



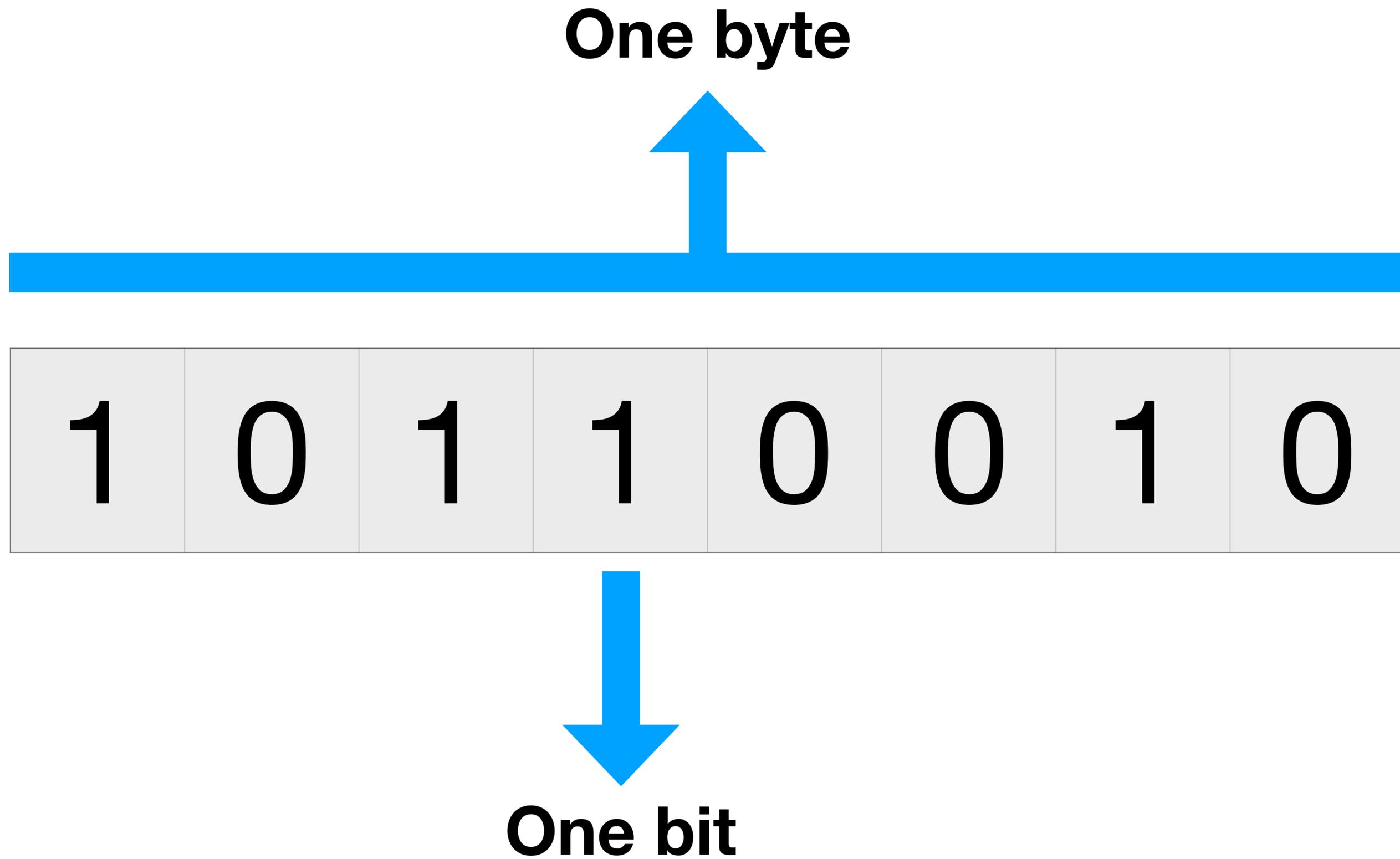
UiO : **University of Oslo**

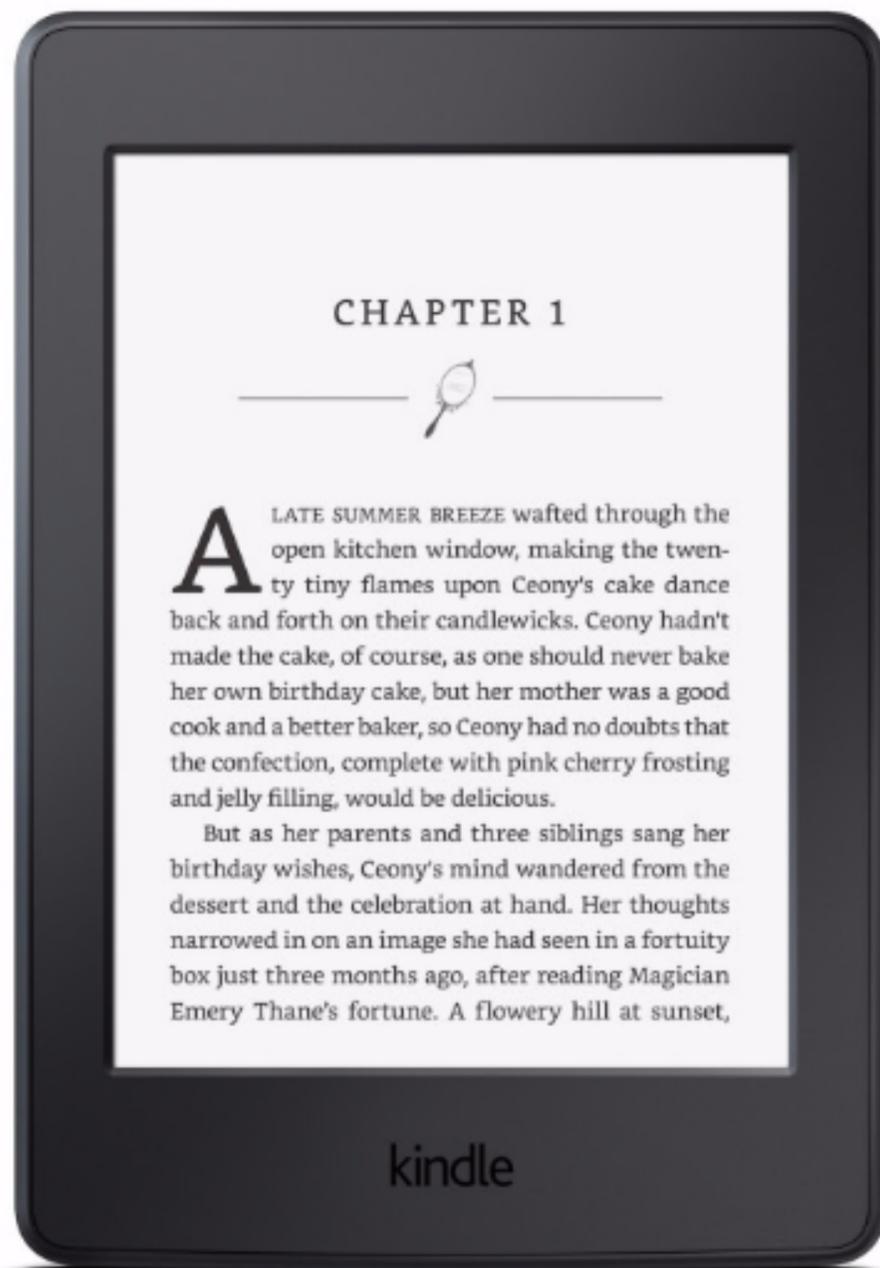
Big Data meets Big Science: How Particle Physicists do computing

James Catmore, University of Oslo
Computing Coordinator for the ATLAS Collaboration

**ATLAS currently stores almost 500 PB of data on disk and tape
(Other experiments store similar volumes)**

How big is that?





