Trigger Happy

Or

How & why we analyse a really, very large amount of data really, very fast

Andy Rose, Imperial College, London

awr01@imperial.ac.uk

www.hep.ph.ic.ac.uk/~awr01

Science: The basics

- Science is the art of knowing what to record, and when

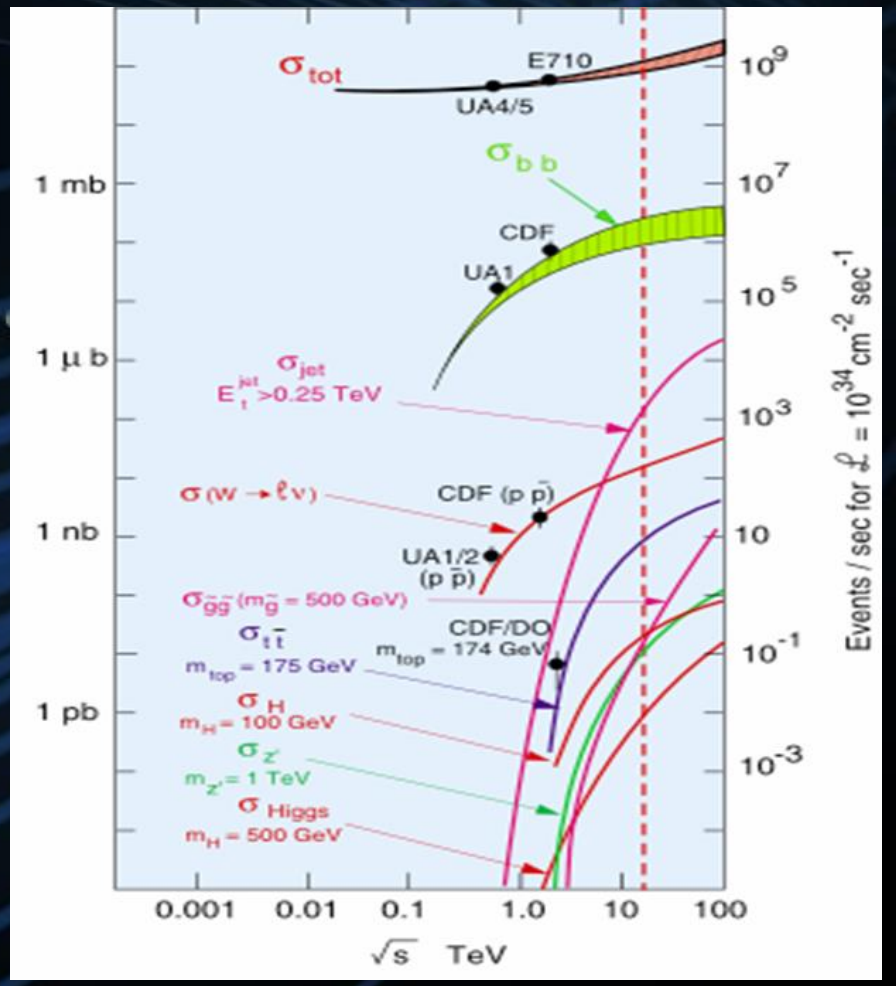
Science: The basics

- Science is the art of knowing what to record, and when

Or being cynical, it's being in the right place at the right time

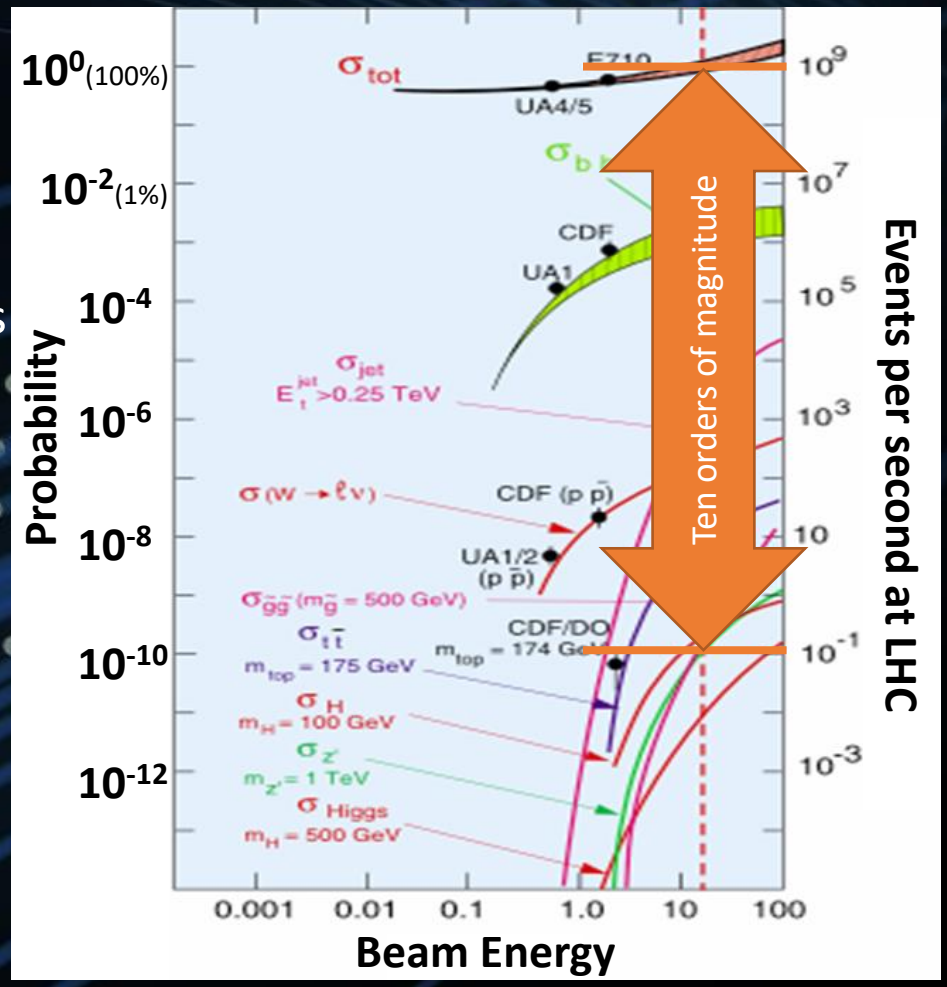
Science: The basics

- Science is the art of knowing what to record, and when
- With CMS & ATLAS in “discovery mode”, we care about the Higgs Boson or rarer
  - Higgs Boson production is ten orders of magnitude below the total interaction rate



Science: The basics

- Science is the art of knowing what to record, and when
- With CMS & ATLAS in “discovery mode”, we care about the Higgs Boson or rarer
  - Higgs Boson production is ten orders of magnitude below the total interaction rate



Science: The basics

- Science is the art of knowing what to record, and when
- With CMS & ATLAS in “discovery mode”, we care about the Higgs Boson or rarer
  - Higgs Boson production is ten orders of magnitude below the total interaction rate
  - That is a needle in a haystack the same mass as the Empire State Building

Science: The basics

- Science is the art of knowing what to record, and when
- With CMS & ATLAS in “discovery mode”, we care about the Higgs Boson or rarer
  - Higgs Boson production is ten orders of magnitude below the total interaction rate
  - That is a needle in a haystack the same mass as the Empire State Building
  - Or if you collided protons once per second, that is one event every 316 years, 324 days (5 hours, 6 minutes and 17 seconds give or take)

Science: The basics

- Science is the art of knowing what to record, and when
- With CMS & ATLAS in “discovery mode”, we care about the Higgs Boson or rarer
  - Higgs Boson production is ten orders of magnitude below the total interaction rate
  - That is a needle in a haystack the same mass as the Empire State Building
  - Or if you collided protons once per second, that is one event every 316 years, 324 days (5 hours, 6 minutes and 17 seconds give or take)

