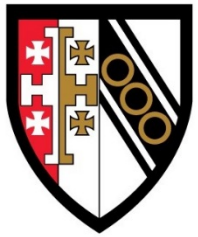
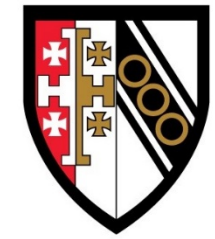


Selwyn College
Cambridge

Big Data and what it means

Anita Faul

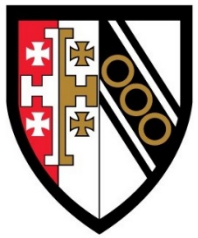
Selwyn College



Selwyn College
Cambridge

How much data do we carry with us?

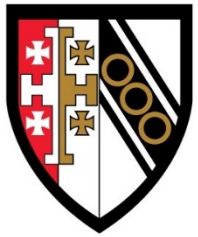
16 Gigabyte on a smart phone is barely enough.



Selwyn College
Cambridge

Actually, it is much more

- In February 2016, it was published that there are 26 different ways a synapsis in the brain can encode its information.
- $26 = 2^{4.7}$, that is 4.7 bits, a little bit more than half a byte.
- It is estimated that an adult has about 10^{14} synapses.
- This makes 50,000 Gigabyte or 50 Terabyte.
- How much of this is used is a different question.



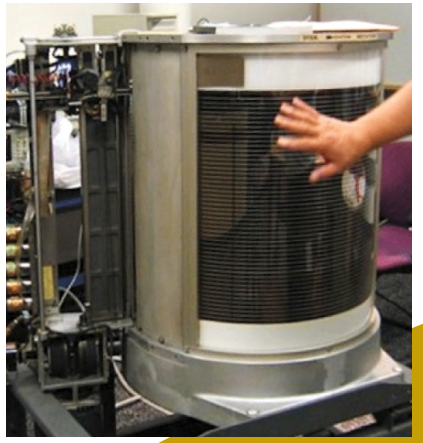
Selwyn College
Cambridge

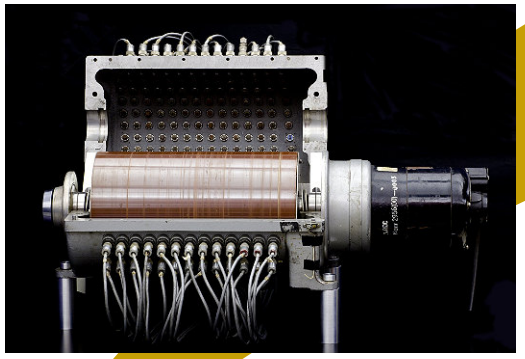
2.5 Quintillion bytes of data are created every day.
That is 2, 500, 000 Terabytes.

How did we get here?

Selwyn College
Cambridge

1956
5 MB

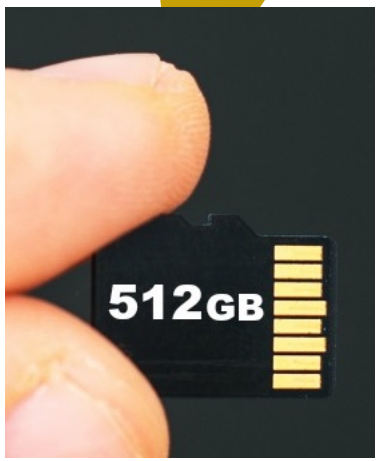




1932
10 KB

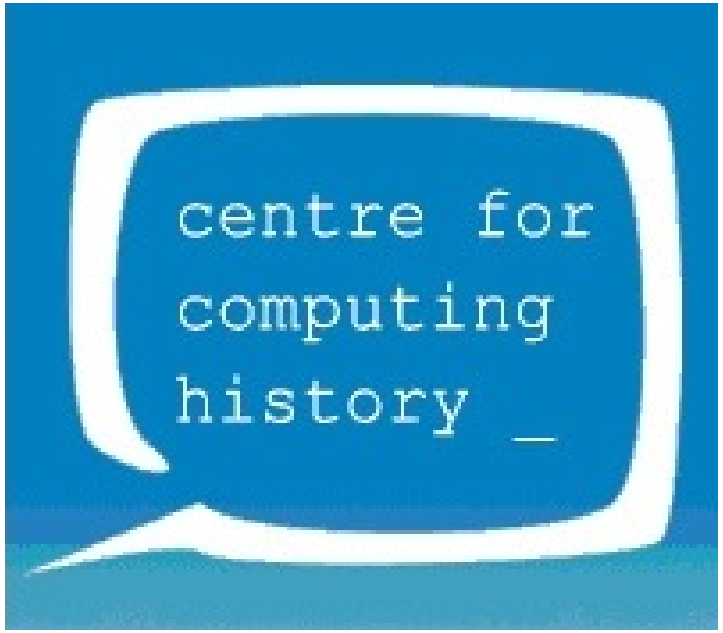
1971
79.7 KB

1982
700 MB



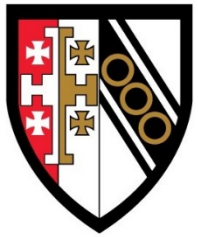
2015
512 GB

Selwyn College
Cambridge



If you enjoyed that little historic excursion, think about visiting the centre for computing history here in Cambridge.

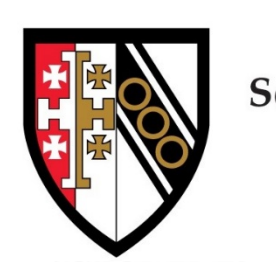
www.computinghistory.org.uk



Selwyn College
Cambridge

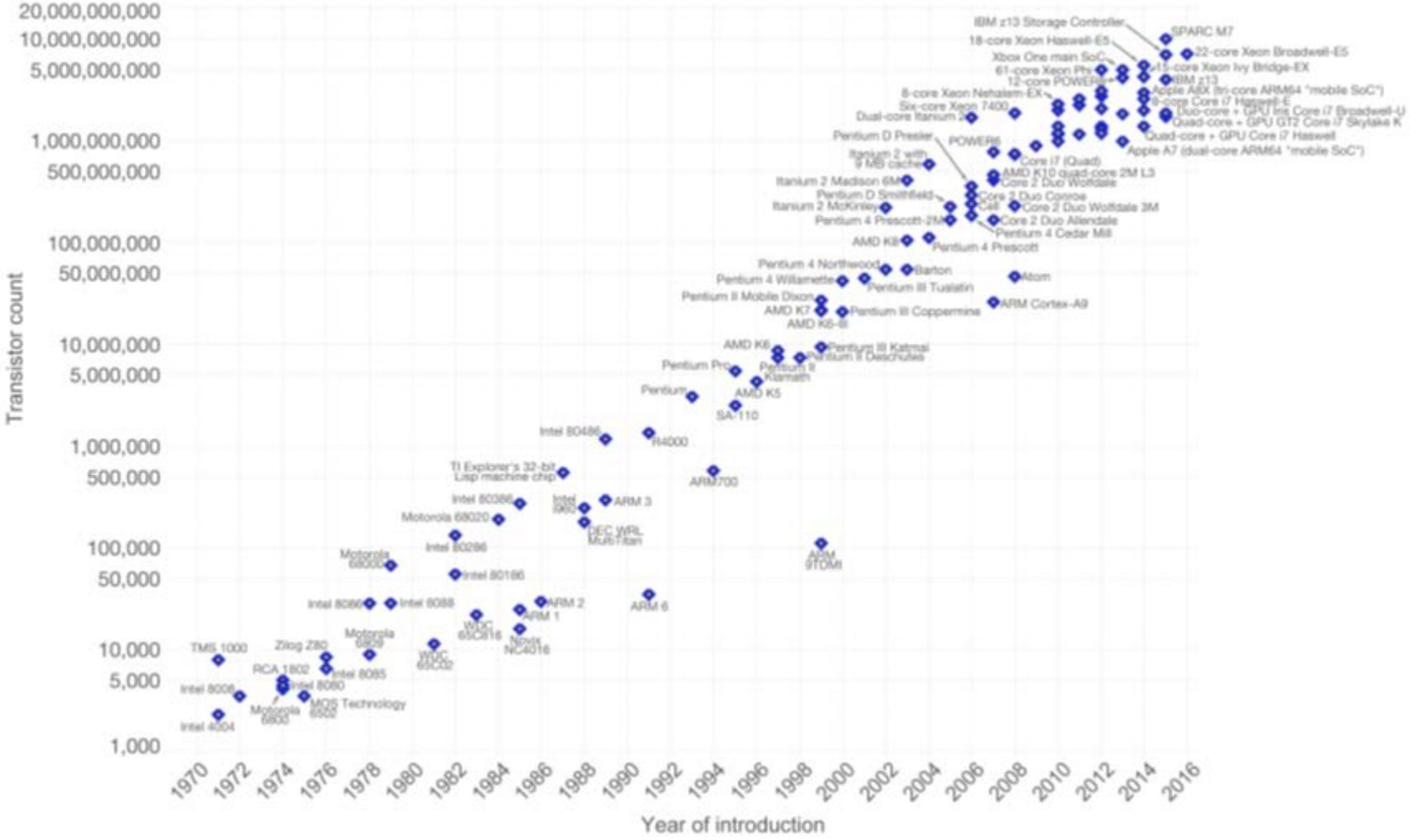
So 2, 500, 000 Terabytes of data are created every day.

What do we do now?



