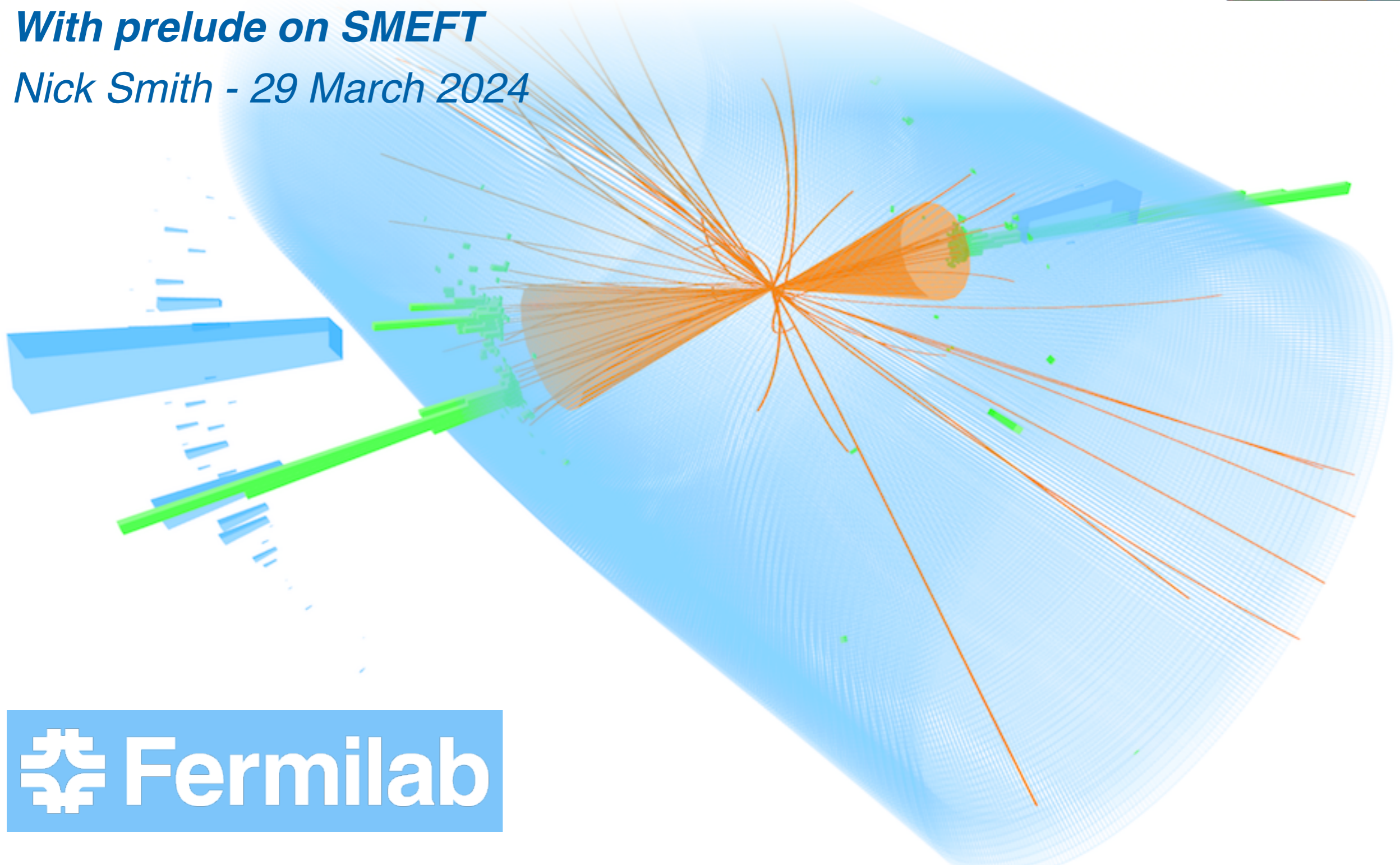


# Inferring nature at scale

*Building sustainable software and computing infrastructure for Particle Physics*

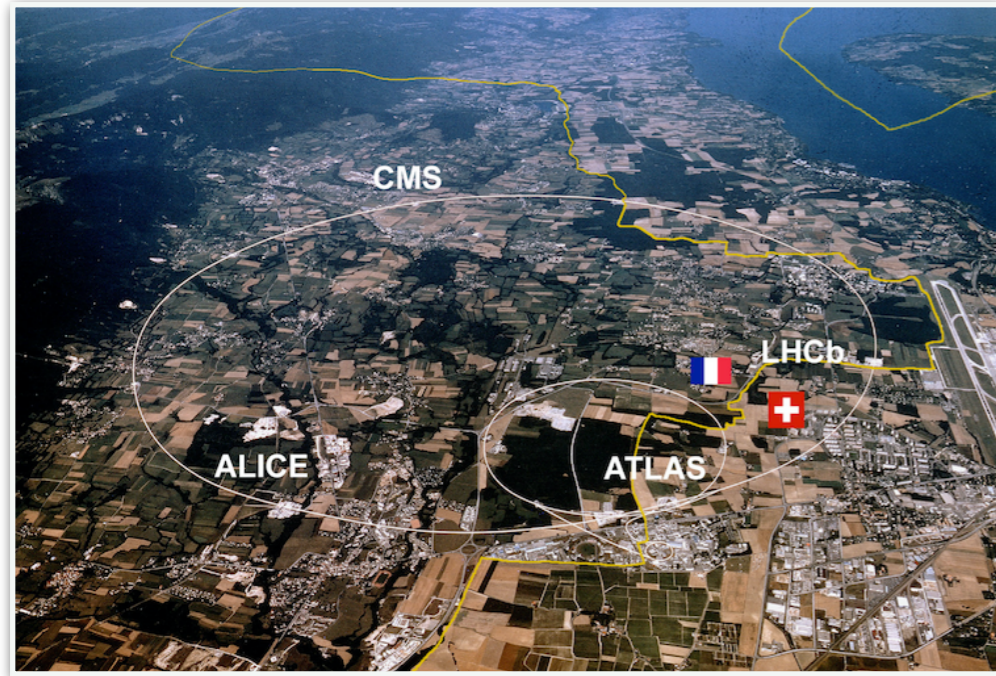
***With prelude on SMEFT***

*Nick Smith - 29 March 2024*



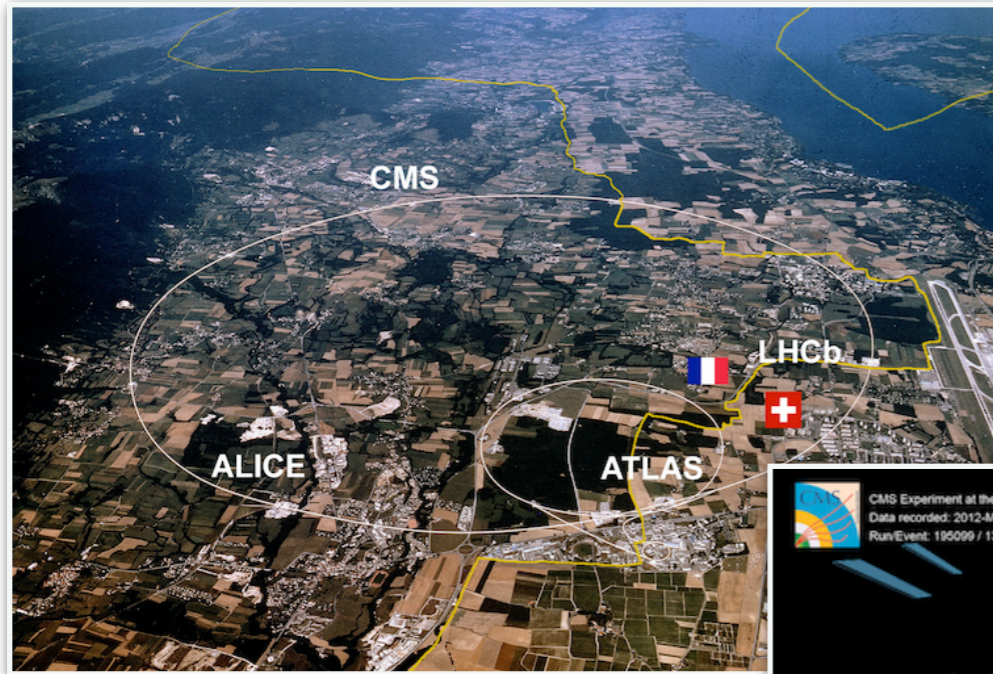
# HEP Experiment: three easy steps

# HEP Experiment: three easy steps



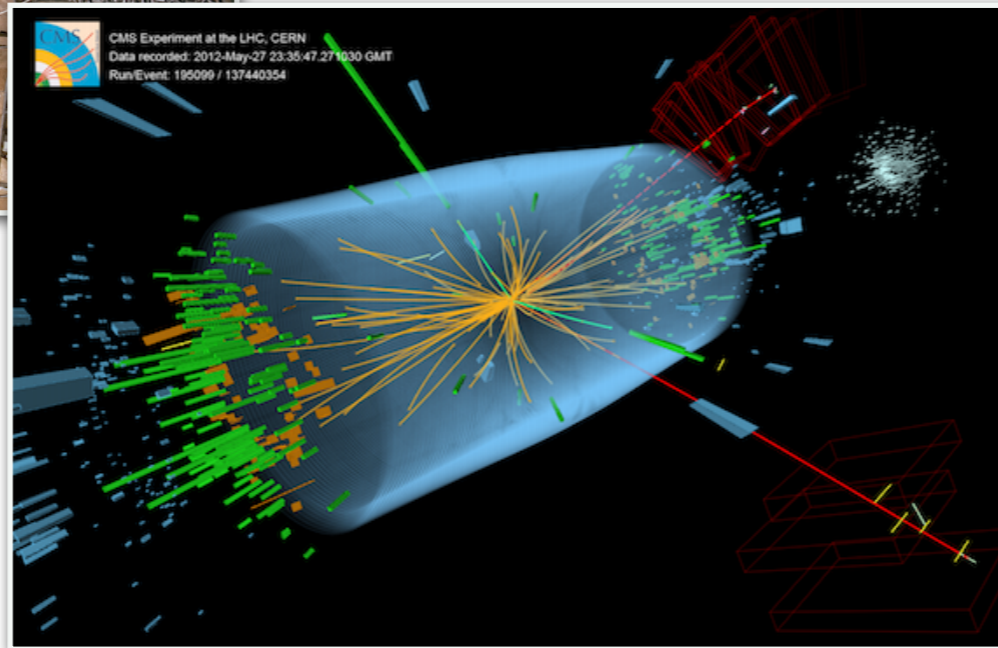
## 1. Collide particles

# HEP Experiment: three easy steps

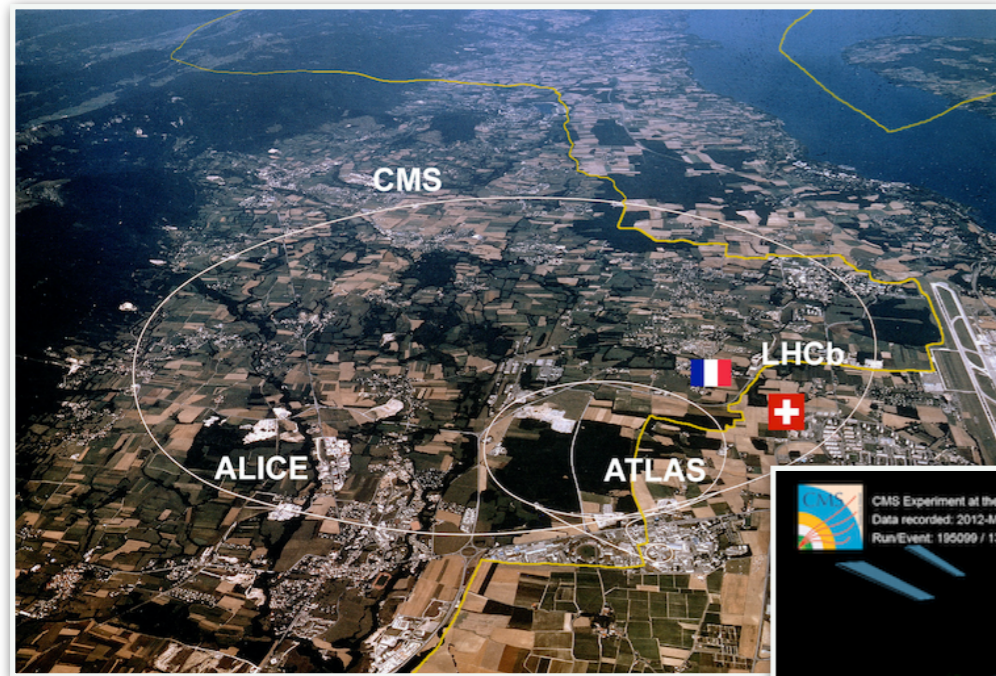


1. Collide particles

2. Take pictures



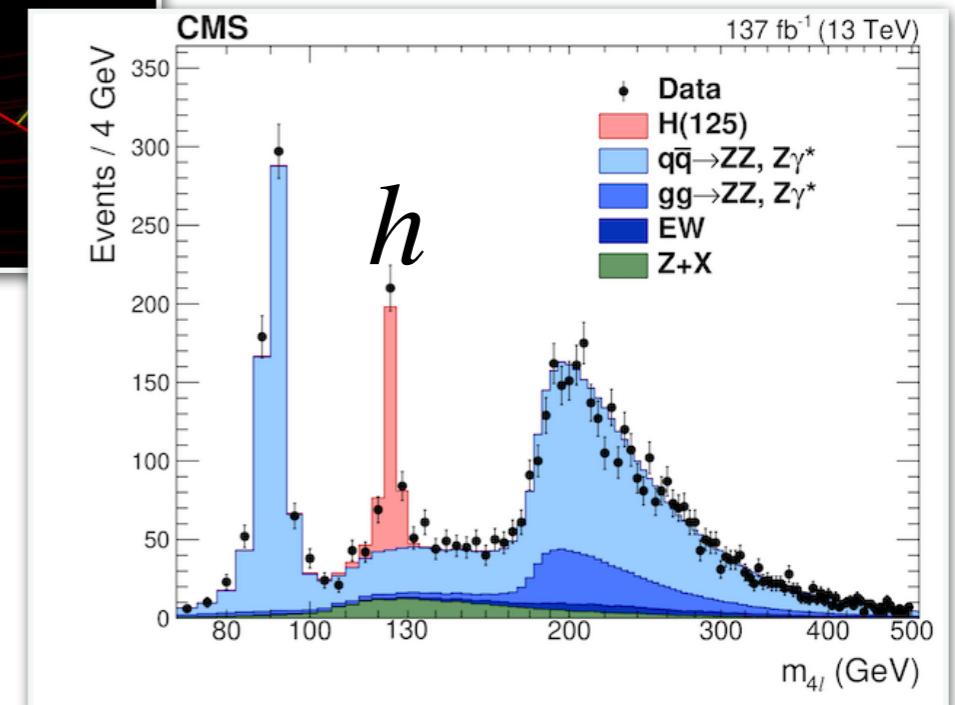
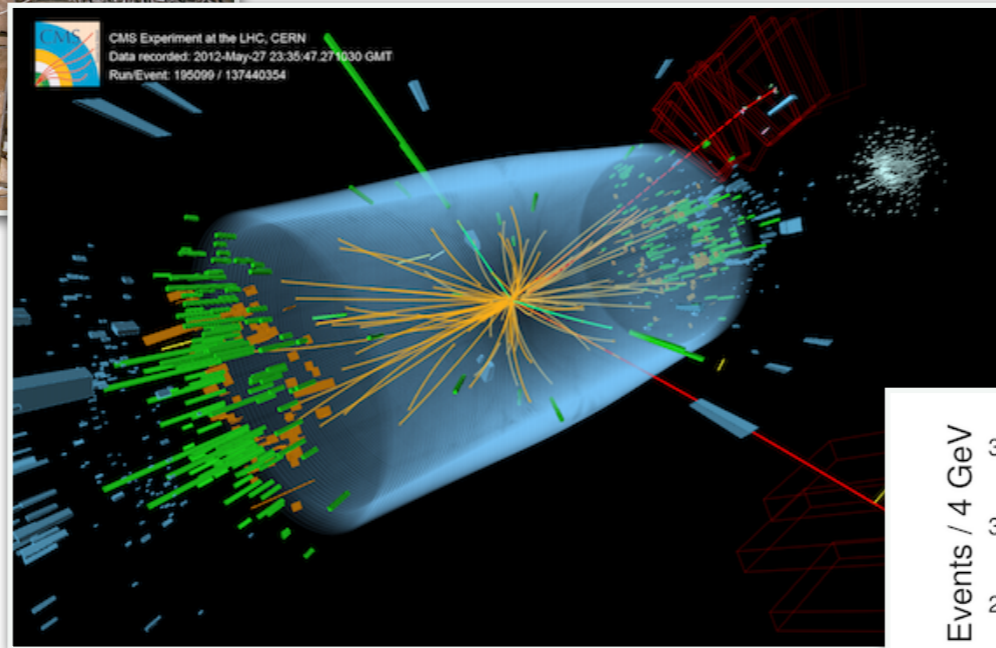
# HEP Experiment: three easy steps