And one Higgs boson is pretty much

USELESS

We need loads of them to be able to study them…

This is why the LHC collides…

~50 protons at a time

×

40 million times a second

This is why the LHC collides…

~50 protons at a time

×

40 million times a second

---

1 Higgs boson every ~5 seconds

This is why the LHC collides…

~50 protons at a time

×

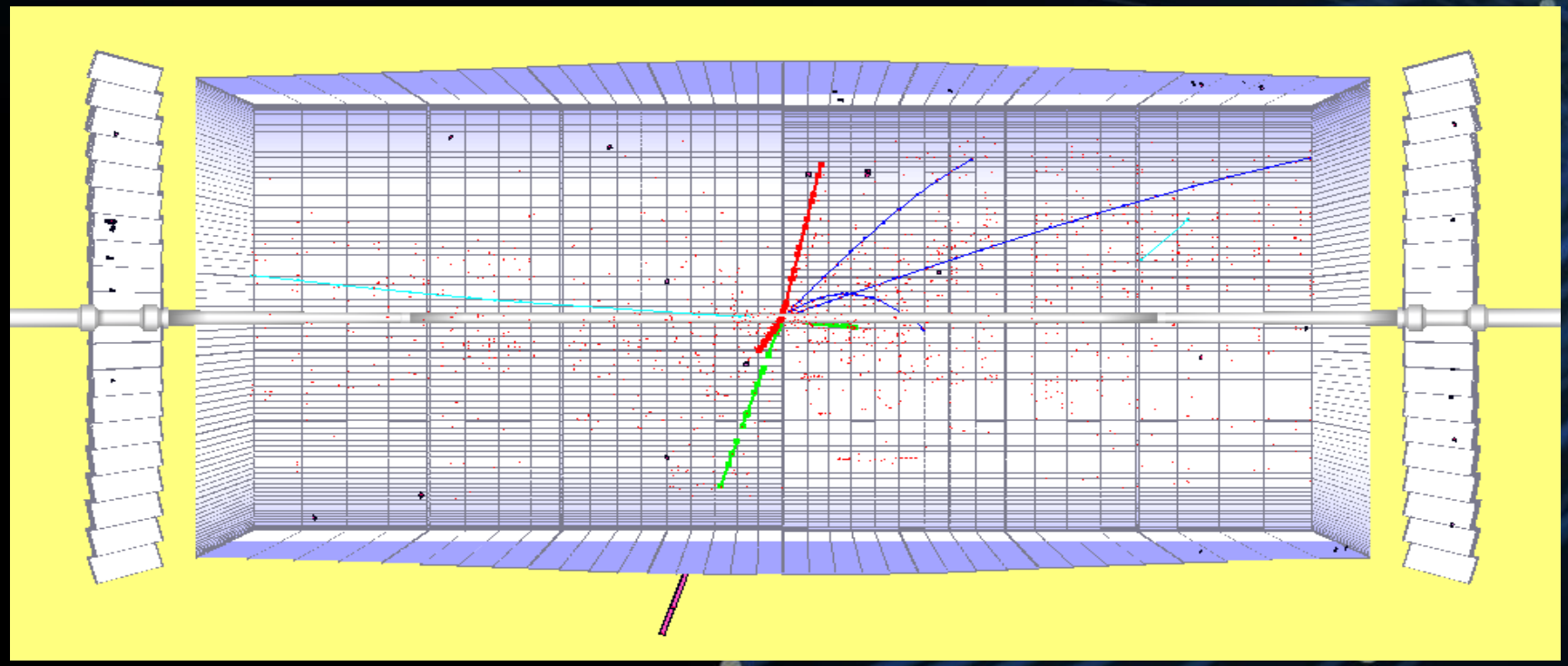
40 million times a second

---

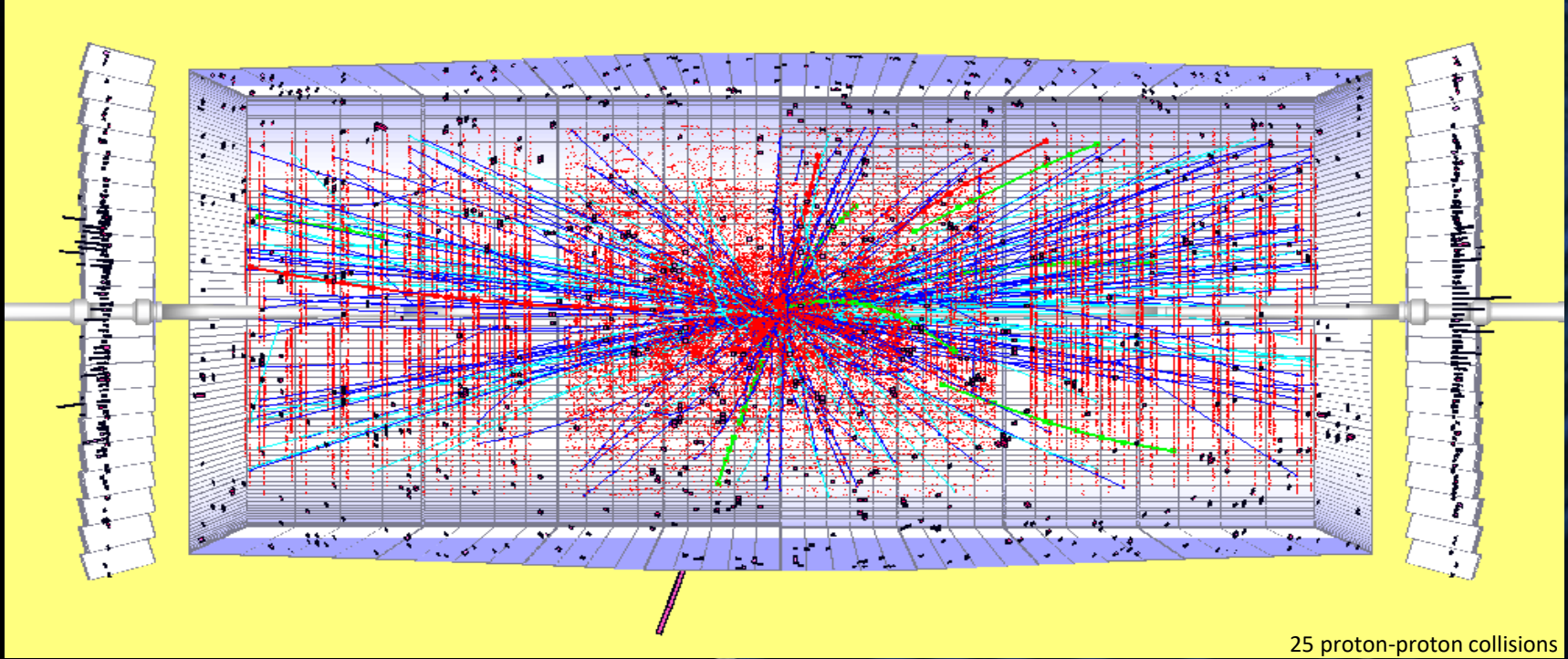
1 Higgs boson every ~5 seconds

But this causes problems

One proton-proton collision

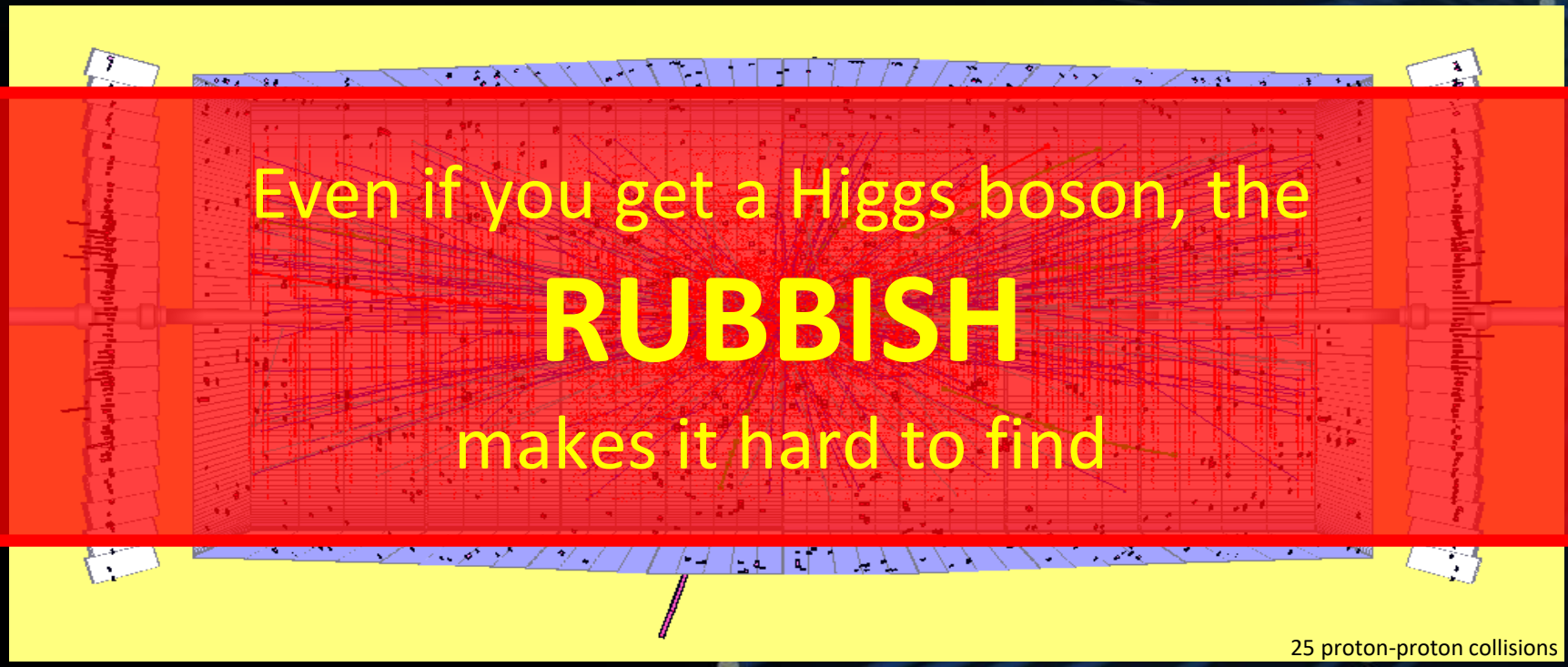


Many proton-proton collisions



25 proton-proton collisions

Many proton-proton collisions



25 proton-proton collisions

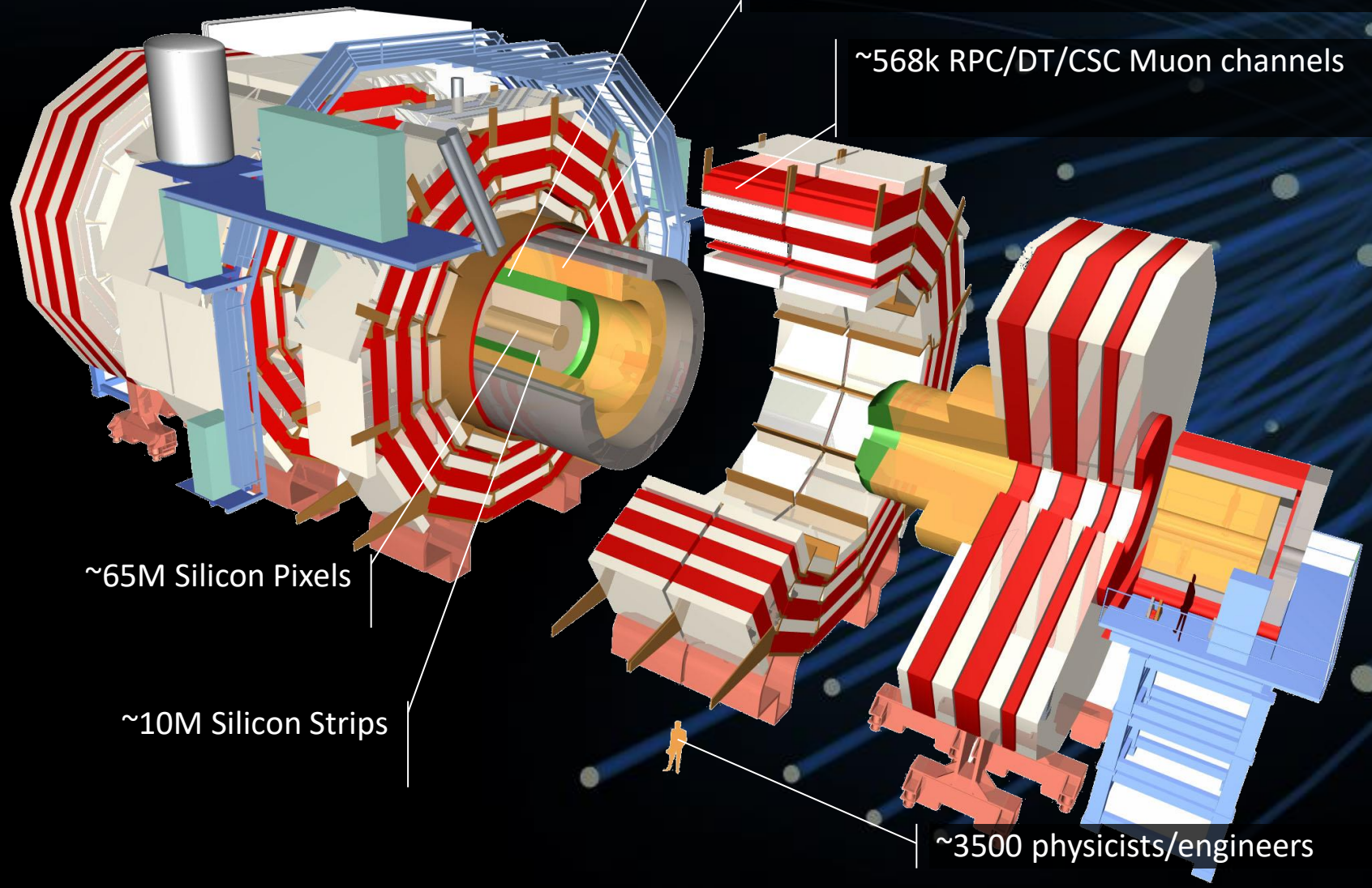
This is why the LHC collides…

$$\frac{\text{~50 protons at a time} \times \text{40 million times a second}}{\text{1 Higgs boson every ~5 seconds}}$$