~2 million bytes (2MB)
Average size of an e-book



**~1 trillion bytes (1TB)
Typical storage on a desktop
computer**



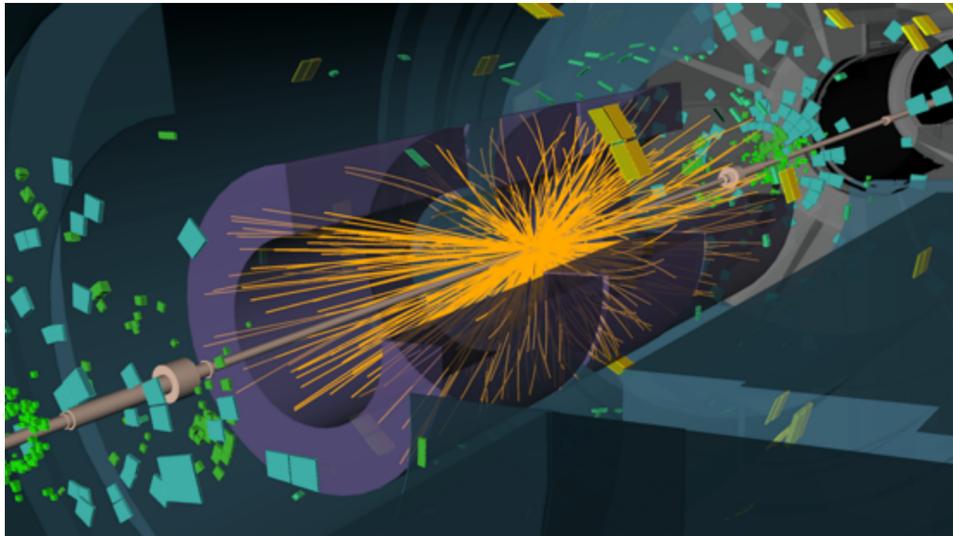
**~1 trillion bytes (1TB)
Typical storage on a desktop
computer**

x 1000 \approx 1PB

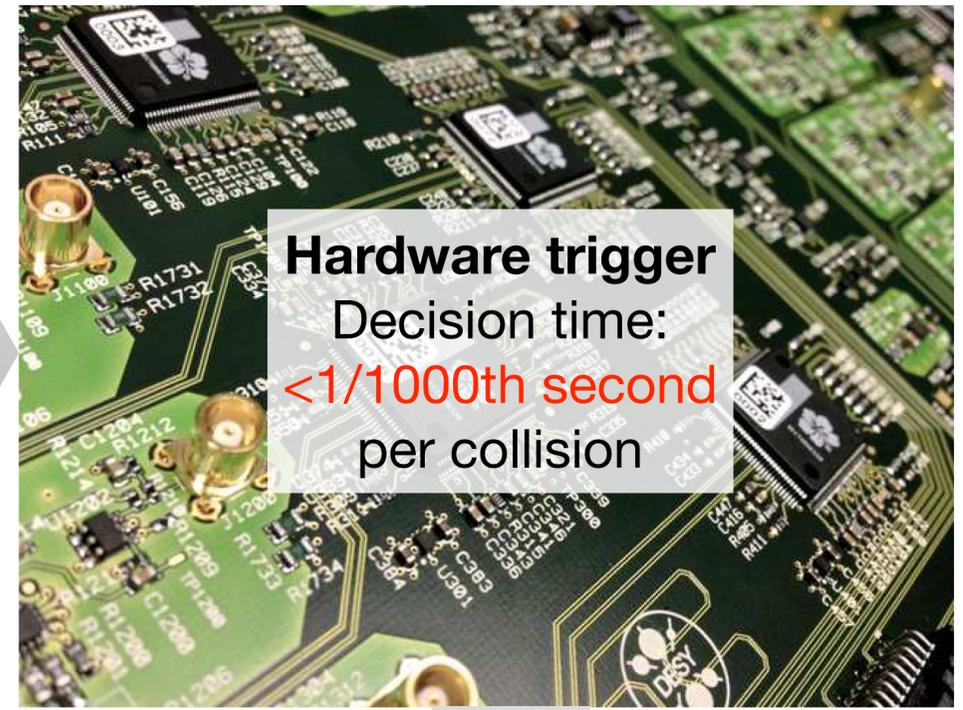


186 quintillion bytes (186 000 PB)
Estimated global monthly internet traffic in 2019

What to keep and what to throw?



40 MILLION COLLISIONS PER SECOND
= **60 TB**/second
= **24 million** 30 Mbps broadband connections



Hardware trigger
Decision time:
<1/1000th second
per collision



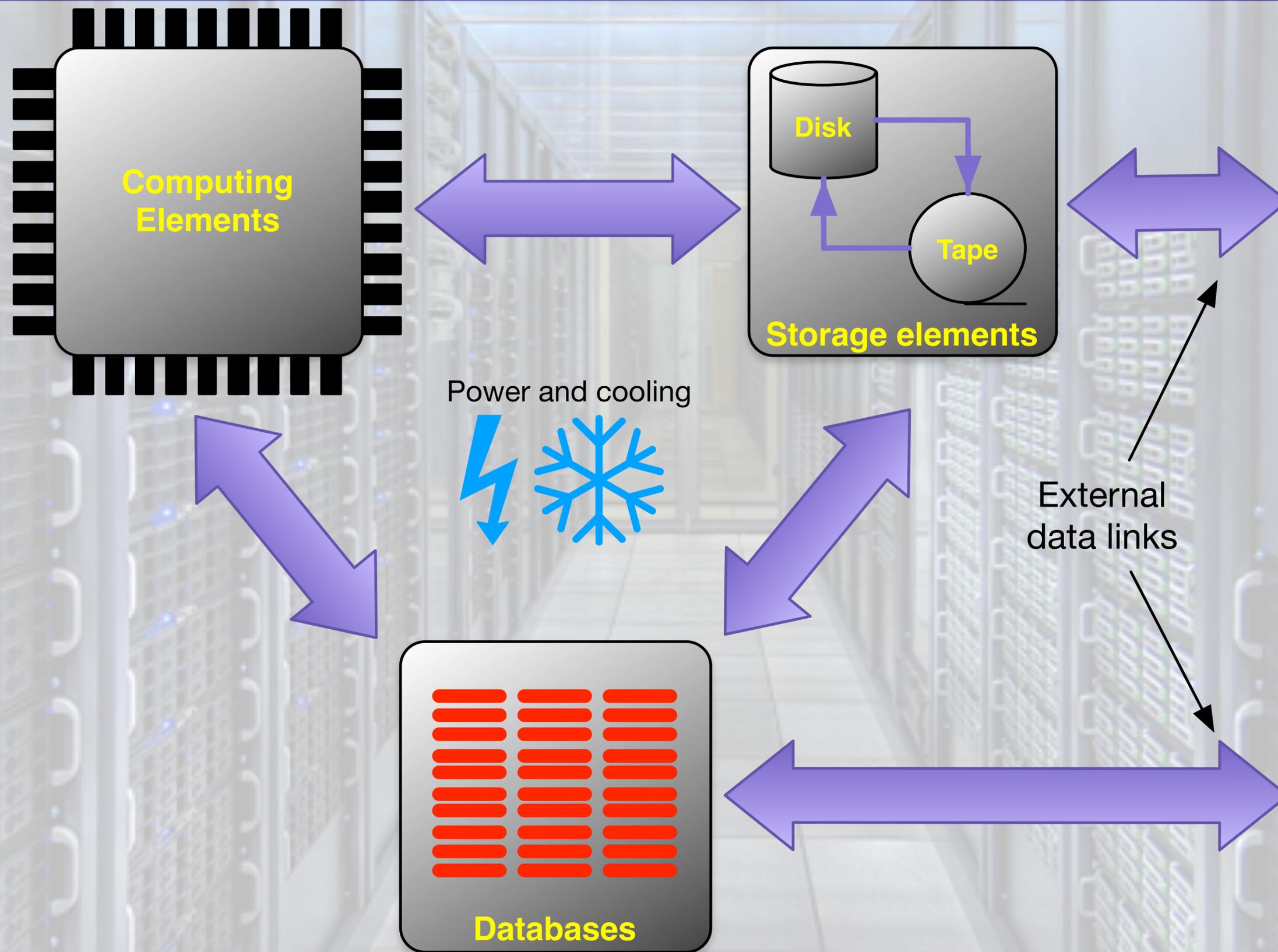
1 000 COLLISIONS PER SECOND
= **1.5 GB**/second
= **400** broadband connections