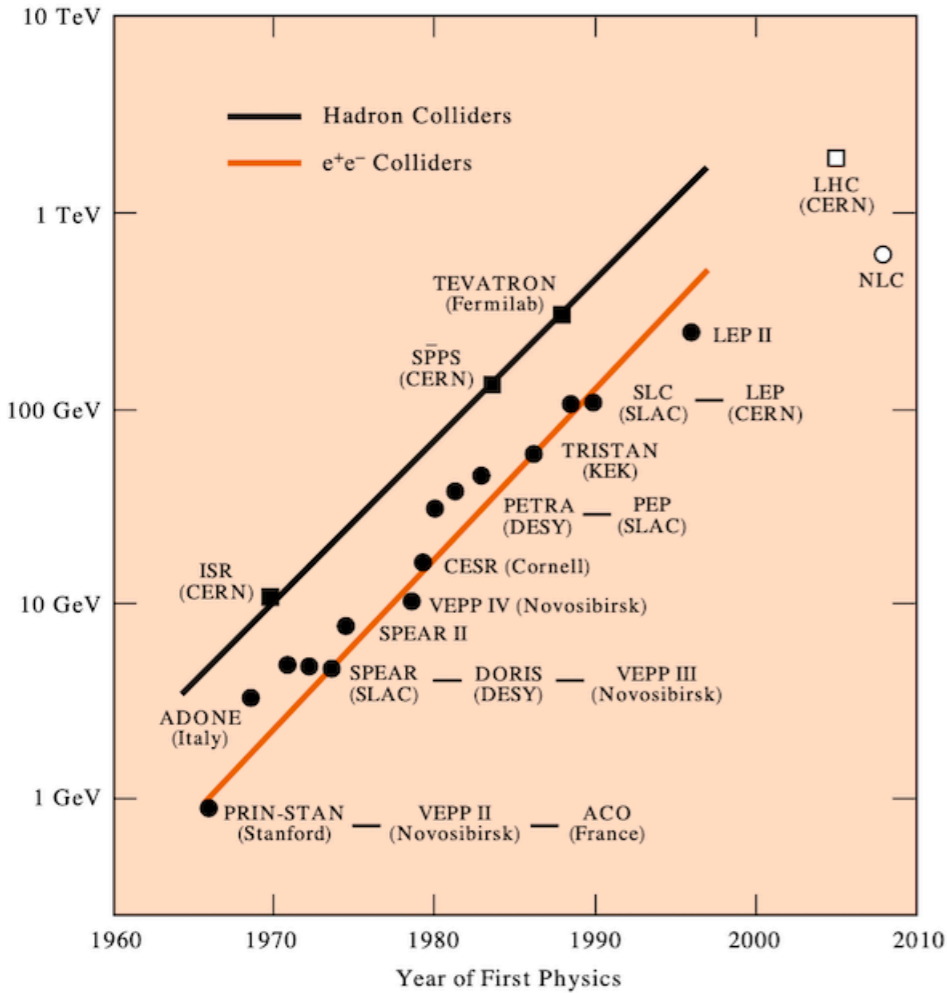
1. Collide particles

2. Take pictures

3. Infer parameters



# Step 1: Collide particles



## Step 2: take pictures



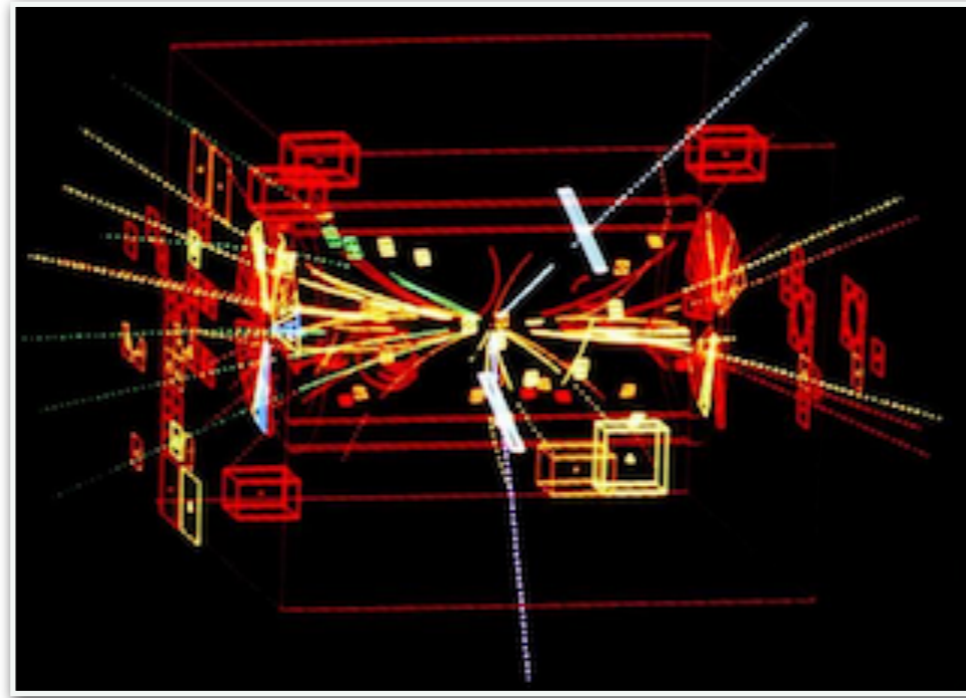
1 photo / event

~6M events [cds:1733654](#)

## Step 2: take pictures



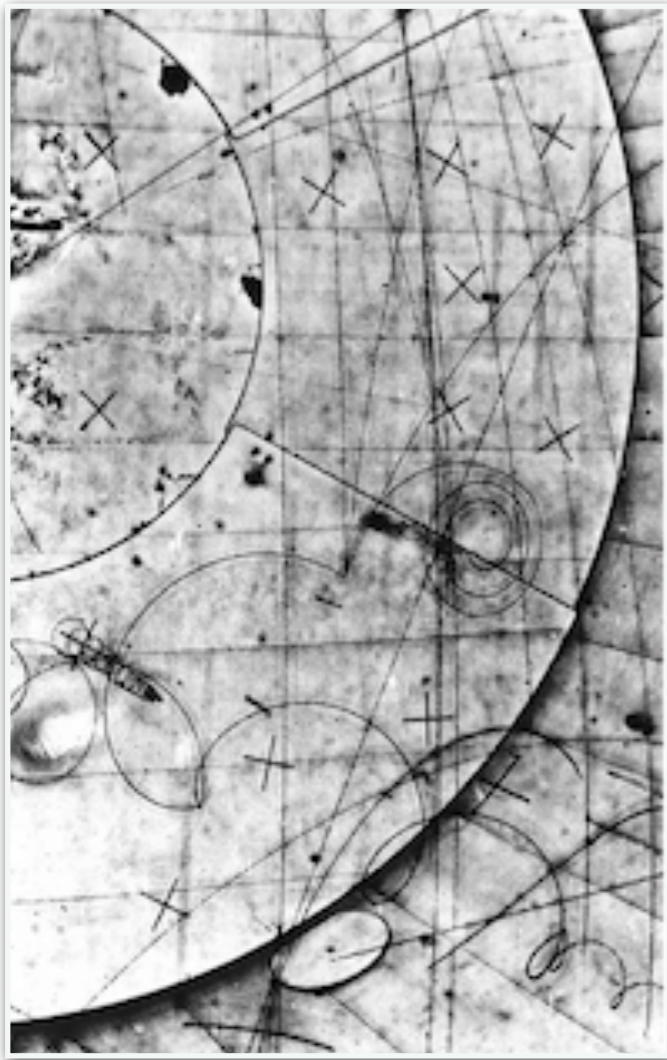
1 photo / event  
~6M events [cds:1733654](https://cds.cern.ch/record/1733654)



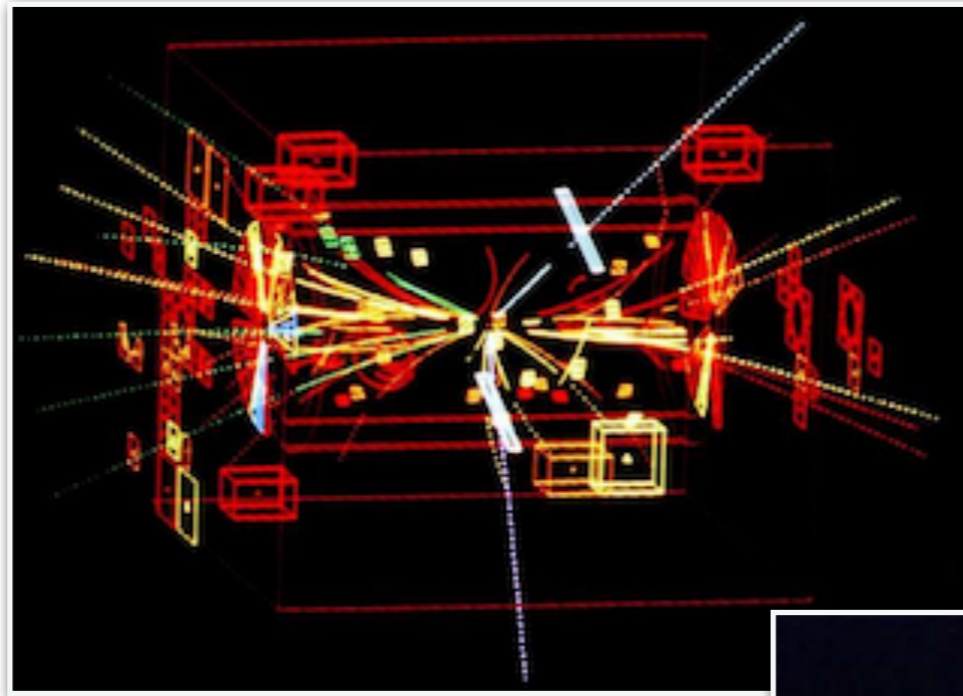
~100kB / event  
~10M events [cds:182190](https://cds.cern.ch/record/182190)



## Step 2: take pictures

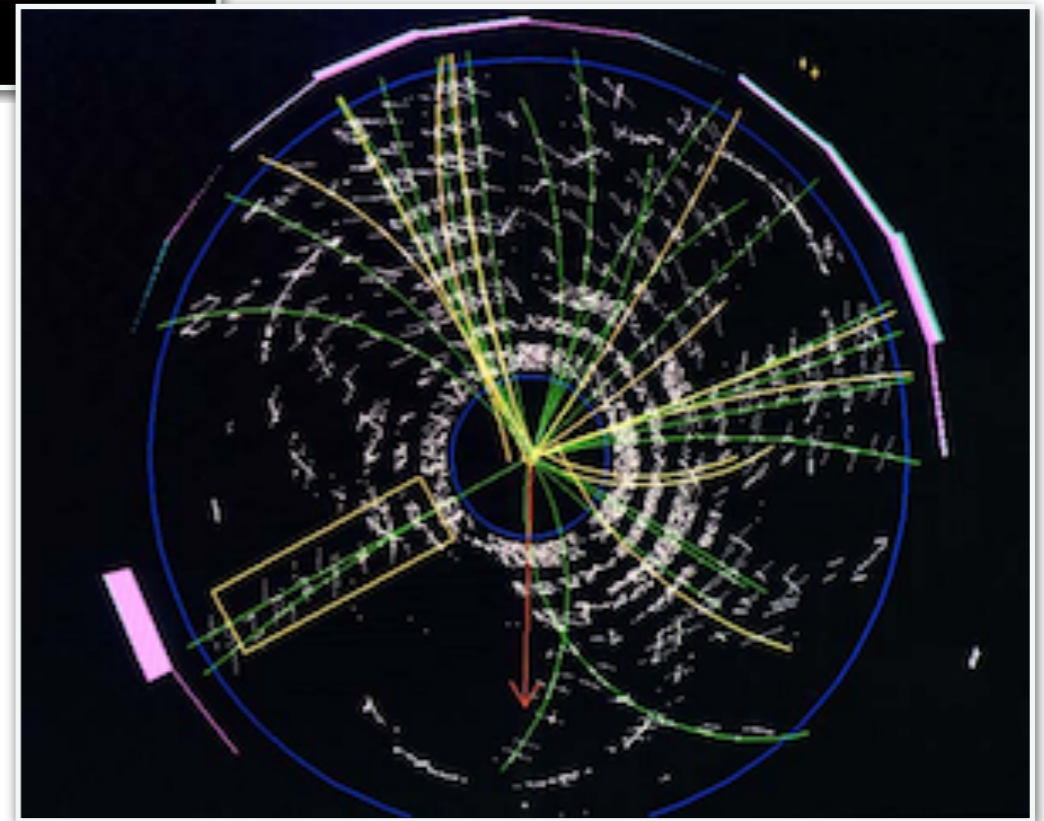


1 photo / event  
~6M events [cds:1733654](https://cds.cern.ch/record/1733654)



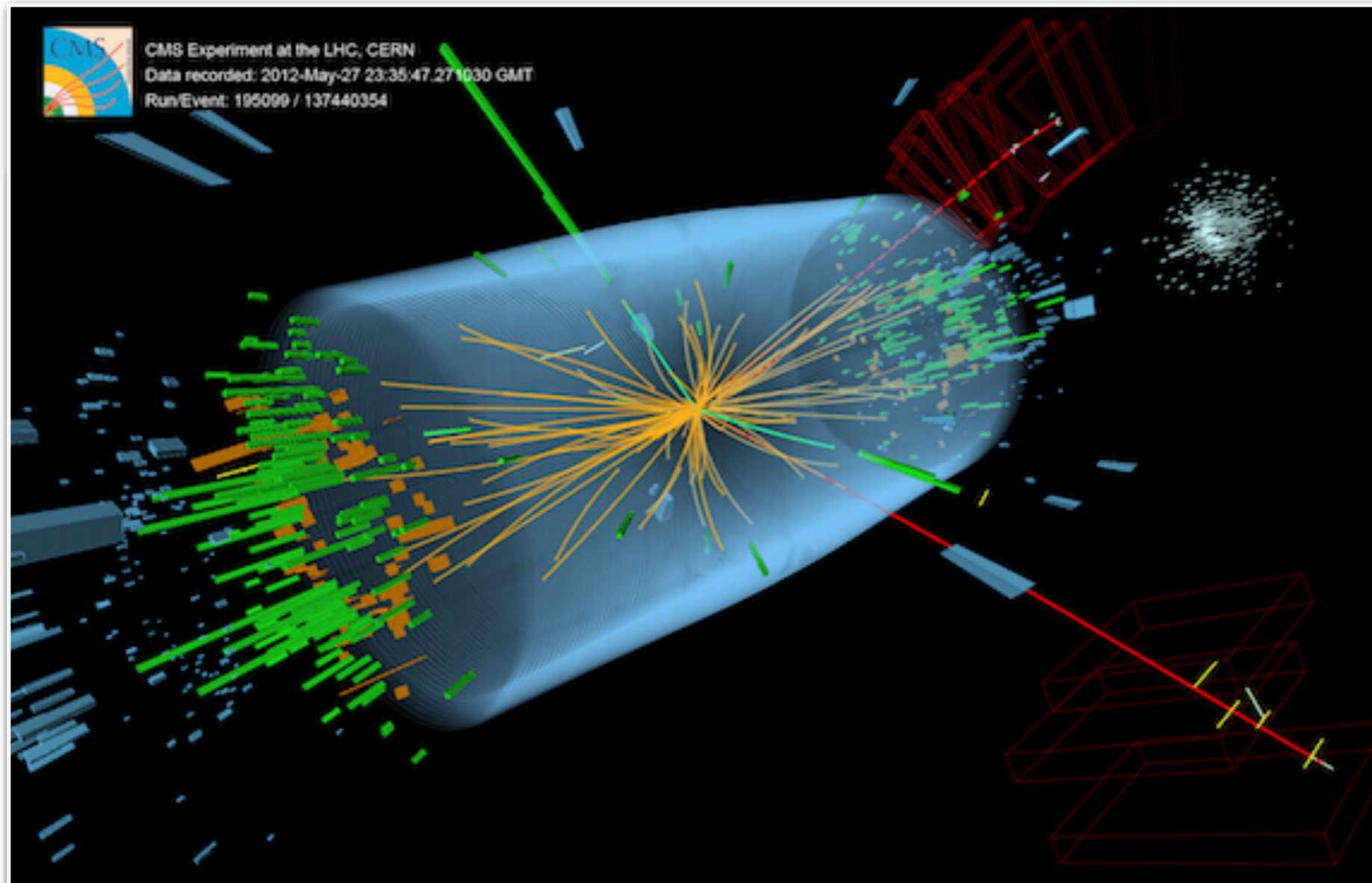
~100kB / event  
~10M events [cds:182190](https://cds.cern.ch/record/182190)

~100kB / event  
~1B events [10.1016/j.nima.2017.01.043](https://cds.cern.ch/record/10.1016/j.nima.2017.01.043)



