it to produce nanoscale features

Another possibility would be to perform *bottom up* assembly of nanostructures from individual atoms

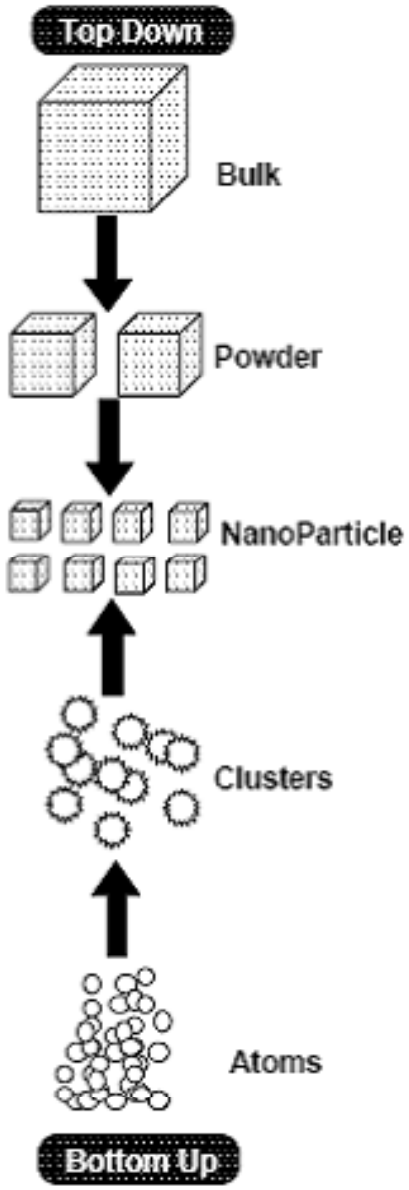
Future high-speed devices

In a bottom up nanostructuring we start with single atoms (~ 0.3nm) or molecules (~few nm) and assemble these together to form nanostructures

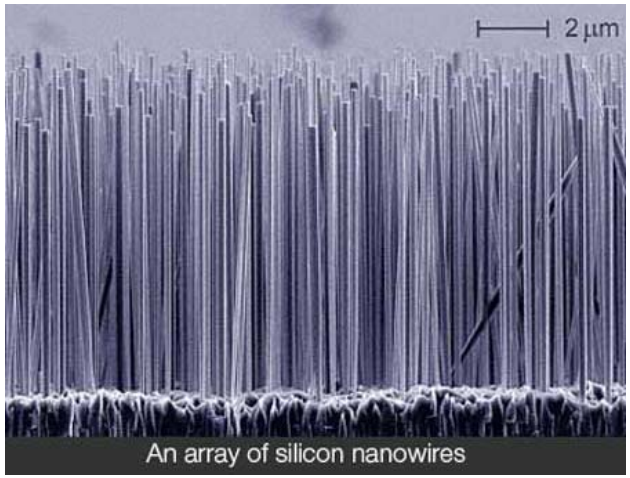
But how to assemble these structures? (we don't have the tools to build structures at the 1nm scale)

We rely on self-assembly processes which may naturally create certain types of structure through energy minimisation

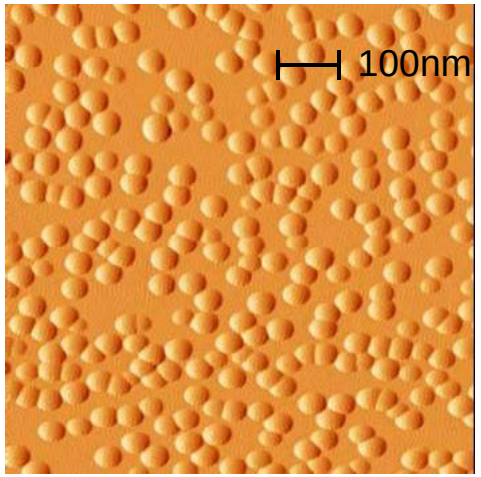
Examples of the types of structure that can be created include nanowires and quantum dots



Future high-speed devices

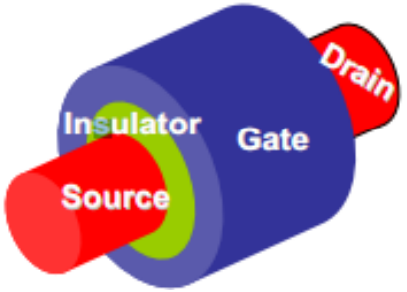


Silicon nanowires



InAs quantum dots

Nanowire (NW) FETs are a natural extension of the FinFET idea which allow the gate to completely wrap around the channel. There is a lot of research in this area and some impressive individual NW results, but fabrication of complex circuits is very difficult with present capabilities



Future high-speed devices

2) Ballistic/Hot carrier effects

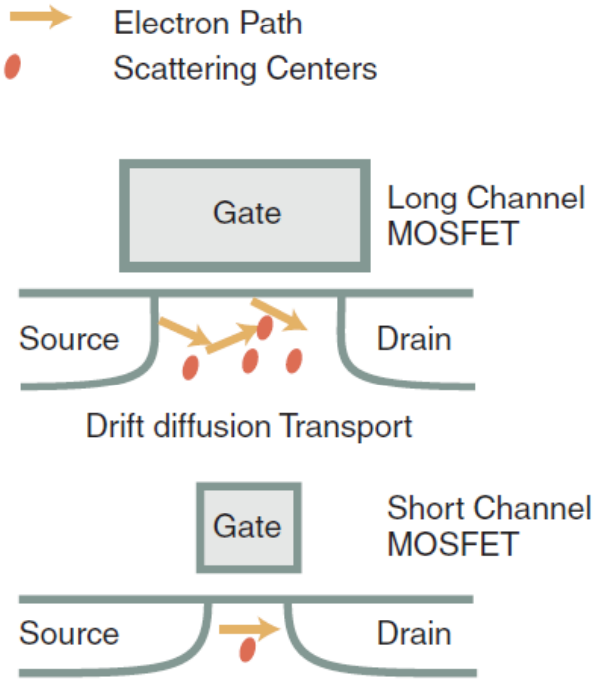
As we shorten the gate lengths the fields increase. For all current CMOS devices we can consider that the carriers have reached their *saturation velocity*. These are highly energetic 'hot' carriers

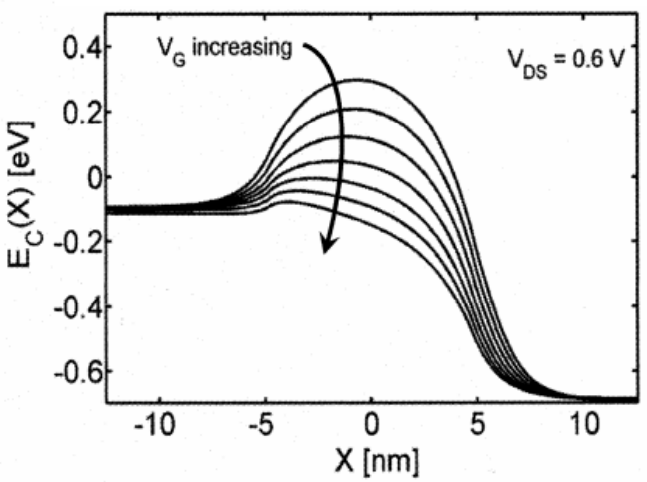
What happens then when the channel length is really small (≤ 20nm) is that its possible for carriers to travel from the source to drain without ever having a scattering event. In this case we say this is **ballistic transport.**