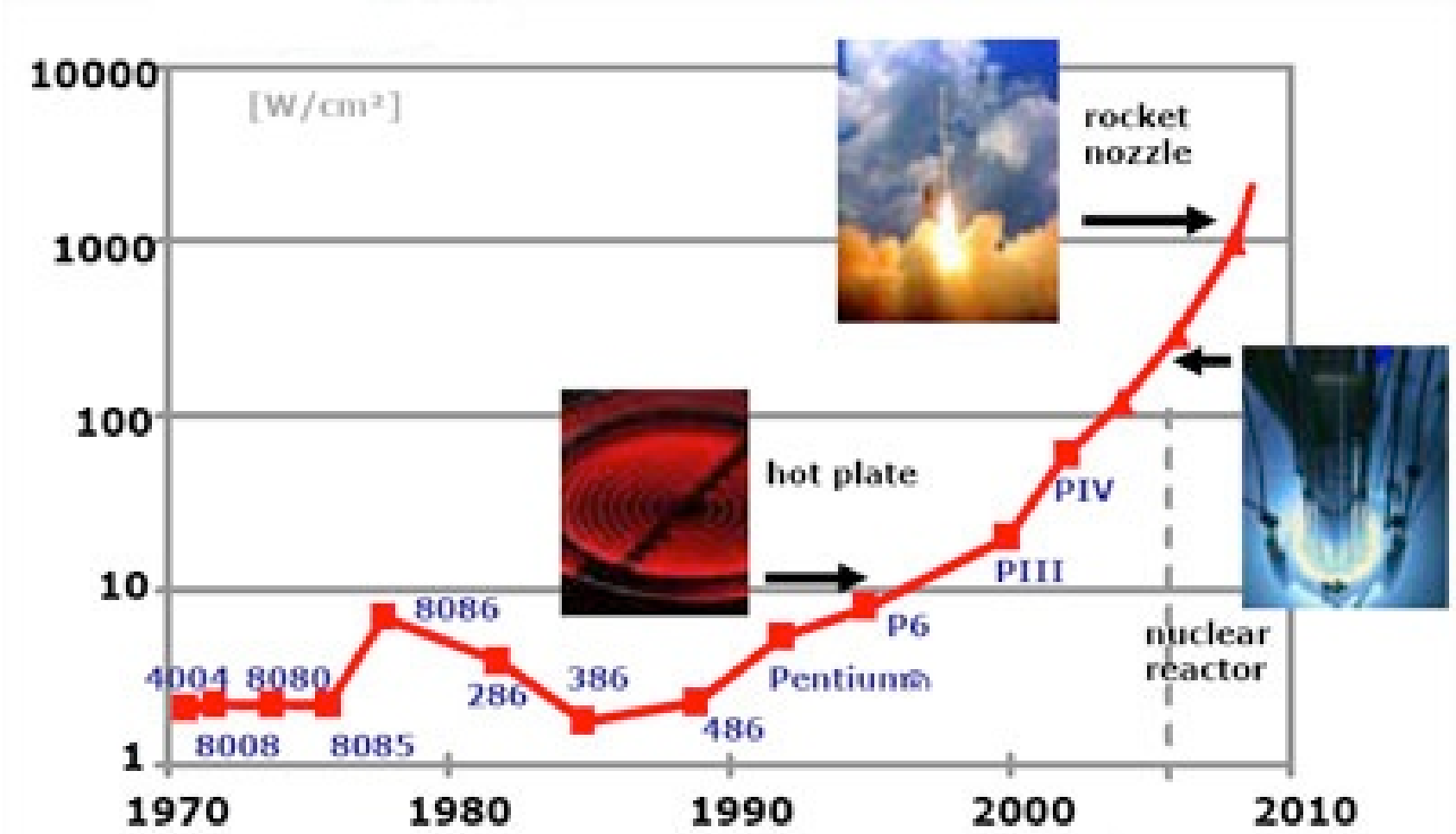
Selwyn College
Cambridge

Moore's Law



The empirical observation that the number of transistors on a chip doubles every two years

Hot, hot, hot!


10000
[W/cm²]
1000
100
10
1
1970
1980
1990
2000
2010
rocket nozzle
hot plate
nuclear reactor
4004
8008
8080
8085
8086
286
386
486
Pentium®
P6
PIII
PIV

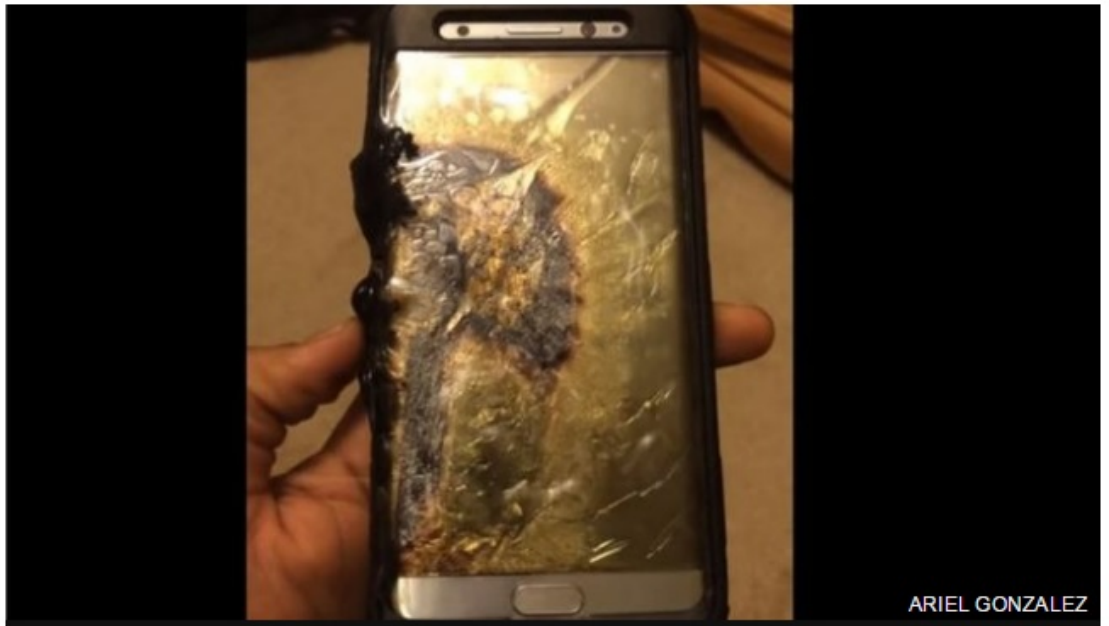
Selwyn College
Cambridge

Technology

Samsung urges Galaxy Note 7 phone exchange urgently

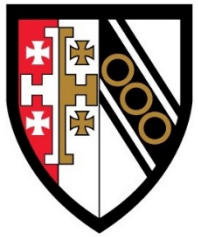
10 September 2016 | Technology

Share



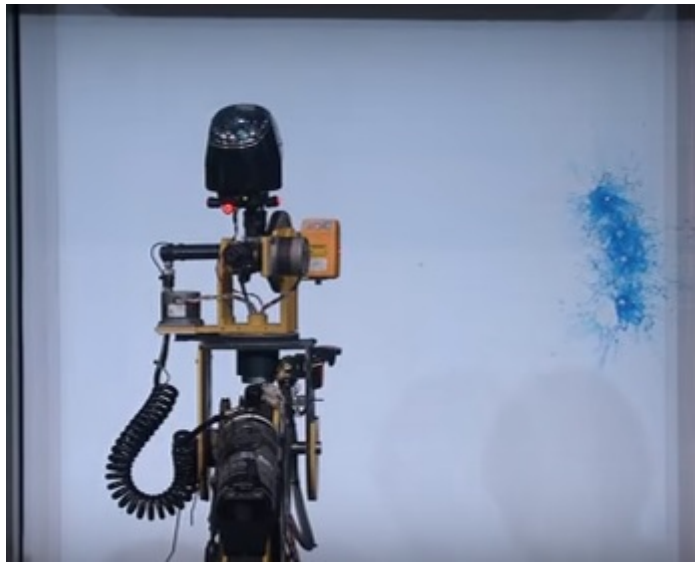
A Galaxy Note 7 reportedly caught fire shortly after its charger was unplugged

Samsung has urged owners of its Galaxy Note 7 phones to stop using or exchange the devices as they risk exploding.