## Step 2: take pictures

~1MB / event  
~100B events



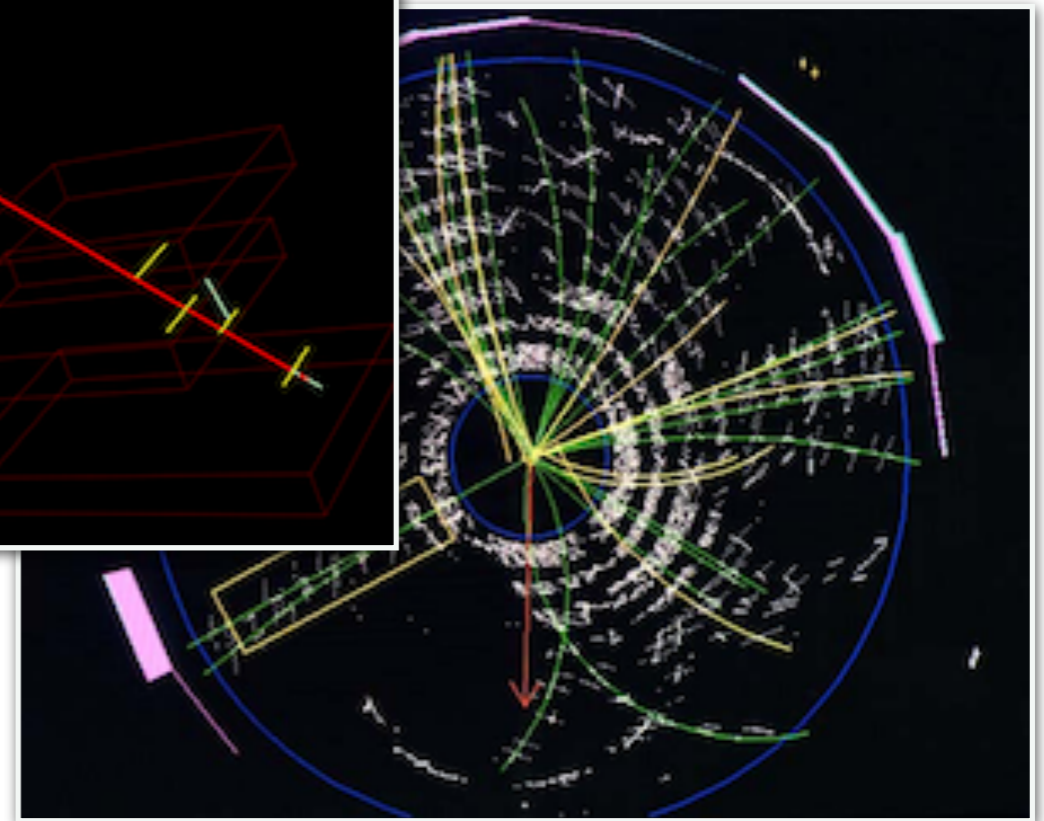
10kB / event

100 events [10.1016/j.nima.2017.01.043](https://doi.org/10.1016/j.nima.2017.01.043)



1 photo / event

~6M events [cds:1733654](https://cds.cern.ch/record/1733654)





# Step 3: infer parameters

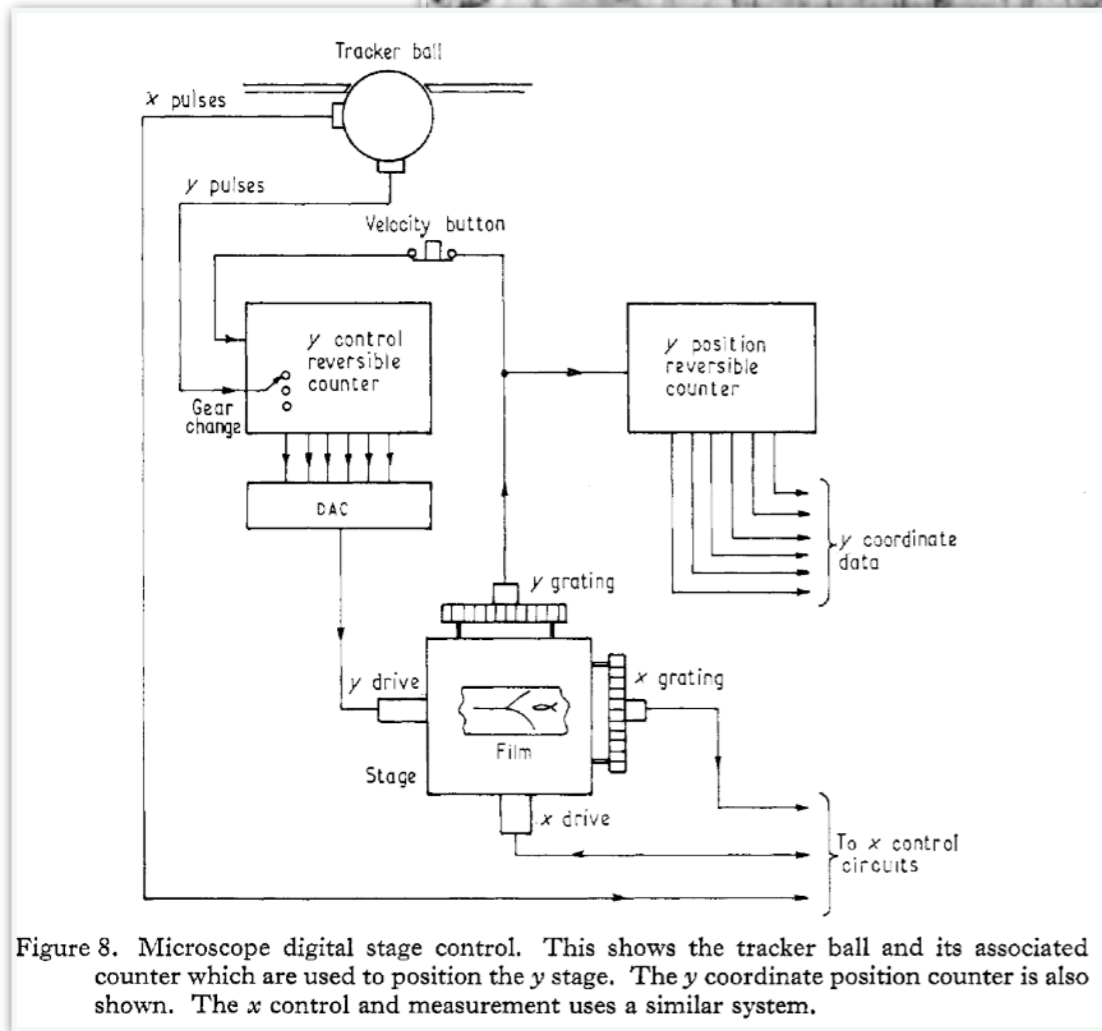
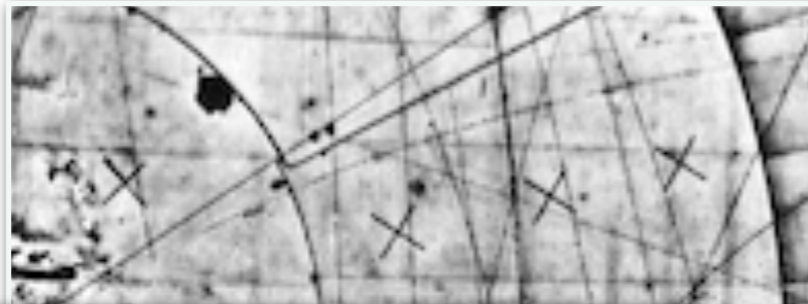
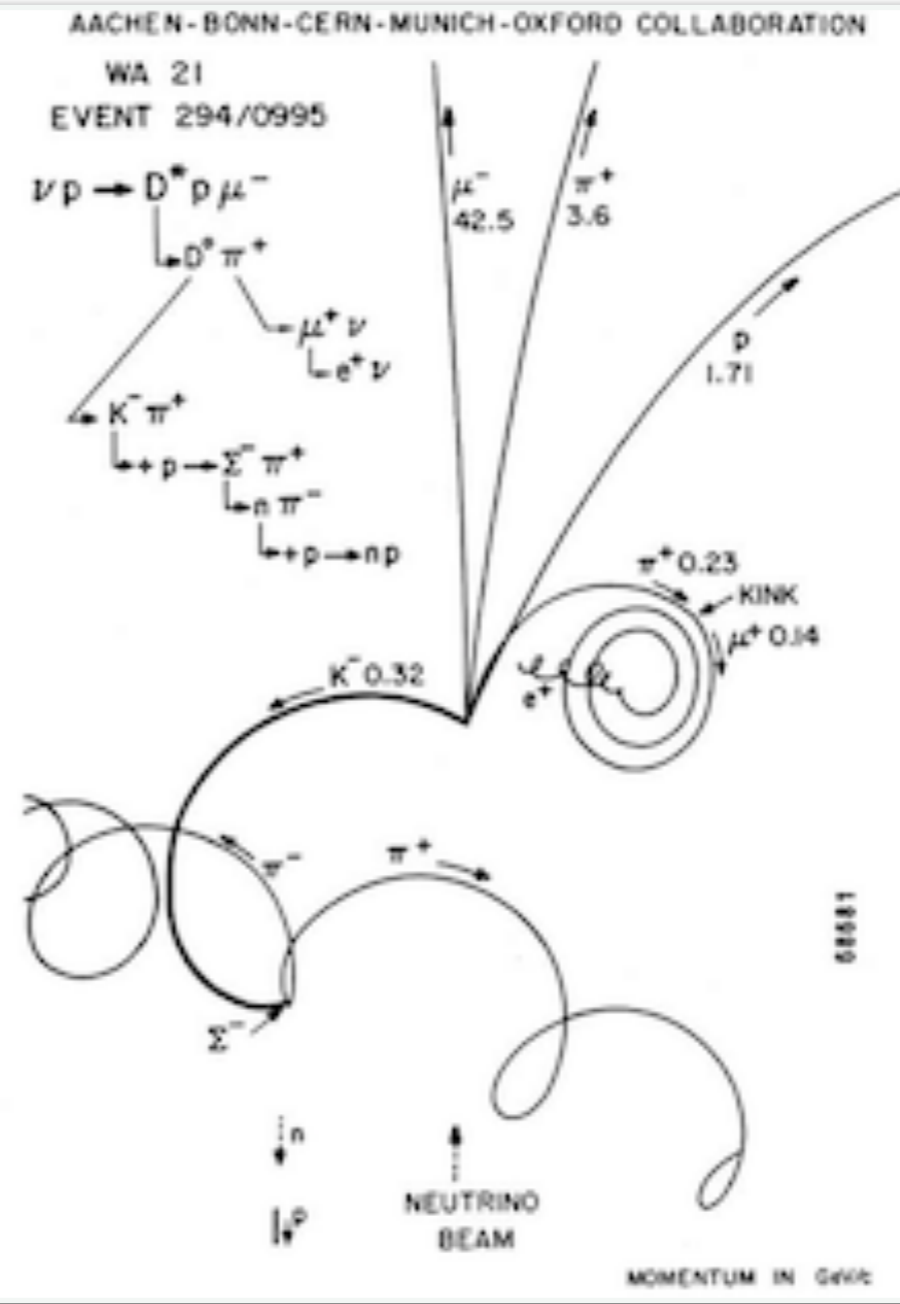
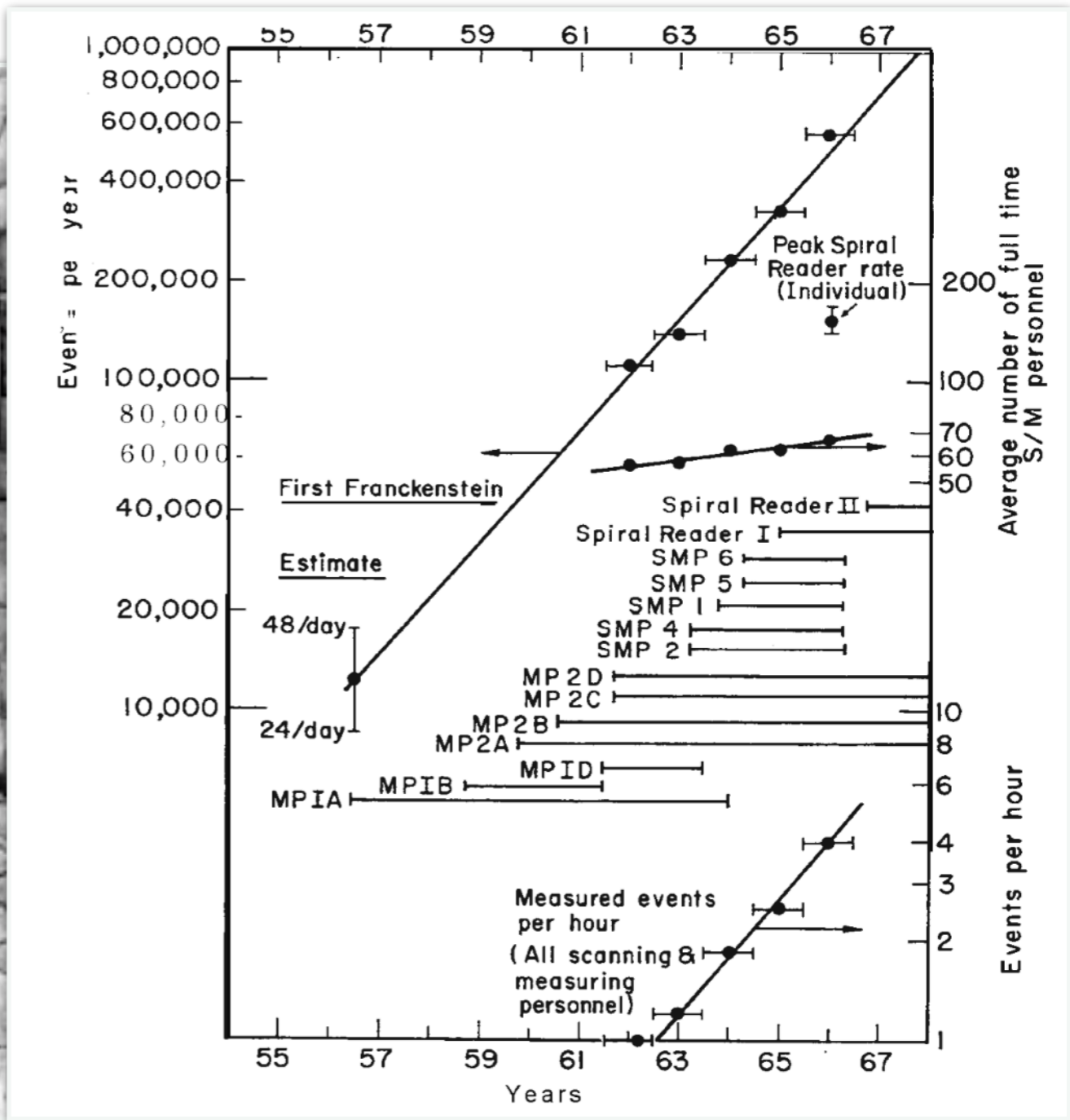
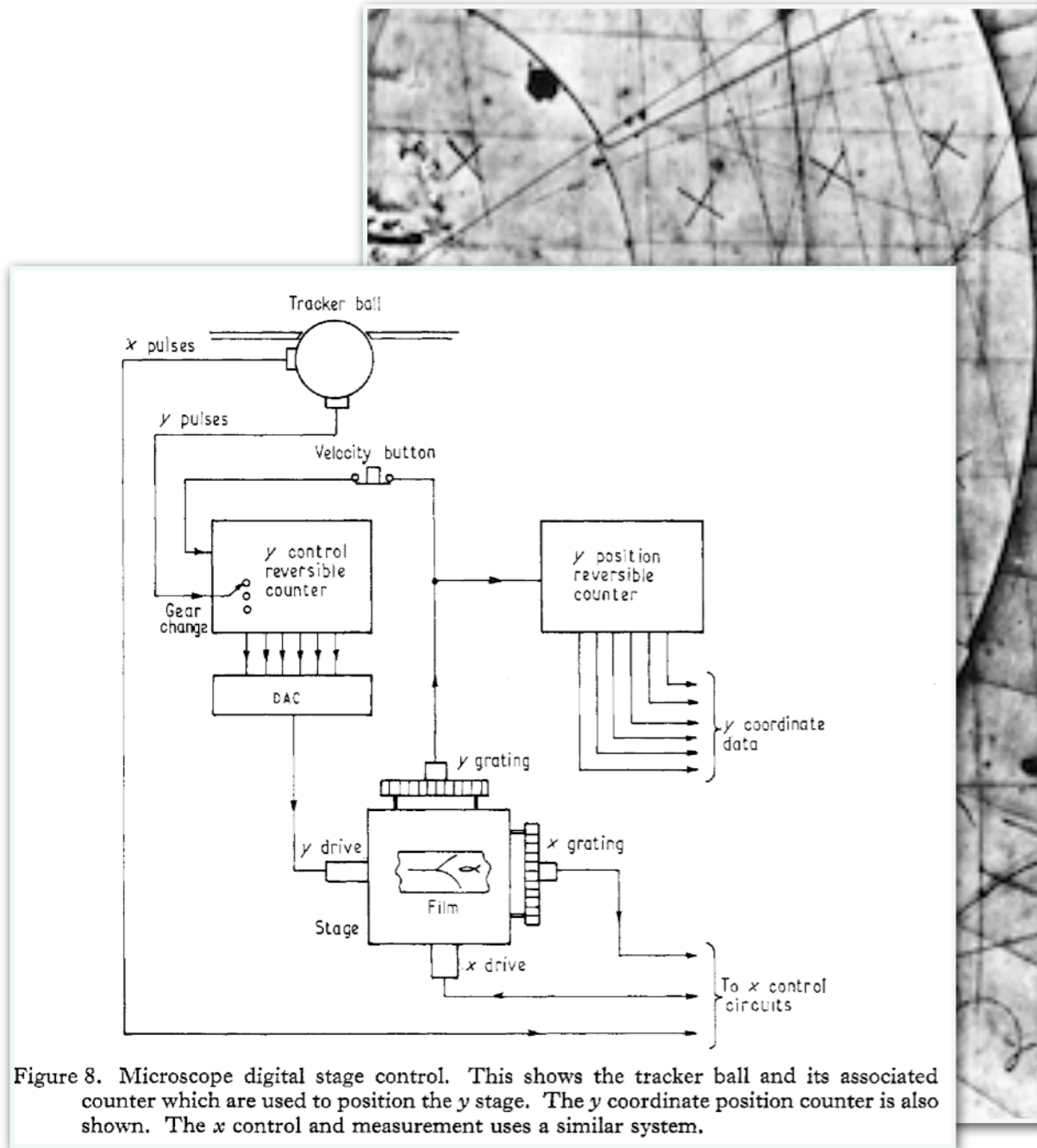


Figure 8. Microscope digital stage control. This shows the tracker ball and its associated counter which are used to position the y stage. The y coordinate position counter is also shown. The x control and measurement uses a similar system.