High level trigger
Decision time:
0.5 seconds
per collision

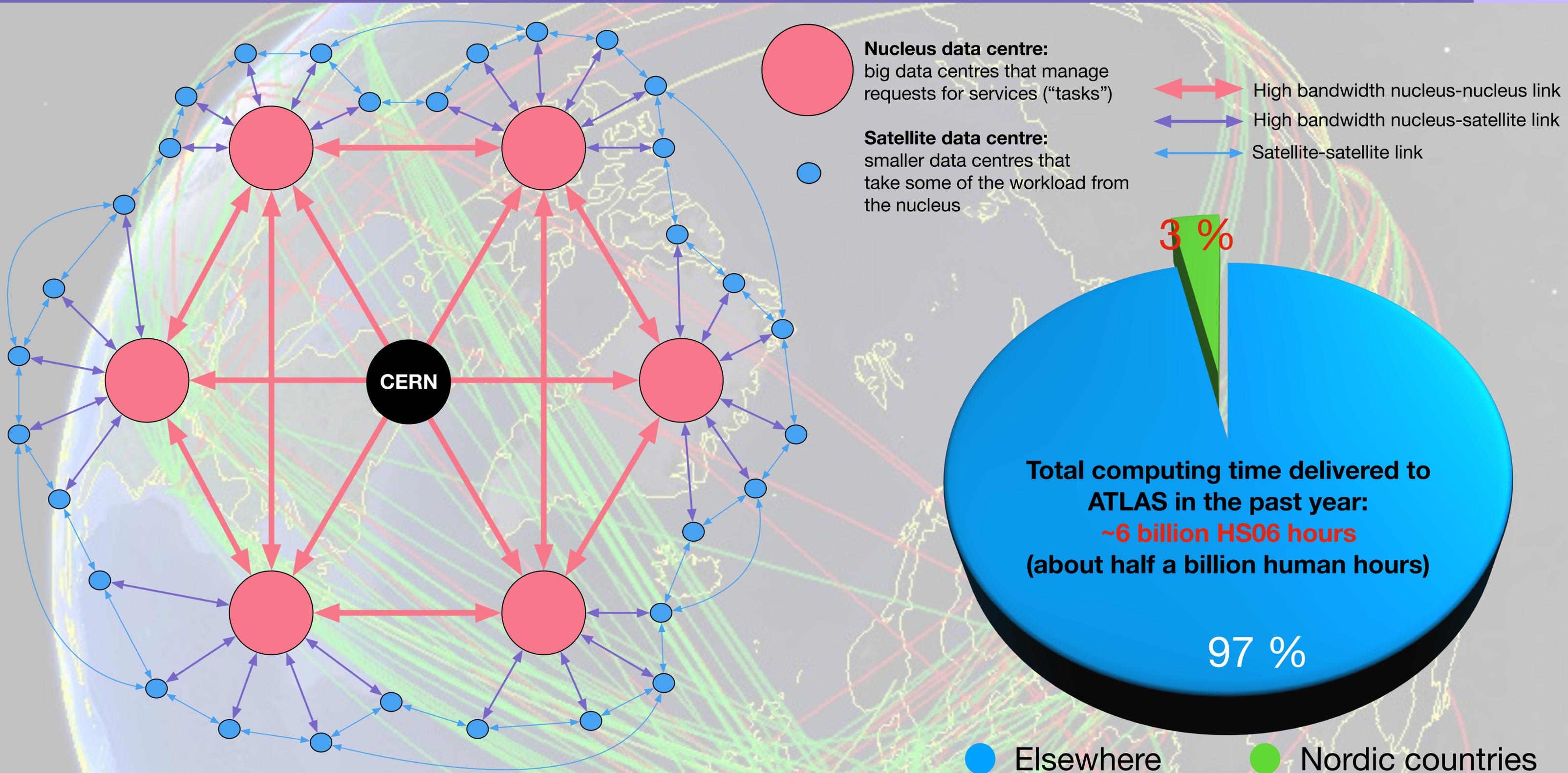
100 000 COLLISIONS PER SECOND
= **160 GB**/second
= **43 000** broadband connections





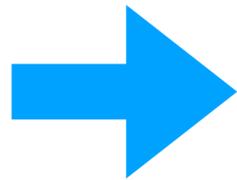
Data Centres and Computing Grids



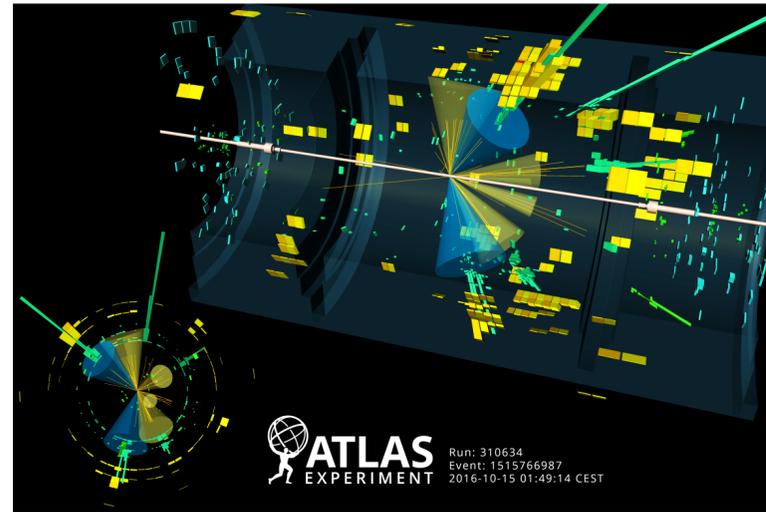
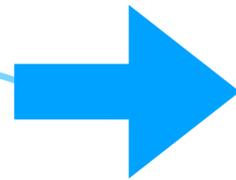
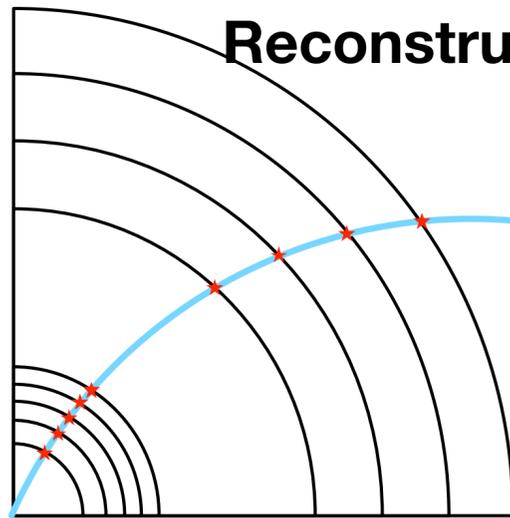


Raw data

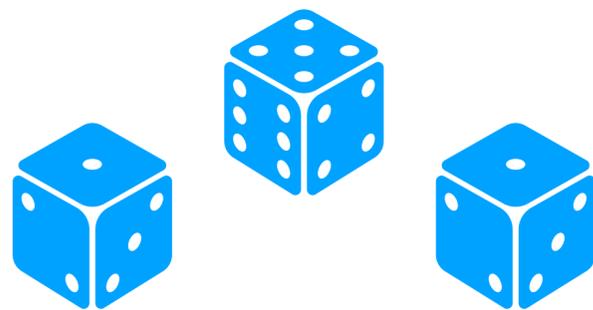
```
01100011010010010100010  
10101000101010100111010  
10101010001001010100010  
00101010101010010101010  
01010111010111101010101  
11101011011010101110101  
01011110000010001010001  
00010010100101101010101  
01101011110100101111010  
11011010101110101010111  
10000010001010001000100  
10100101101010101011010
```



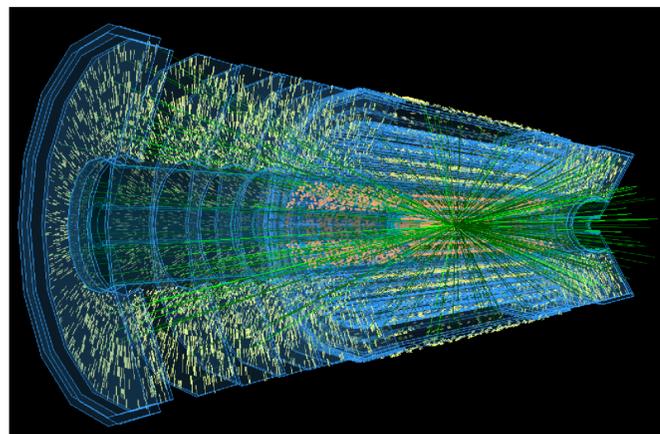
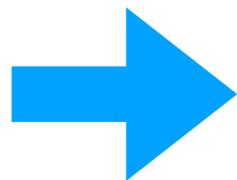
Reconstruction



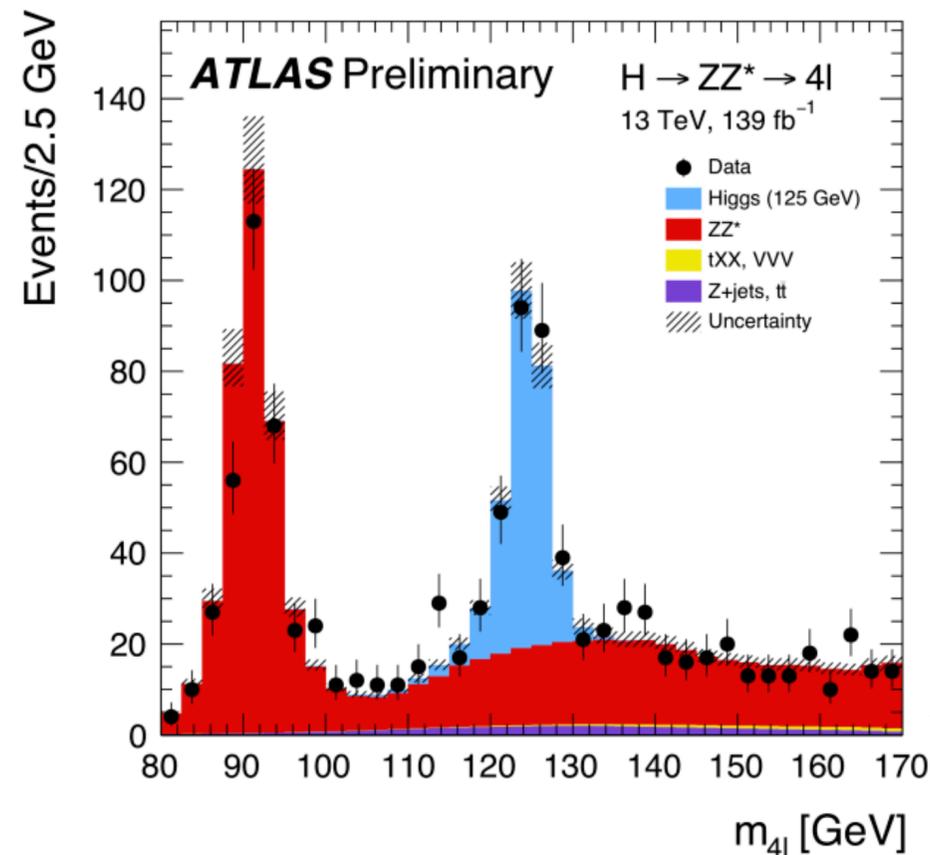
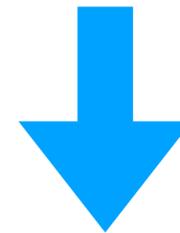
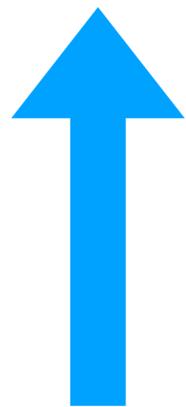
Events