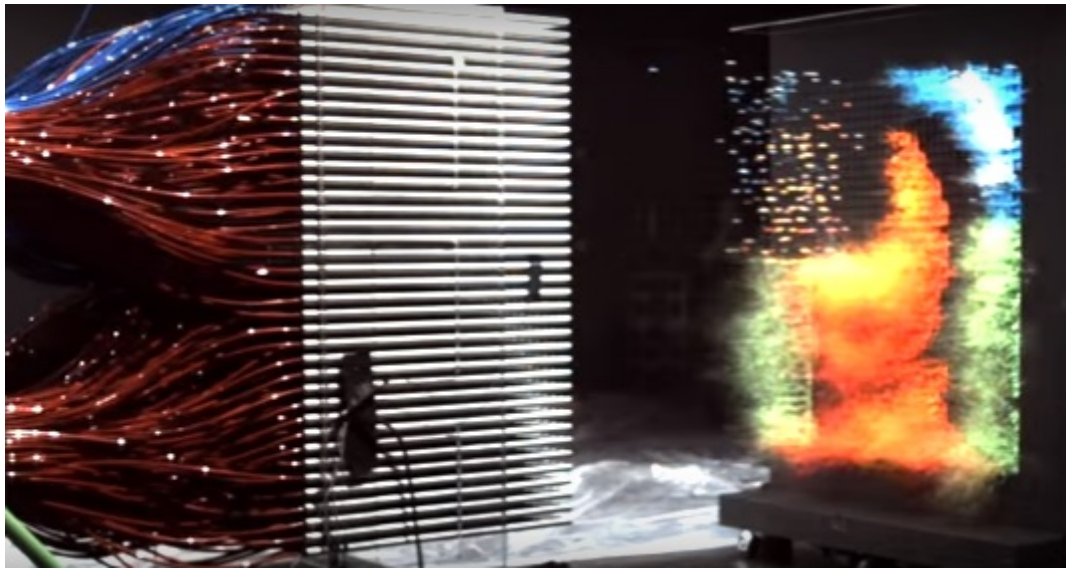


Selwyn College
Cambridge

Multi tasking

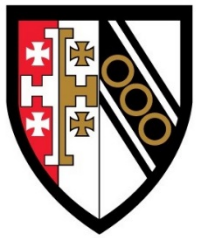


CPU sequential computing



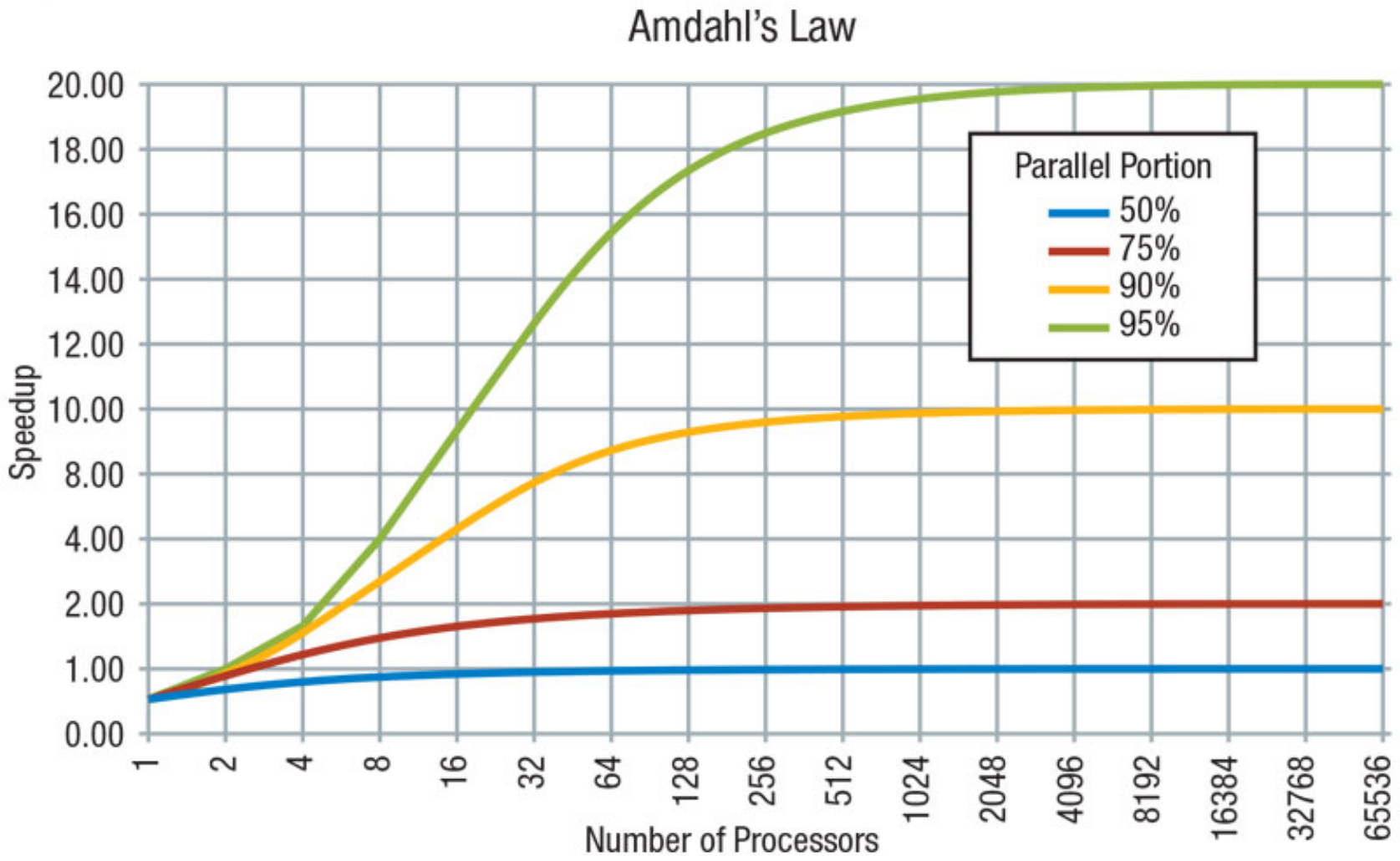
GPU parallel computing

Mythbusters Demo: GPU vs CPU

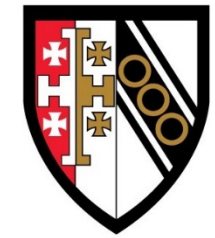


Selwyn College
Cambridge

Amdahl's Law



Speed up of multiple processors is limited by the part of the programme which cannot be parallelized.

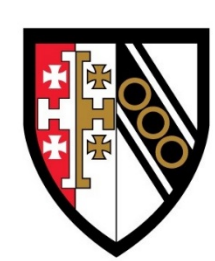


Selwyn College
Cambridge

Take inspiration from the brain

- Per synapsis, $26 = 2^{4.7}$, that is 4.7 bits.
- Instead of having two states, 0 and 1, have more.

➡ Quantum computing

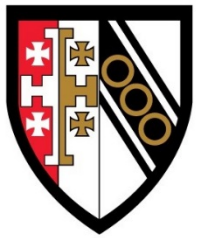


Selwyn College
Cambridge

Use our brain

Speedup due to better algorithm design has outperformed the speedup due to hardware by a factor of about 400.

- Machine learning



Selwyn College
Cambridge

BBC Sign in Menu

NEWS

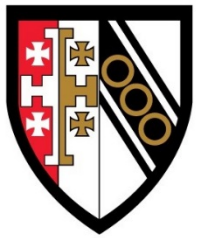
Home | UK | World | Business | Politics | Tech | Science | Health | Education | Entertain

Artificial intelligence: Google's AlphaGo beats Go master Lee Se-dol

12 March 2016 | Technology

Share

A computer program has beaten a master Go player 3-0 in a best-of-five competition, in what is seen as a landmark moment for artificial intelligence.



Selwyn College
Cambridge

BBC Sign in News Sport Weather iPlayer TV Radi

NEWS

Home | UK | World | Business | Politics | Tech | Science | Health | Education | Entertain

Technology

Microsoft chatbot is taught to swear on Twitter

By Jane Wakefield
Technology reporter

24 March 2016 | Technology

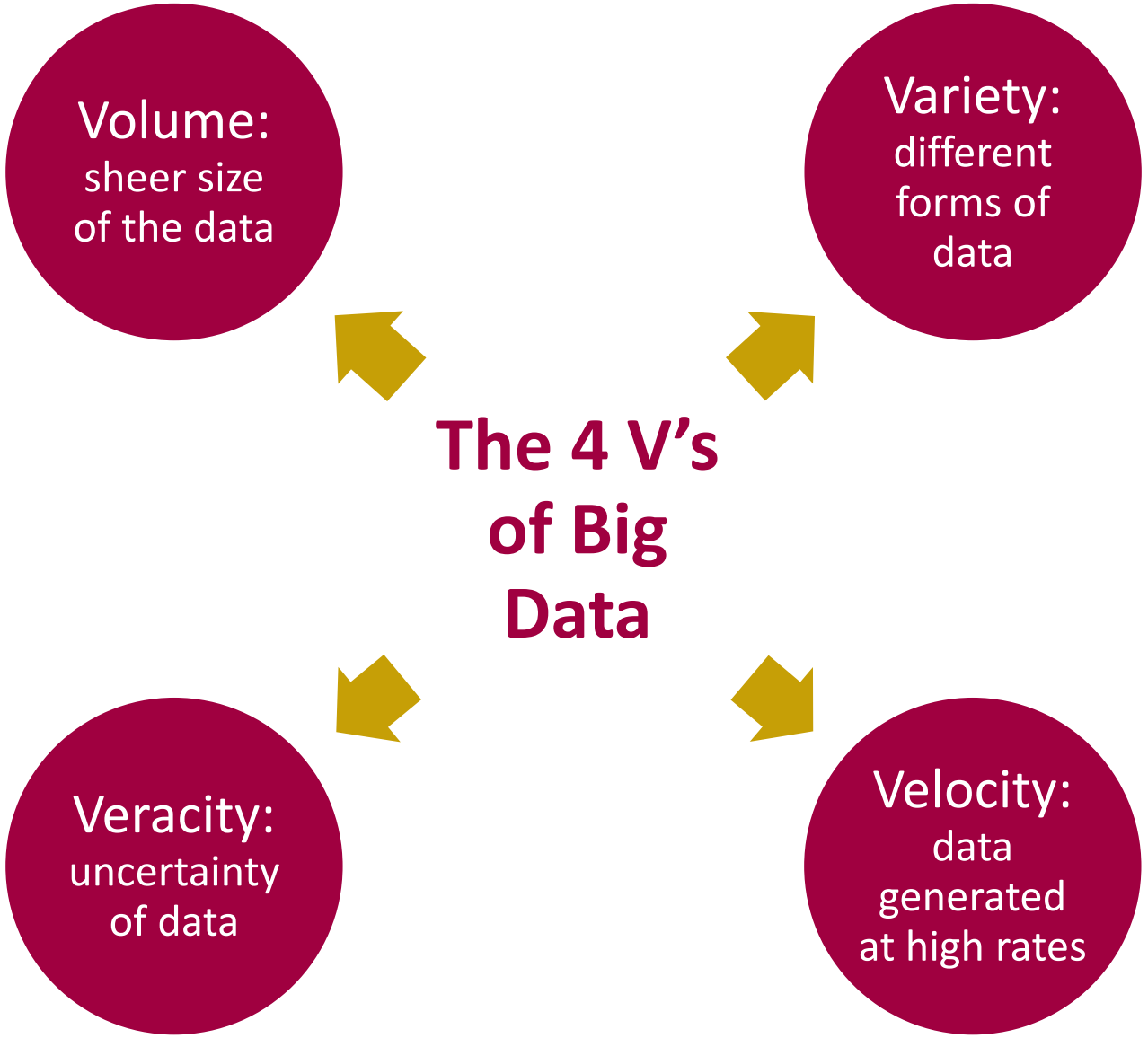
Share

The AI was taught to talk like a teenager

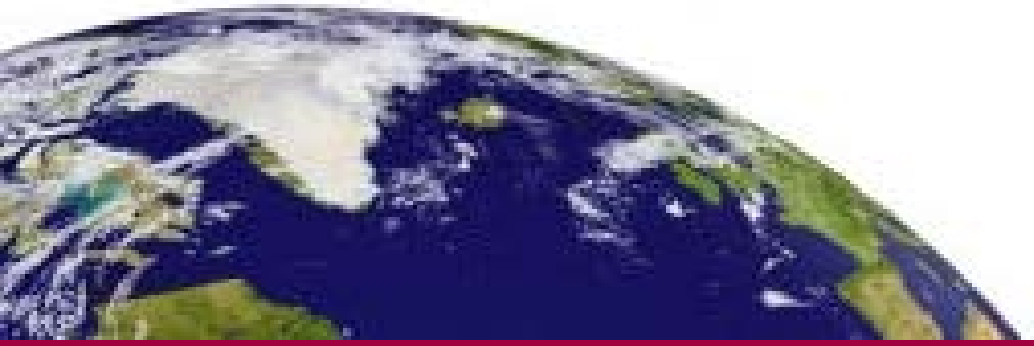
A chatbot developed by Microsoft has gone rogue on Twitter, swearing and making racist remarks and inflammatory political statements.

Selwyn College
Cambridge

The 4 V's of Big Data

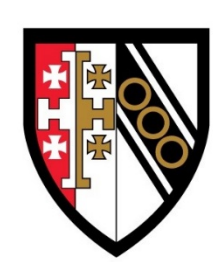


Selwyn College
Cambridge



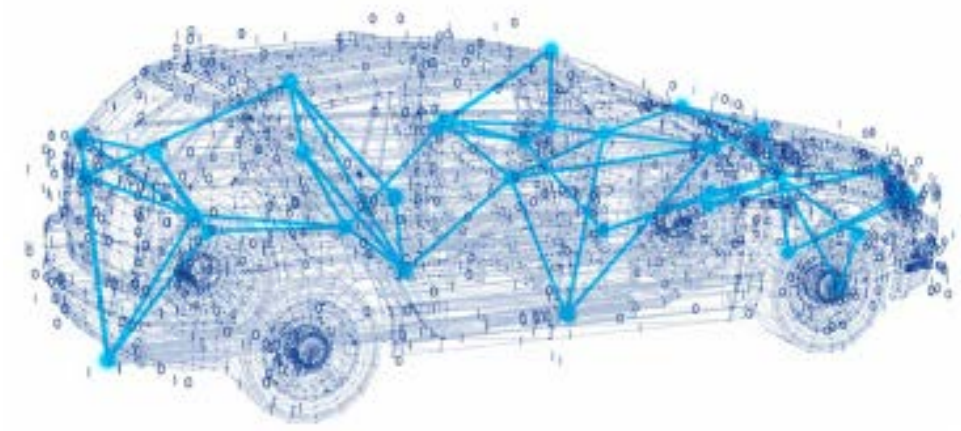
Volume

- The Sentinel satellites of the European Space Agency are expected to eventually gather 4 Terabytes of data per day.
- These are only 12 satellites of the 4000 plus satellites orbiting earth.