# Step 3: infer parameters



# Step 3: infer parameters

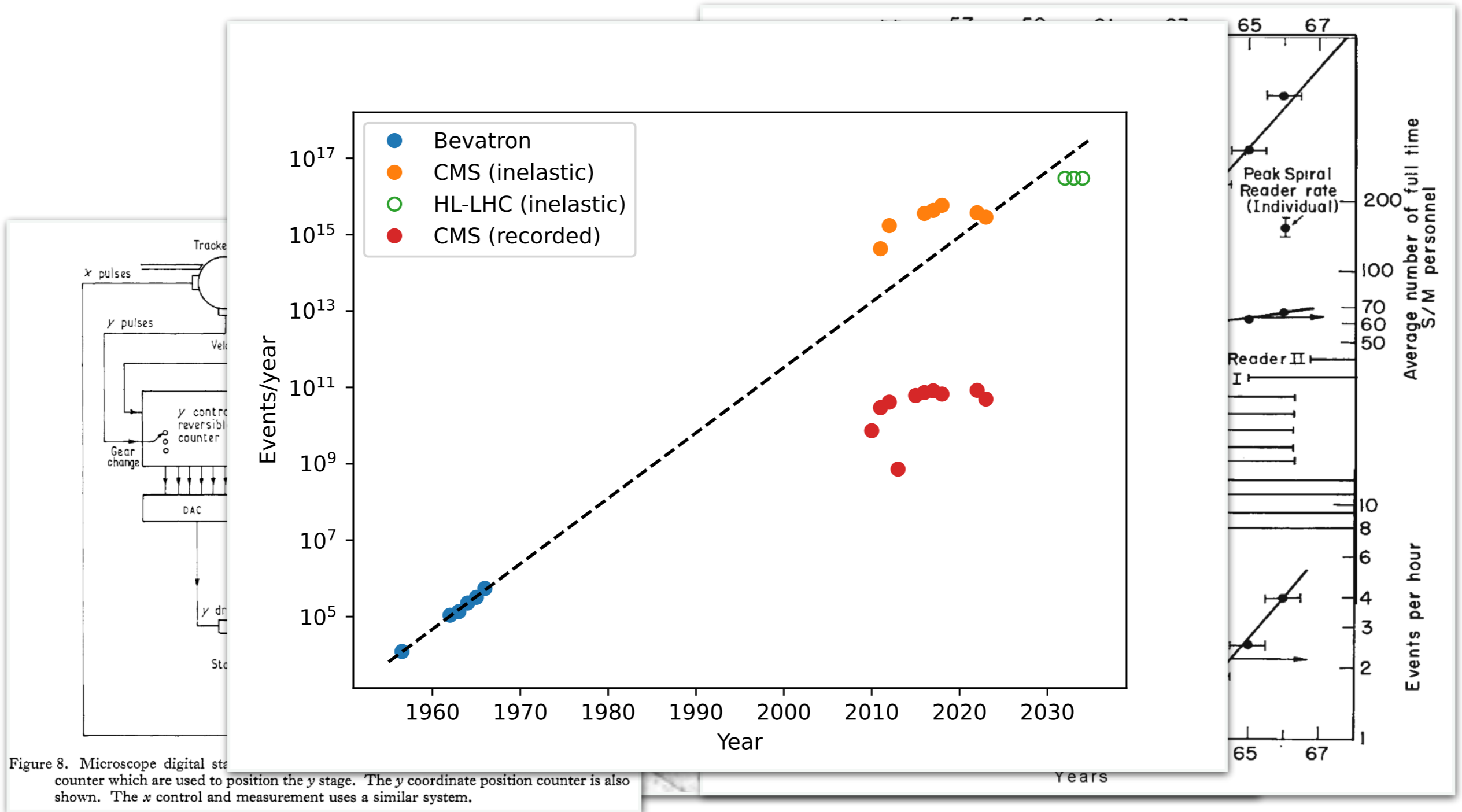
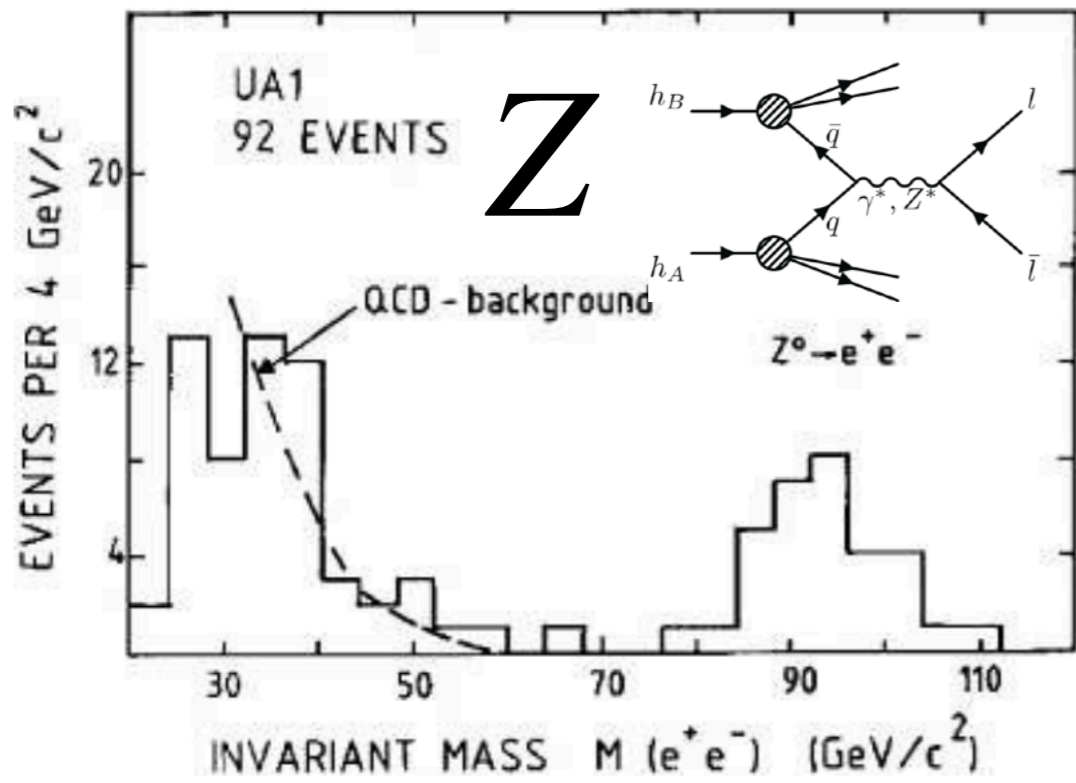


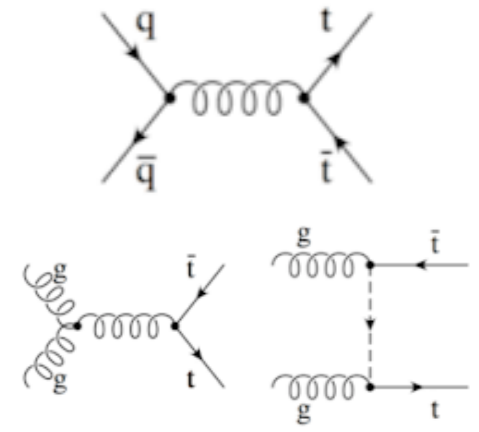
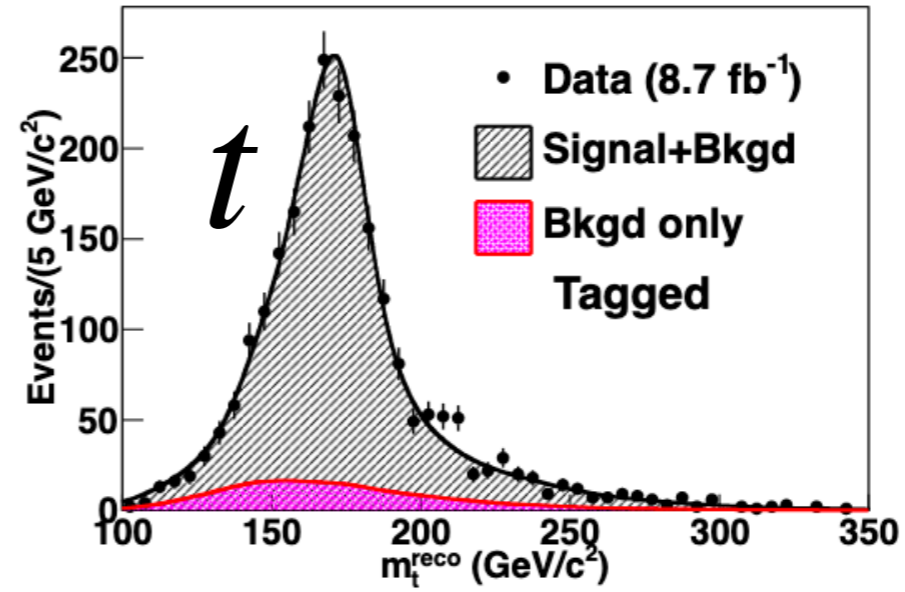
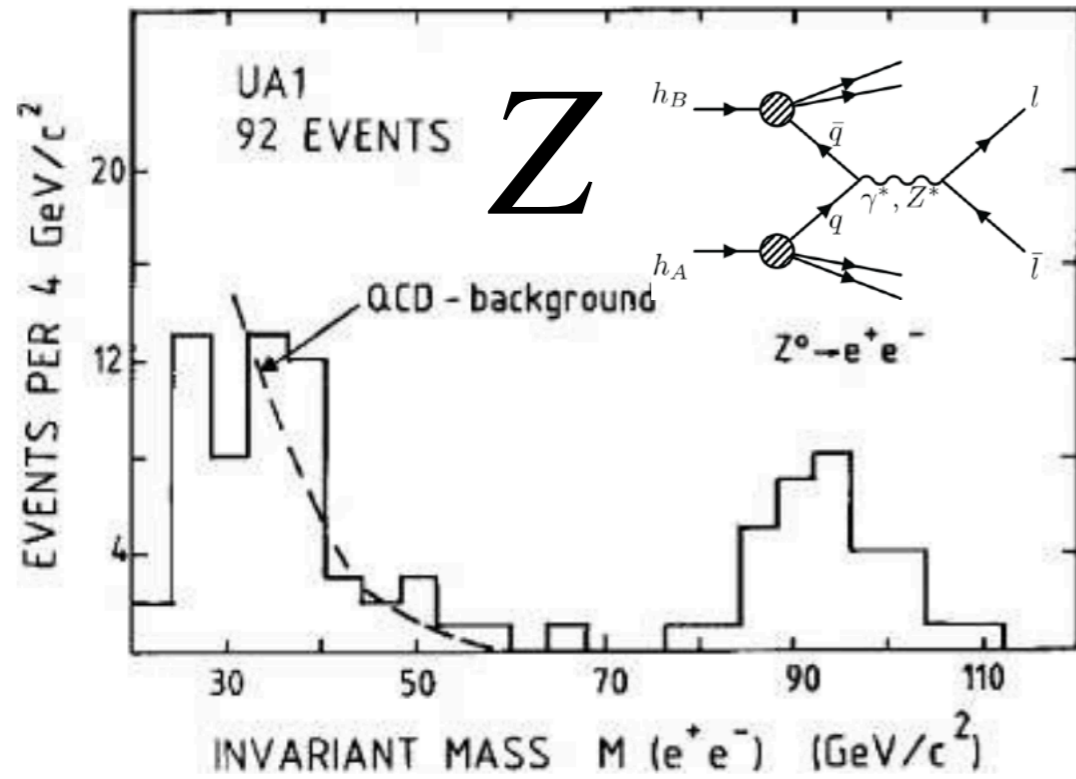
Figure 8. Microscope digital stage. The x control and measurement uses a similar system. The y coordinate position counter is also shown.



# Step 3: infer parameters

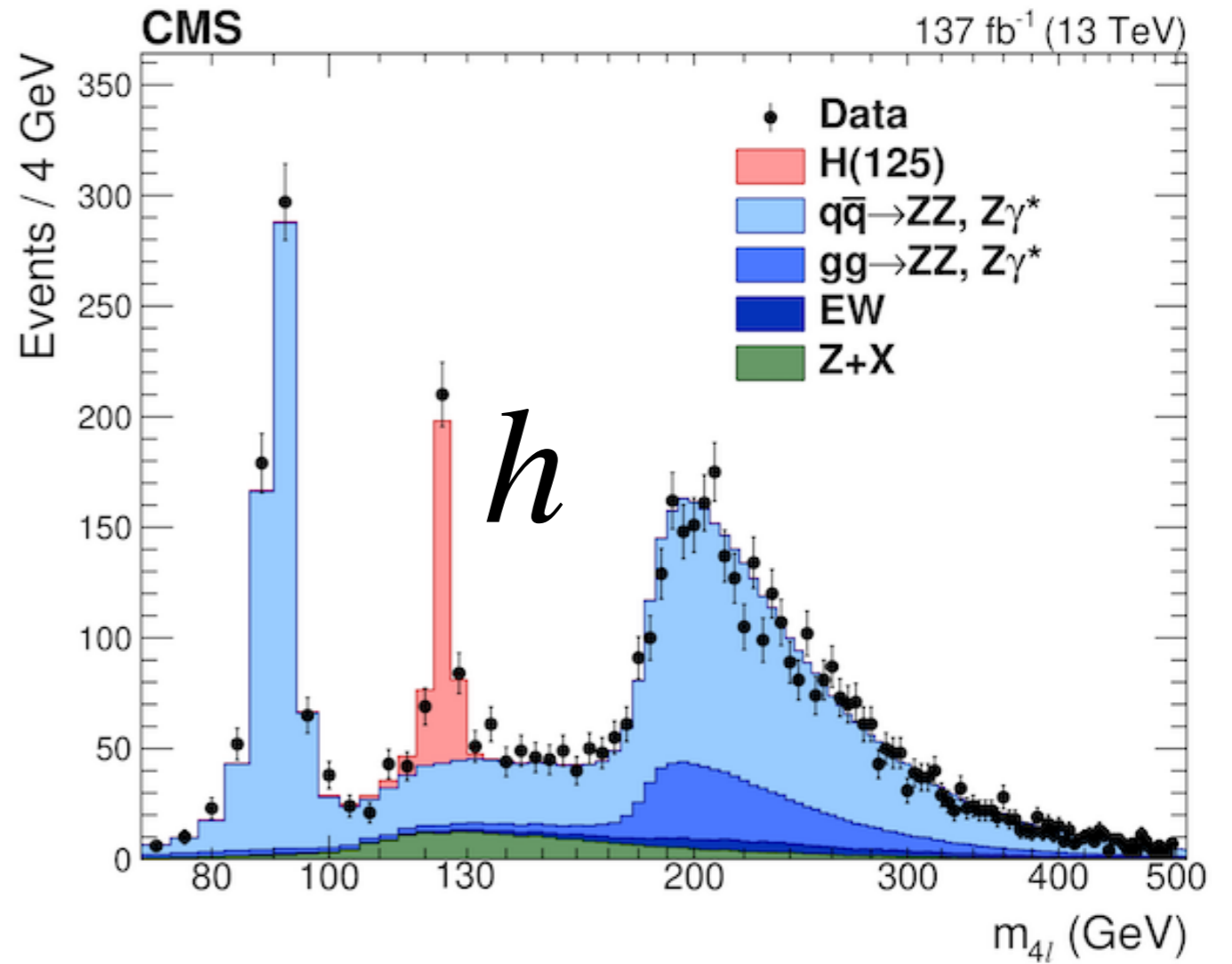
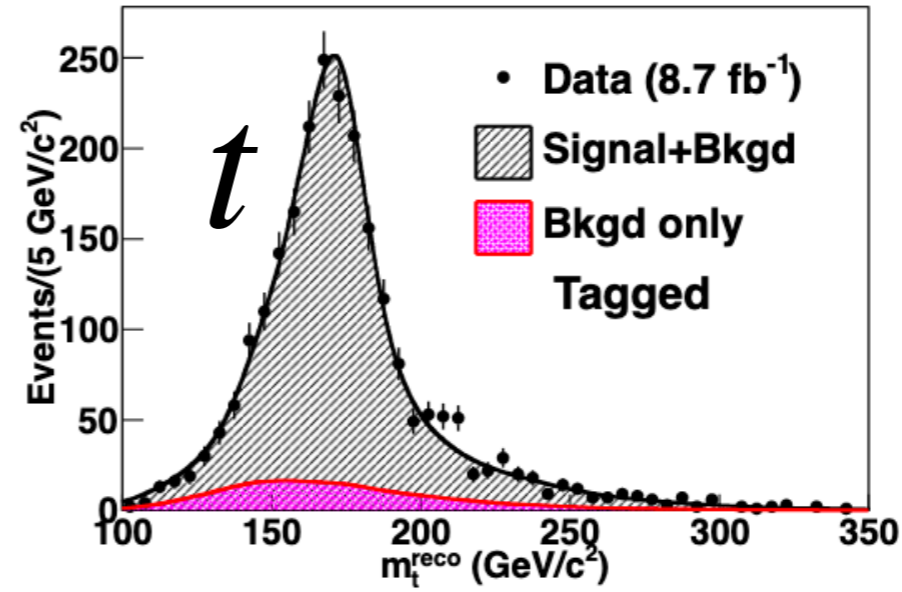
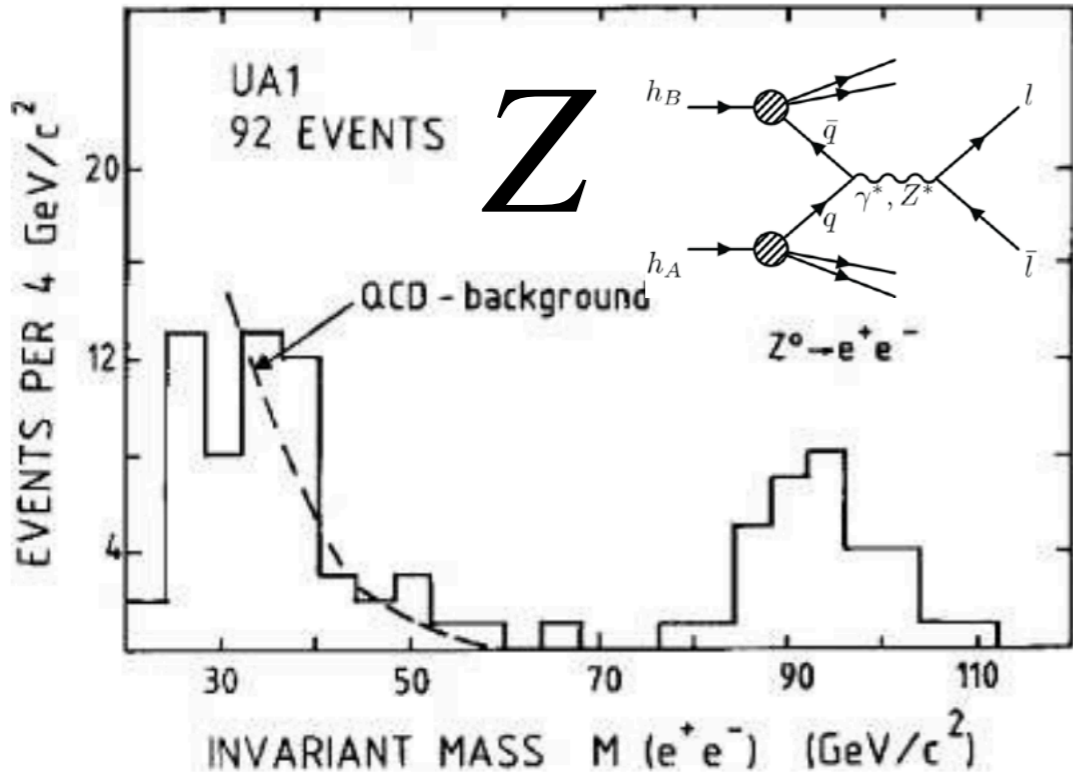