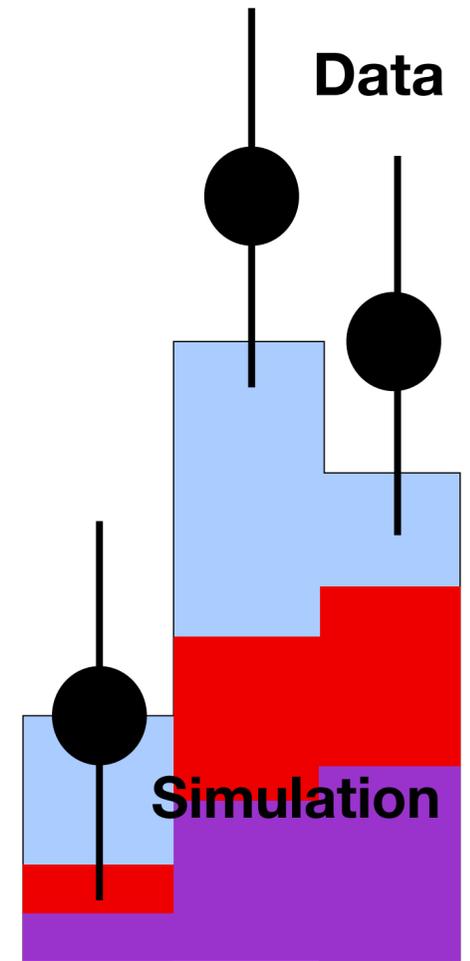
$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} + i\bar{\psi}\not{D}\psi + h.c. + \bar{\psi}_i \gamma_{ij} \psi_j \phi + h.c. + |\mathcal{D}_\mu \phi|^2 - V(\phi)$$