Selwyn College
Cambridge

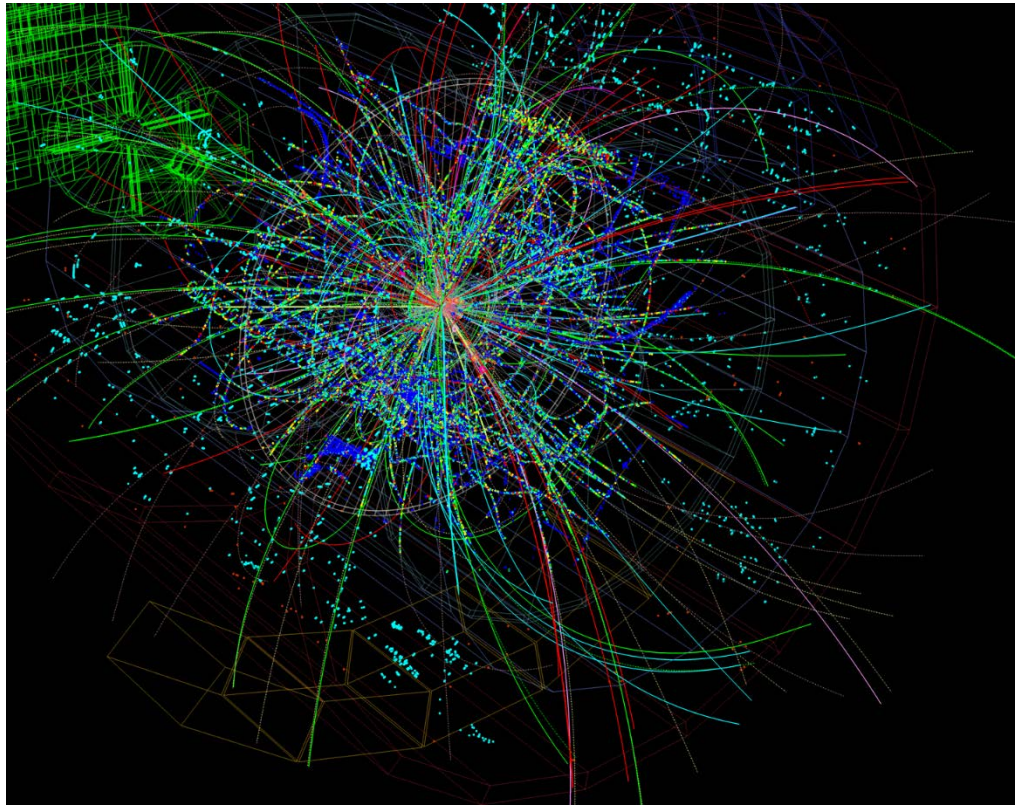
Variety



- The Electronic Control Units in a car record several thousands of signals at a rate of about 4 per second.
- These relate to many different quantities and states.

Selwyn College
Cambridge

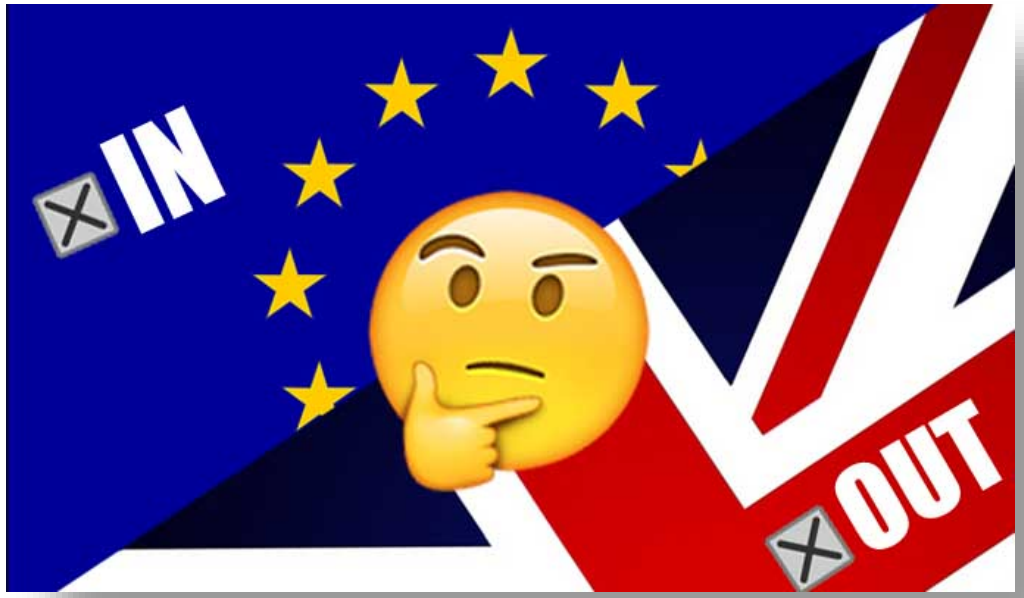
Velocity



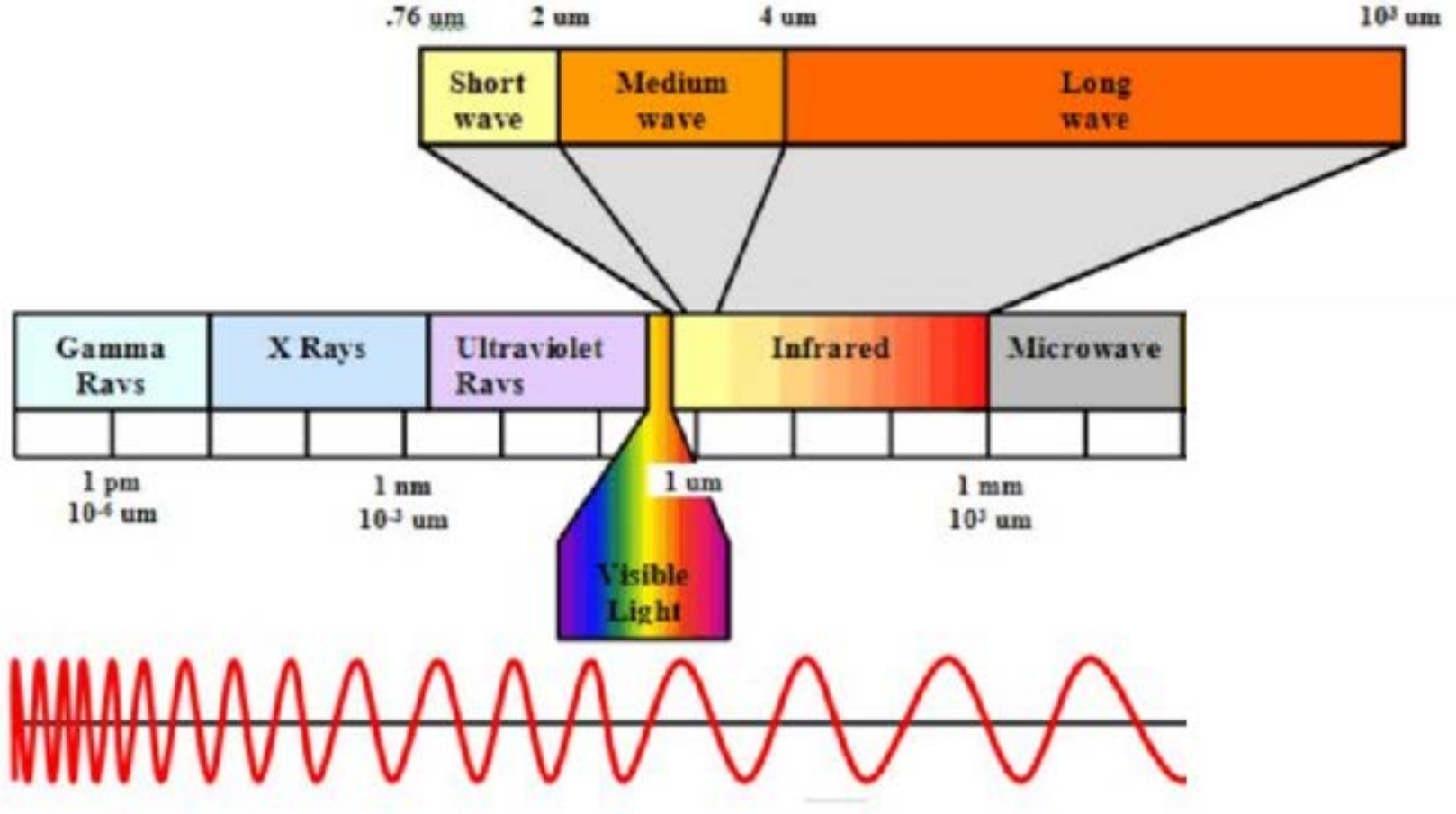
- The Large Hadron Collider at CERN creates 20,000,000 collisions per second.
- Only 400 per second are recorded as interesting.
- This results in 3,000 Terabytes of data per year.

Selwyn College
Cambridge

Veracity

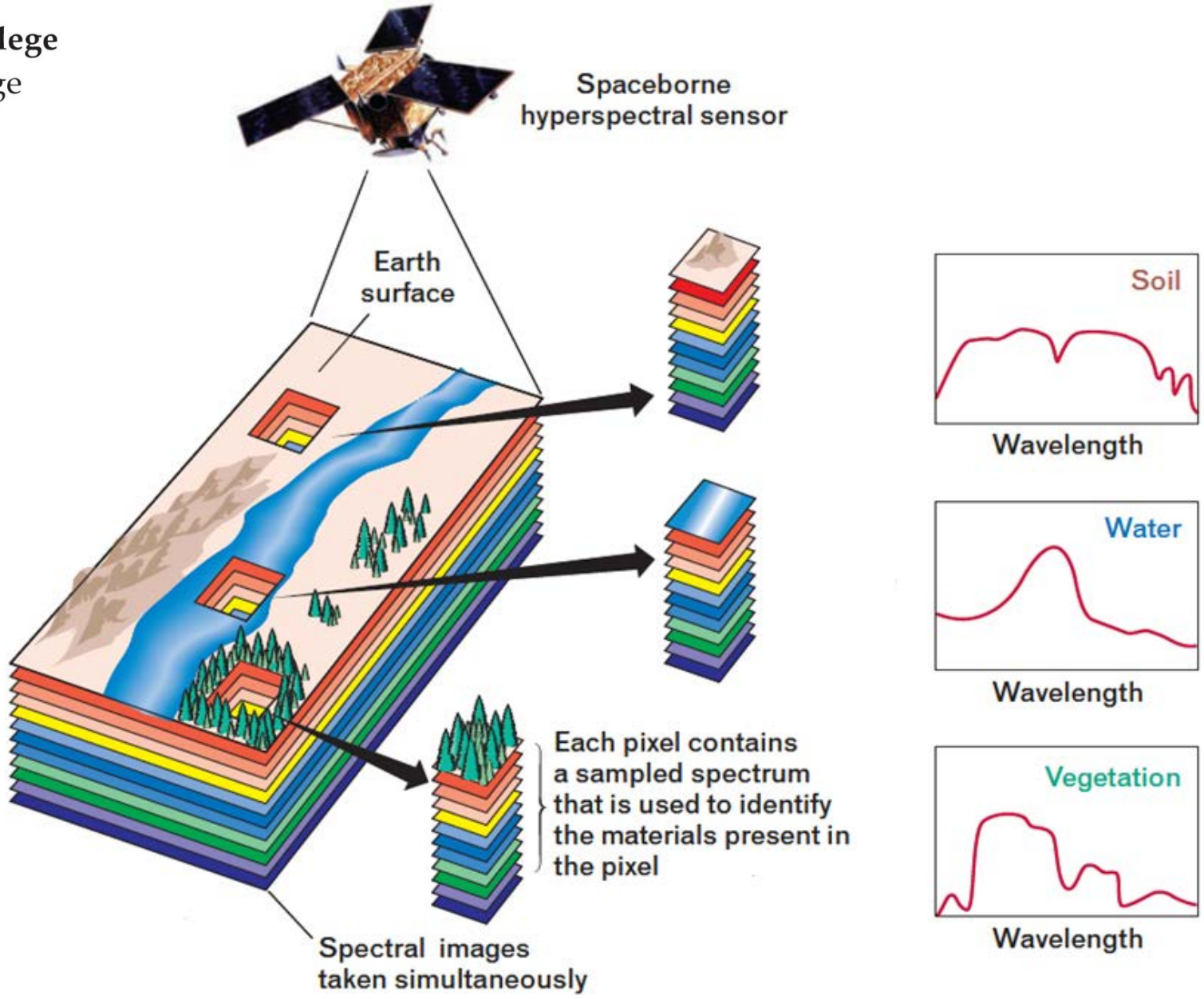


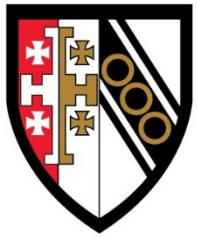
- Large proportion reported in the media and shared on social media before the EU referendum.