Future high-speed devices

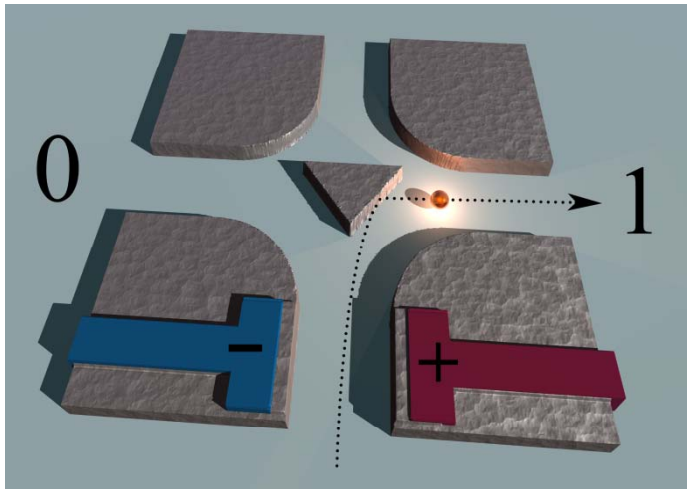
This transport is very fast, which is potentially good for high speed operation. However the device operation is somewhat changed. We can no longer think of the FET gate as restricting a current flow through depletion of the channel. Instead the gate becomes an electrostatic barrier to ballistic particles.

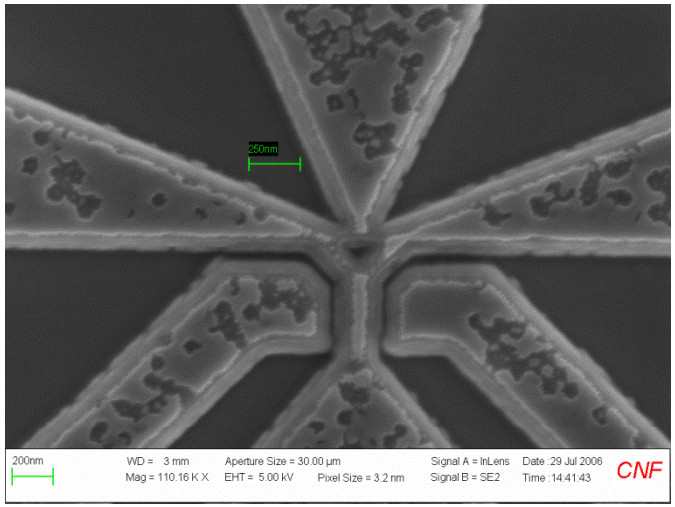
A variation on of this approach is to use a potential barrier to deflect electrons rather than impede them. This approach is called the ballistic deflection transistor.





Future high-speed devices





The concept is attractive because it does not impede the motion of electrons and instead the gate simply imparts a small kinetic energy to deflect them it to path which can represent a digital ‘1’ or a ‘0’. It can therefore offer very low power consumption and a very high speed capability.

Future high-speed devices

3) Quantum mechanical tunnelling

We have already discussed the issue of quantum mechanical tunnelling through the gate dielectric which causes leakage and static power dissipation.

However as we reduce the dimensions below ~10nm we begin to see quantum mechanical tunnelling between the source and drain and what is in between (the gate) becomes ineffective.

4) Ultimate limit- the material itself.

The channel of the 22nm is only 70 Si atoms thick!.

The oxide thickness is currently 3 atomic layers!

Much smaller and we enter the physics of molecular electronics

Future high-speed devices

(few atoms) rather than crystals (many atoms) and at this point our conventional semiconductor models are useless. We enter a whole new type of physics.

It could be possible to use a single atom or molecule as a switch, but its operation would be very much different to conventional high-speed devices

Carbon-based transistors

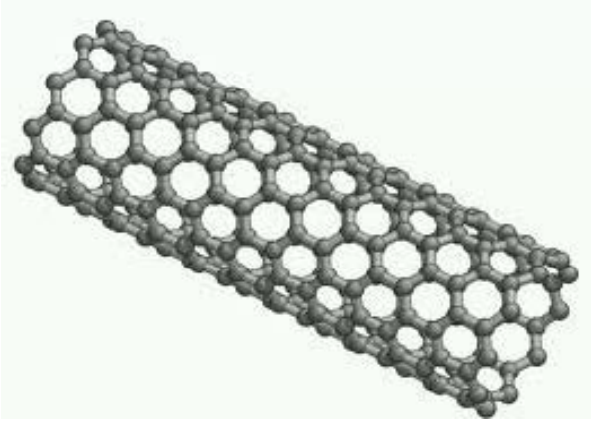
Carbon is in the same column of the periodic table as Si & Ge and has semiconducting properties as one would expect. But carbon is rather special in that it exists in many structural types (allotropes). Graphite & diamond are the two most common, but we consider here carbon nanotubes and graphene (not strictly different allotropes, but forms of single layer graphite)

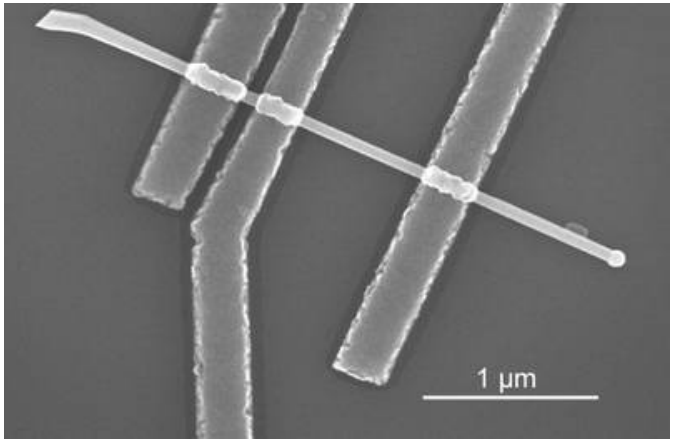
Future high-speed devices

Carbon nanotubes

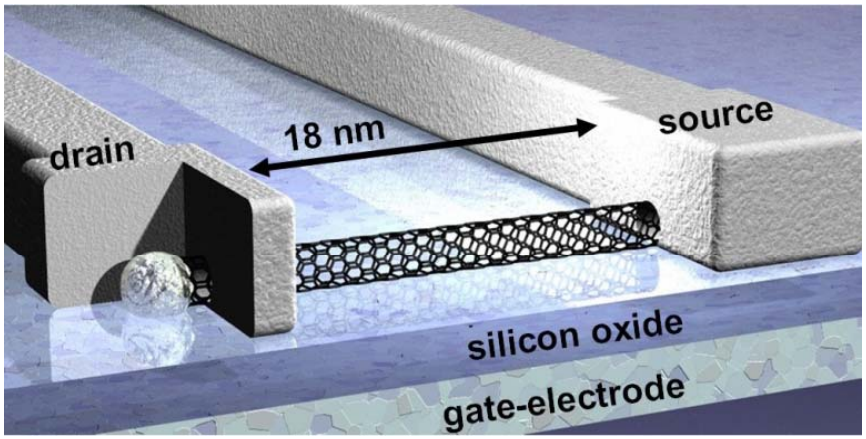
These form by self-assembly and are produced by a variety of techniques, the simplest of which uses a methane plasma.