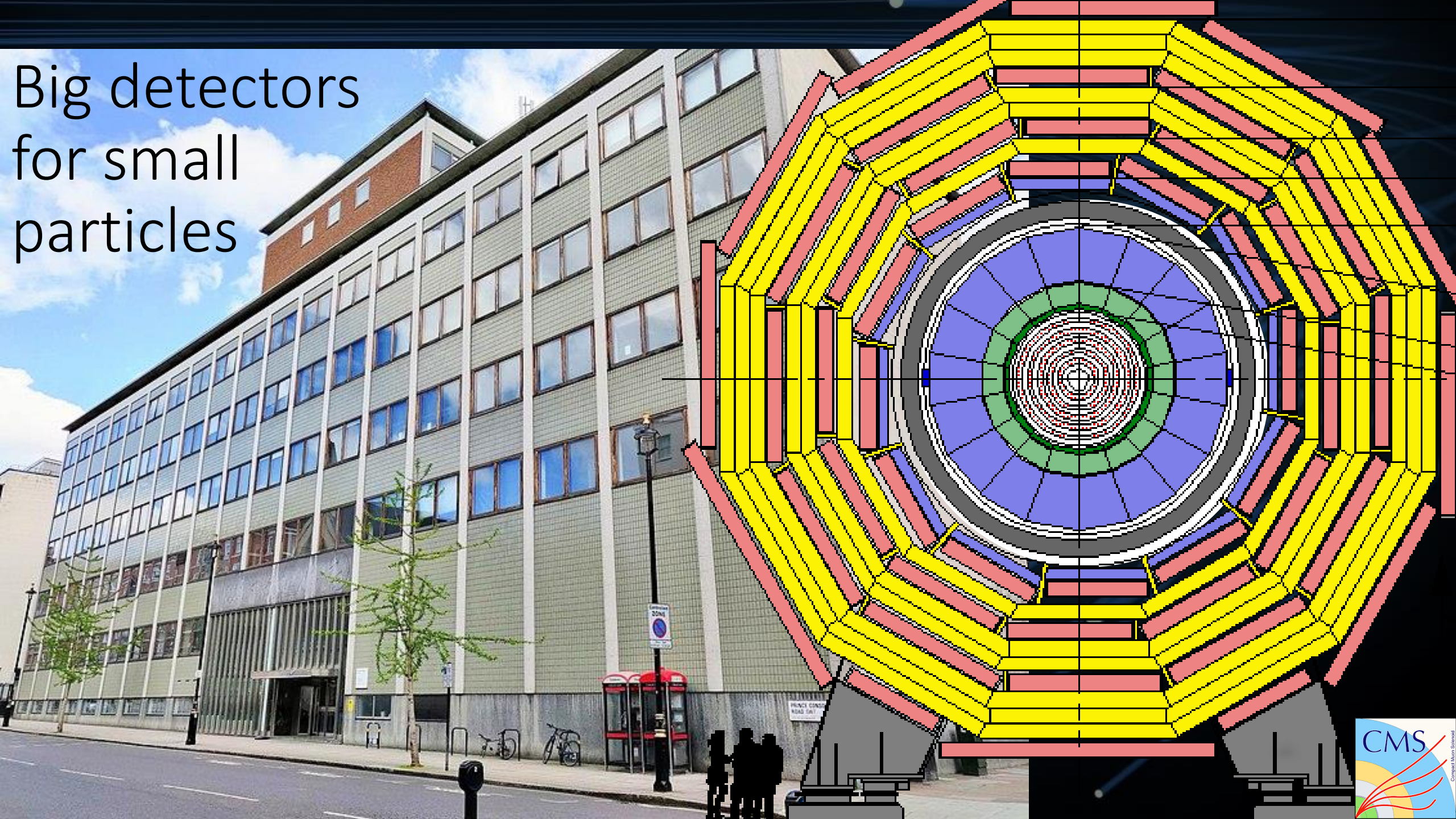
But this causes problems

This causes problems too

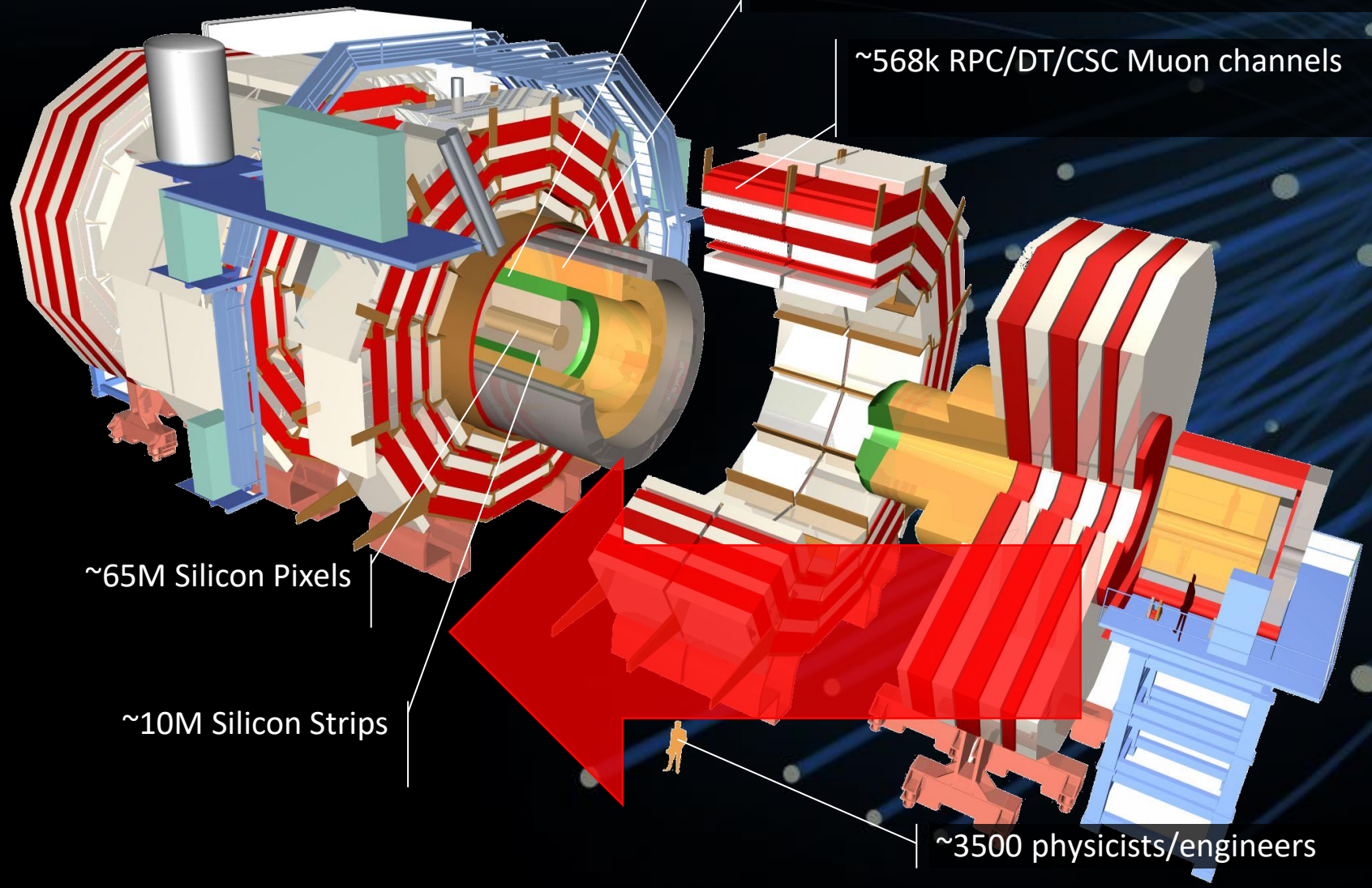
The CMS detector



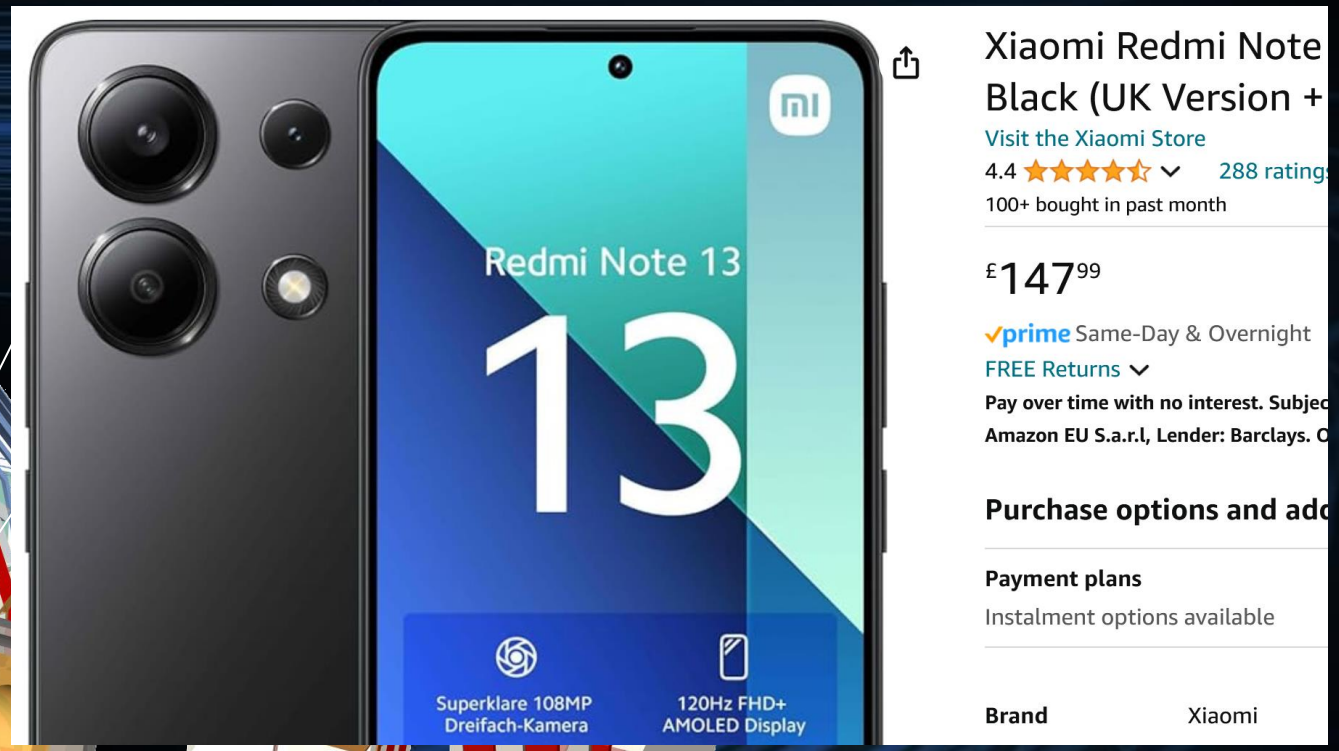
Big detectors for small particles

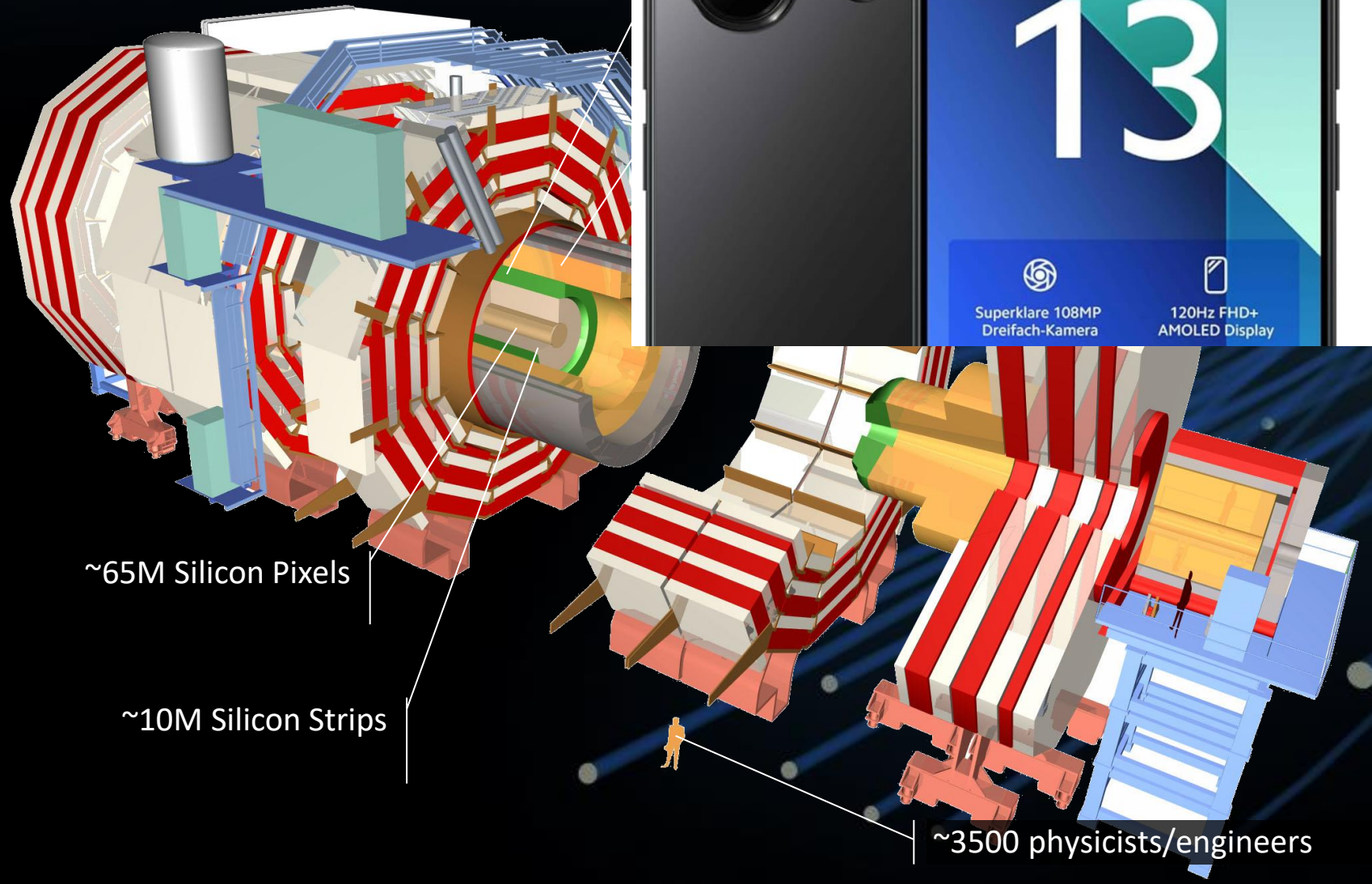


The CMS detector



The CMS detector



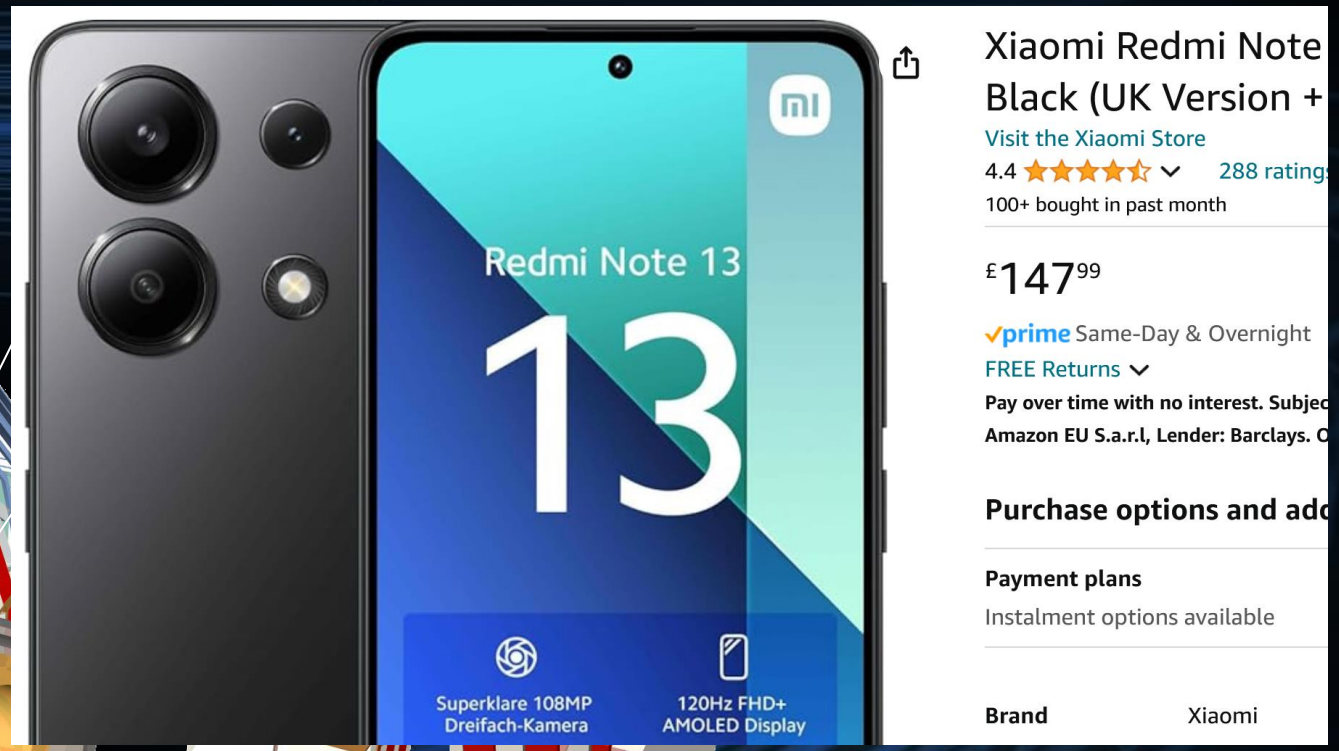


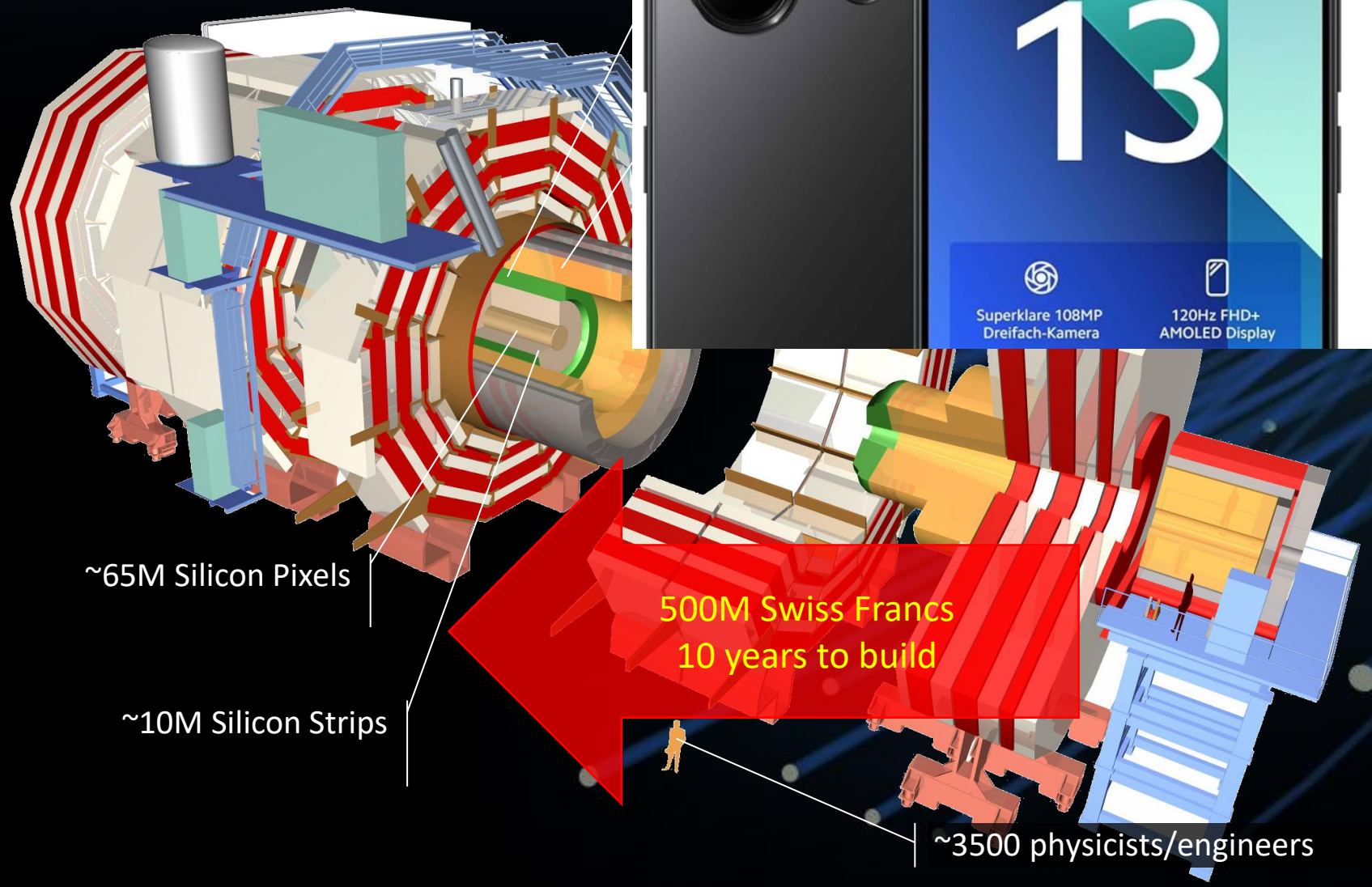