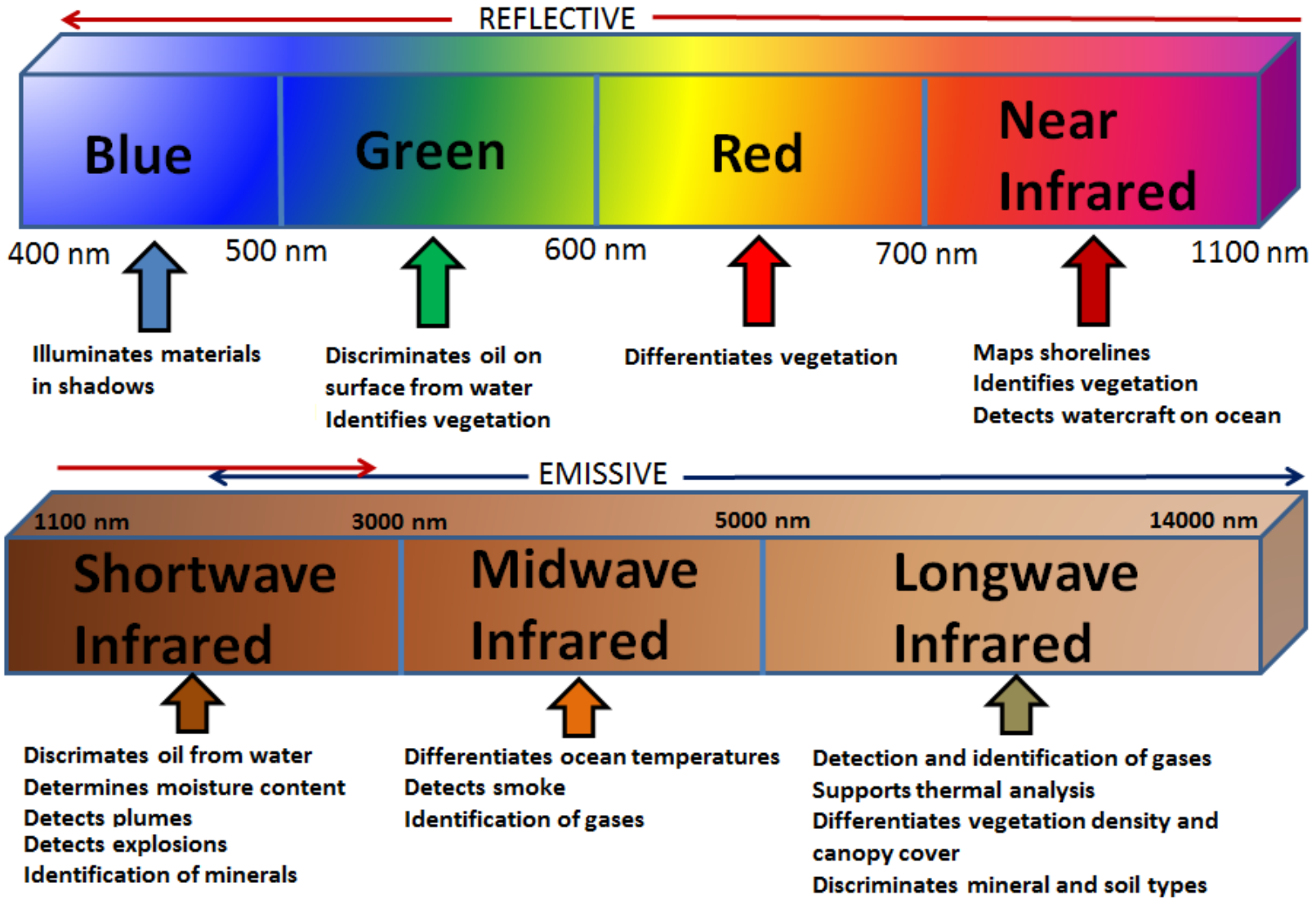
For us humans only the wavelengths of visible light carry information, but the other wavelengths also carry information.