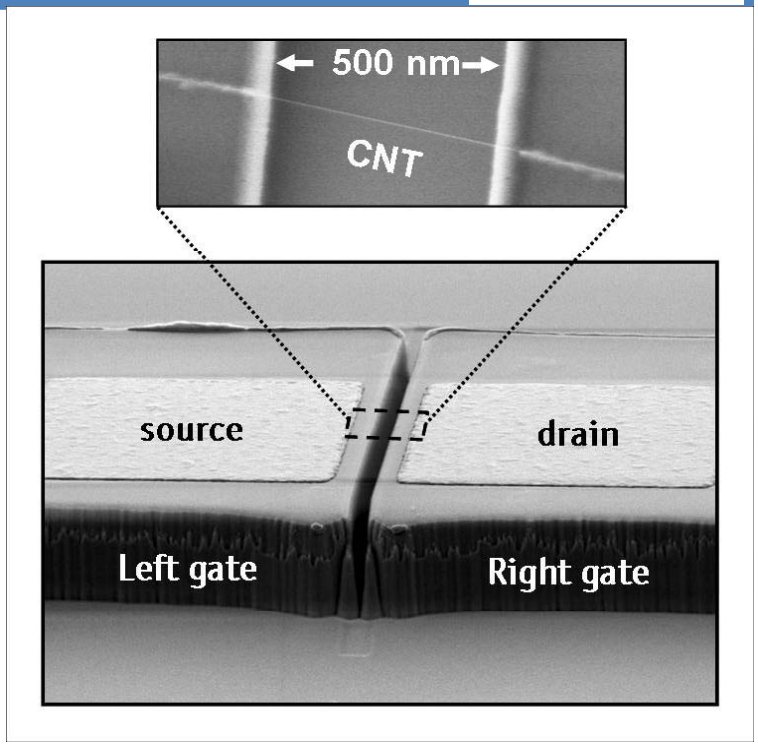
They occur in dimensions down to the nm range. Controlling the dimensions is a major challenge. On the laboratory scale we may produce simple transistor structures by depositing source, drain and gate contacts





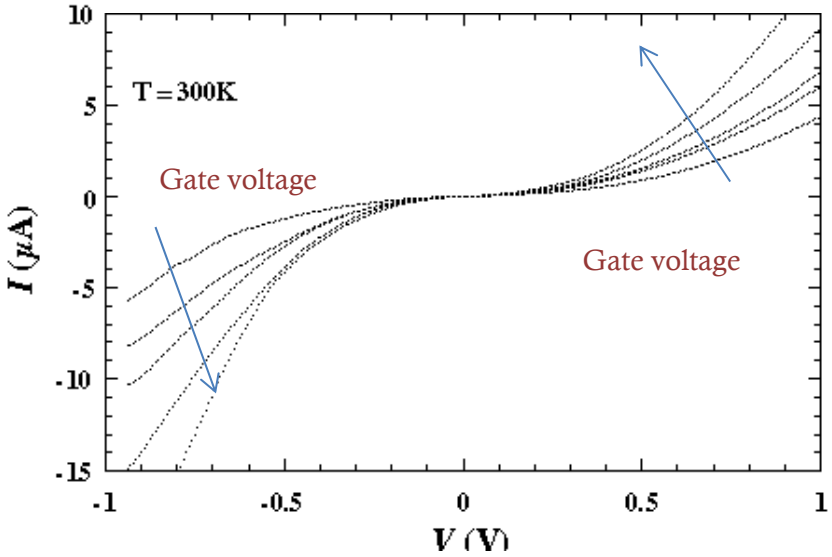
Future high-speed devices





The nanotube conducts electrons or holes depending on the direction of the gate voltage

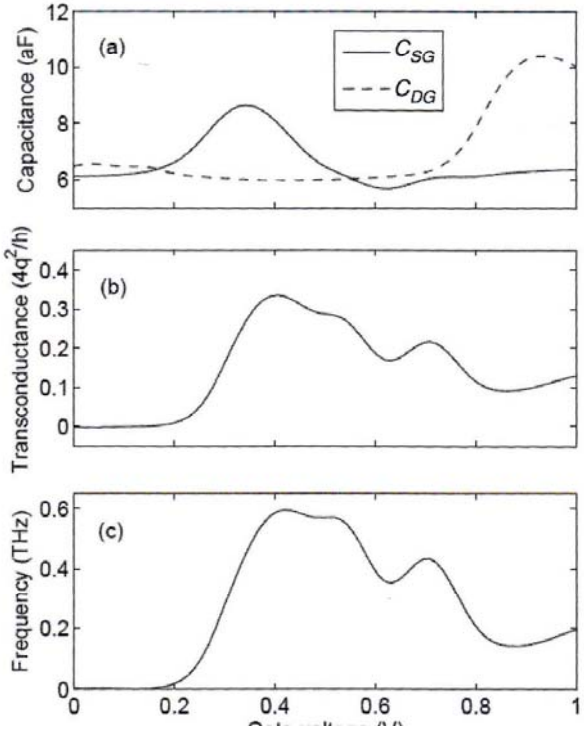
Hard to make a gate contact directly to the carbon nanotube- need to use back gating, which is not very efficient.



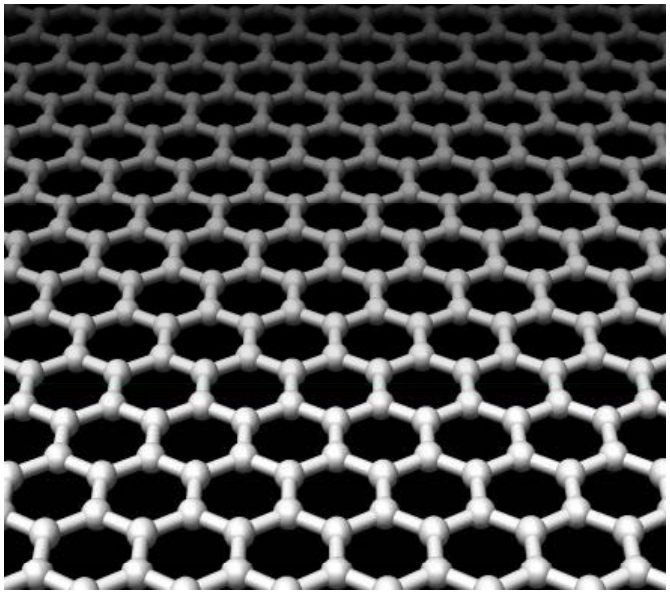
Future high-speed devices

Carbon nanotubes offer a very high mobility and saturation velocity. Carriers move ballistically and there are no collisions with the lattice. Rather than treat this as a convention transistor we need to consider the quantum mechanical transmission and reflection of electron waves.