# Step 3: infer parameters

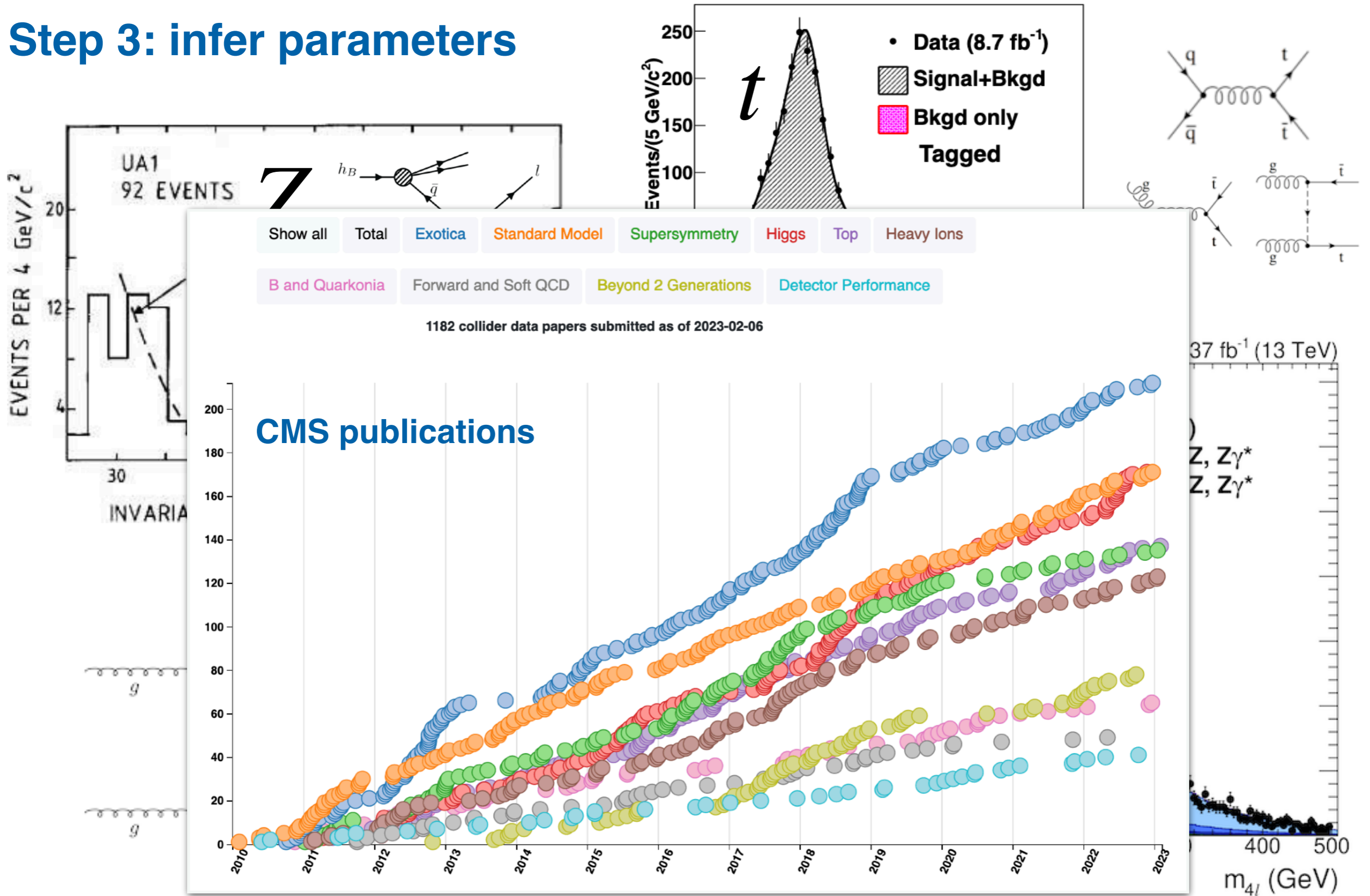




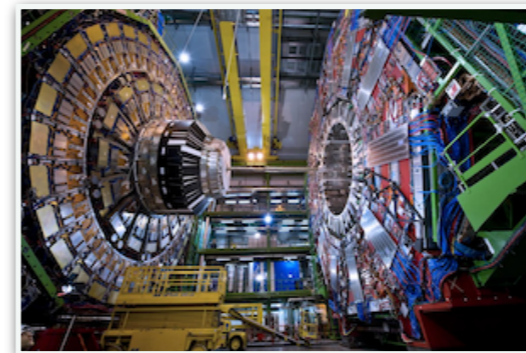
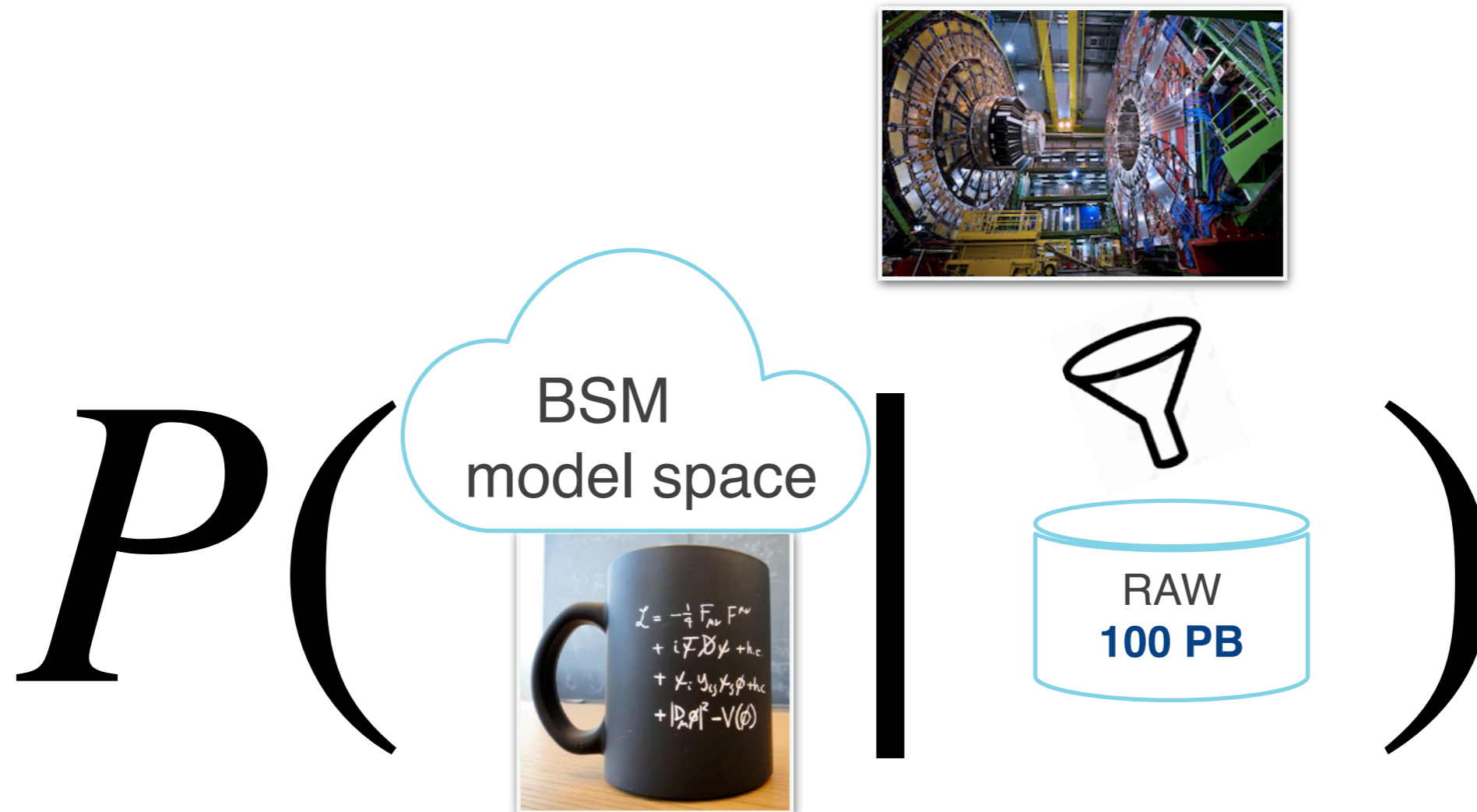
# Step 3: infer parameters



# Step 3: infer parameters



# Inference: the dream

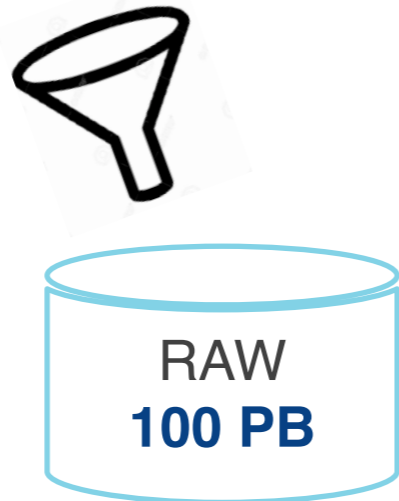


# Inference: the dream

<sup>^</sup> frequentist



*P*(



# Inference: the dream

<sup>^</sup> frequentist

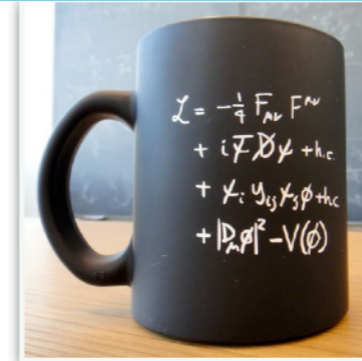


*P*



RAW  
100 PB

BSM  
model space

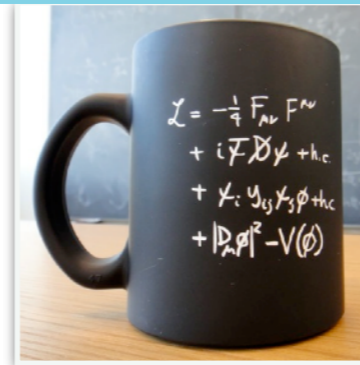


Need to reduce:

- Model space
- Data space

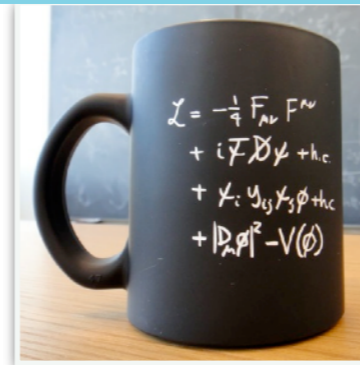
# Reducing model space

BSM  
model space



# Reducing model space

BSM  
model space



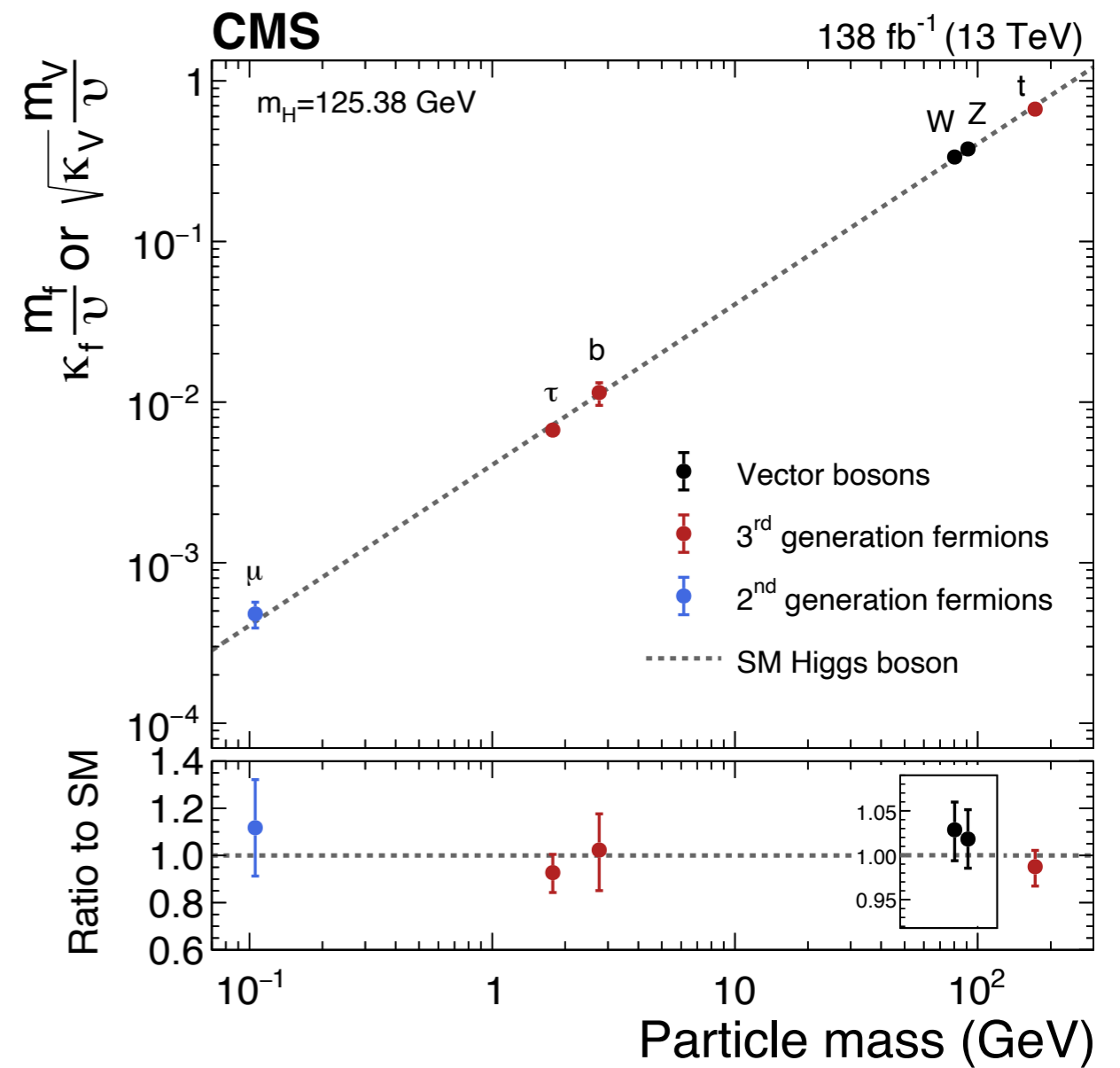
Precision measurements  
Validating the Standard Model