~65M Silicon Pixels

~10M Silicon Strips

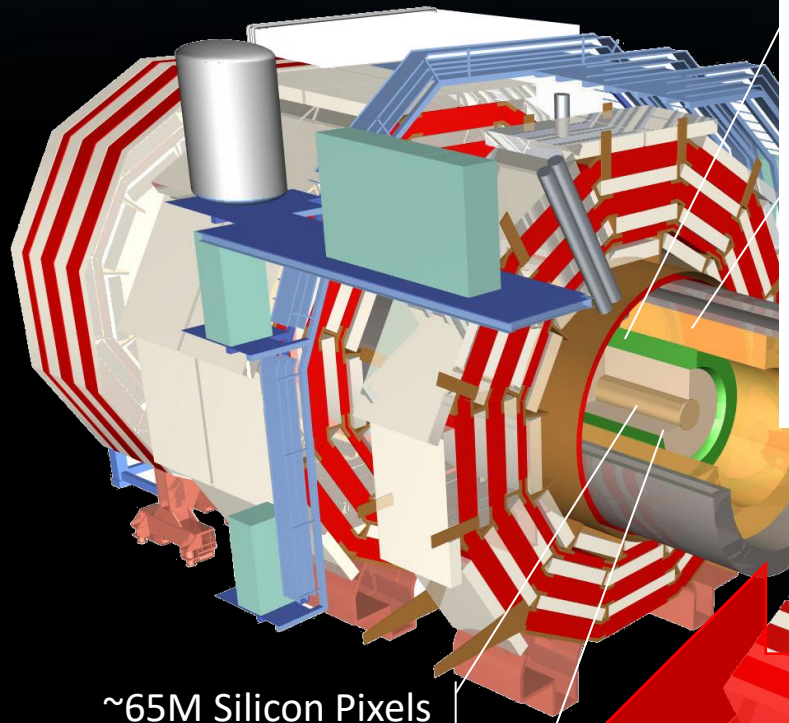
~3500 physicists/engineers

The CMS detector





The CMS detector



Xiaomi Redmi Note
Black (UK Version +
Visit the Xiaomi Store
4.4 ★★★★½ 288 ratings
100+ bought in past month

£147.99

prime Same-Day & Overnight
FREE Returns
Pay over time with no interest. Subject
Amazon EU S.a.r.l, Lender: Barclays. C