The transmission of electrons from the source and drain is no longer a smooth function of the gate voltage but is highly peaked. These ‘resonances’ are a consequence of quantum mechanical reflections.



Future high-speed devices



Graphene is a single sheet of carbon atoms which have the structure found in the 'hard' plane of graphite. It can also be thought of as an un-folded carbon nanotube.

Andre Geim and Konstantin Novoselov from the Univ. Manchester won the nobel prize for the discovery of graphene in 2010

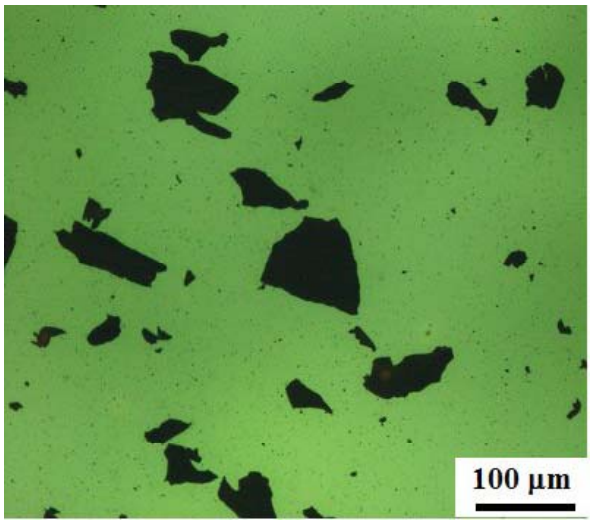
Electrical properties of graphene

- High room temperature mobility(~ 40,000 $cm^2V^{-1}\ s^{-1}$)
- Very high saturation velocity (>$10^6 M.s^{-1}$)

These values are a factor of 10 higher than the conventional semiconductors like Si and GaAs and are very promising for transistors

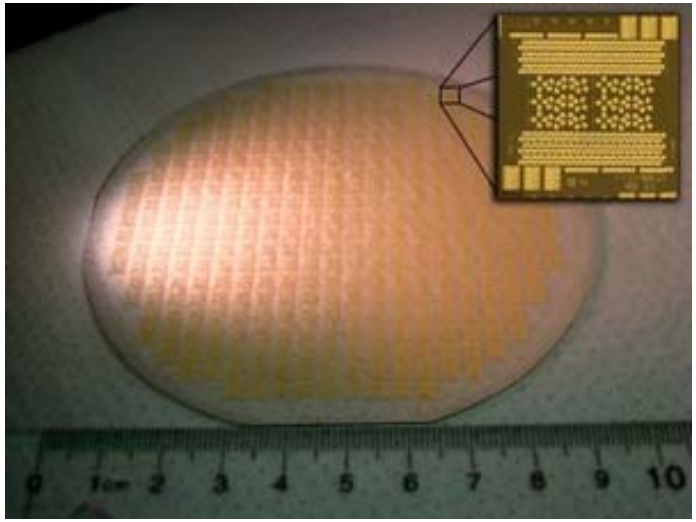
Future high-speed devices

But these very high mobilities are for a single atomic sheet. They have only been observed in graphene particles made by exfoliating layers from graphite with sticky tape!



There are however a number of attempts to create thin film graphene. One of these uses the decompostion of SiC wafers are high temperature (1600^{o}c) to leave a thin carbon film on the surface.

Another approach grows a thin film of graphene on a copper film.



Future high-speed devices

The copper is then etched away and the graphene film is transferred to a diamond substrate

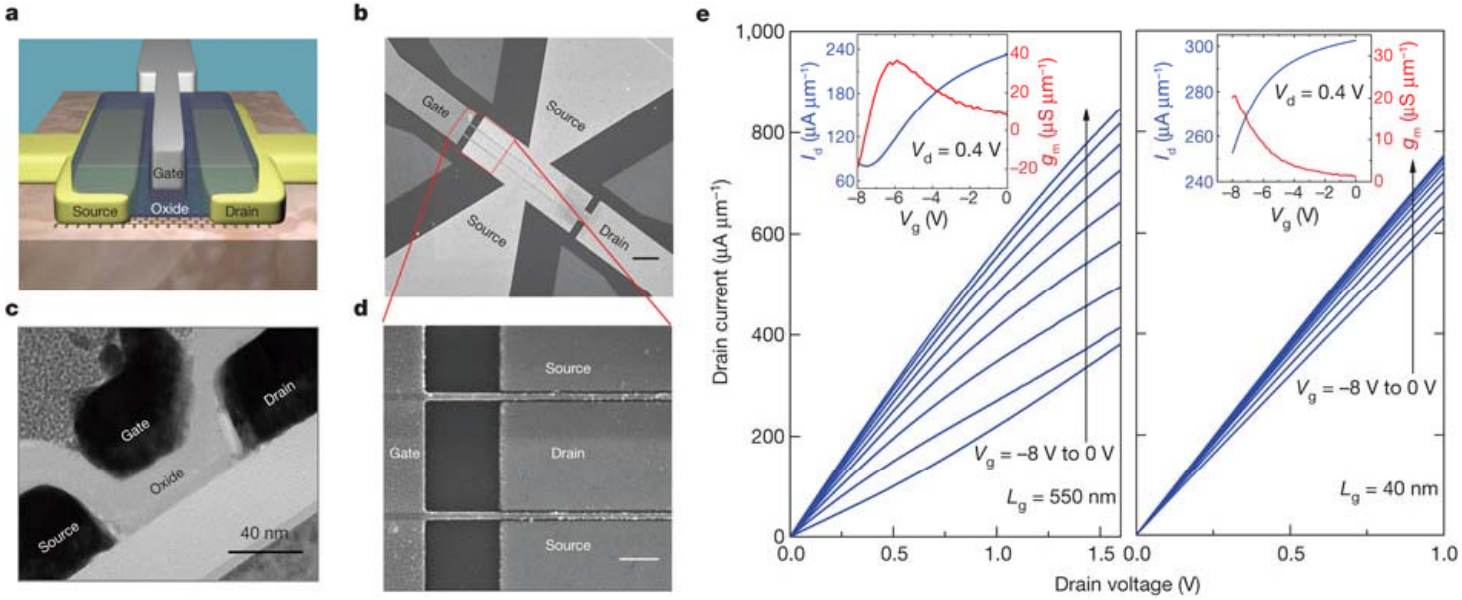
The properties of these forms are not as good as the exfoliated graphene

Graphene in any form is very difficult to incorporate into 'traditional' device schemes. It is also very hard to contact to.

Despite these issues, there has been some very spectacular progress in device performance.

IBM Research Labs:
MESFET-type structure with a single layer of graphene comprising the channel.
Recent result: f_T~ 155 GHz for a 40mm gate length

Future high-speed devices



Issues

Very poor gate action. As a result this device has a very poor on-off ratio. Also a very severe short channel effect, so the best f_T is seen from devices which are hardly functioning as a FET.