# Higgs Precision

- 12 years on: reaching 10% precision



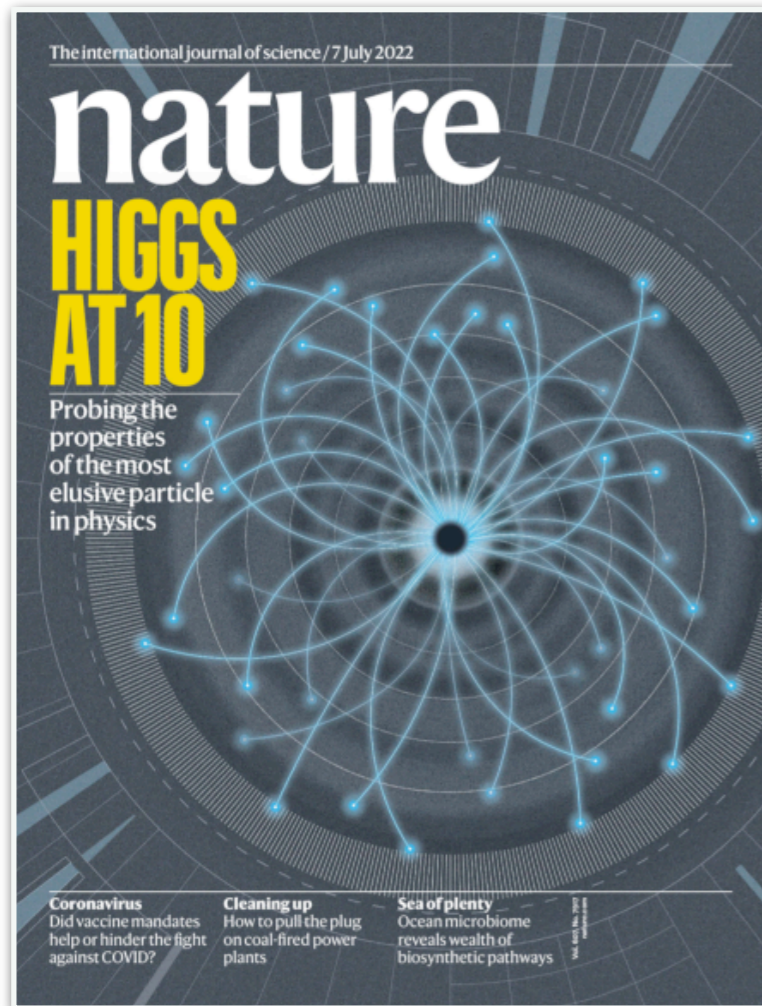
HIG-22-001



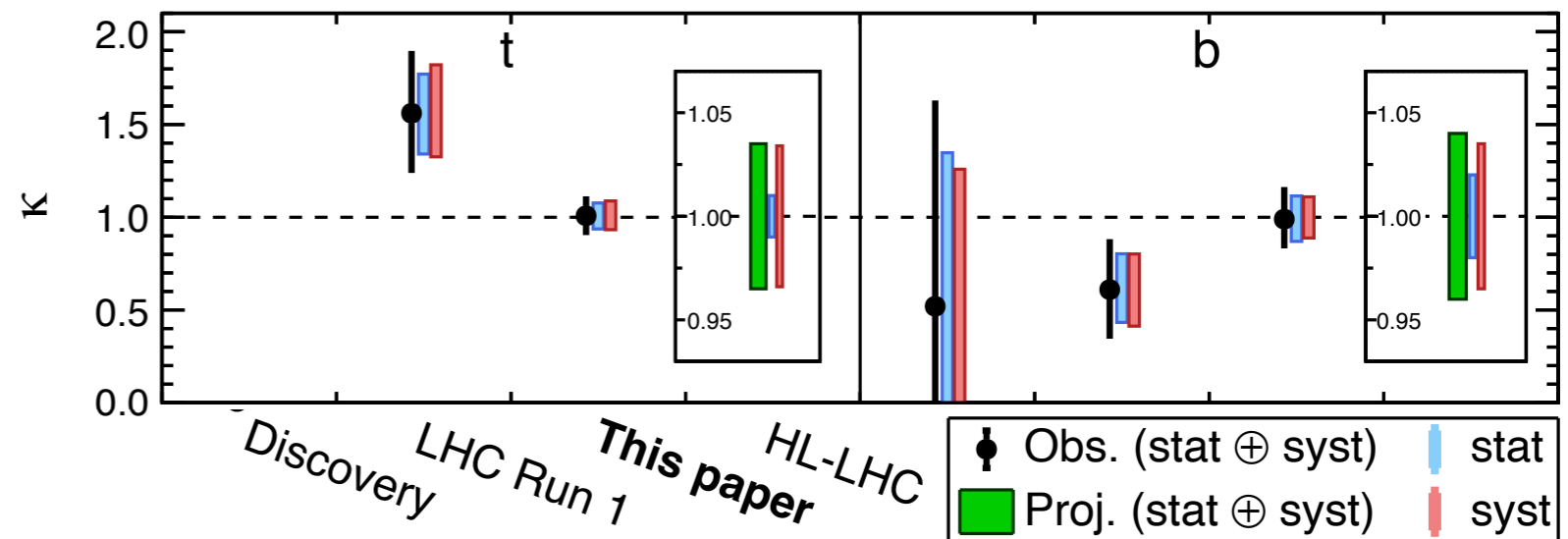
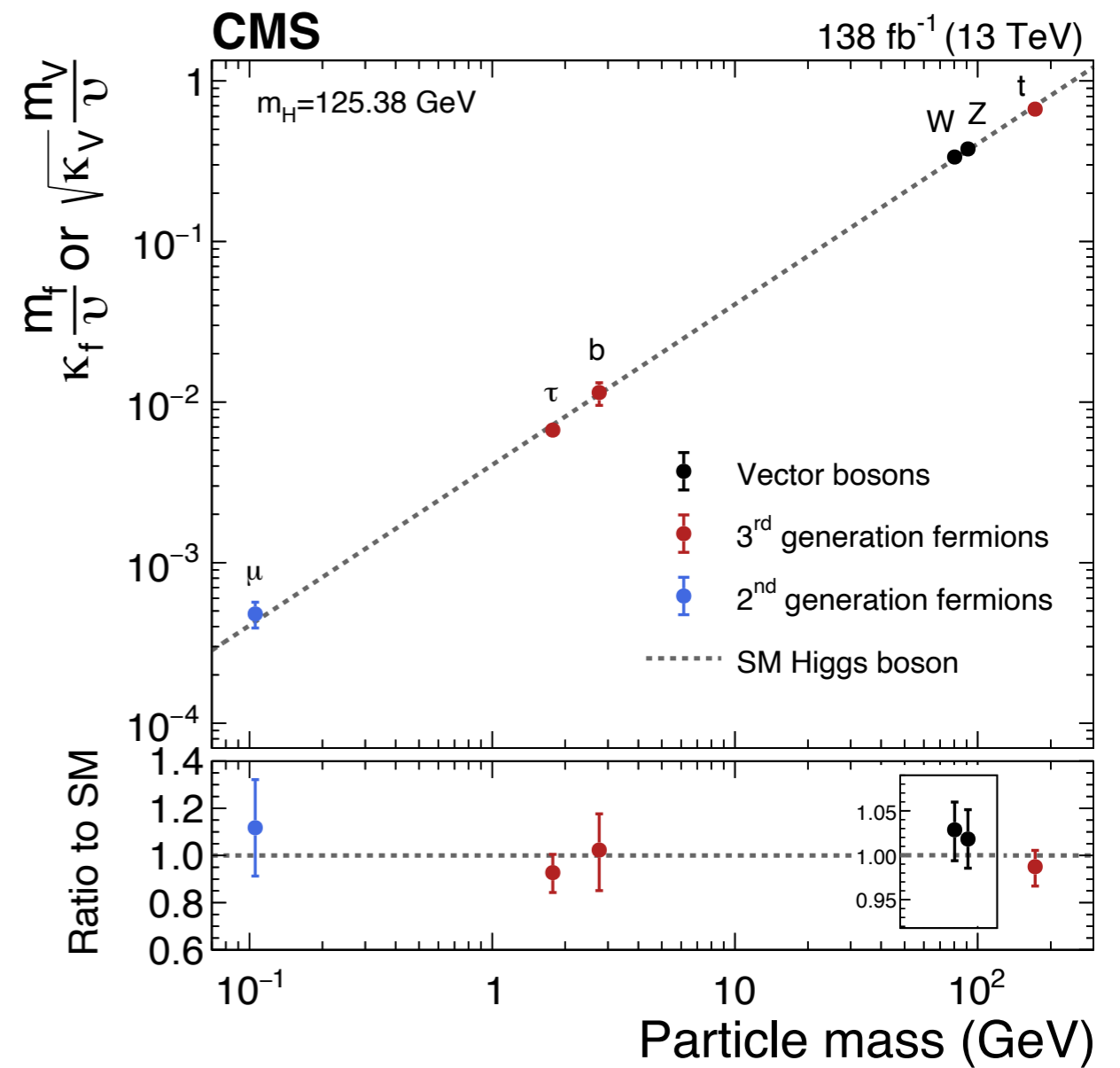


# Higgs Precision

- 12 years on: reaching 10% precision
- Projecting few % precision in HL-LHC

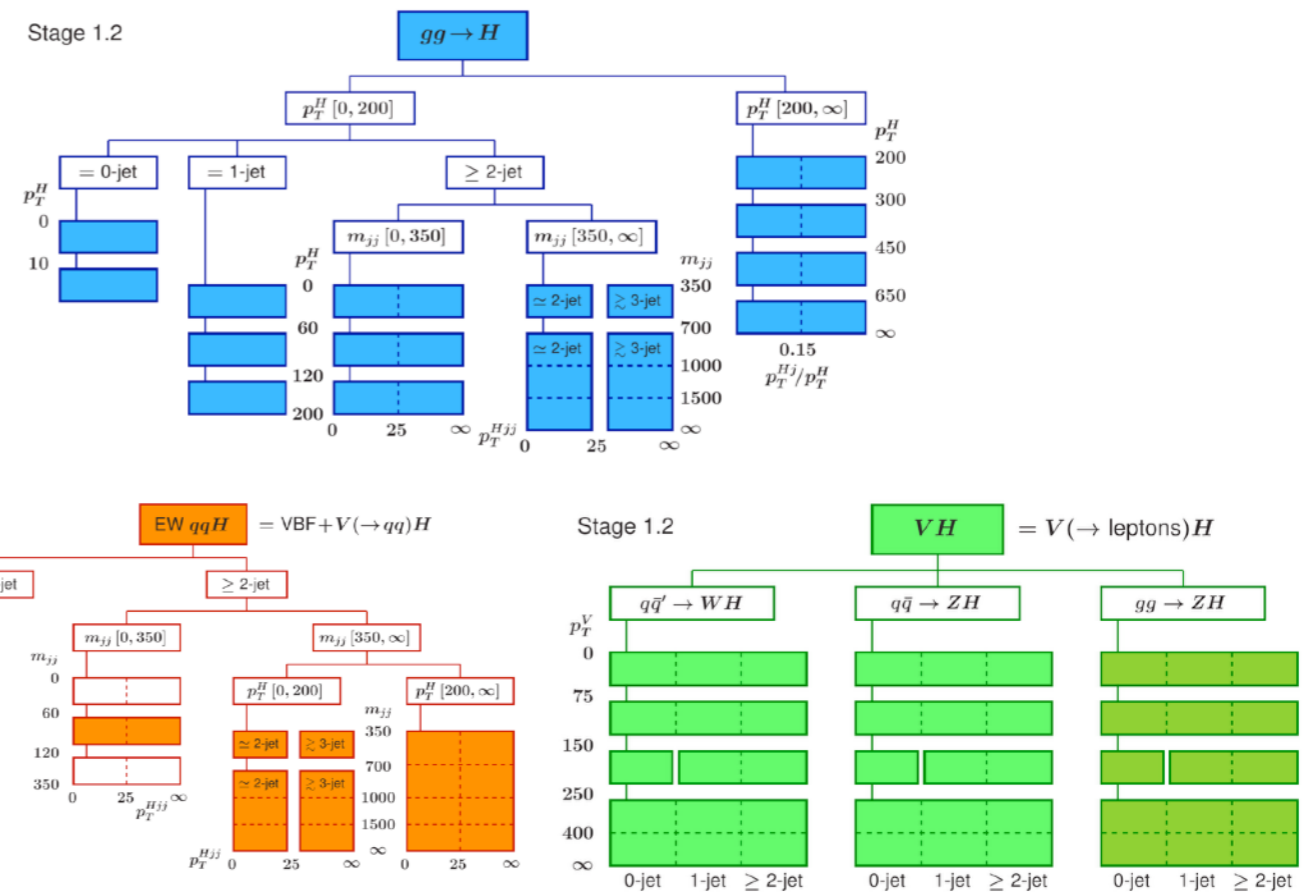
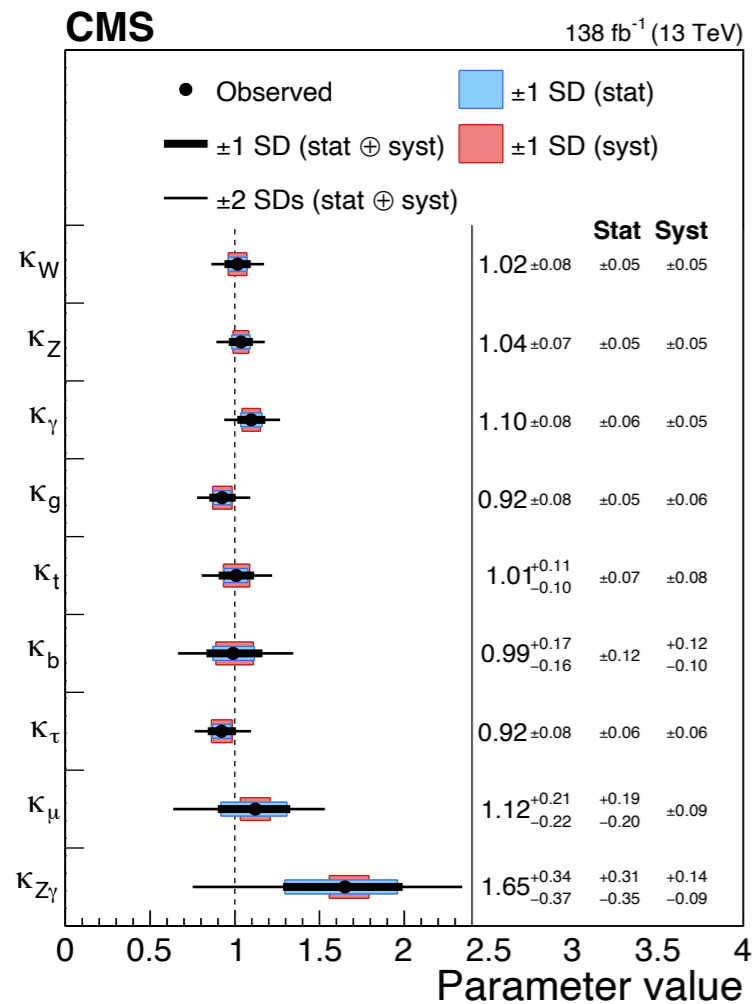


[HIG-22-001](#)



# Higgs results are becoming differential

- Move from couplings ( $\kappa$ ) to STXS framework
  - Simplified Template Cross/X Sections: unfolding into well-designed bins



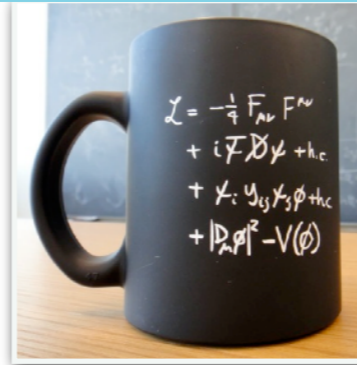
<https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWGFiducialAndSTXS>