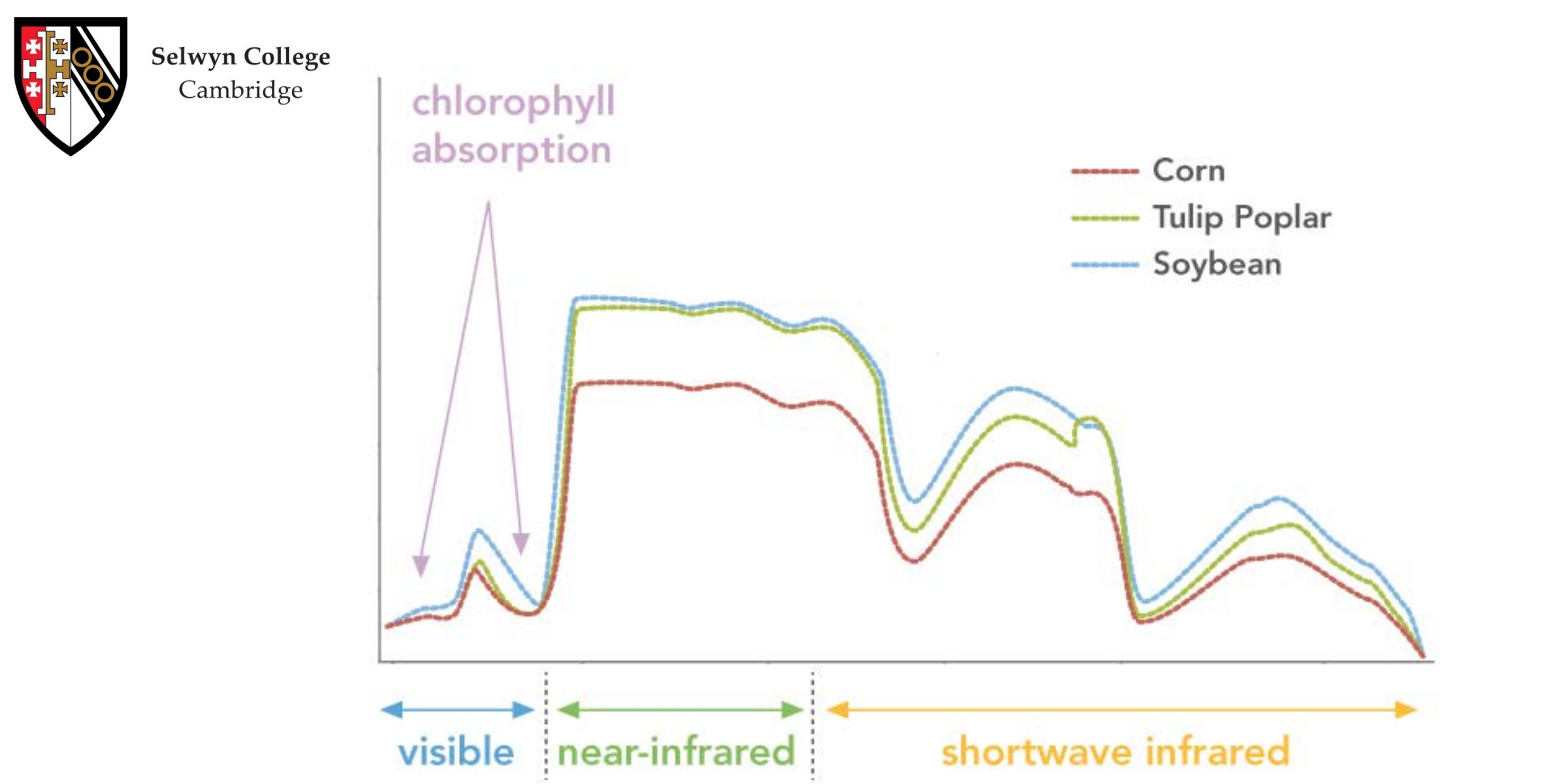
Selwyn College
Cambridge

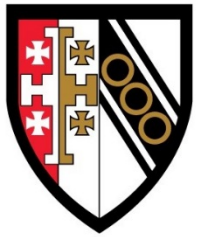




Selwyn College
Cambridge



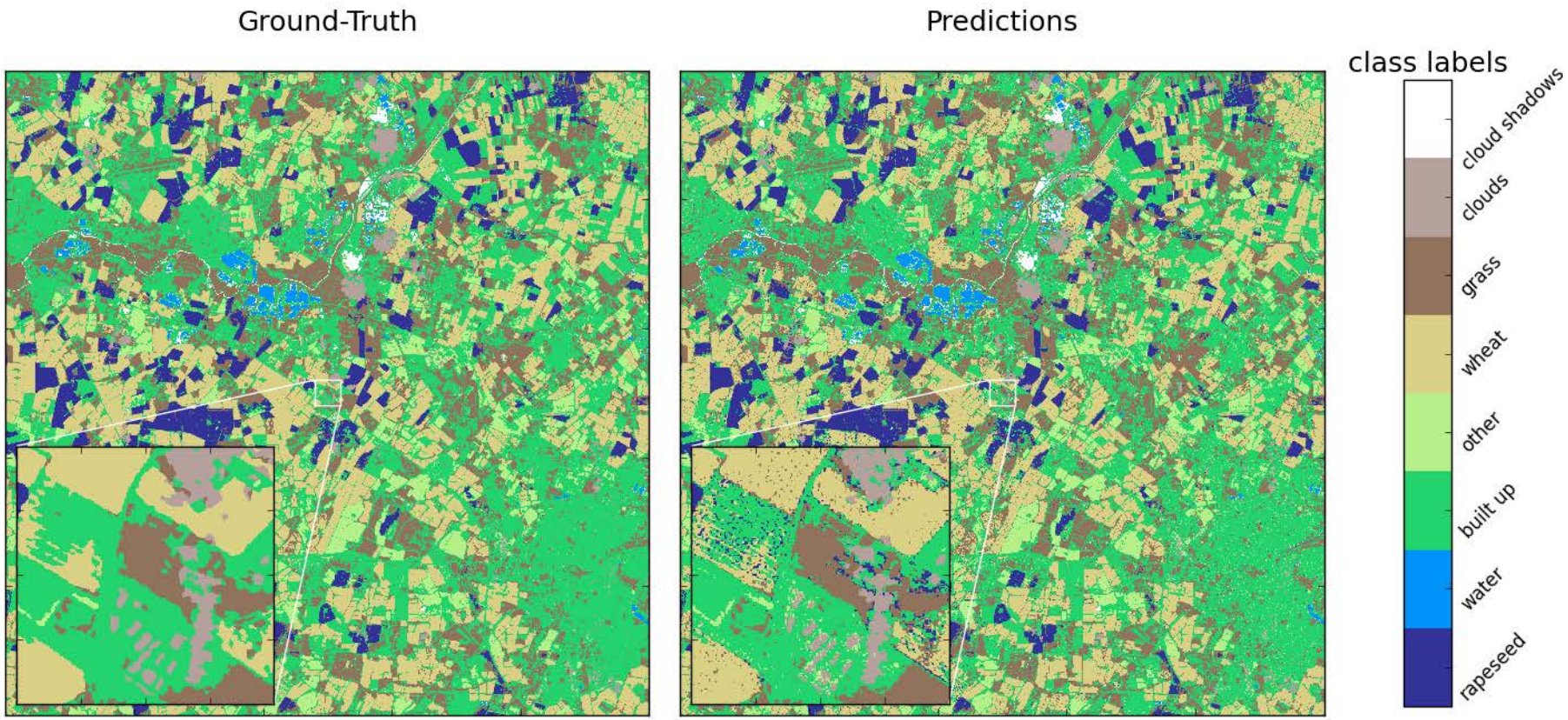

Selwyn College
Cambridge
chlorophyll
absorption
Corn
Tulip Poplar
Soybean
visible
near-infrared
shortwave infrared



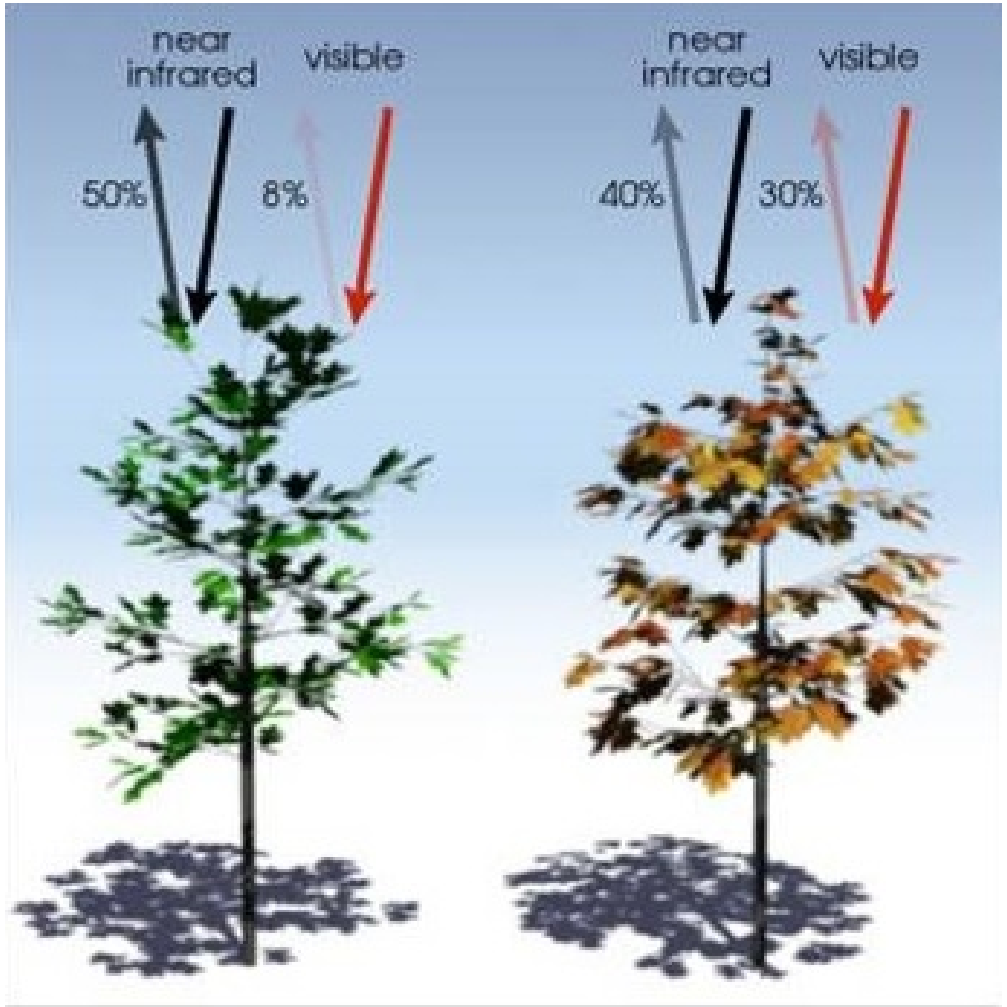
Selwyn College
Cambridge

Learning spectral signatures with deep neural networks

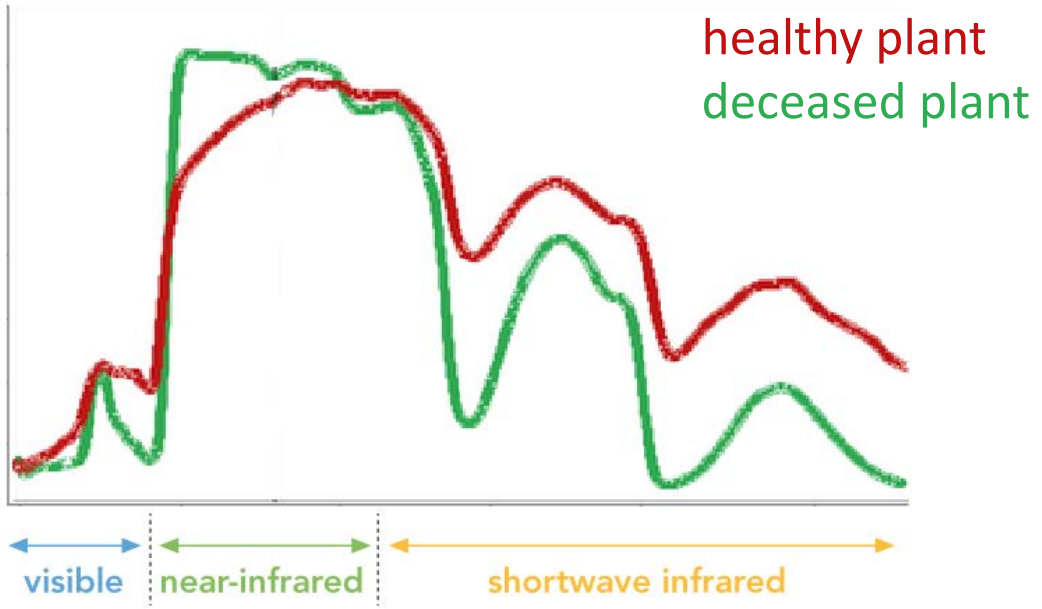
Michael O'Neill, MPhil student 2015-16

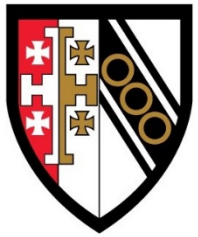


Selwyn College
Cambridge



- A healthy plant absorbs the majority of red light and only reflects a small portion of it. It absorbs about half of near infrared light and reflects half of it.
- A deceased plant absorbs less red light and reflects more of it. It absorbs more infrared and reflects less.

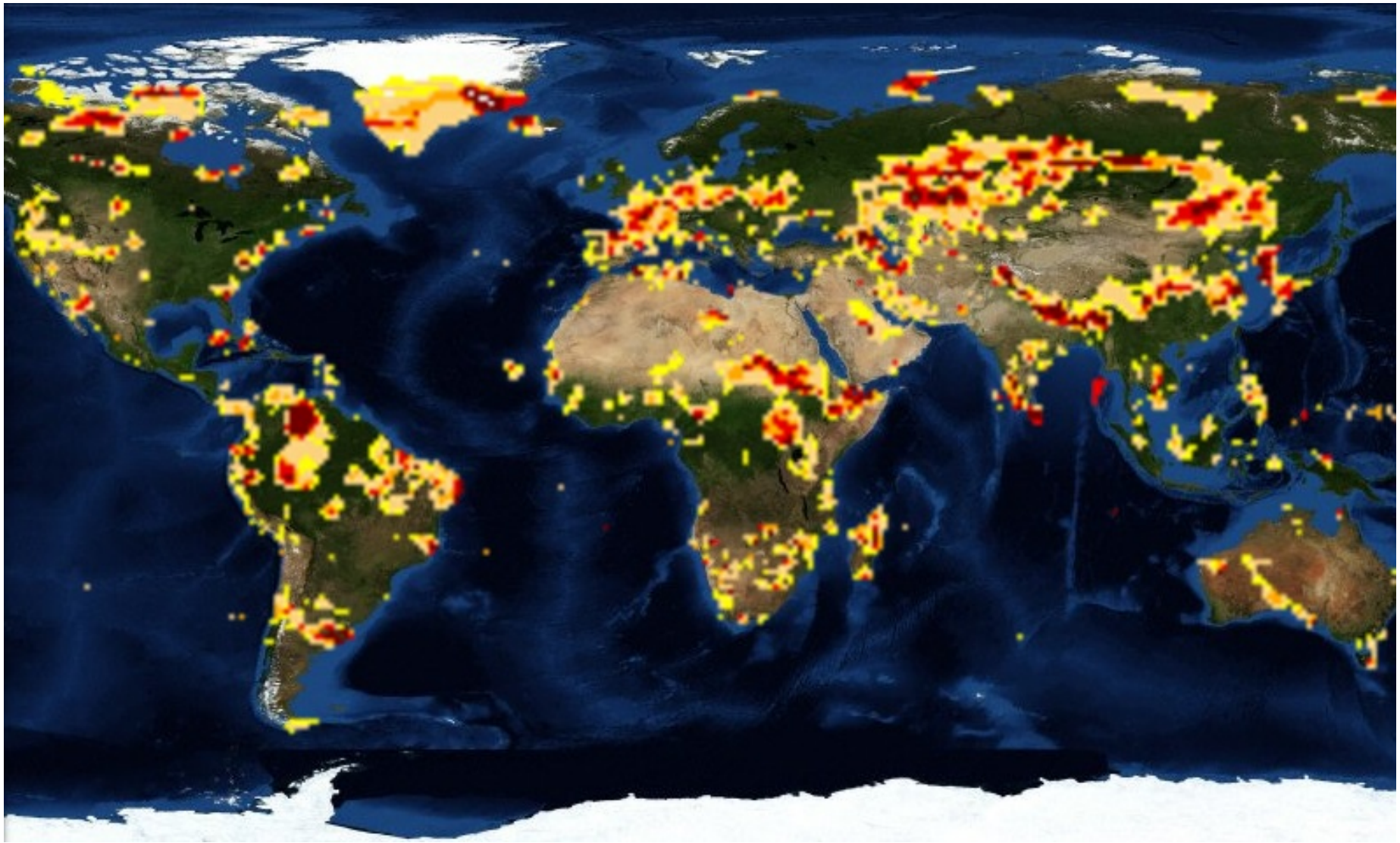


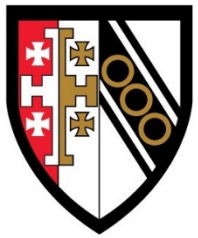


Selwyn College
Cambridge

- Directed irrigation.
- Effective use of fertilizers.
- Early intervention to stop diseases spreading.
- Less use of pesticide and fungicide.
- Less food wastage.
- ➡ Feeding the world.