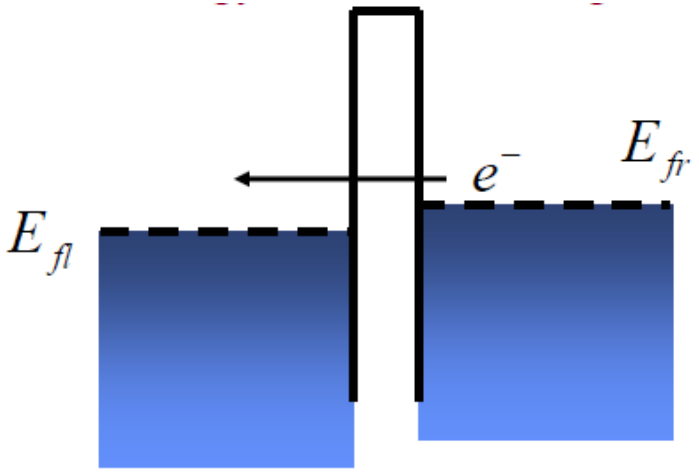
The main issue is contact resistance. It is very hard to contact to this single sheet of atoms. Resistances can be 100-1000Ω

Future high-speed devices

Single electron transistor (SET)

What if we made a transistor switch which uses only one electron. Wouldn't that be the fastest and most efficient device?

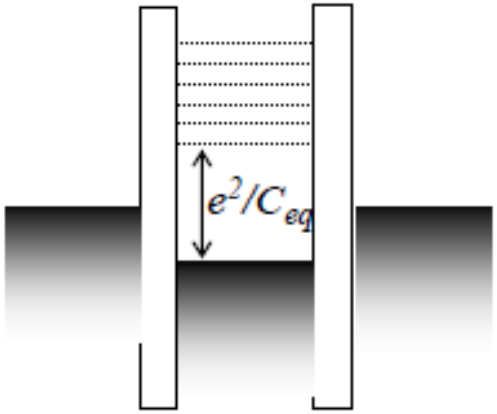
To understand the SET, we need to go back to quantum mechanical tunnelling. Consider two n-type semiconductor regions with an insulator between them. The insulator prevents classical current flow, but there will be a finite tunnelling rate



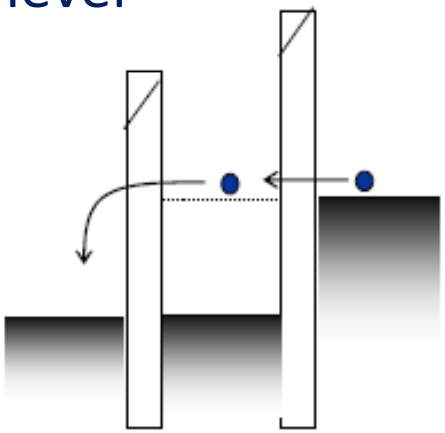
Now consider a double barrier device with semiconductor in the middle

Future high-speed devices

The inner semiconductor region is made sufficiently thin so it has quantum confined energy levels. Electrons cannot tunnel in to this region unless these line up with the Fermi level



No Bias

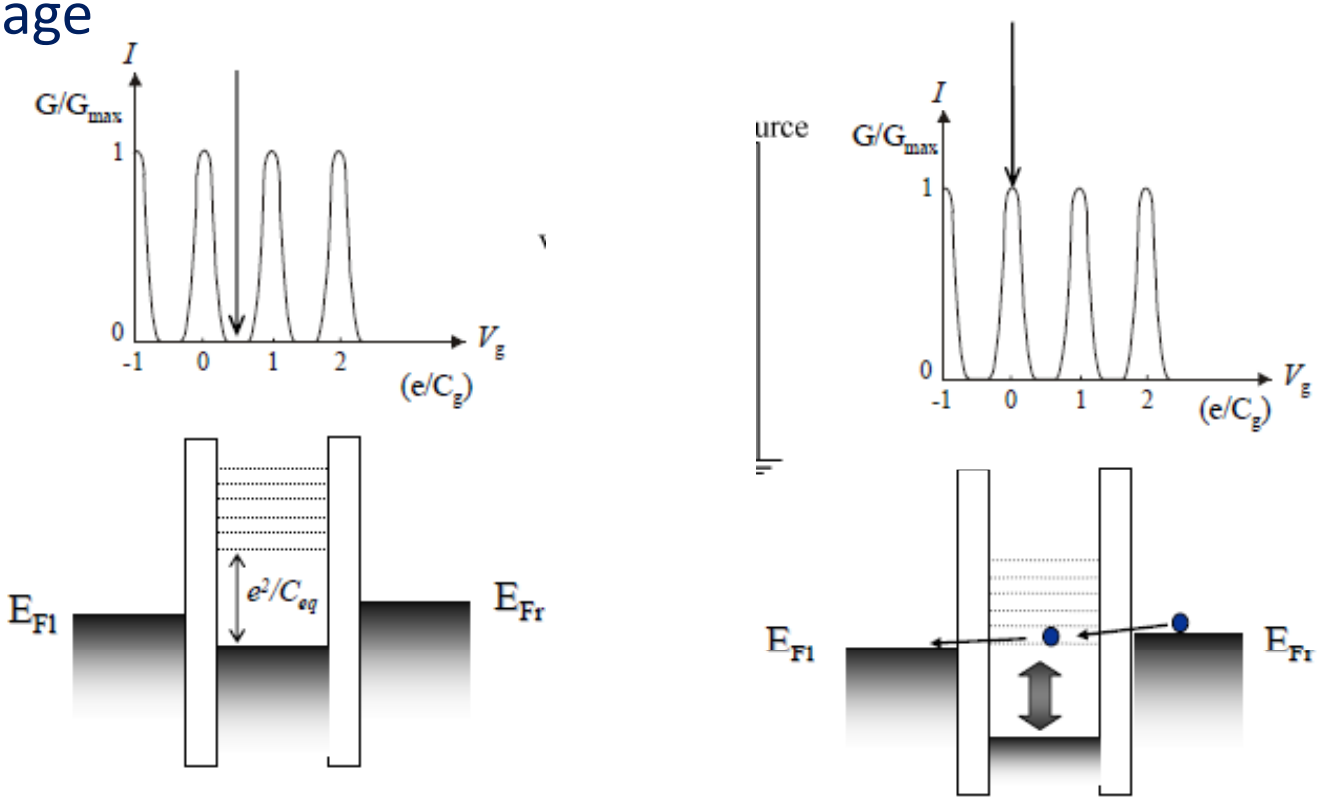


Biassed until tunnelling starts

When 1 electron tunnels into the middle region this raises the Fermi energy of that region and the bands become misaligned. In a quantum well this is not an issue as the number of quantum well states is high.