# Reducing model space

Direct searches  
New resonances

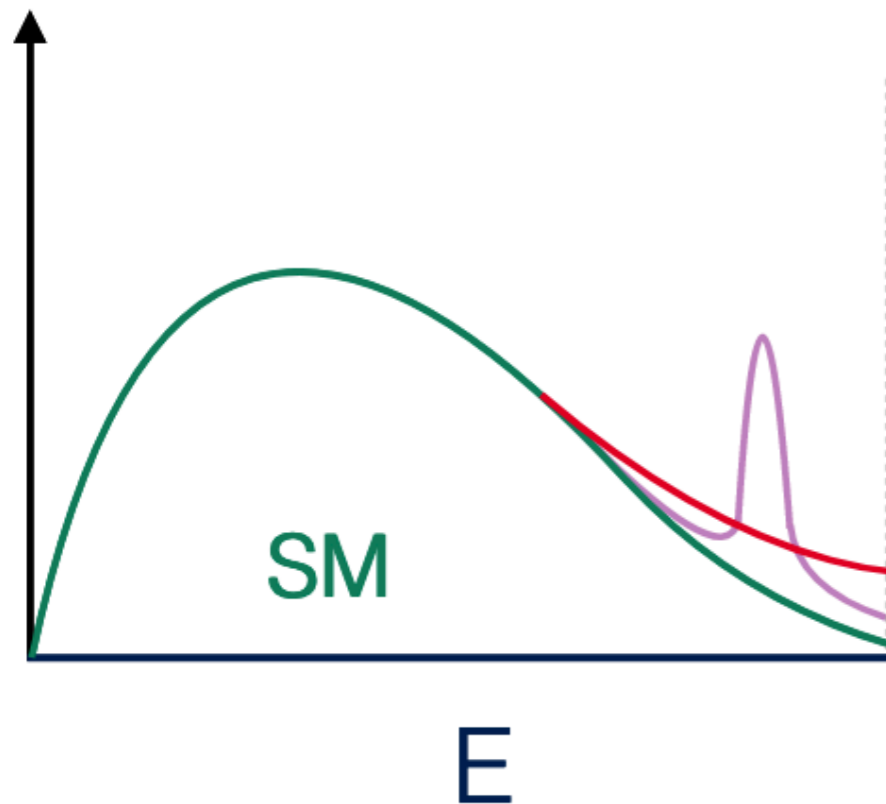
BSM  
model space



Precision measurements  
Validating the Standard Model

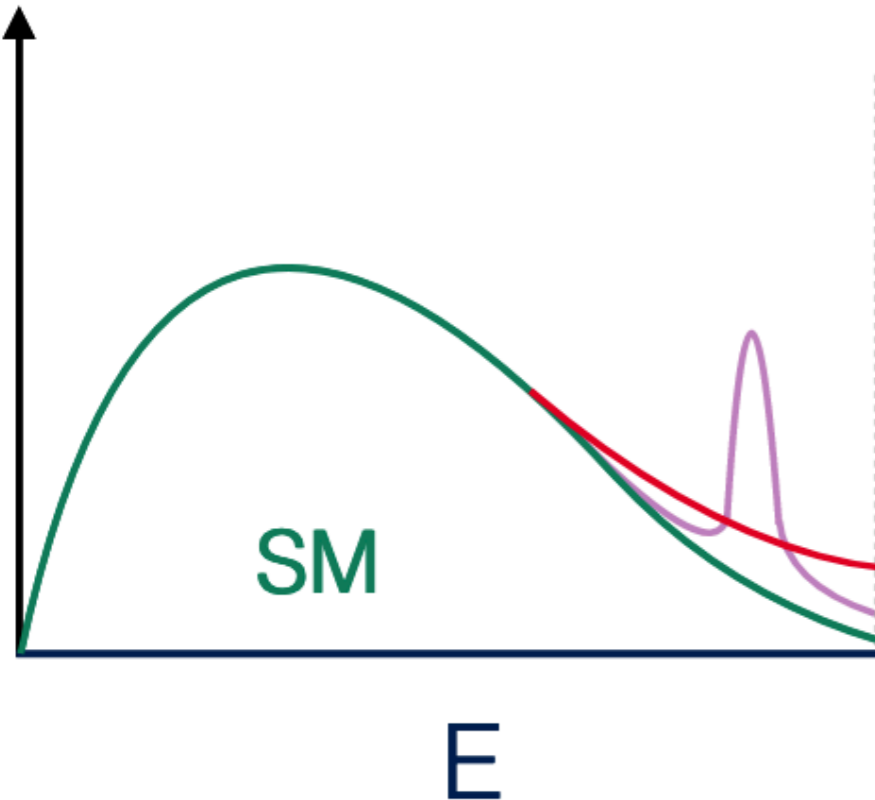
# Reducing model space

Direct searches  
New resonances



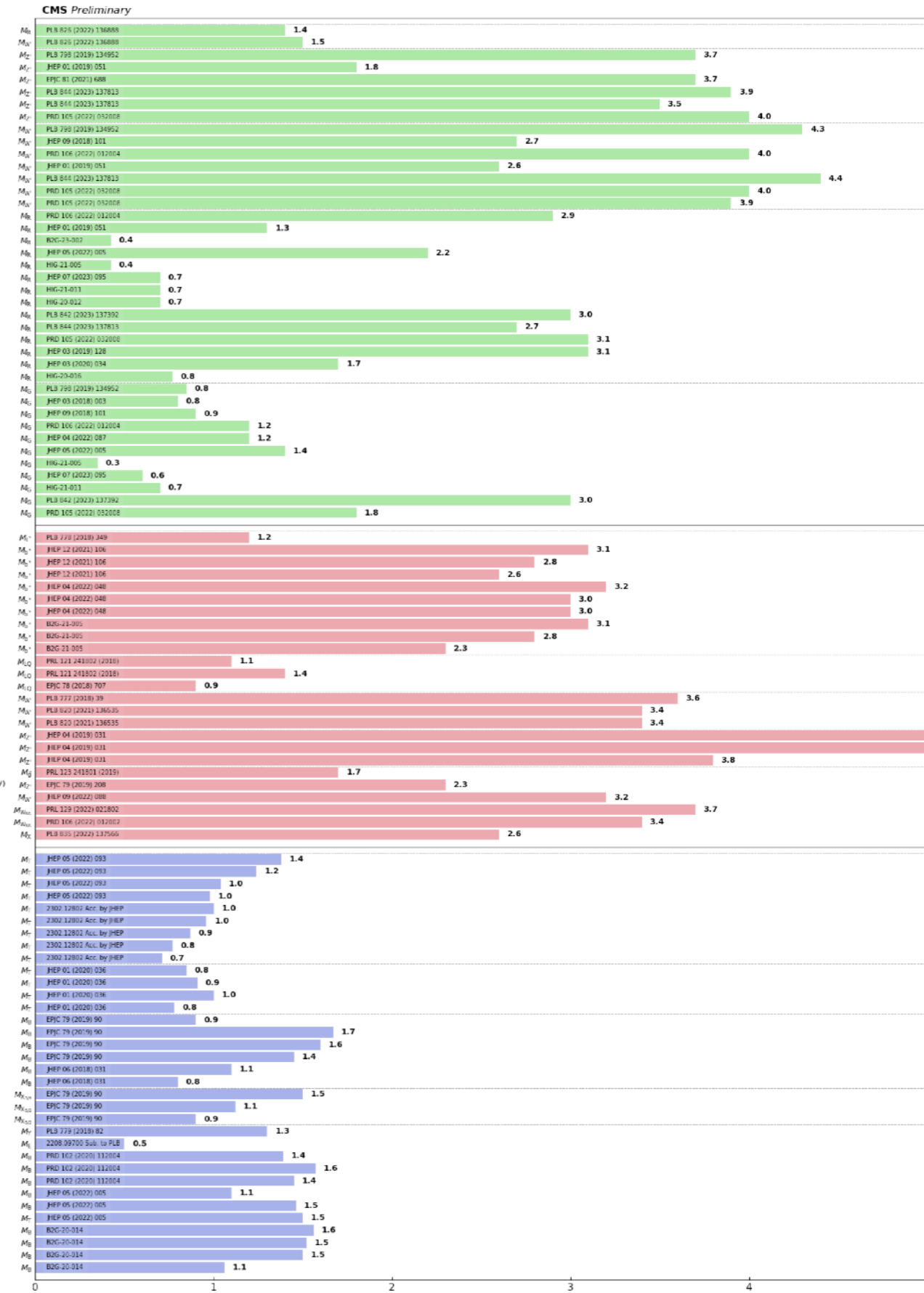
# Reducing model space

Direct searches  
New resonances



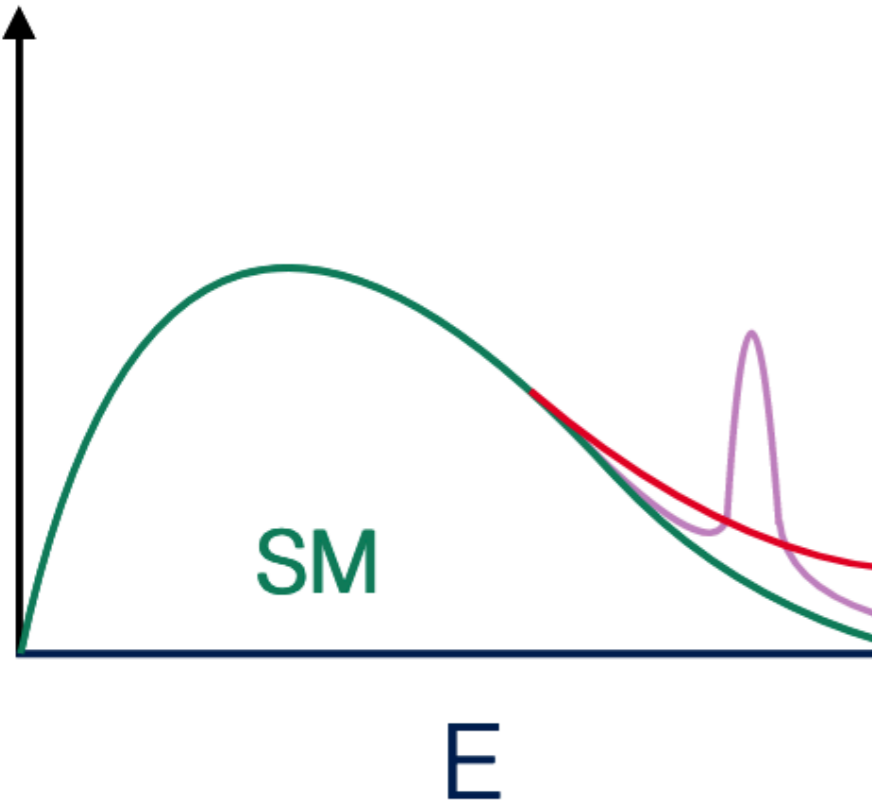
- WW/WH/HH/VV resonances**
  - HST**
    - $R \rightarrow q\bar{q} \rightarrow W\gamma$  ( $g_{Wq} = 0.1, A = 4M_W$ )
    - $W \rightarrow q\bar{q} \rightarrow W\gamma$  ( $g_{Wq} = 0.1, A = 4M_W$ )
    - $Z' (2016 \text{ combination})$
    - $Z' \rightarrow ZH \rightarrow q\bar{q}\tau\tau$
    - $Z' \rightarrow ZH \rightarrow \ell\ell, \nu\ell b\bar{b}$
    - $Z' \rightarrow ZH \rightarrow q\bar{q}q\bar{q}$
    - $Z' \rightarrow WW \rightarrow q\bar{q}q\bar{q}$
    - $Z' \rightarrow WW \rightarrow \ell\nu q\bar{q}$
    - $W (2016 \text{ combination})$
    - $W \rightarrow WZ \rightarrow Hq\bar{q}$
    - $W \rightarrow WZ \rightarrow \nu\nu q\bar{q}$
    - $W \rightarrow WH \rightarrow q\bar{q}\tau\tau$
    - $W \rightarrow WZ \rightarrow q\bar{q}q\bar{q}$
    - $W \rightarrow WH \rightarrow \ell\nu q\bar{q}$
    - $W \rightarrow WZ \rightarrow \ell\nu q\bar{q}$
    - $R \rightarrow ZZ \rightarrow \nu\nu q\bar{q}$
    - $R \rightarrow HH \rightarrow q\bar{q}\tau\tau$
    - $R \rightarrow HH$
    - $R \rightarrow HH \rightarrow b\bar{b}W$  (lep.) merged-jet
    - $R \rightarrow HH \rightarrow b\bar{b}WW$  (lep.)
    - $R \rightarrow HH \rightarrow \text{multi-leptons}$
    - $R \rightarrow HH \rightarrow \gamma\gamma b\bar{b}$
    - $R \rightarrow HH \rightarrow b\bar{b}b\bar{b}$
    - $R \rightarrow HH \rightarrow b\bar{b}b\bar{b}$  merged-jet
    - $R \rightarrow VV \rightarrow q\bar{q}q\bar{q}$
    - $R \rightarrow WW \rightarrow \ell\nu q\bar{q}$
    - $R \rightarrow WW$
    - $G (2016 \text{ combination})$
    - $G \rightarrow ZZ \rightarrow \ell\ell\nu\nu$
    - $G \rightarrow ZZ \rightarrow Hq\bar{q}$
    - $G \rightarrow ZZ \rightarrow \nu\nu q\bar{q}$
    - $G \rightarrow ZZ \rightarrow Hq\bar{q}$
    - $G \rightarrow HH \rightarrow b\bar{b}WW$  (lep.) merged-jet
    - $G \rightarrow HH \rightarrow b\bar{b}WW$  (lep.)
    - $G \rightarrow HH \rightarrow \text{multi-leptons}$
    - $G \rightarrow HH \rightarrow \gamma\gamma b\bar{b}$
    - $G \rightarrow HH \rightarrow b\bar{b}b\bar{b}$  merged-jet
    - $G \rightarrow WW \rightarrow \ell\nu q\bar{q}$
  - Radion,  $\Lambda_{UV} = 3\text{TeV}$** 
    - $\tau^+\tau^- \rightarrow \ell\nu b\bar{b} + \text{jets}$  (R-S model,  $\theta = 1$ )
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH+RH)
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (RH)
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH)
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH+RH)
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (RH)
    - $b^* \rightarrow tW \rightarrow b\bar{\nu}q$  (LH+RH)
    - $b^* \rightarrow tW \rightarrow b\bar{\nu}q$  (RH)
    - $b^* \rightarrow tW \rightarrow b\bar{\nu}q$  (LH)
    - $b^* \rightarrow tW \rightarrow b\bar{\nu}q$  (RH)
    - $LQ\bar{L}Q \rightarrow b\bar{b}\nu\nu$
    - $LQ\bar{L}Q \rightarrow \ell\ell\nu\nu$
    - $LQ\bar{L}Q \rightarrow \ell\ell\tau\tau$
    - $W \rightarrow tb, \ell\ell$  (RH)  $M_{W'} > M_W$
    - $W \rightarrow tb, \ell\ell$  (LH)
    - $W \rightarrow tb, \ell\ell$  (RH)
    - $Z' \rightarrow \ell\ell$  ( $\Gamma/M_Z = 30\%$ )
    - $Z' \rightarrow \ell\ell$  ( $\Gamma/M_Z = 10\%$ )
    - $Z' \rightarrow \ell\ell$  ( $\Gamma/M_Z = 1\%$ )
    - Stealth  $\tilde{g} \rightarrow \tilde{q}\bar{q}$  ( $\gamma + 2\text{jets}, M_{\tilde{g}} = 0.2\text{TeV}$ )
    - $Z' \rightarrow \ell\ell \rightarrow \ell Z/H/A \rightarrow \ell\nu + \text{jets}$  ( $M_{Z'} = 1.5\text{TeV}$ )
    - $W \rightarrow T\bar{b}/\bar{T}t \rightarrow W\nu + \text{jets}$  ( $M_T = 2/3M_W$ )
    - $W_{KK} \rightarrow RW \rightarrow WWW$  (of + 1 $\ell$ )
    - $W_{KK} \rightarrow RW \rightarrow WWW$  (L)
    - $X \rightarrow aa \rightarrow b\bar{b}b\bar{b}$  ( $M_X = 0.1\text{TeV}, M_{UV}/f = \theta$ )
  - Bulk  $G, K/M_{Pl} = 0.5$** 
    - $\tau^+\tau^- \rightarrow \ell\nu b\bar{b} + \text{jets}$  (R-S model,  $\theta = 1$ )
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH+RH)
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (RH)
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH)
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH+RH)
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (RH)
    - $b^* \rightarrow tW \rightarrow b\bar{\nu}q$  (LH+RH)
    - $b^* \rightarrow tW \rightarrow b\bar{\nu}q$  (RH)
    - $b^* \rightarrow tW \rightarrow b\bar{\nu}q$  (LH)
    - $b^* \rightarrow tW \rightarrow b\bar{\nu}q$  (RH)
    - $LQ\bar{L}Q \rightarrow b\bar{b}\nu\nu$
    - $LQ\bar{L}Q \rightarrow \ell\ell\nu\nu$
    - $LQ\bar{L}Q \rightarrow \ell\ell\tau\tau$
    - $W \rightarrow tb, \ell\ell$  (RH)  $M_{W'} > M_W$
    - $W \rightarrow tb, \ell\ell$  (LH)
    - $W \rightarrow tb, \ell\ell$  (RH)
    - $Z' \rightarrow \ell\ell$  ( $\Gamma/M_Z = 30\%$ )
    - $Z' \rightarrow \ell\ell$  ( $\Gamma/M_Z = 10\%$ )
    - $Z' \rightarrow \ell\ell$  ( $\Gamma/M_Z = 1\%$ )
    - Stealth  $\tilde{g} \rightarrow \tilde{q}\bar{q}$  ( $\gamma + 2\text{jets}, M_{\tilde{g}} = 0.2\text{TeV}$ )
    - $Z' \rightarrow \ell\ell \rightarrow \ell Z/H/A \rightarrow \ell\nu + \text{jets}$  ( $M_{Z'} = 1.5\text{TeV}$ )
    - $W \rightarrow T\bar{b}/\bar{T}t \rightarrow W\nu + \text{jets}$  ( $M_T = 2/3M_W$ )
    - $W_{KK} \rightarrow RW \rightarrow WWW$  (of + 1 $\ell$ )
    - $W_{KK} \rightarrow RW \rightarrow WWW$  (L)
    - $X \rightarrow aa \rightarrow b\bar{b}b\bar{b}$  ( $M_X = 0.1\text{TeV}, M_{UV}/f = \theta$ )
- Resonances**
  - Excited quarks**
    - $\tau^+\tau^- \rightarrow \ell\nu b\bar{b} + \text{jets}$  (R-S model,  $\theta = 1$ )
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH+RH)
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (RH)
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH)
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH+RH)
    - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (RH)
    - $b^* \rightarrow tW \rightarrow b\bar{\nu}q$  (LH+RH)
    - $b^* \rightarrow tW \rightarrow b\bar{\nu}q$  (RH)
    - $b^* \rightarrow tW \rightarrow b\bar{\nu}q$  (LH)
    - $b^* \rightarrow tW \rightarrow b\bar{\nu}q$  (RH)
  - LQ**
    - $LQ\bar{L}Q \rightarrow b\bar{b}\nu\nu$
    - $LQ\bar{L}Q \rightarrow \ell\ell\nu\nu$
    - $LQ\bar{L}Q \rightarrow \ell\ell\tau\tau$
  - $W'-tb$** 
    - $W \rightarrow tb, \ell\ell$  (RH)  $M_{W'} > M_W$
    - $W \rightarrow tb, \ell\ell$  (LH)
    - $W \rightarrow tb, \ell\ell$  (RH)
  - $Z'-\ell\ell$** 
    - $Z' \rightarrow \ell\ell$  ( $\Gamma/M_Z = 30\%$ )
    - $Z' \rightarrow \ell\ell$  ( $\Gamma/M_Z = 10\%$ )
    - $Z' \rightarrow \ell\ell$  ( $\Gamma/M_Z = 1\%$ )
  - other**
    - Stealth  $\tilde{g} \rightarrow \tilde{q}\bar{q}$  ( $\gamma + 2\text{jets}, M_{\tilde{g}} = 0.2\text{TeV}$ )
    - $Z' \rightarrow \ell\ell \rightarrow \ell Z/H/A \rightarrow \ell\nu + \text{jets}$  ( $M_{Z'} = 1.5\text{TeV}$ )
    - $W \rightarrow T\bar{b}/\bar{T}t \rightarrow W\nu + \text{jets}$  ( $M_T = 2/3M_W$ )
    - $W_{KK} \rightarrow RW \rightarrow WWW$  (of + 1 $\ell$ )
    - $W_{KK} \rightarrow RW \rightarrow WWW$  (L)
    - $X \rightarrow aa \rightarrow b\bar{b}b\bar{b}$  ( $M_X = 0.1\text{TeV}, M_{UV}/f = \theta$ )
- Very heavy fermions**
  - (qb)T**
    - $b Z' (Z \rightarrow \nu\nu)$  ( $\Gamma/m = 0.3$ , Singlet)
    - $b Z' (Z \rightarrow \nu\nu)$  ( $\Gamma/m = 0.2$ , Singlet)
    - $b Z' (Z \rightarrow \nu\nu)$  ( $\Gamma/m = 0.1$ , Singlet)
    - $b Z' (Z \rightarrow \nu\nu)$  ( $\Gamma/m = 0.05$ , Singlet)
    - $b tH (H \rightarrow \gamma\gamma)$ , ( $\Gamma/m = 0.05$ , Singlet)
    - $b tH (H \rightarrow \gamma\gamma)$ , ( $\Gamma/m = 0.04$ , Singlet)
    - $b tH (H \rightarrow \gamma\gamma)$ , ( $\Gamma/m = 0.03$ , Singlet)
    - $b tH (H \rightarrow \gamma\gamma)$ , ( $\Gamma/m = 0.02$ , Singlet)
    - $b tH (H \rightarrow \gamma\gamma)$ , ( $\Gamma/m = 0.01$ , Singlet)
  - (qt)T**
    - $t tH (H \rightarrow b\bar{b})$  ( $\Gamma/m = 0.3$ , Doublet)
    - $t tH (H \rightarrow b\bar{b})$  ( $\Gamma/m = 0.3$ , Singlet)
    - $t tH (H \rightarrow b\bar{b})$  ( $\Gamma/m = 0.1$ , Singlet)
    - $t tH (H \rightarrow b\bar{b})$  ( $\Gamma/m = 0.05$ , Singlet)
  - (qt)(qb)B**
    - $t Wt \rightarrow \ell\nu + \text{jets}$  ( $\Gamma/m = 0.1$ , LH)
    - $b Wt \rightarrow \ell\nu + \text{jets}$  ( $\Gamma/m = 0.3$ , LH)
    - $b Wt \rightarrow \ell\nu + \text{jets}$  ( $\Gamma/m = 0.2$ , LH)
    - $b Wt \rightarrow \ell\nu + \text{jets}$  ( $\Gamma/m = 0.1$ , LH)
    - $b tH (H \rightarrow b\bar{b})$  ( $\Gamma/m = 0.3$ , Doublet)
    - $b tH (H \rightarrow b\bar{b})$  ( $\Gamma/m = 0.2$ , Doublet)
    - $t Wt \rightarrow \ell\nu + \text{jets}$  ( $\Gamma/m = 0.3$ , LH)
  - (qt)X**
    - $t Wt \rightarrow \ell\nu + \text{jets}$  ( $\Gamma/m = 0.2$ , LH)
    - $t Wt \rightarrow \ell\nu + \text{jets}$  ( $\Gamma/m = 0.1$ , LH)
    - $Y_{-q}Y_{-q} \rightarrow bW bW \rightarrow \ell\nu q\bar{q}q\bar{q}$
  - Pair prod.**
    - $q\bar{q} \rightarrow 4b$   $t\bar{t}/\nu\nu\nu$
    - $RR \rightarrow b\bar{q}q b\bar{q}q$  ( $B(Z) = 1$ )
    - $BB \rightarrow b\bar{q}q b\bar{q}q$  ( $B(t) = 1$ )
    - $RR \rightarrow b\bar{q}q b\bar{q}q$  (Singlet)
    - $RR \rightarrow \ell\nu + \text{jets}$  (Doublet)
    - $BB \rightarrow \ell\nu + \text{jets}$  (Singlet)
    - $TT \rightarrow \ell\nu + \text{jets}$  (Singlet and Doublet)
    - $BB \rightarrow \ell\nu + \text{jets}$  ( $B(H) = 1$ )
    - $BB \rightarrow \ell\nu + \text{jets}$  ( $B(tZ) = 1$ )
    - $BB \rightarrow \ell\nu + \text{jets}$  (Doublet)
    - $BB \rightarrow \ell\nu + \text{jets}$  (Singlet)