Detector simulation

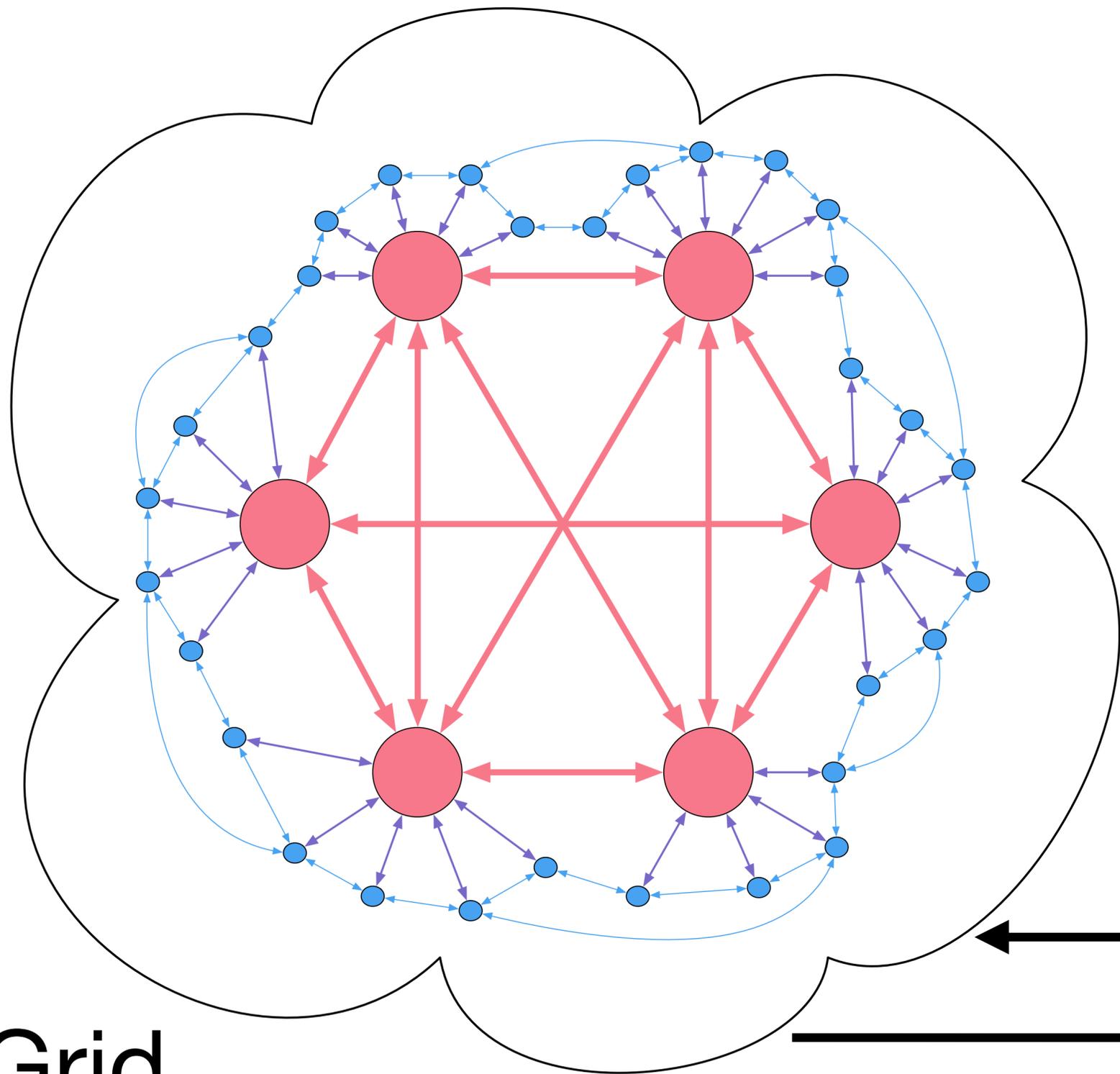


Analysis



Event generation

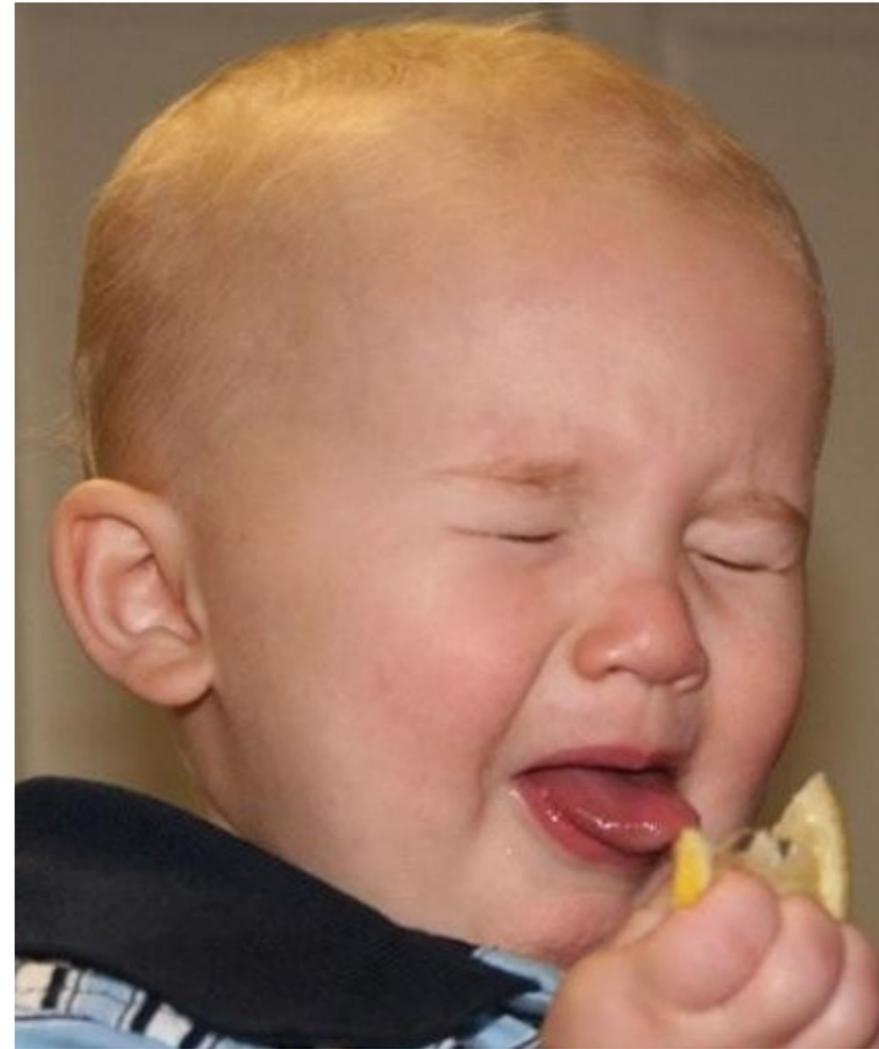
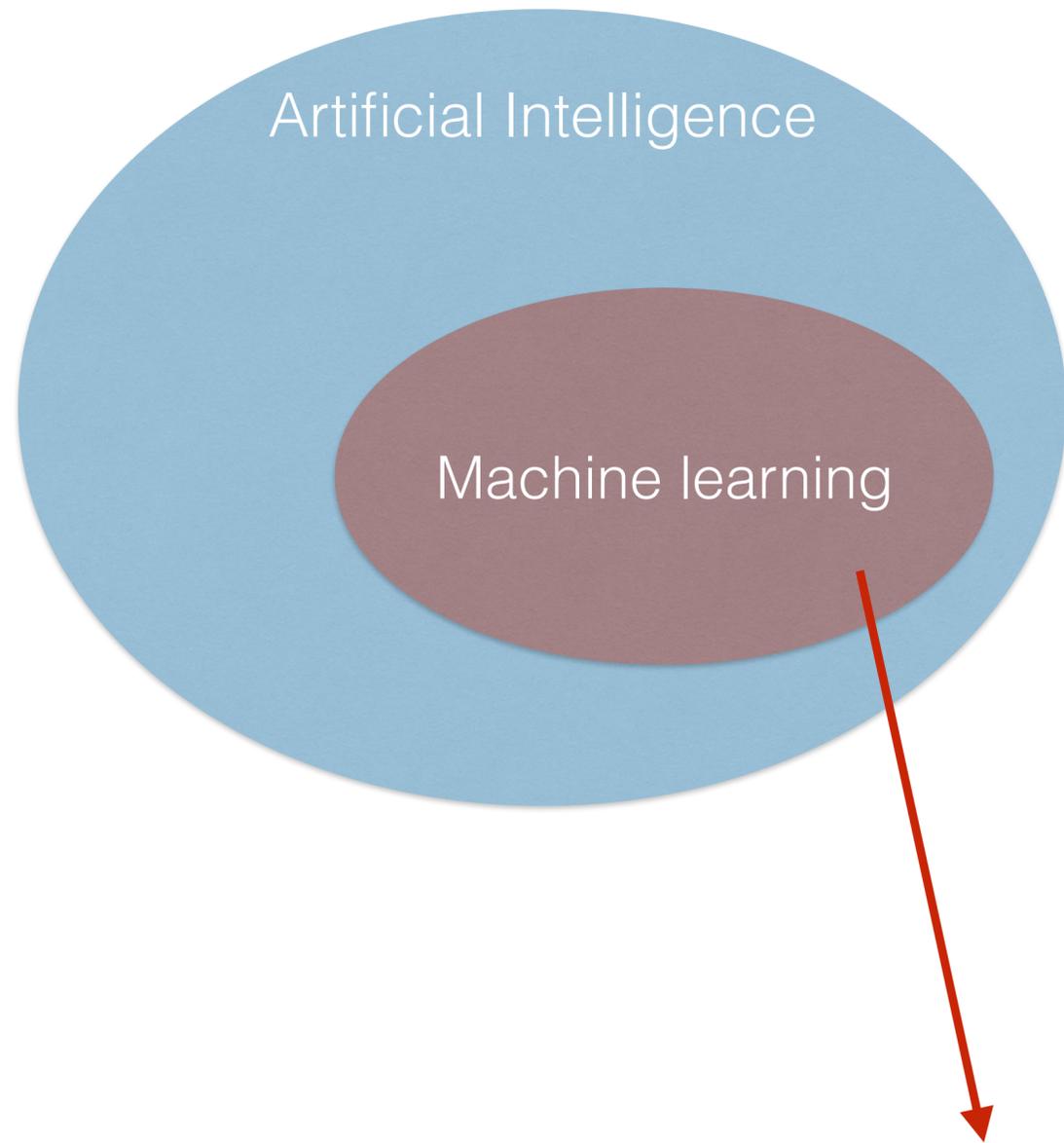
How physicists work