Purchase options and add

Payment plans
Instalment options available

Brand Xiaomi

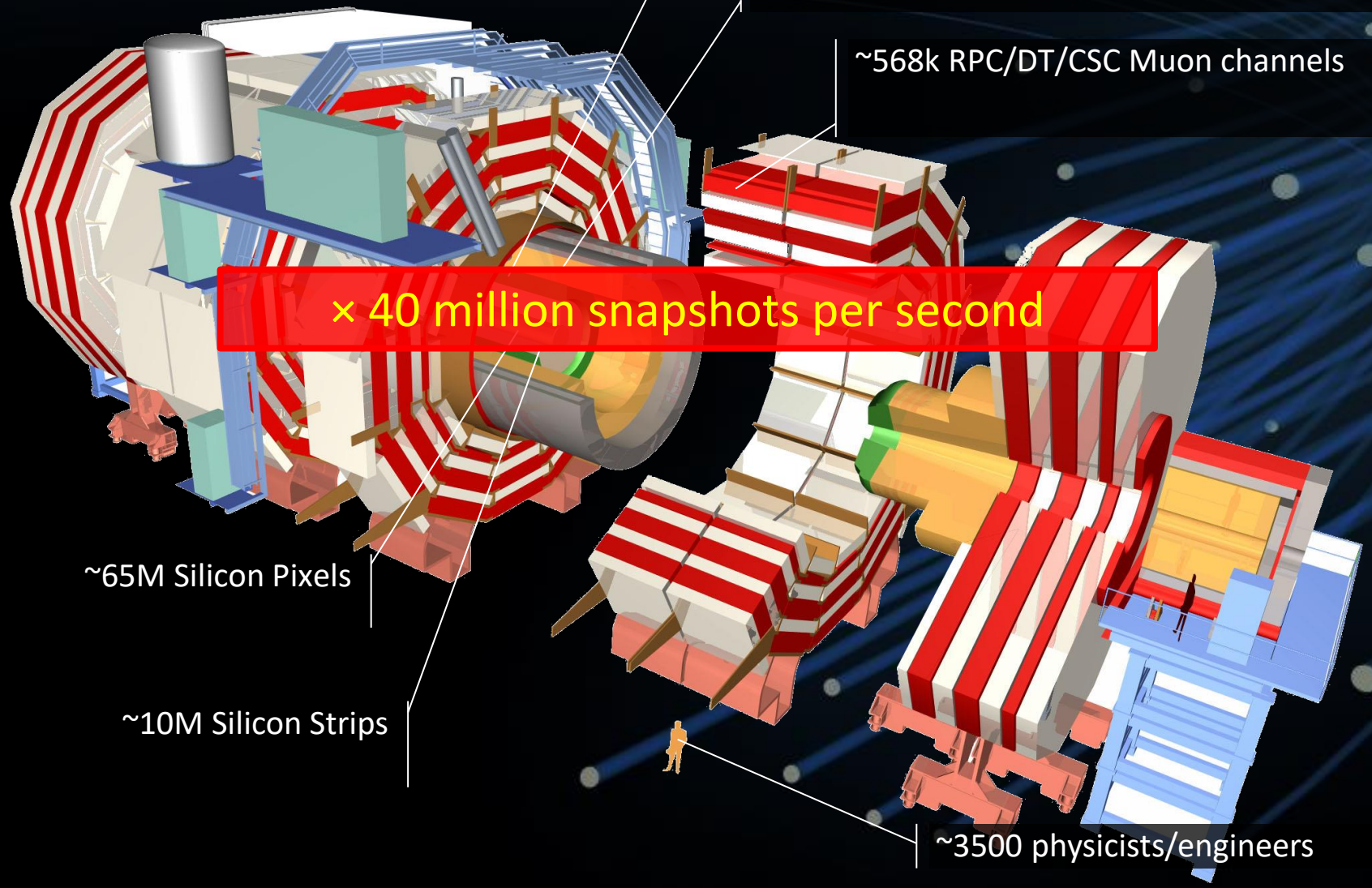
~65M Silicon Pixels

~10M Silicon Strips

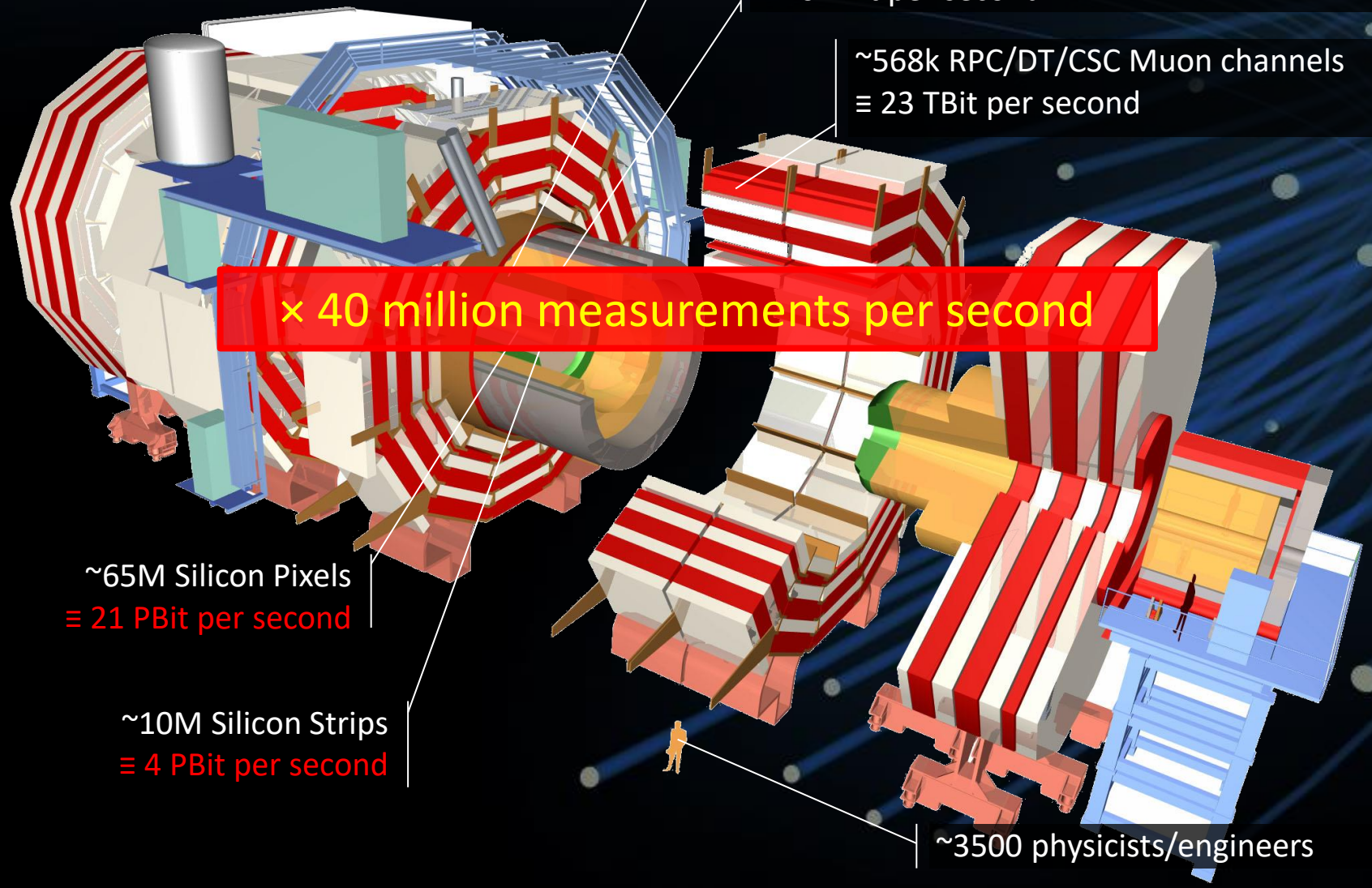
Covers 205m2
Most intense radiation environment on the planet
Installation completed before release of 1st iPhone
And…

~3500 physicists/engineers

The CMS detector



The CMS detector

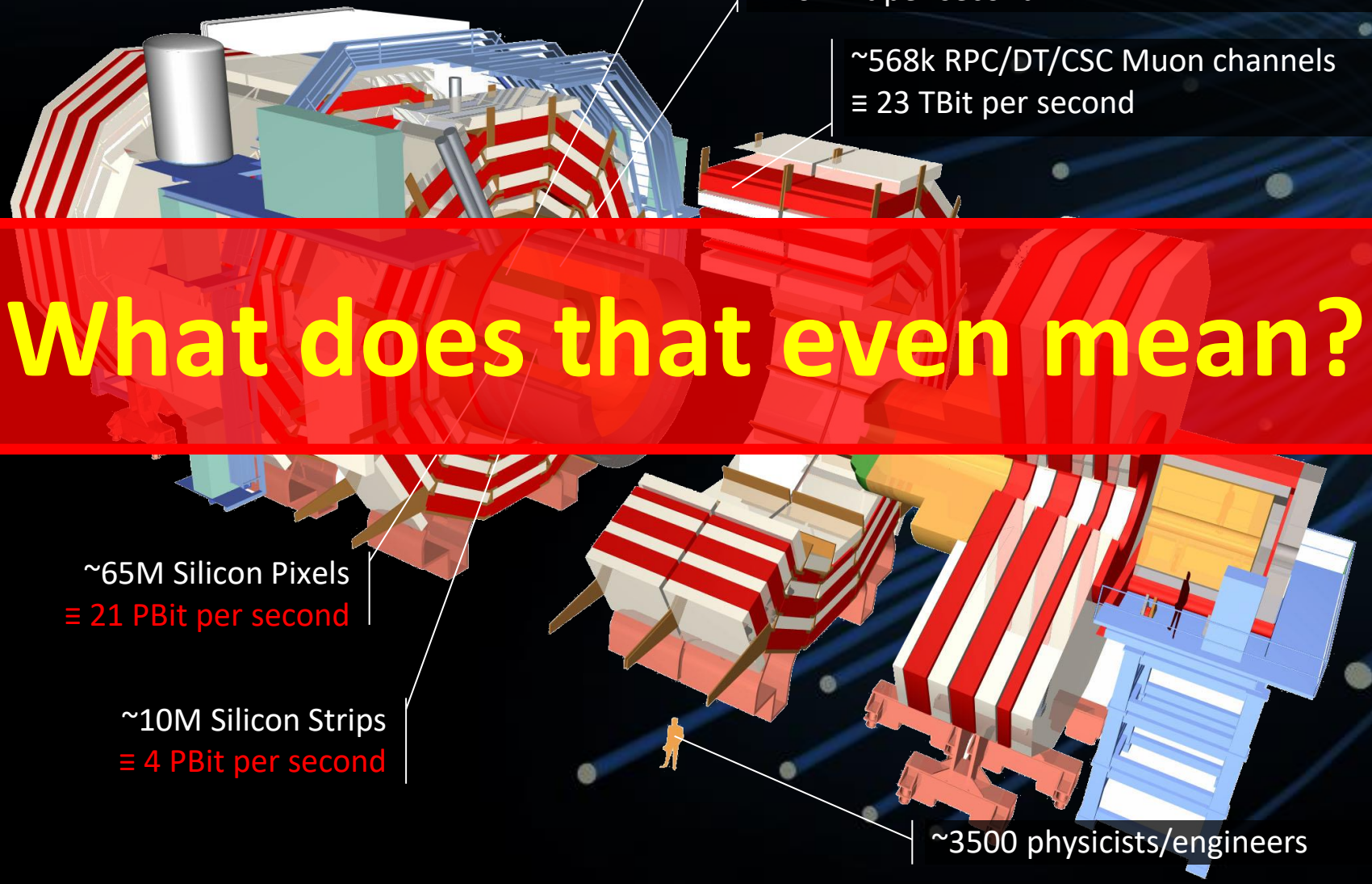


The CMS detector

~80k $PbWO_4$ Ecal Crystals
≡ 40 TBit per second

~15k channel Brass/Plastic sampling HCAL
≡ 10 TBit per second

~568k RPC/DT/CSC Muon channels
≡ 23 TBit per second



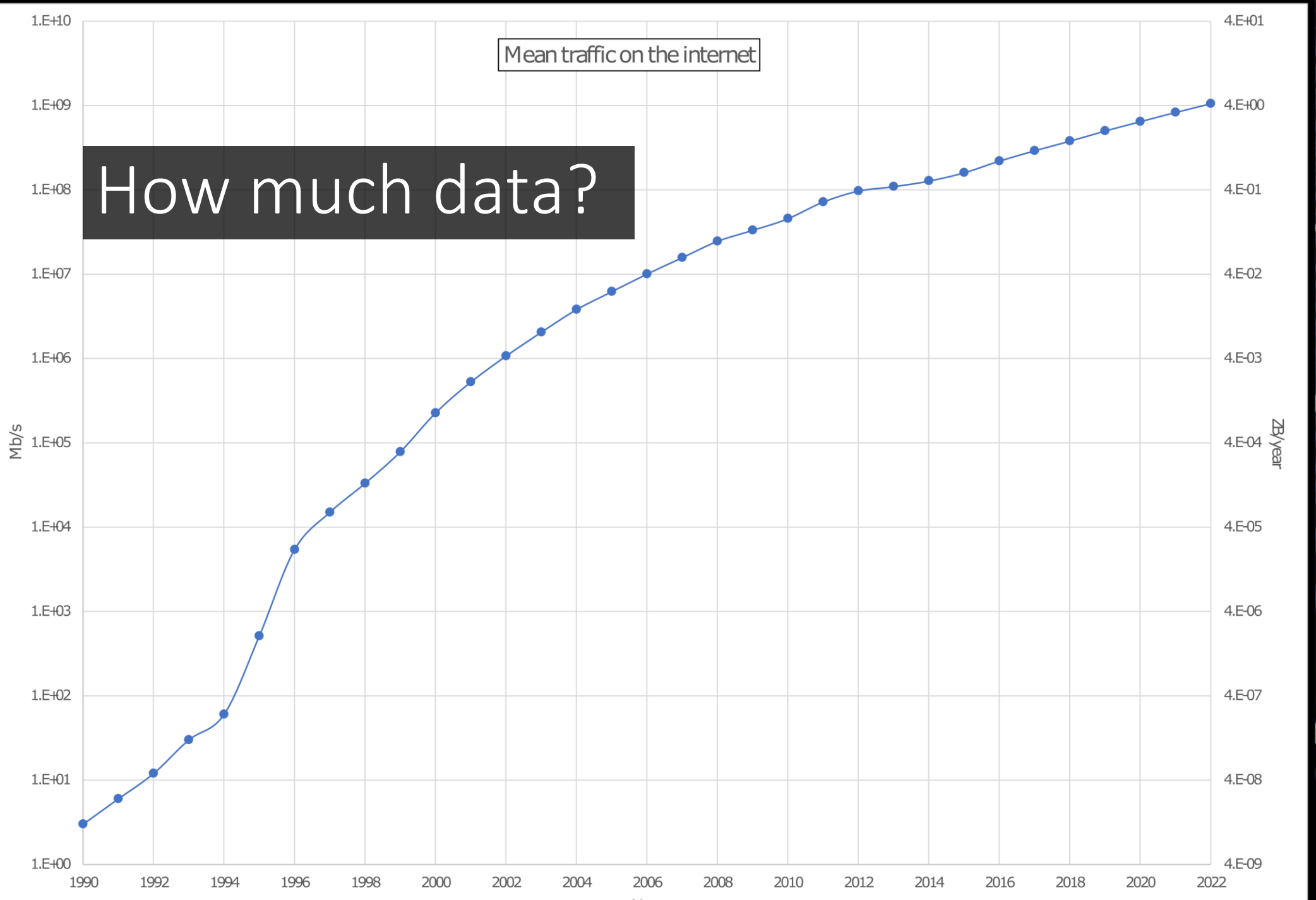
What does that even mean?

~65M Silicon Pixels
≡ 21 PBit per second

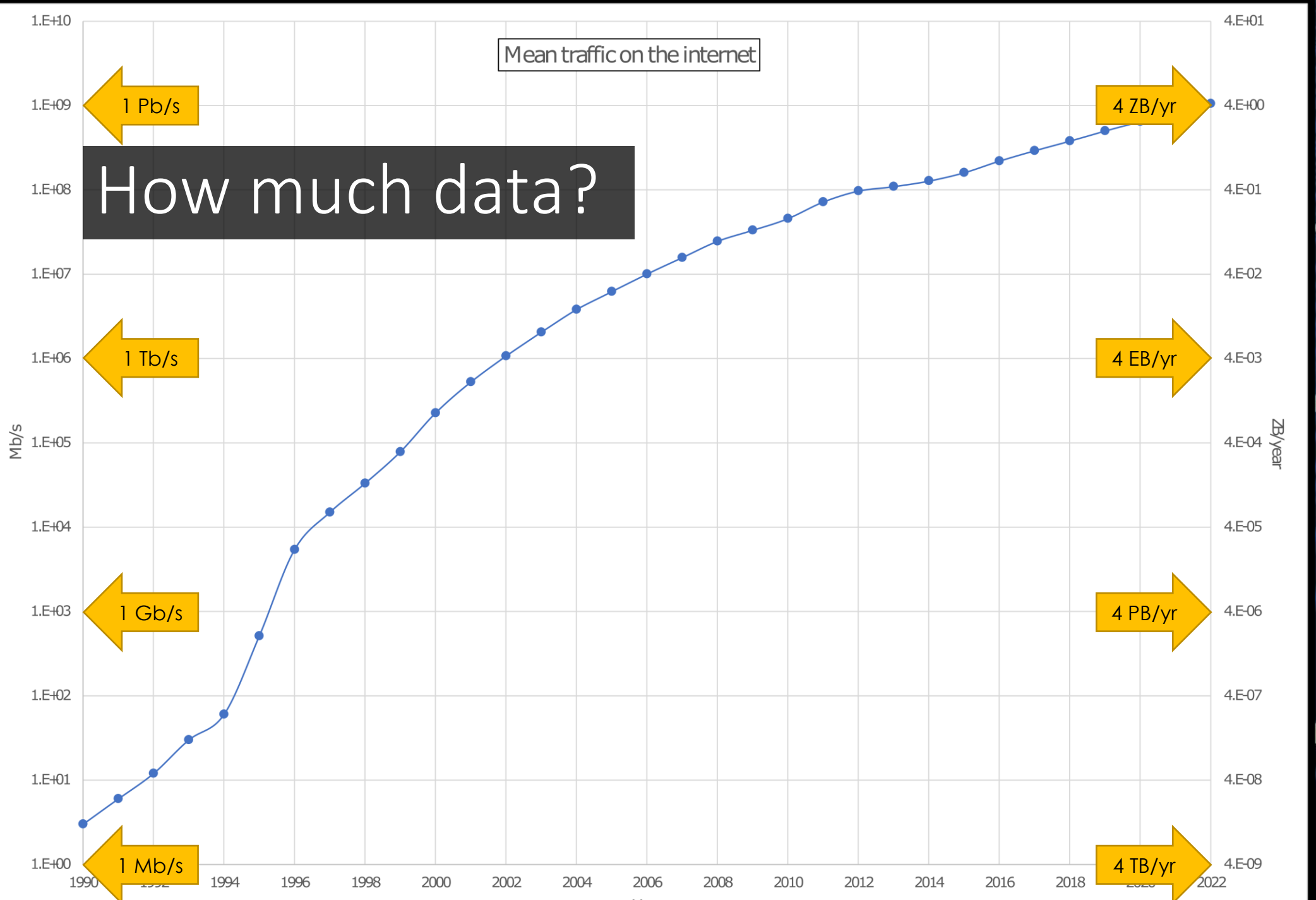
~10M Silicon Strips
≡ 4 PBit per second

~3500 physicists/engineers

How much data?