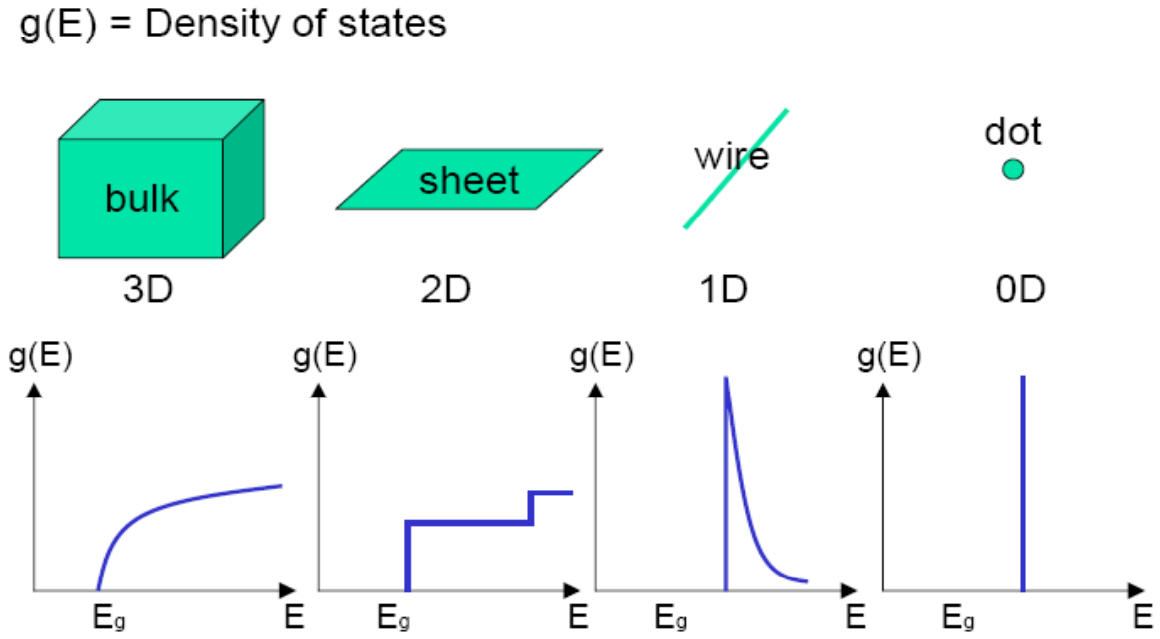
Future high-speed devices

But if we use a structure than has very few states then the change in Fermi level can block the tunnelling. In this way the single electron transistor can control the flow of electrons one-by one, resulting in a conductance oscillation as a function of the gate voltage



Future high-speed devices

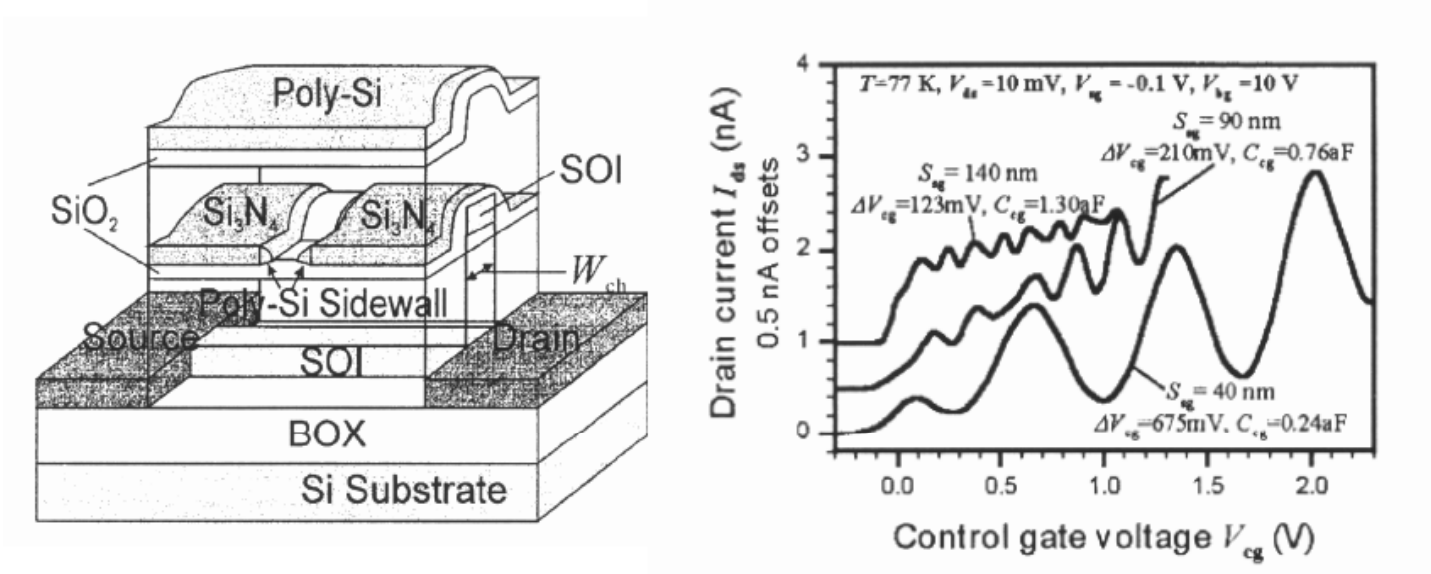
For the single electron transistor to work the central or ‘bridge’ material really needs to be a quantum dot with only 1 electron state at a particular energy



In this case only one electron can tunnel through at a certain time and further electrons are prevented from tunnelling by what is called the **coulomb blockade**

Future high-speed devices

Semiconductor quantum dots may be produced by self-assembled epitaxial techniques.
They can also be produced by electrostatically isolating bulk materials by means of lithographically defined gates



D. H. Kim *et al., IEEE Trans. ED 49, 2002*

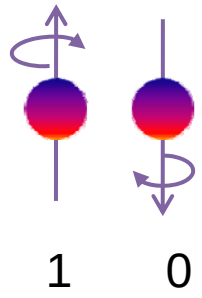
Future high-speed devices

Positive points: Very high speed (femtosecond tunnelling times), very low drive voltages, almost zero leakage

Negative points: Low voltage gain, poor reproducibility, needs a complete change in logic control compared to FET based devices

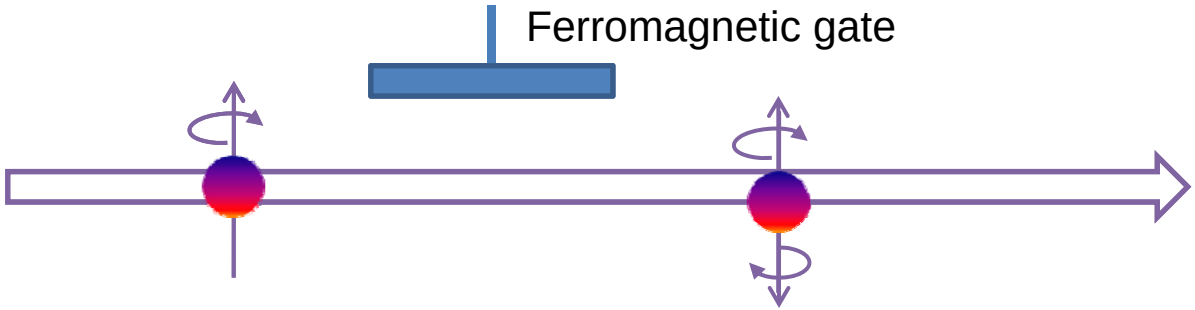
Spin-transistor

Electrons have two states- spin up and spin down.



We could use a ferromagnetic gate to change the spin state. Energy required to flip spins is a fraction of that needed to move 1 electron (~ 0.0005eV) and the flip operation is ultra fast (fsec).