Request for data processing

Results

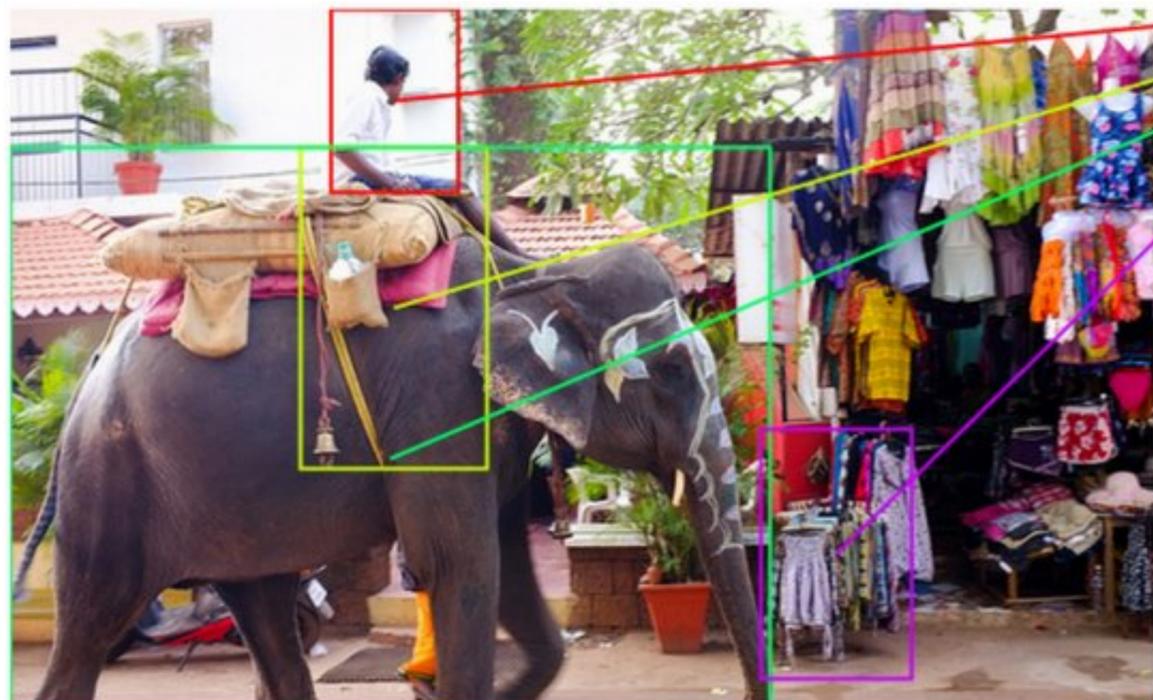
Grid



uses statistical inference to extract generalities from “training” data

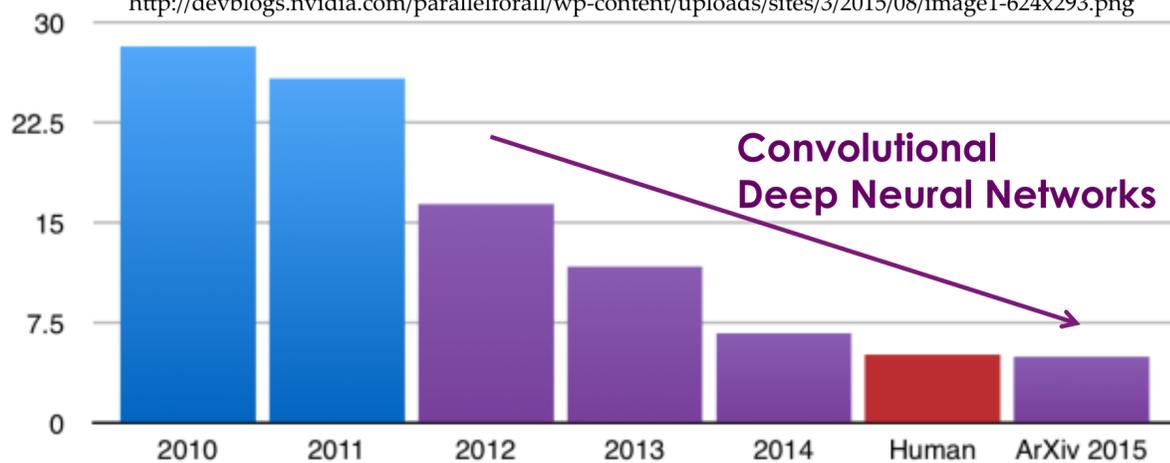
→ **“learns” from the training data**

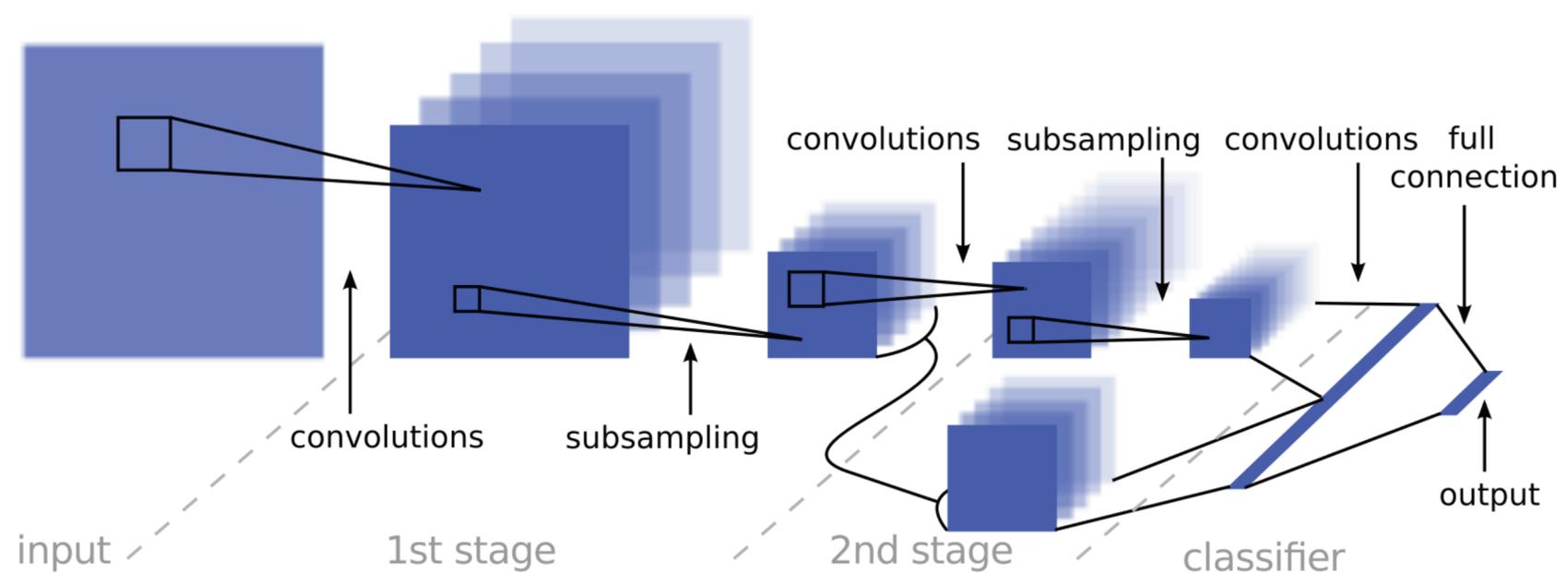
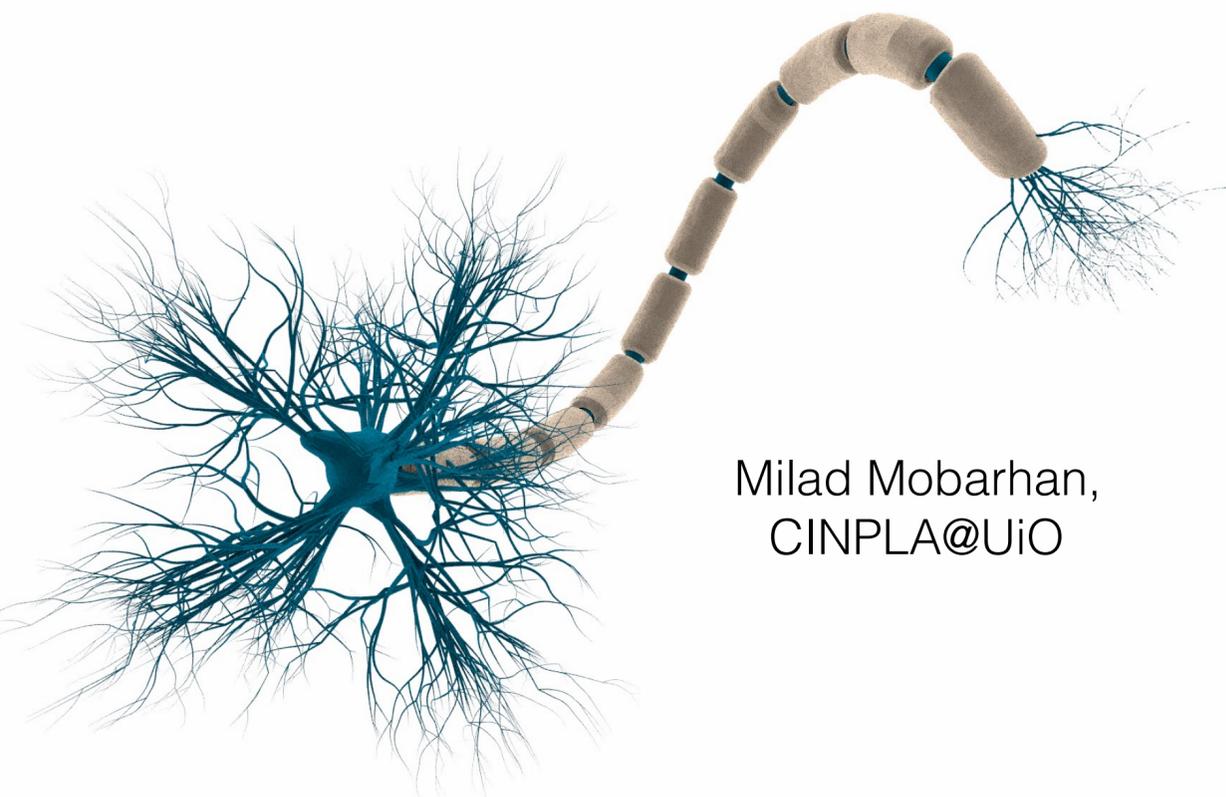
→ when exposed to new data, ***demonstrates behaviours that have not been explicitly programmed***