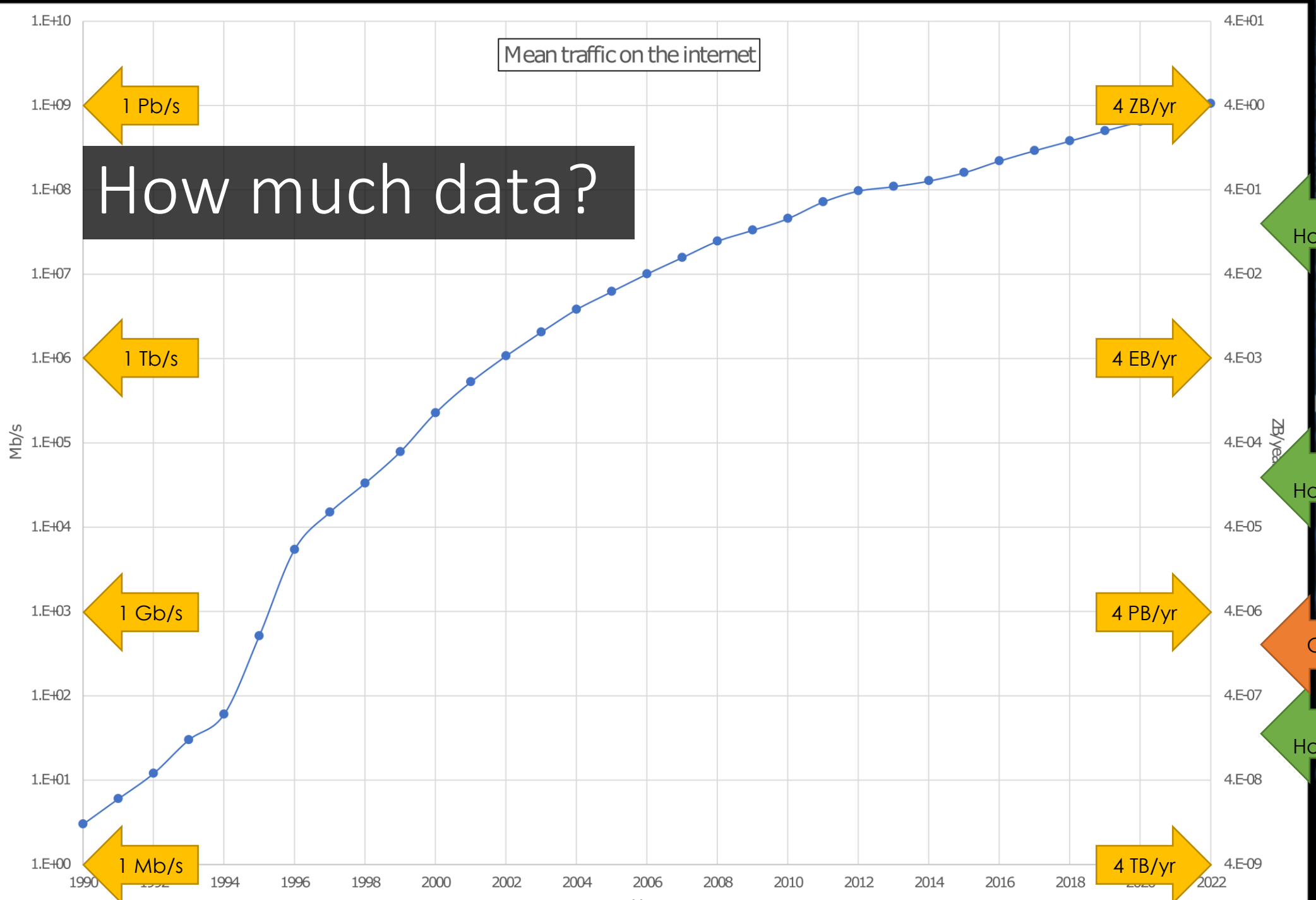


How much data?



CMS Raw

How much data?



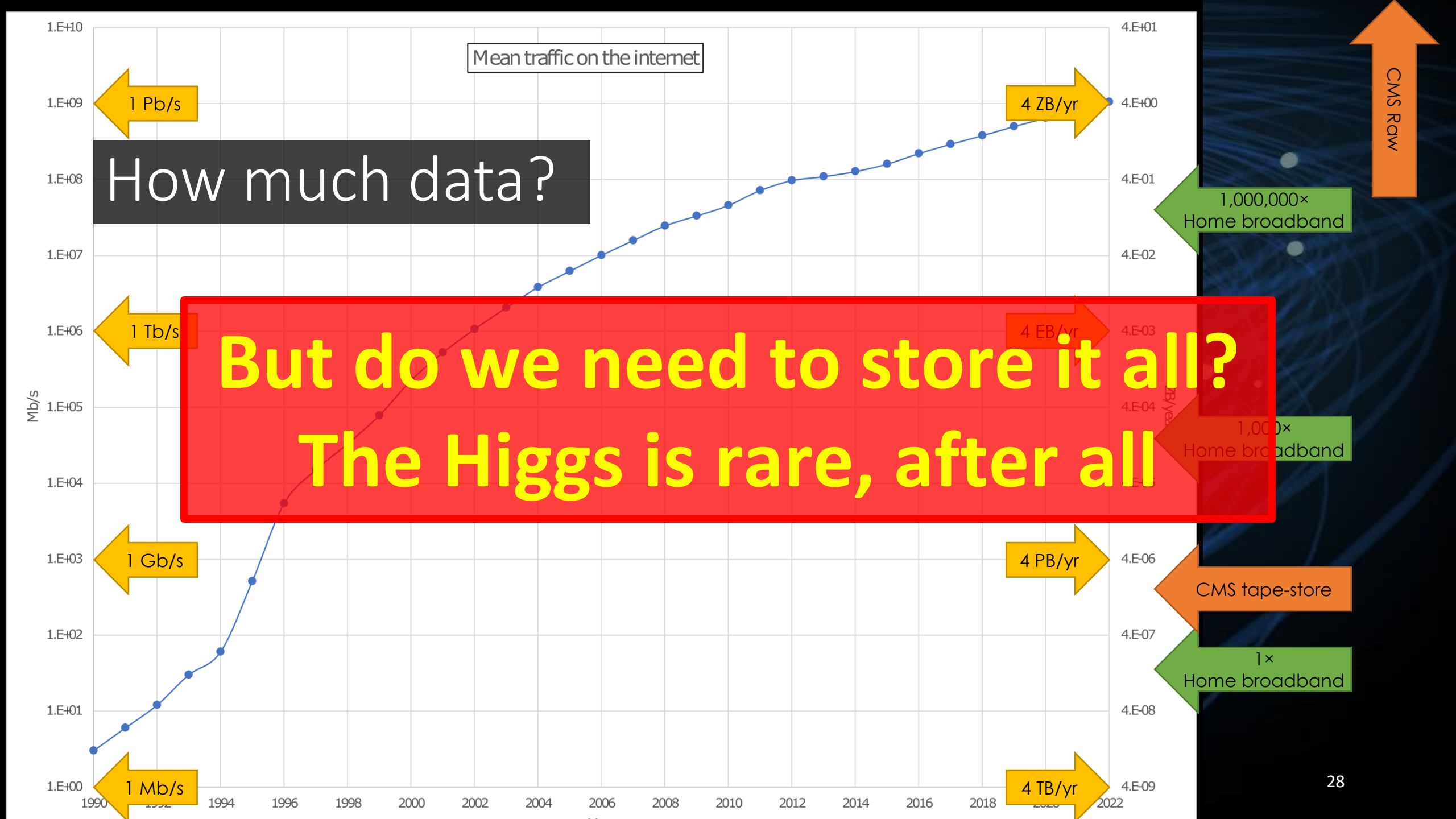
CMS Raw

1,000,000× Home broadband

1,000× Home broadband

CMS tape-store

1× Home broadband



How much data?

But do we need to store it all?

The Higgs is rare, after all

So what do we do?

- We keep the data safe on the detector for several microseconds
- We quickly analyse some fraction of the data and select the interesting crossings to keep

So what do we do?

- We keep the data safe on the detector for several microseconds
- We quickly analyse some fraction of the data and select the interesting crossings to keep

Called Triggering

So what do we do?

- We keep the data safe on the detector for several microseconds
- We quickly analyse some fraction of the data and select the ~~interesting crossings to keep~~ uninteresting crossings to discard

The LHC is a discovery machine, after all

Called Triggering

But!

- If you get it wrong
- If your system fails
- If you take too long

} you throw away your valuable physics

But!

- If you get it wrong
- If your system fails
- If you take too long

} you throw away your valuable physics

And that is a *really, really* expensive mistake

This is why the LHC collides…

~50 protons at a time

×

40 million times a second

---

1 Higgs boson every ~5 seconds

But this causes problems

This causes other problems too

A note on timescales

- At 40MHz BX rate, a 4GHz CPU could perform 100 CPU operations before the next data arrives
  - Not enough to be useful
- Compare that to the O(100M) detector channels

Sequential processing

	6pm	7pm	8pm	9pm	10pm	11pm	12pm	01am	02am	03am
										
										
										
										

Sequential processing

	6pm	7pm	8pm	9pm	10pm	11pm	12pm	01am	02am	03am
										
										
										
										

Sequential processing

	6pm	7pm	8pm	9pm	10pm	11pm	12pm	01am	02am	03am
										
										
										
										