Future high-speed devices



Very similar concept to a MOSFET, but the gate imparts a magnetic field rather than an electric field

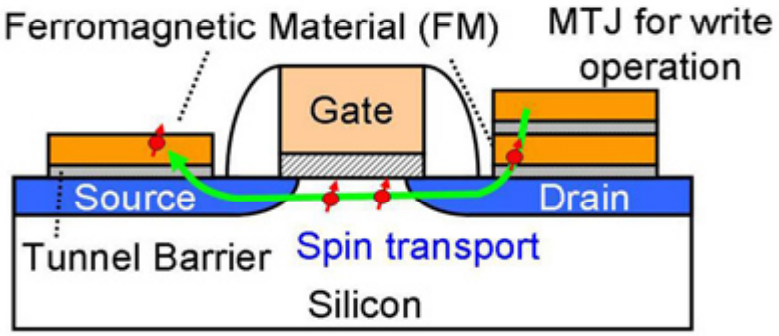


Diagram of Toshiba's spintronics-based MOS field-effect transistor

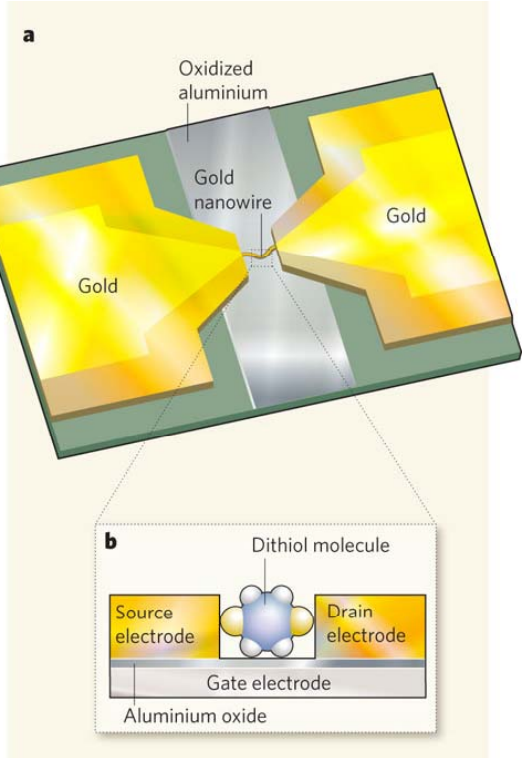
We need some mechanism to analyse the spins (rather than To measure charge). One approach is to use a magnetic junction which favours the tunnelling of one particular spin type

Future high-speed devices

This approach could lead to ultra low power fast devices in the future, but in many semiconductors at room temperature the spin state is coherent only for a short time (typically nsec). Present focus is on nitride semiconductors where spin lifetimes ~1µs have been observed

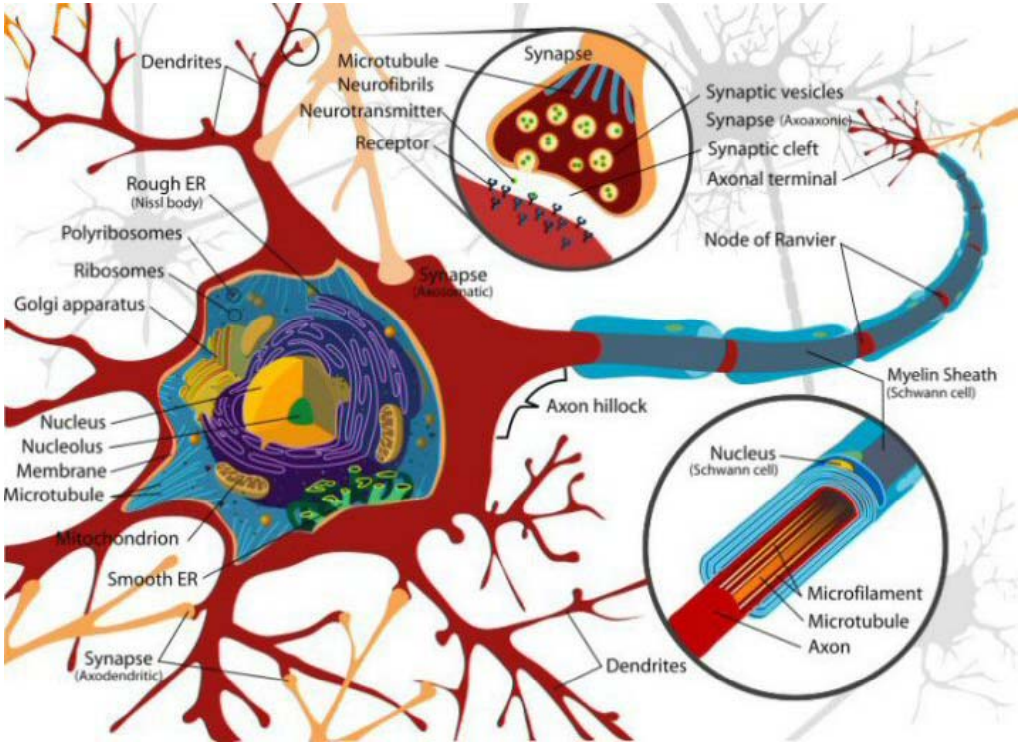
Molecular Transistor

This relies on changes in conductivity through a molecule (typical dimensions ~5-10nm) in response to an electrostatic field. Similar concepts to the single electron transistor.