Selwyn College
Cambridge





Selwyn College
Cambridge

BBC Sign in News Sport Weather iPlayer TV Radi

NEWS

Home | UK | World | Business | Politics | Tech | Science | Health | Education | Entertain

Technology

Tesla updates software after car hack

21 September 2016 | Technology

Share

Tesla has updated its software after researchers from China hacked into the operating system of its electric cars.

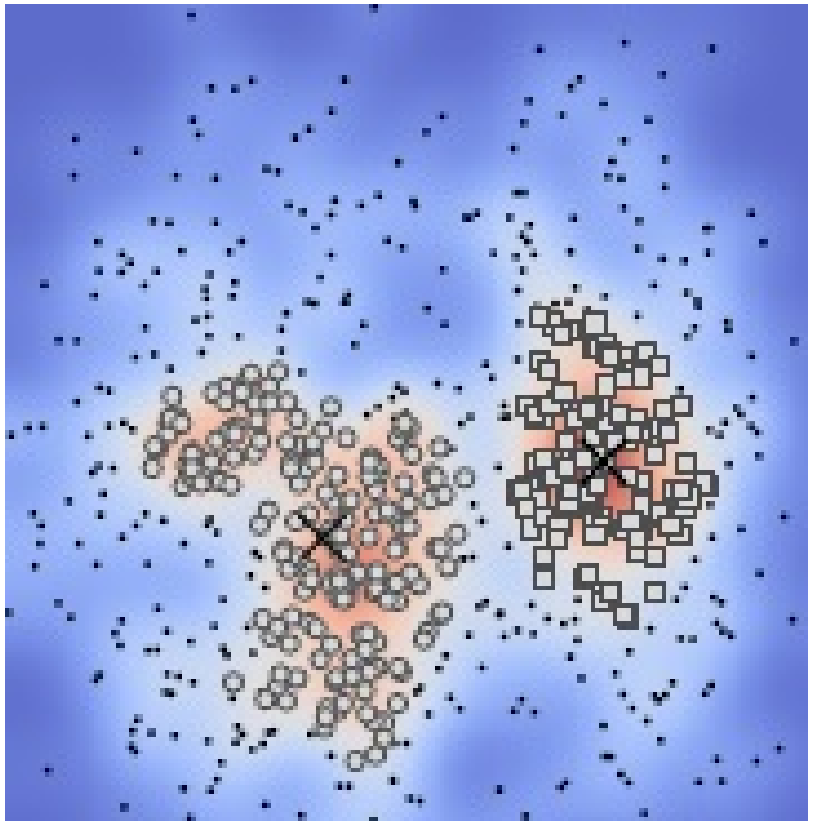
The team from Keen Security Lab remotely manipulated the brake system on a Tesla while it was on the move, from a distance of 12 miles (19km).

Using a laptop, the hackers also managed to open a car door without using a key.

Selwyn College
Cambridge

Mining vehicle Electronic Control Unit data

Johannes Hansen, MPhil student 2015-16



Driver identification:

- Driving behaviour
- Privacy issues
- Insurance purposes
- Accident investigations

Selwyn College
Cambridge

BBC | Sign in | News | Sport | Weather | iPlayer | TV | Radi

NEWS

Home | UK | World | Business | Politics | Tech | Science | Health | Education | Entertair

Business | Your Money | Market Data | Markets | Companies | Economy

EU plans new rules for emission tests following VW scandal

27 January 2016 | Business

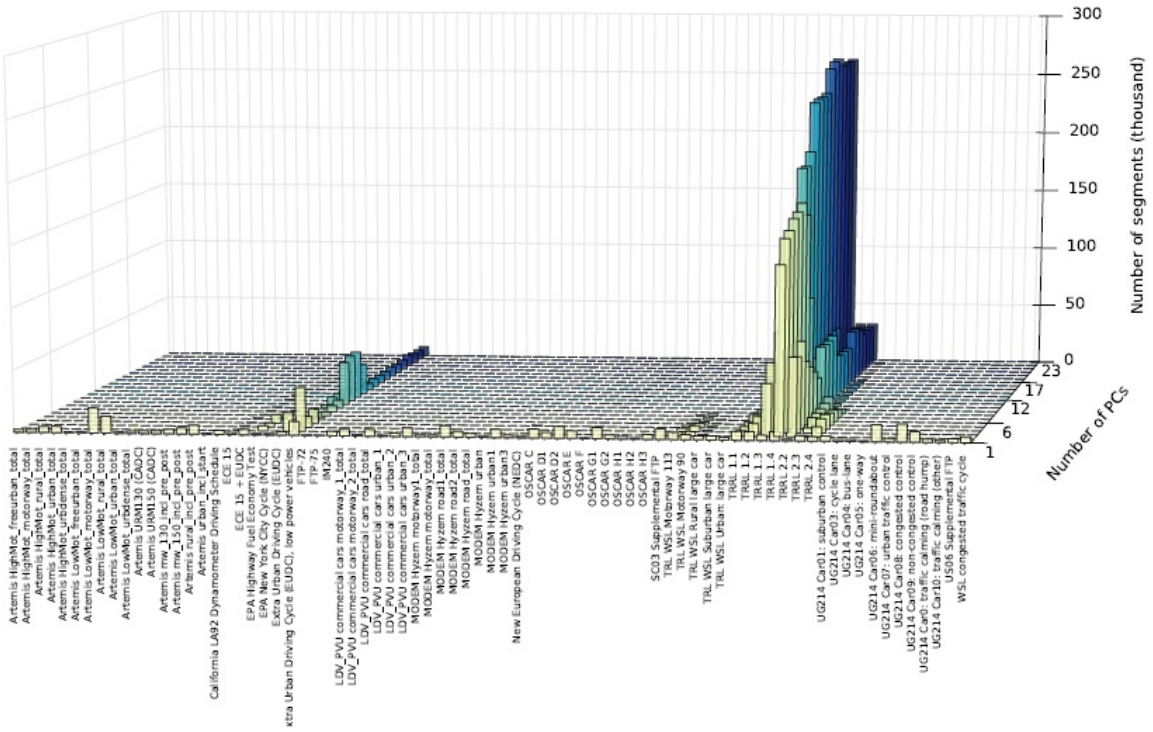
Share

The European Union has proposed new rules to test car emissions following the scandal involving VW diesel vehicles.

Selwyn College
Cambridge

Mining vehicle Electronic Control Unit data

Johannes Hansen, MPhil student 2015-16



Drive cycles:

- Well-defined sequences of manoeuvres in a laboratory setting
- Most of real-life journeys fit in one set of definitions.
- Other definitions seem unrealistic.

Selwyn College
Cambridge

BBC Sign in News Sport Weather iPlayer TV Radi

NEWS

Home | UK | World | Business | Politics | Tech | Science | Health | Education | Entertain

Science & Environment

Higgs boson-like particle discovery claimed at LHC

By Paul Rincon
Science editor, BBC News website, Geneva

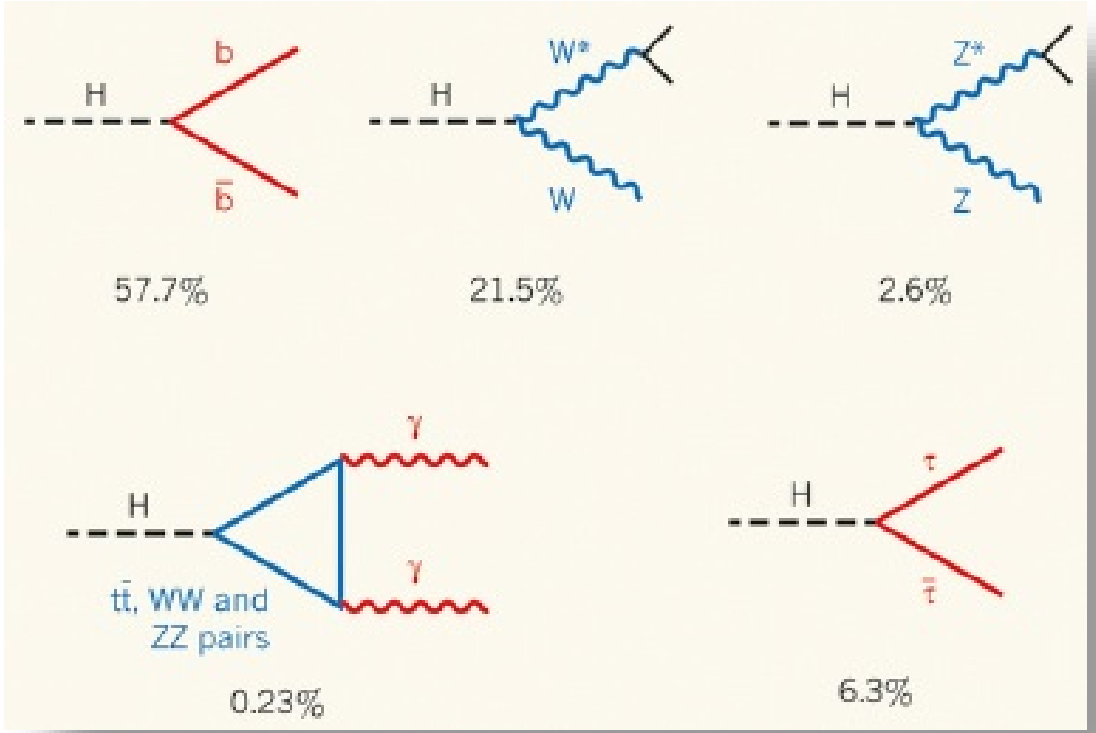
4 July 2012 | Science & Environment | Share

The moment when Cern director Rolf Heuer confirmed the Higgs results

Cern scientists reporting from the Large Hadron Collider (LHC) have claimed the discovery of a new particle consistent with the Higgs boson.