That would just be stupid

Sequential processing

Pipelined processing

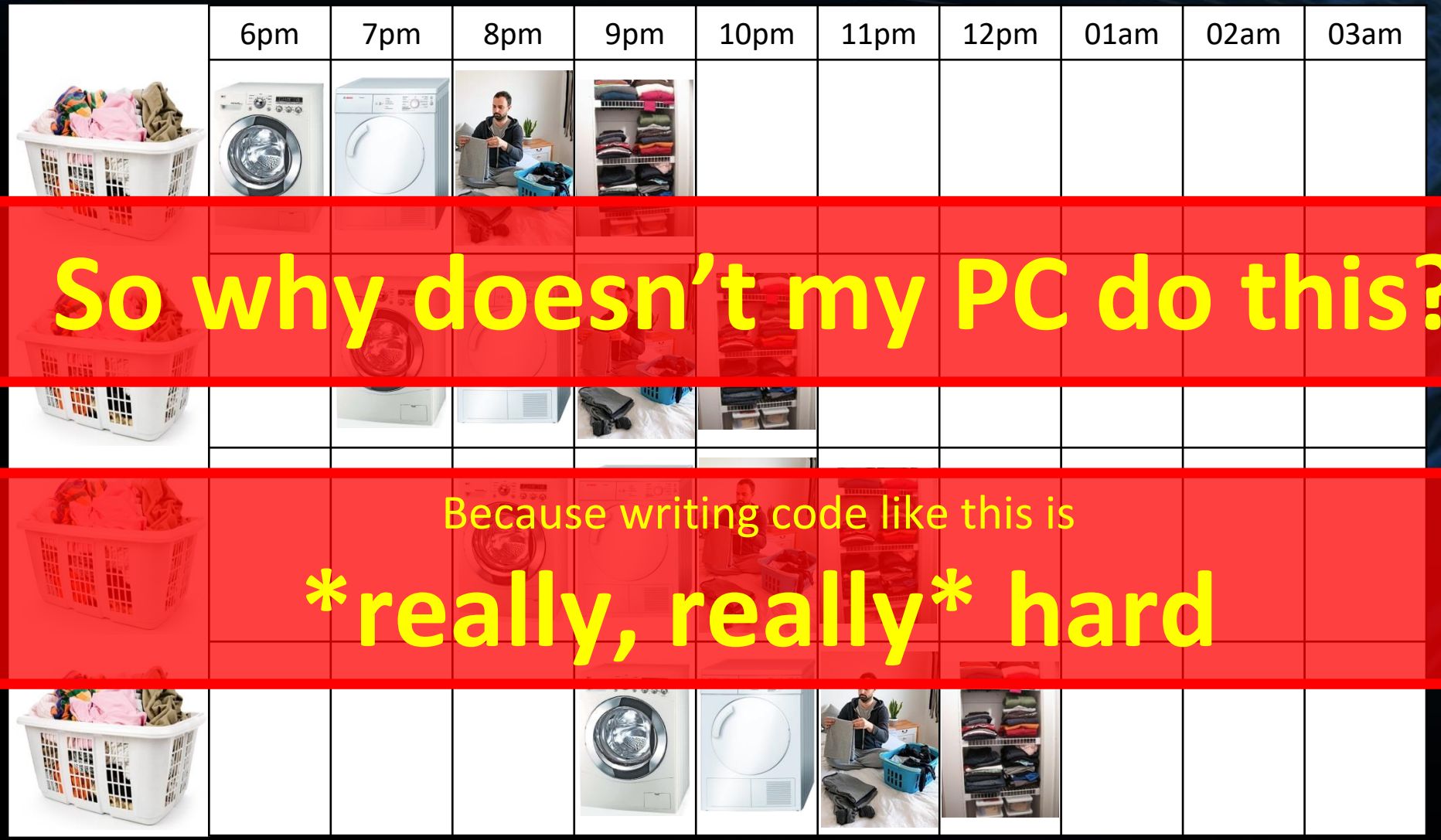
	6pm	7pm	8pm	9pm	10pm	11pm	12pm	01am	02am	03am
										
										
										
										

Pipelined processing

	6pm	7pm	8pm	9pm	10pm	11pm	12pm	01am	02am	03am
										
										
										

So why doesn't my PC do this?

Pipelined processing



So why doesn't my PC do this?

Because writing code like this is

***really, really* hard**

Field-Programmable Gate Arrays

- Programmable circuit-on-chip
- Upwards of 8.9 million logic cells & up to 12,000 “math” cells
  - All clocked at ~500MHz
- Over O(10^{15}) operations/second
  - That is PetaBOps
- Fully parallel and fully pipelineable

Field-Programmable Gate Arrays

- Programmable circuit-on-chip
- Upwards of 8.9 million logic cells & up to 12,000 “math” cells
  - All clocked at ~500MHz
- Over O(10^{15}) operations/second
  - That is PetaBOps
- Fully parallel and fully pipelineable

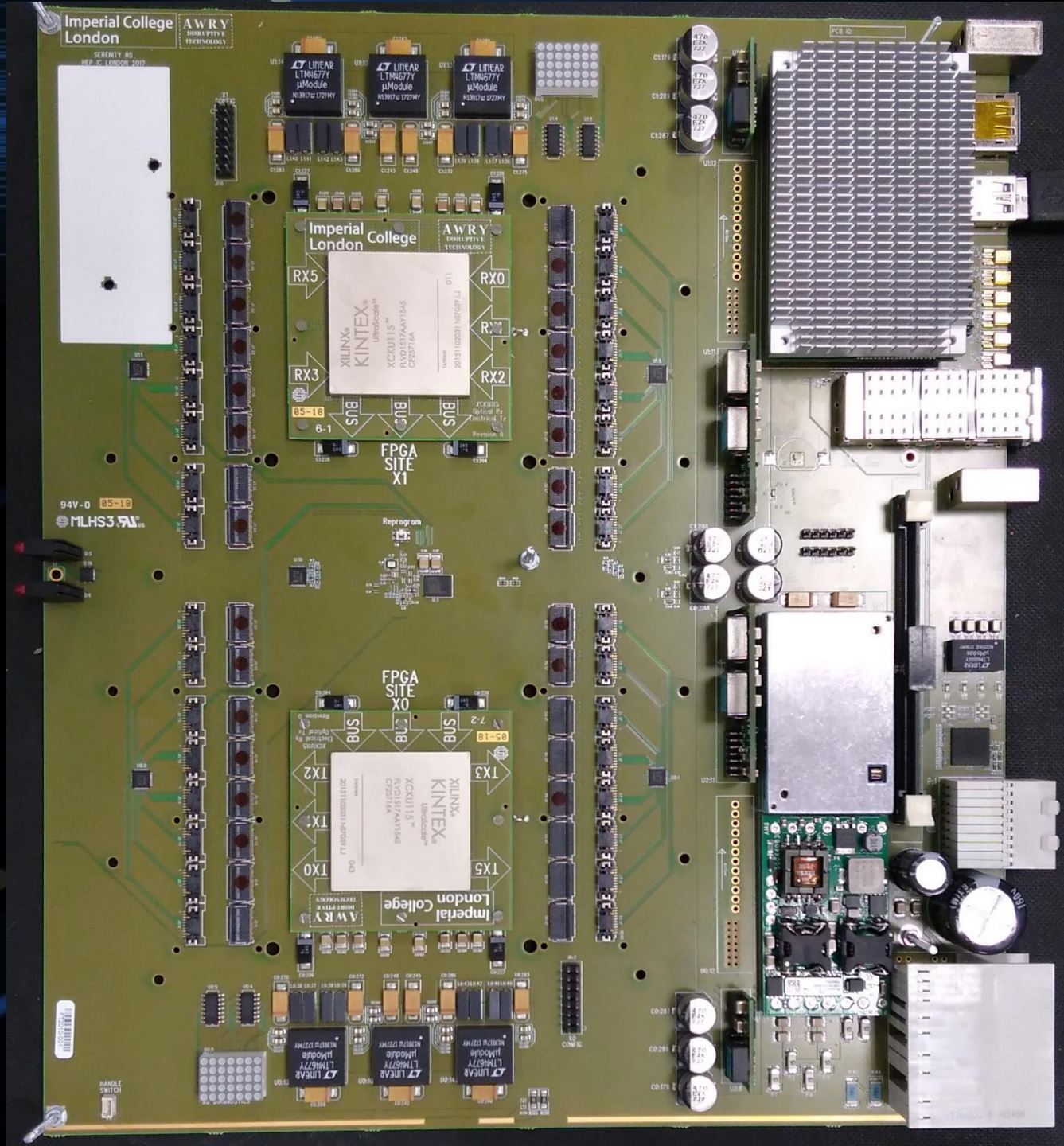
and *really, really* hard to program efficiently

Field-Programmable Gate Arrays

- A chip needs to be attached to something!
- Imperial's Serenity-Z platform
  - 208 optical transmit links &
    208 optical receive links
    @ 28.5Gbps =

    5.9 + 5.9 Tbps
  - Skewable to

    7.0 + 4.8 Tbps

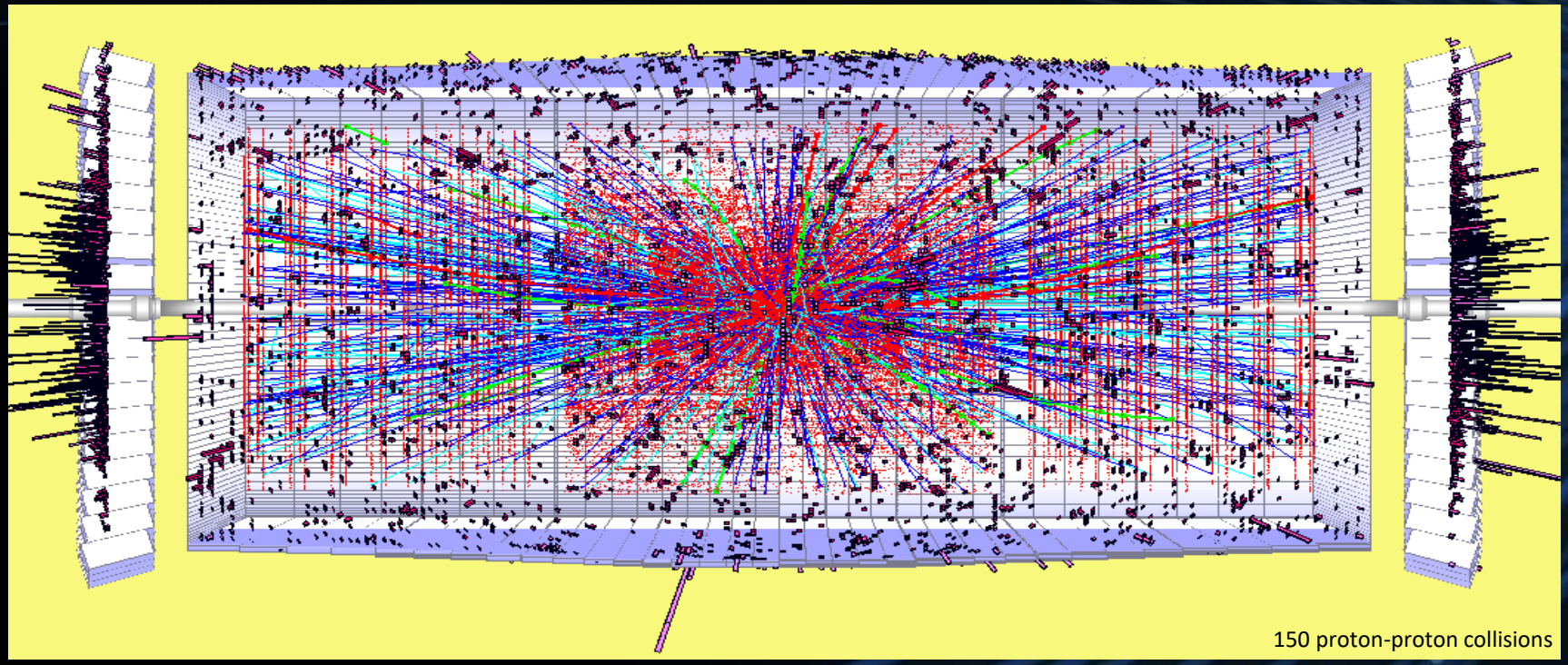


Field-Programmable Gate Arrays

- A chip needs to be attached to something!
- Imperial's Serenity-S platform
  - 124 optical transmit links &
    124 optical receive links
    @ 28.5Gbps =

    3.5 + 3.5 Tbps
  - Easier to manufacture

The future: High-luminosity LHC



- Averaging 250 proton-proton collisions per event
- Really, really hard to find the interesting event under all that rubbish
- Need to sift through 300Tbps to find the interesting events

But this is the 21st century…

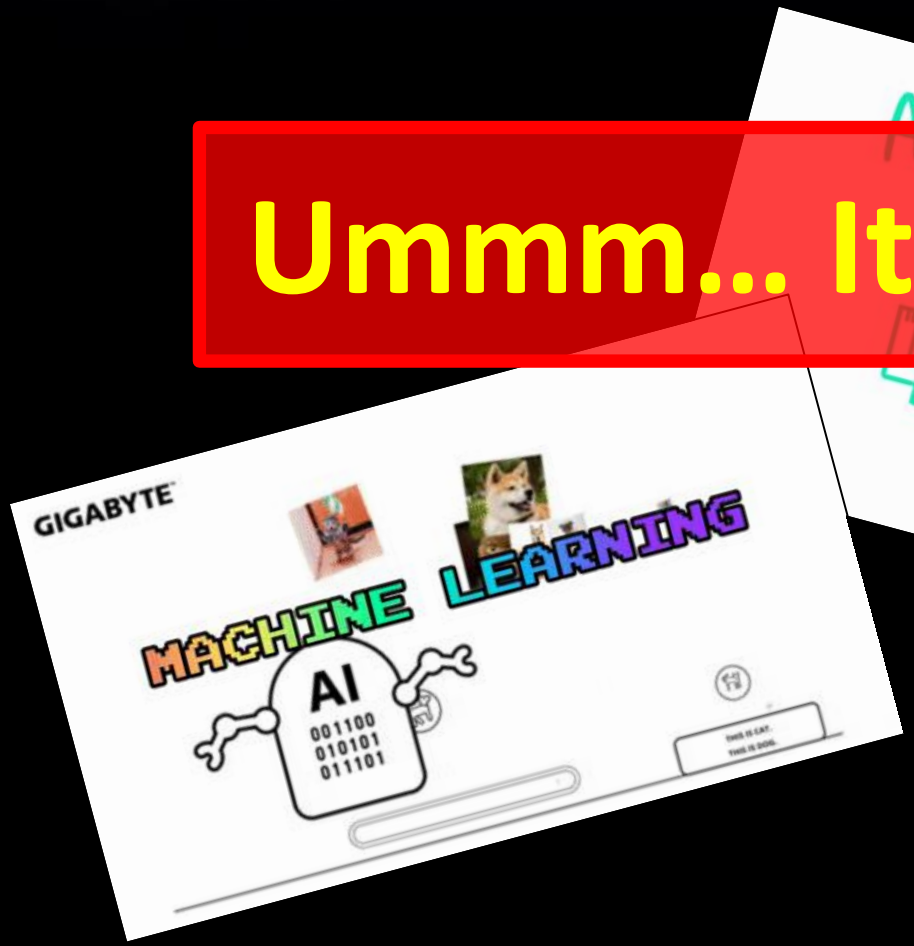
- How do you deal with hard problems and big data?