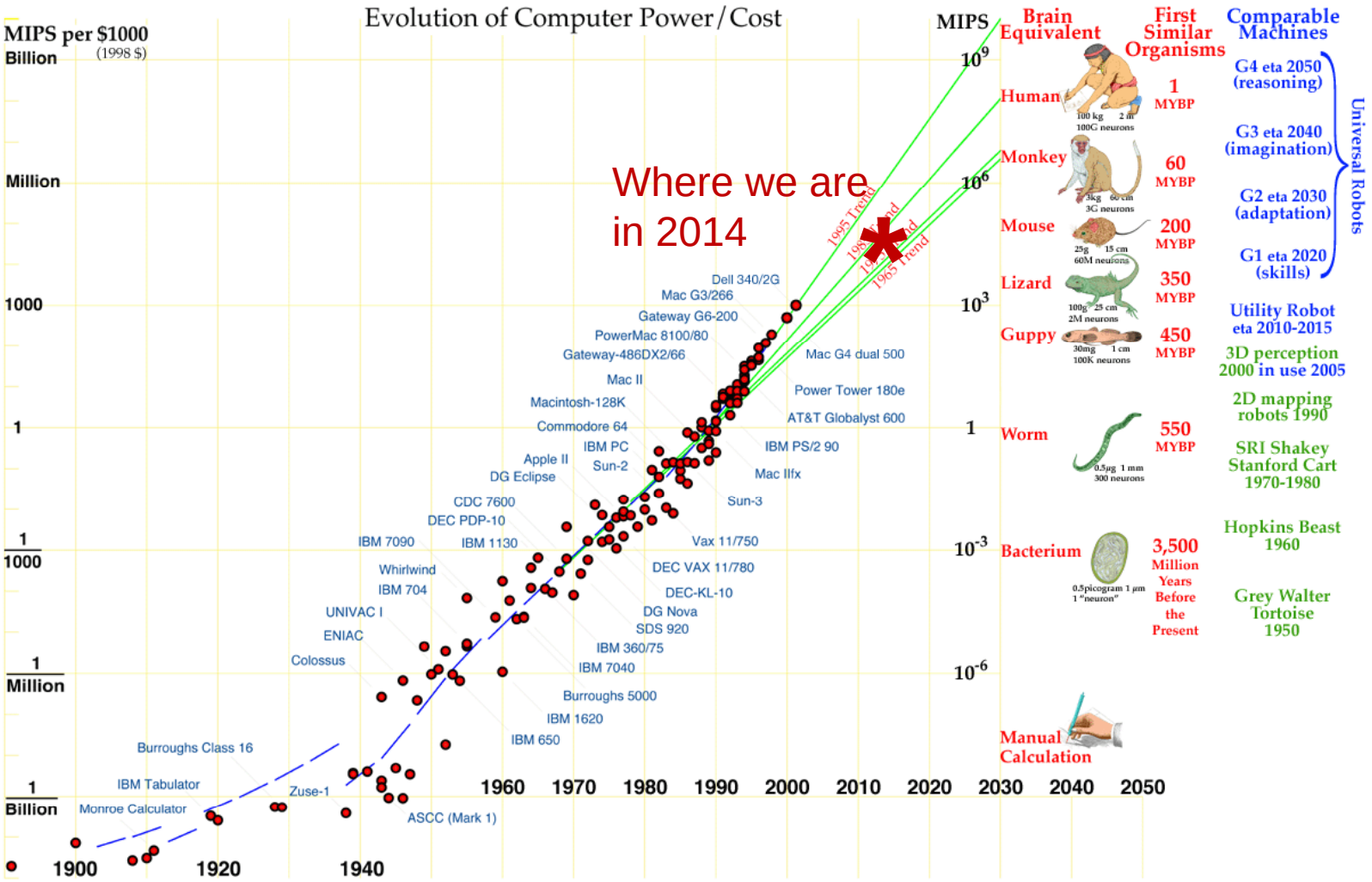
It is very hard to contact to a single molecule and this will limit practical application of this technology. However the long term future of computing is almost certainly a self-assembled biological structure

A self-assembled molecular supercomputer



More powerful than any man-made system by a considerable margin

Moore's law brain equivalent



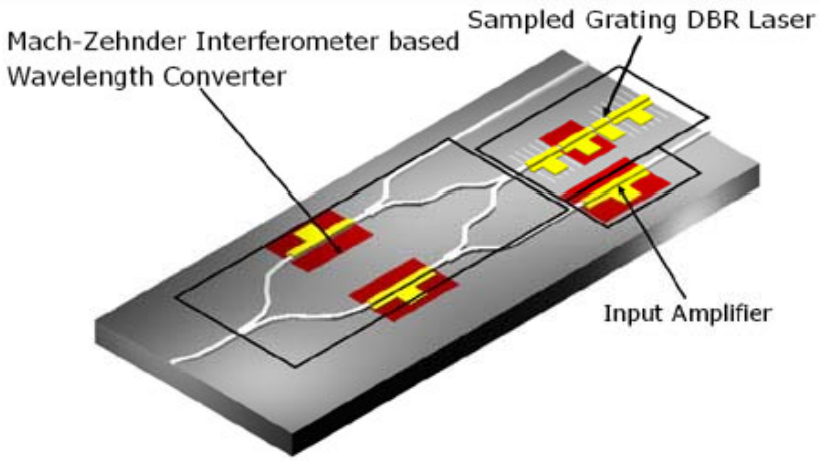
Future high-speed devices

Optical (photonic) switch

Instead of switching electrons, why not switch photons?

Its possible to generate **ultrafast pulses of photons** (in fact down to 10^{-16}s timescales). Its also possible to switch light with electric fields (inducing a change in band gap) or with light making use of **optical non-linearities.** Its also possible to guide light through circuits using **optical waveguides.**

One such approach is based on an optical interferometer.
Light from a laser is split into two paths of the same length. The light normally combines constructively at the output.



Future high-speed devices

However if a second (control) beam is added to one arm, it changes the refractive index. This changes the path length and the light begins to interfere.

Inducing a 180^{o} phase change in one arm would be sufficient to cancel the output beam. So with the addition of an external control beam we are able to turn the output on or off. This can be done at very high speed (picoseconds)

The long term applications of optical waveguides may come as quantum computing replaces classical computing methods

Classical computing: Information is stored and manipulated as bits, which take the discrete values 0 and 1. In conventional transistor this would be ‘low current’ or high current’. In a single electron transistor this would be ‘no electron’ or ‘1 electron’. The logic that we perform on these bits is the same as we would do on paper.

The University Of Sheffield.

Quantum computing

Information is stored in quantum bits, or *qbits*. A qbit can be in measured as a $|0\}$ or a $|1\}$, but exists in a *superposition* of these states, $a|0\} + b|1\}$, where a and b are complex numbers.

Superposition is a quantum mechanical effect. Its states that an electron can exist partly in all its possible states simultaneously but when measured or observed, it gives a result corresponding to only one of the possible configurations. We can think of the state of a qbit as a vector, then the superposition of states is just vector addition.

Consider three classical bits. These can store 2^3 numbers: 0, 1, 2, 3, 4, 5, 6, 7 (001, 010, 011, 100, 101, 110, 111)

3 Quantum qbits can do the same. They could for example store

Future high-speed devices

the numbers 3 and 7 in much the same way as classical bits

$$|0\rangle \otimes |1\rangle \otimes |1\rangle \equiv |011\rangle \equiv |3\rangle$$
$$|1\rangle \otimes |1\rangle \otimes |1\rangle \equiv |111\rangle \equiv |7\rangle$$

But qbits can also store a 3 & 7 *simultaneously* due to superposition. Consider that the first bit exists as a vector sum of 1 and 0 $\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$

We then have 3 bits as:

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |1\rangle \otimes |1\rangle \equiv \frac{1}{\sqrt{2}}(|011\rangle + |111\rangle) \equiv \frac{1}{\sqrt{2}}(|3\rangle + |7\rangle)$$

This ability to store multiple values is the major advantage of quantum computing. Manipulating these states to perform logic requires so called unitary operators which are quite different from normal logic

Future high-speed devices

However in simplistic terms we can find logic-like operations by performing through interactions with other quantum states. So we could perform an AND type operation with a known 3 bit state of 7

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |1\rangle \otimes |1\rangle \bullet |111\rangle \equiv |111\rangle$$

These operations are performed by process known as quantum entanglement.

This is a phenomenon which occurs when pairs of particles interact such that the quantum state of each particle cannot be described independently and instead, a quantum state is only known for the system as a whole.

Future high-speed devices

The particles can be quantised in such position, spin, polarisation etc of electrons or photons. Much current work is focussed on quantum computers based on the interactions between single photons.

This only gives a very brief introduction to a complex area which is well beyond the course

In this section I have tried to illustrate the challenges and potential for future high speed devices. It may be that through new materials and structures that the existing ‘Moore’s Law’ approach continues at least from the next 10 years.

However there are also new physical phenomena which may lead to completely new device approaches

End of course