Selwyn College
Cambridge

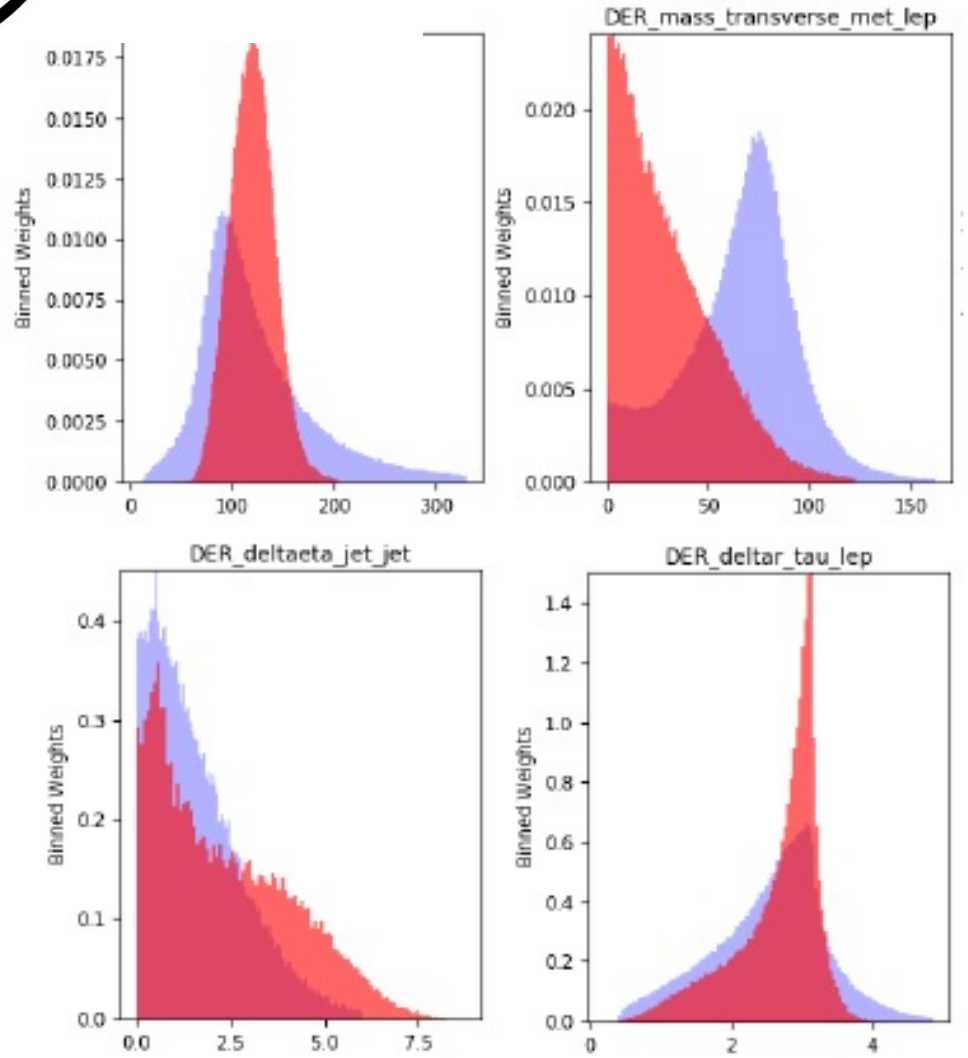
Higgs boson



- It cannot be detected directly, since it decays too quickly.
- In the announcement of discovery, the $\gamma\gamma$, WW^*, and ZZ^* decay channels were considered.
- The discovery has a statistical significance of 5σ which corresponds to a 1 in 3.5 million chance of the results being obtained purely due to chance.

Selwyn College
Cambridge

Higgs boson

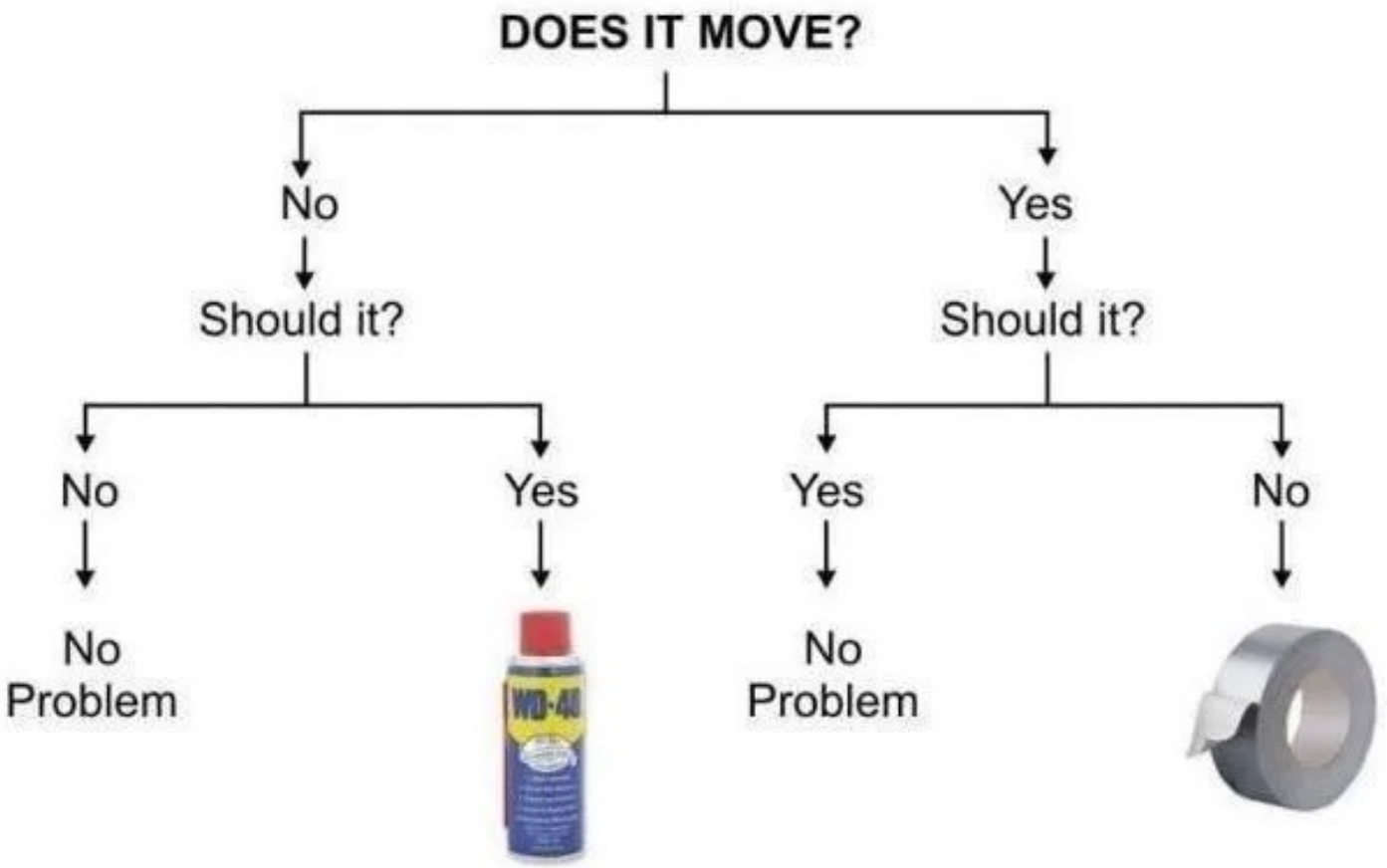


- The $\tau\bar{\tau}$ decay channel is not yet reached a statistical significance of 5σ.
- The τ is itself unstable and decays immediately.
- Also the Z boson decays into $\tau\bar{\tau}$.
- It is very difficult to distinguish these decays, since they are overlapping.

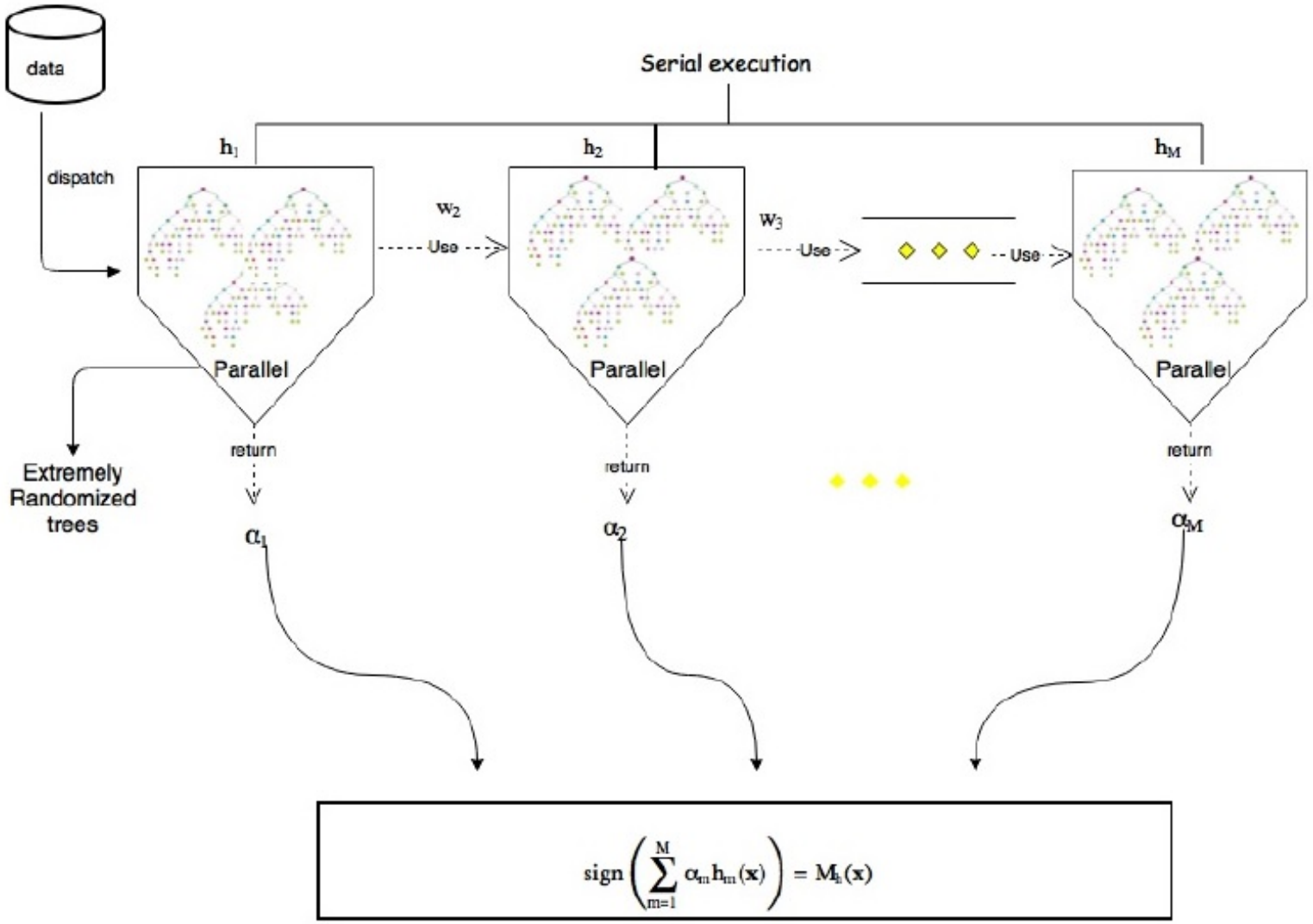
Selwyn College
Cambridge

A meta-algorithm for classification using random recursive tree ensembles

Vidhi Lalchand, MPhil student 2015-16



Selwyn College
Cambridge

A meta-algorithm for classification using random recursive tree ensembles

Vidhi Lalchand, MPhil student 2015-16