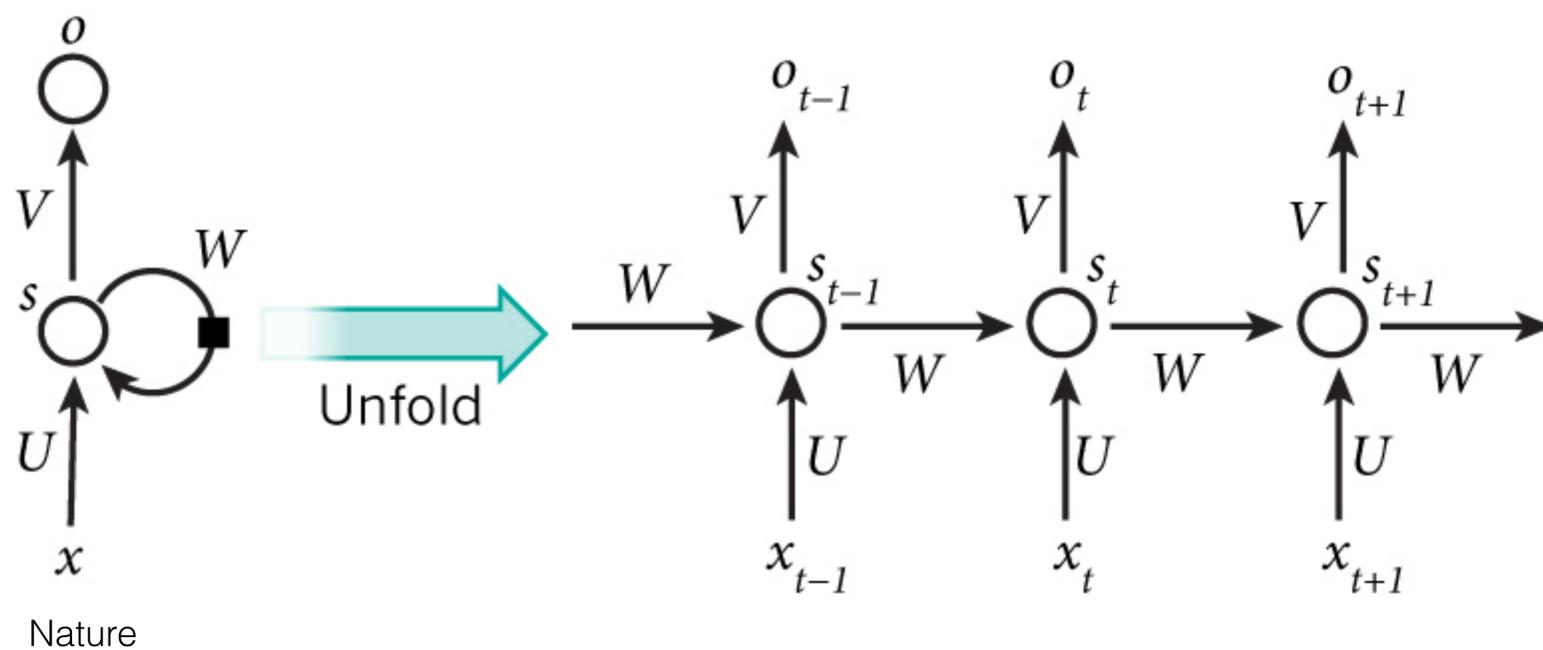
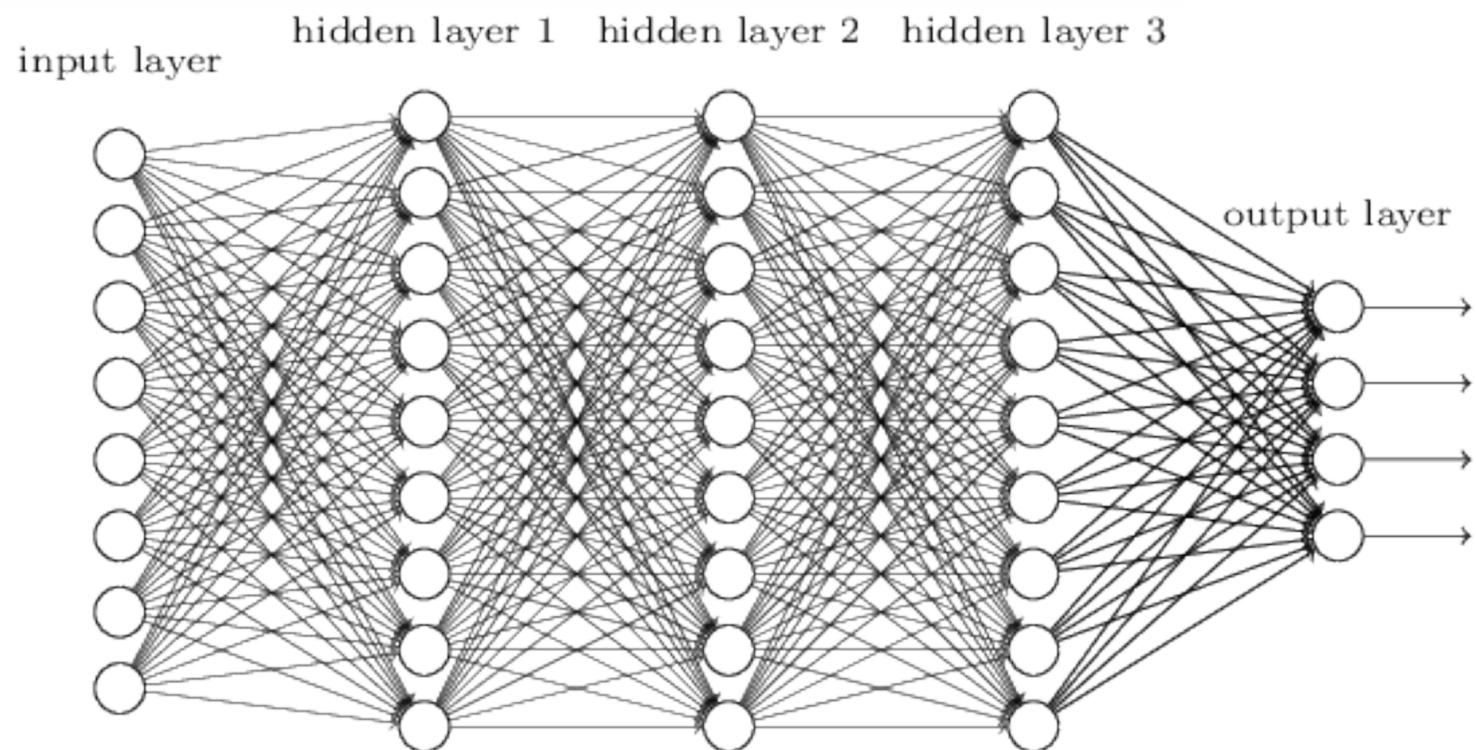
ILSVRC top-5 error on ImageNet