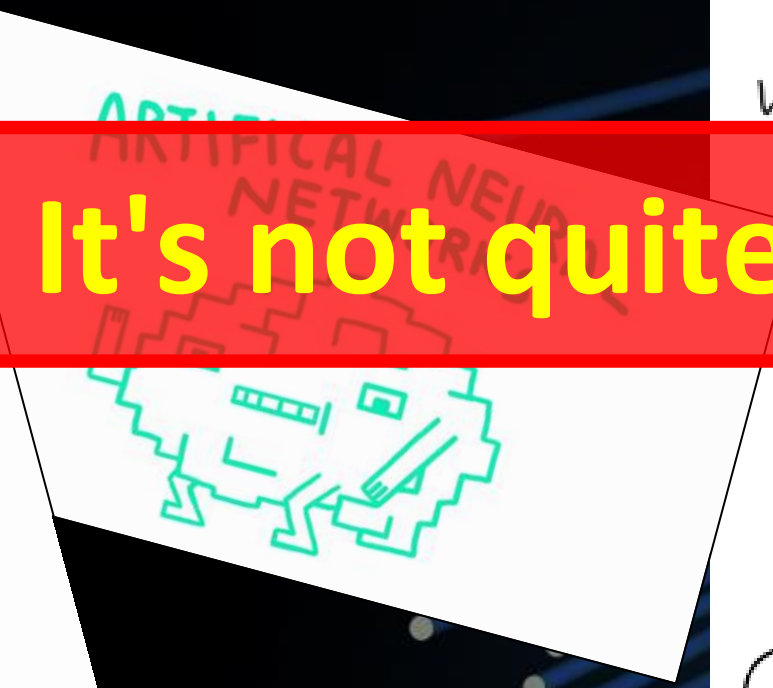
But this is the 21st century...

- How do you deal with hard problems and big data?

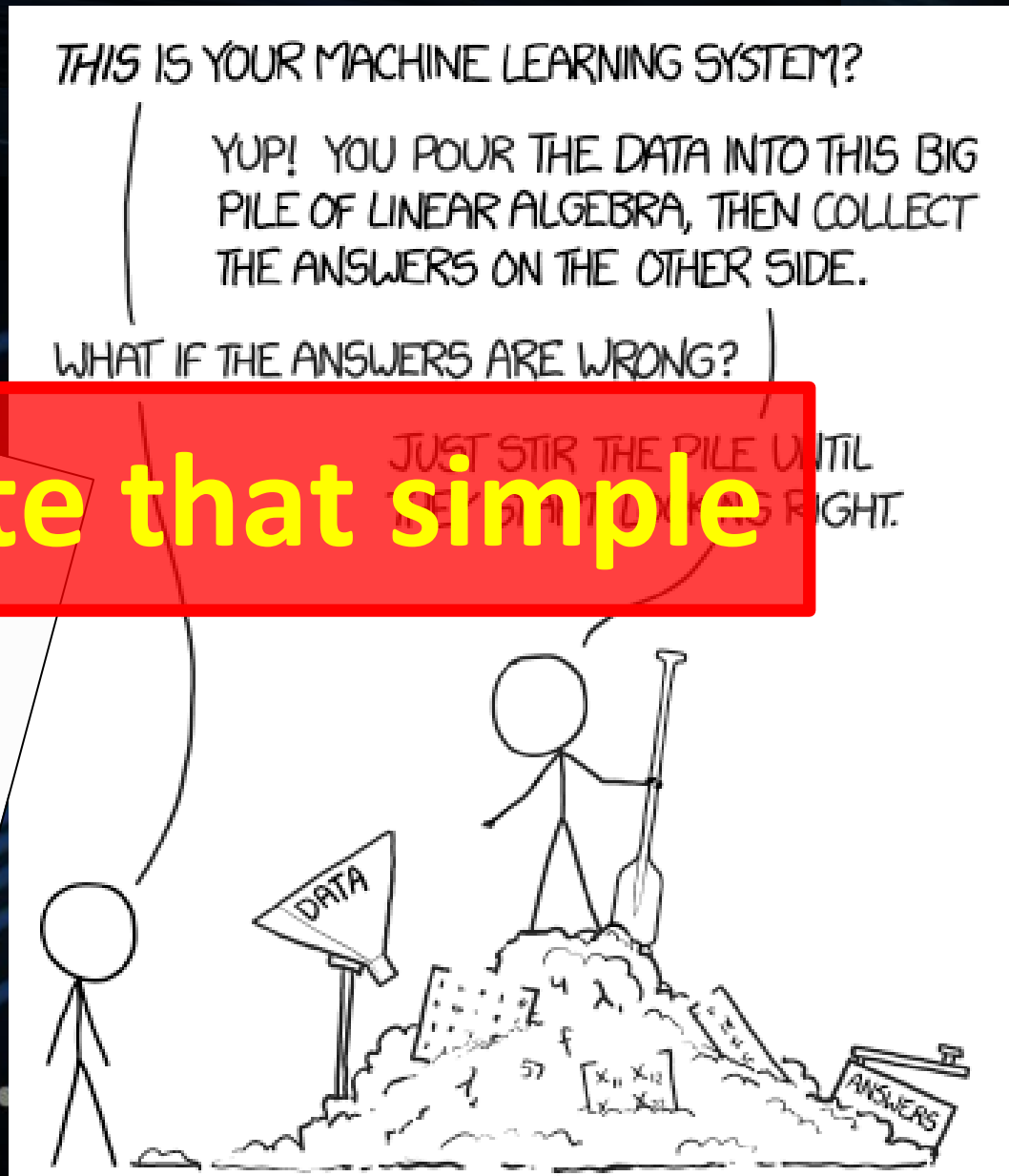
But this is the 21st century...

- How do you deal with hard problems and big data?



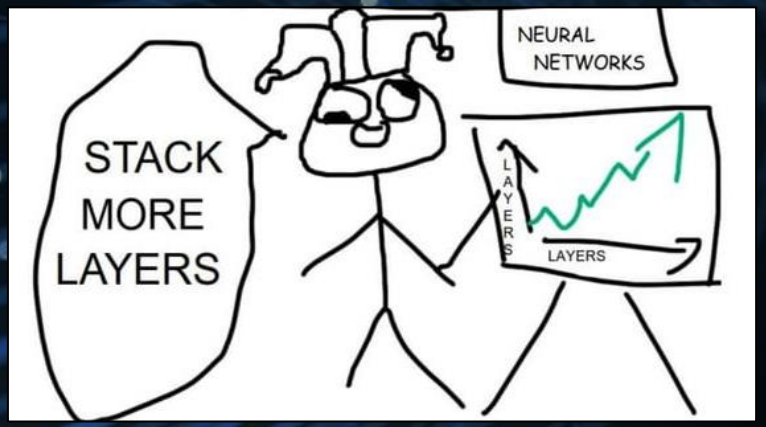


Ummm... It's not quite that simple



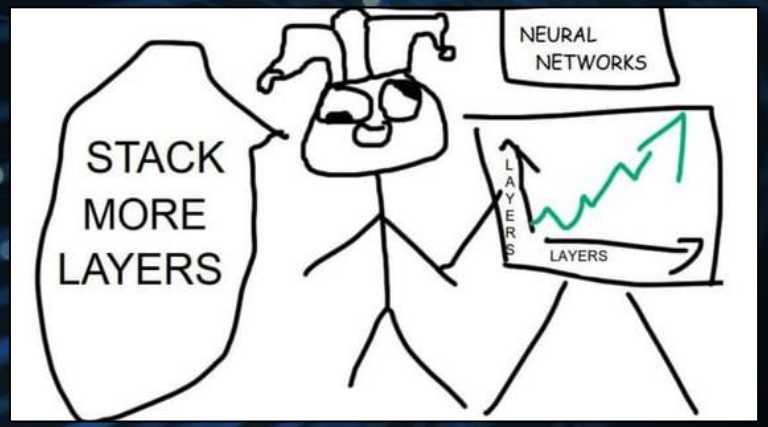
Why can't you just use Machine Learning at HL-LHC?

- The data is still only safe on the detector for ~10 microseconds
  - Not much time
- The trigger is receiving ~100Tbps from the tracker, ~100Tbps from the barrel calorimeter and ~100Tbps from the endcap calorimeter
  - AKA 1 hellavalotta data
- Recall, that you can't see the data the trigger throws away…
  - You need to be *REALLY* sure you know what your algorithm is doing.
- The detector environment is always changing, inference may not look anything like your training data



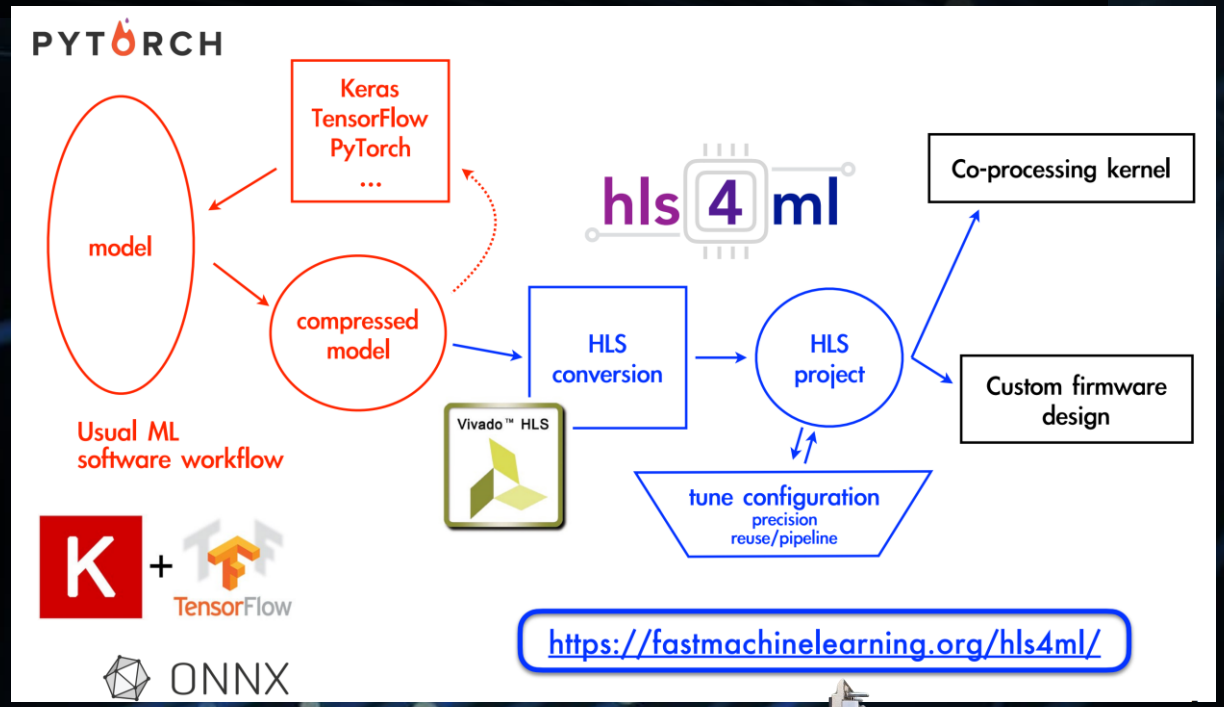
Why can't you just use Machine Learning at HL-LHC?

- The data is still only safe on the detector for ~10 microseconds
  - Not much time
- The trigger is receiving ~100Tbps from the tracker, ~100Tbps from the barrel calorimeter and ~100Tbps from the endcap calorimeter
  - AKA 1 hellavalotta™ data
- Recall, that you can't see the data the trigger throws away…
  - You need to be *REALLY* sure you know what your algorithm is doing.
- The detector environment is always changing, inference may not look anything like your training data



That said… if it works, why wouldn't you use it?

- For certain, well-defined sections of the trigger chain, machine-learning is a useful and powerful tool
  - Tools developed for translating Neural Networks for use in high-throughput and low-latency systems
- Work being done in the area of understandable AI
- Work being done in the area of continual learning



Conclusion

- The hardware triggers at the LHC are some of the most computationally challenging problems on the planet
- They are a great deal of responsibility

Conclusion

- The hardware triggers at the LHC are some of the most computationally challenging problems on the planet
- They are a great deal of responsibility

And, if you like puzzles,
a great deal of fun ☺