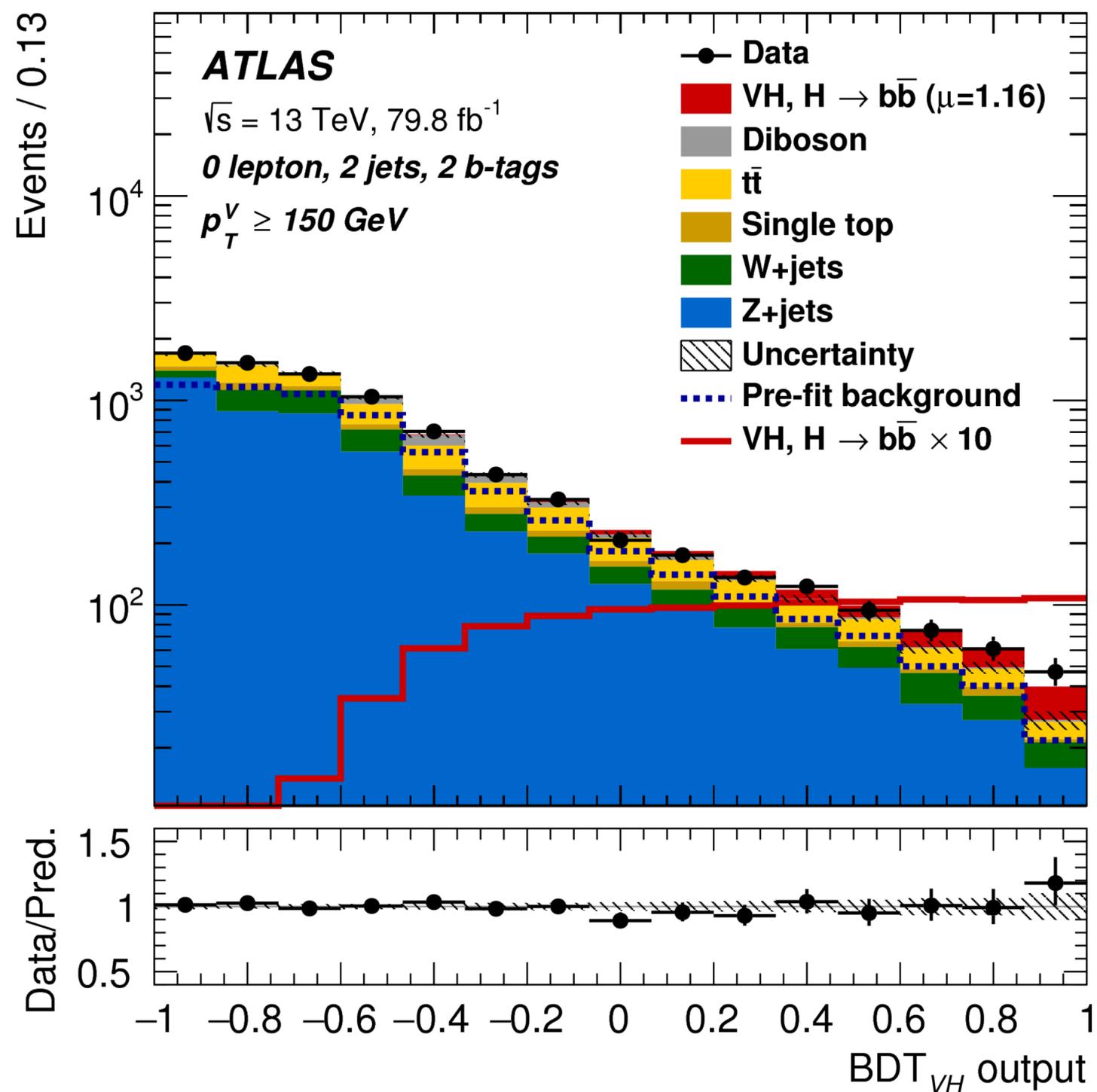
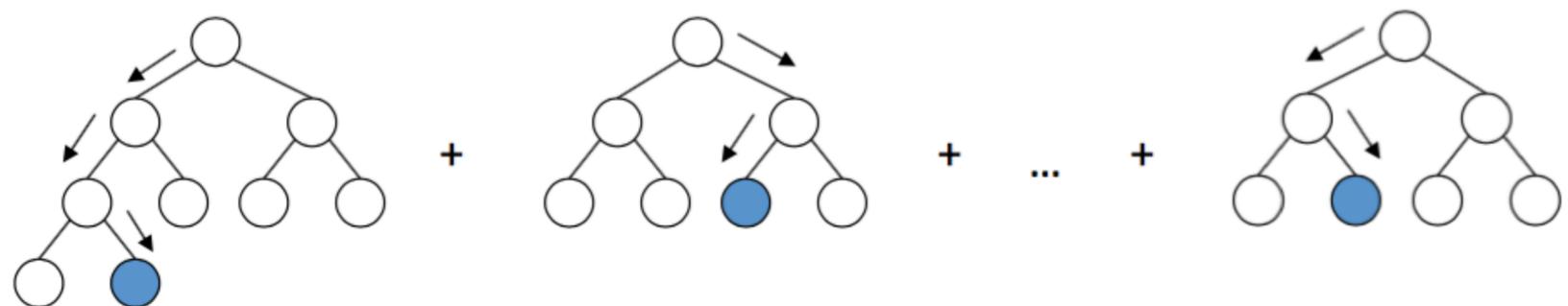
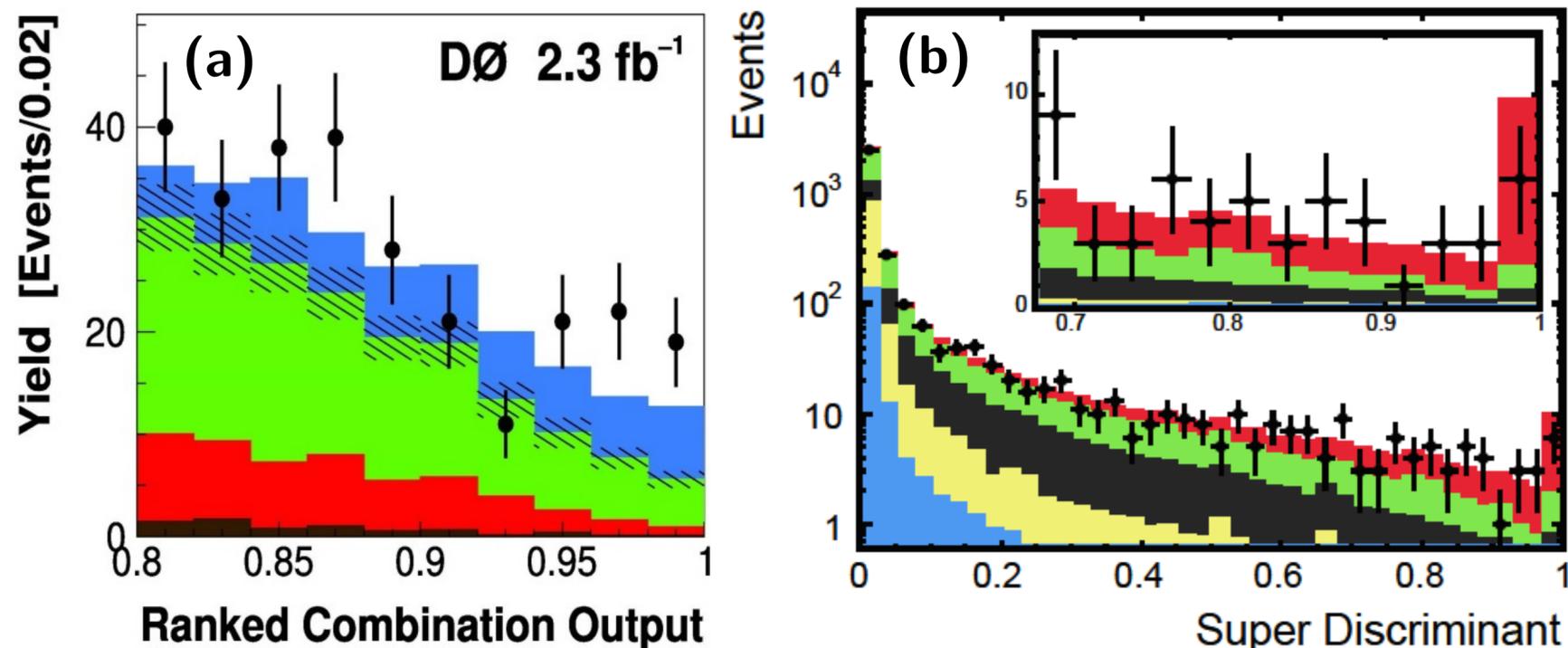
<http://devblogs.nvidia.com/paralleforall/wp-content/uploads/sites/3/2015/08/image1-624x293.png>

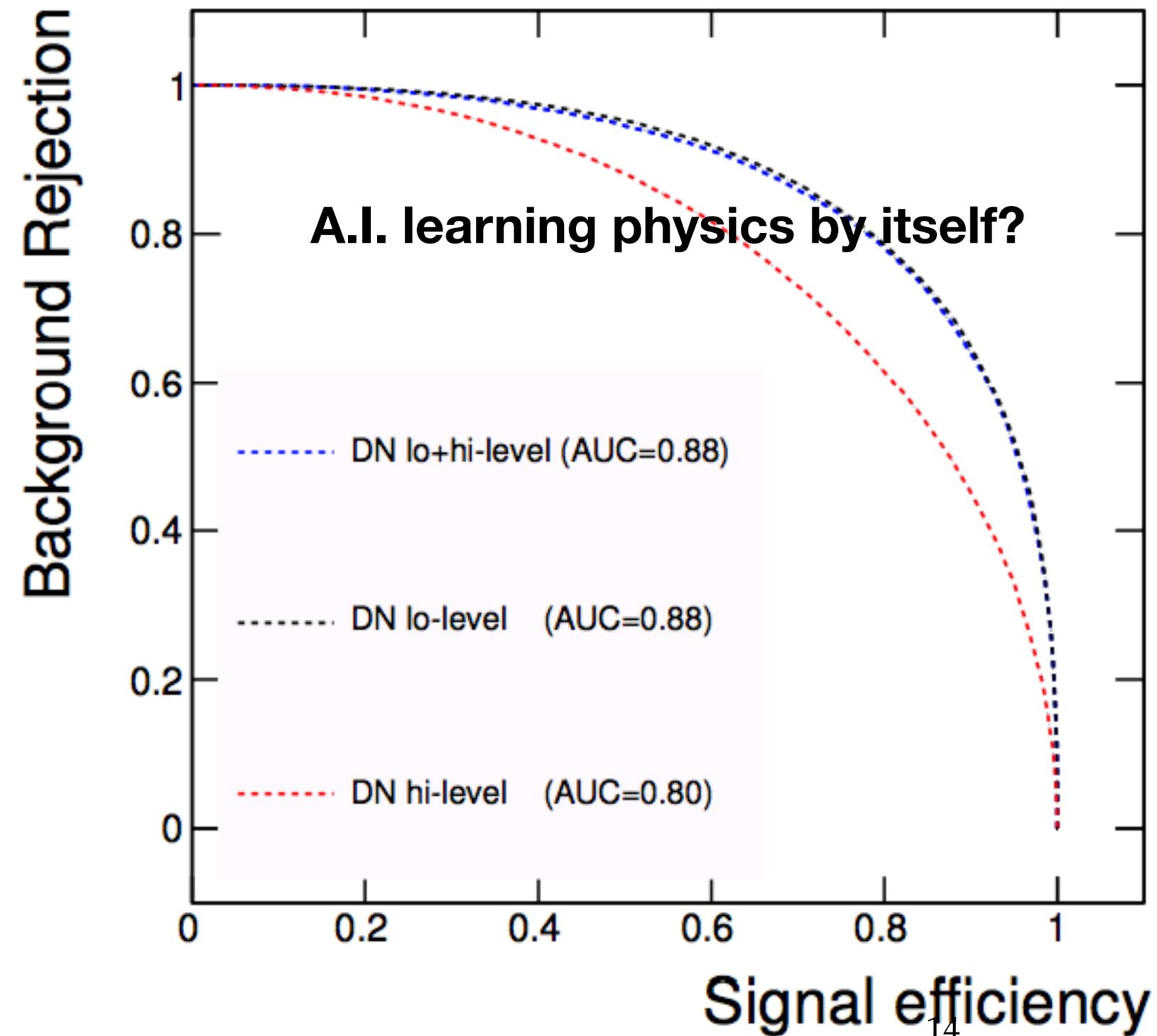




P. Sermanet & Y. LeCun, DOI: 10.1109/IJCNN.2011.6033589.







Data archival with DNA?