## Overview of CMS B2G Results



# Reducing model space

Direct searches  
New resonances



## WW/WH/HH/VV resonances

- HST**
  - $R \rightarrow q\bar{q} \rightarrow W\gamma$  ( $g_W = 0.1, A = 4M_W$ )
  - $W \rightarrow q\bar{q} \rightarrow W\gamma$  ( $g_W = 0.1, A = 4M_W$ )
  - Z' (2016 combination)
- Z', HVT B**
  - $Z' \rightarrow ZH \rightarrow q\bar{q}\tau\tau$
  - $Z' \rightarrow ZH \rightarrow \ell\ell, \nu\ell b\bar{b}$
  - $Z' \rightarrow ZH \rightarrow q\bar{q}q\bar{q}$
  - $Z' \rightarrow WW \rightarrow q\bar{q}q\bar{q}$
  - $Z' \rightarrow WW \rightarrow \ell\nu q\bar{q}$
- W', HVT B**
  - W (2016 combination)
  - $W \rightarrow WZ \rightarrow Hq\bar{q}$
  - $W \rightarrow WZ \rightarrow \nu\nu q\bar{q}$
  - $W \rightarrow WH \rightarrow q\bar{q}\tau\tau$
  - $W \rightarrow WZ \rightarrow q\bar{q}q\bar{q}$
  - $W \rightarrow WH \rightarrow \ell\nu q\bar{q}$
  - $W \rightarrow WZ \rightarrow \ell\nu q\bar{q}$
  - $R \rightarrow ZZ \rightarrow \nu\nu q\bar{q}$
  - $R \rightarrow HH \rightarrow q\bar{q}\tau\tau$
- Radion,  $\Lambda_{UV} = 3TeV$** 
  - $R \rightarrow HH \rightarrow b\bar{b}W$  (lep.) merged-jet
  - $R \rightarrow HH \rightarrow b\bar{b}WW$  (lep.)
  - $R \rightarrow HH \rightarrow \text{multi-leptons}$
  - $R \rightarrow HH \rightarrow \gamma\gamma b\bar{b}$
  - $R \rightarrow HH \rightarrow b\bar{b}b\bar{b}$
  - $R \rightarrow HH \rightarrow b\bar{b}b\bar{b}$  merged-jet
  - $R \rightarrow VV \rightarrow q\bar{q}q\bar{q}$
  - $R \rightarrow WW \rightarrow \ell\nu q\bar{q}$
- Bulk G,  $K/M_{Pl} = 0.5$** 
  - G (2016 combination)
  - $G \rightarrow ZZ \rightarrow \ell\nu\nu$
  - $G \rightarrow ZZ \rightarrow Hq\bar{q}$
  - $G \rightarrow ZZ \rightarrow \nu\nu q\bar{q}$
  - $G \rightarrow ZZ \rightarrow Hq\bar{q}$
  - $G \rightarrow HH \rightarrow b\bar{b}W$  (lep.) merged-jet
  - $G \rightarrow HH \rightarrow b\bar{b}WW$  (lep.)
  - $G \rightarrow HH \rightarrow \text{multi-leptons}$
  - $G \rightarrow HH \rightarrow \gamma\gamma b\bar{b}$
  - $G \rightarrow HH \rightarrow b\bar{b}b\bar{b}$  merged-jet
  - $G \rightarrow WW \rightarrow \ell\nu q\bar{q}$

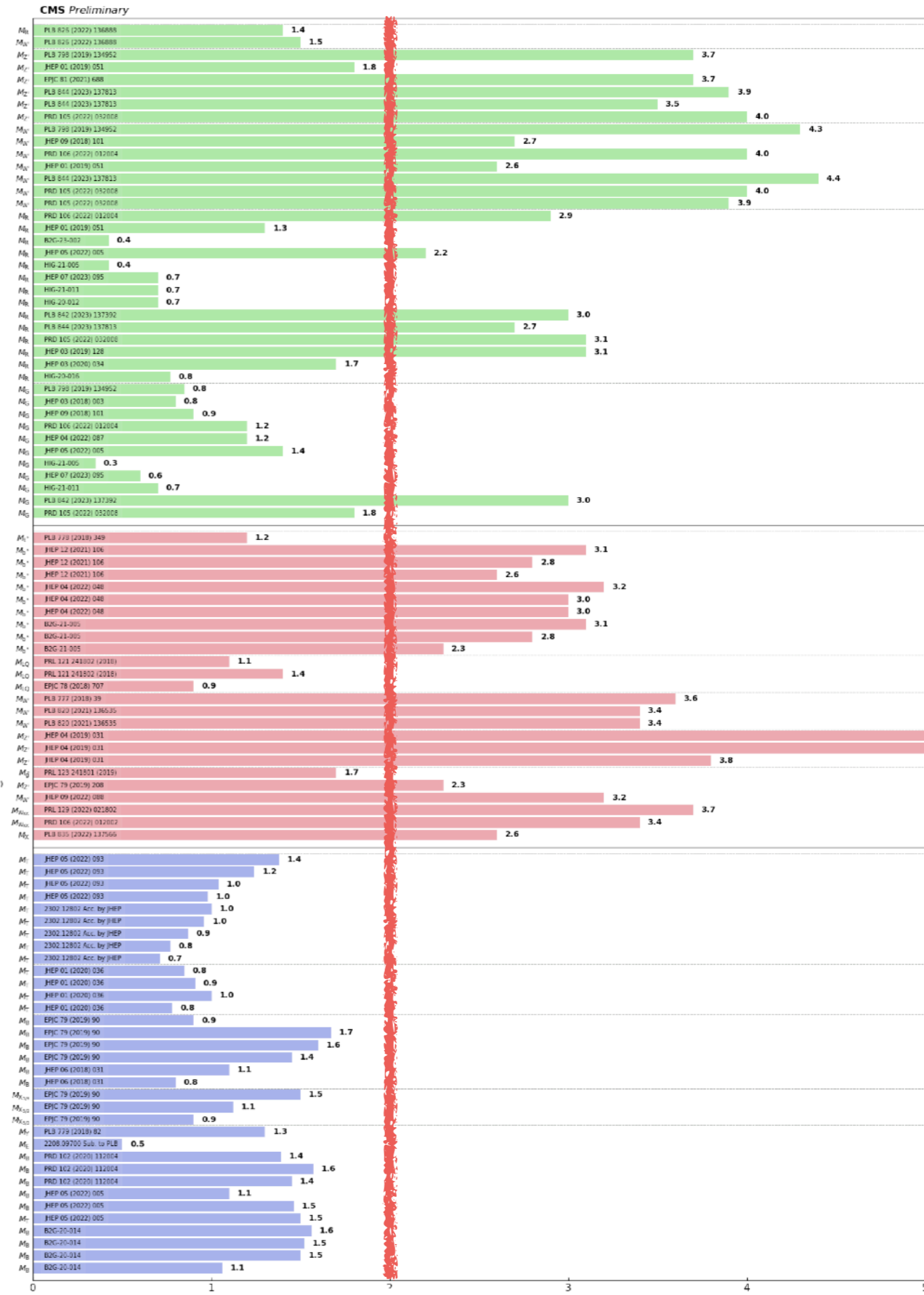
## Resonances

- Excited quarks**
  - $t^* \rightarrow t + jets$  (R-5 model,  $D = 1$ )
  - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH+RH)
  - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (RH)
  - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH)
  - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH+RH)
  - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (RH)
  - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH+RH)
  - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (RH)
  - $b^* \rightarrow tW \rightarrow b\bar{q}q$  (LH)
- LQ**
  - $LQ \rightarrow b + jets$
  - $LQ \rightarrow t + jets$
- W-ib**
  - $W \rightarrow tb, \ell\ell$  (LH)
  - $W \rightarrow tb, \ell\ell$  (RH)
- Z-tt**
  - $Z' \rightarrow t\bar{t}$  ( $\Gamma/M_Z = 30\%$ )
  - $Z' \rightarrow t\bar{t}$  ( $\Gamma/M_Z = 10\%$ )
  - $Z' \rightarrow t\bar{t}$  ( $\Gamma/M_Z = 1\%$ )
- other**
  - Stealth  $\tilde{g} \rightarrow \tilde{g}q\bar{q}$  ( $\gamma + 2jets, M_{\tilde{g}} = 0.2TeV$ )
  - $Z' \rightarrow t\bar{t} \rightarrow ZH + jets$  ( $M_{Z'} = 1.5 TeV$ )
  - $W \rightarrow T\bar{b} + jets$  ( $M_W = 2/3 M_{Pl}$ )
  - $W_{KK} \rightarrow RW \rightarrow WWW$  (of + 1f)
  - $W_{KK} \rightarrow RW \rightarrow WWW$  (1f)
  - $X \rightarrow aa \rightarrow b\bar{b}b\bar{b}$  ( $M_X = 0.1 TeV, M_{UV} = 0$ )

## Very heavy fermions

- (q)T**
  - $bT$  ( $Z \rightarrow \nu\nu$ ) ( $\Gamma/m = 0.3$ , Singlet)
  - $bT$  ( $Z \rightarrow \nu\nu$ ) ( $\Gamma/m = 0.2$ , Singlet)
  - $bT$  ( $Z \rightarrow \nu\nu$ ) ( $\Gamma/m = 0.1$ , Singlet)
  - $bT$  ( $Z \rightarrow \nu\nu$ ) ( $\Gamma/m = 0.05$ , Singlet)
  - $bT$  ( $H \rightarrow \gamma\gamma$ ) ( $\Gamma/m = 0.05$ , Singlet)
  - $bT$  ( $H \rightarrow \gamma\gamma$ ) ( $\Gamma/m = 0.04$ , Singlet)
  - $bT$  ( $H \rightarrow \gamma\gamma$ ) ( $\Gamma/m = 0.03$ , Singlet)
  - $bT$  ( $H \rightarrow \gamma\gamma$ ) ( $\Gamma/m = 0.02$ , Singlet)
  - $bT$  ( $H \rightarrow \gamma\gamma$ ) ( $\Gamma/m = 0.01$ , Singlet)
- (q)T**
  - $tH$  ( $H \rightarrow b\bar{b}$ ) ( $\Gamma/m = 0.3$ , Doublet)
  - $tH$  ( $H \rightarrow b\bar{b}$ ) ( $\Gamma/m = 0.3$ , Singlet)
  - $tH$  ( $H \rightarrow b\bar{b}$ ) ( $\Gamma/m = 0.1$ , Singlet)
  - $tH$  ( $H \rightarrow b\bar{b}$ ) ( $\Gamma/m = 0.05$ , Singlet)
- (q)T(q)B**
  - $tW$  ( $H \rightarrow \ell\nu + jets$ ) ( $\Gamma/m = 0.1$ , LH)
  - $bW$  ( $H \rightarrow \ell\nu + jets$ ) ( $\Gamma/m = 0.3$ , LH)
  - $bW$  ( $H \rightarrow \ell\nu + jets$ ) ( $\Gamma/m = 0.2$ , LH)
  - $bW$  ( $H \rightarrow \ell\nu + jets$ ) ( $\Gamma/m = 0.1$ , LH)
  - $bH$  ( $H \rightarrow b\bar{b}$ ) ( $\Gamma/m = 0.3$ , Doublet)
  - $bH$  ( $H \rightarrow b\bar{b}$ ) ( $\Gamma/m = 0.2$ , Doublet)
  - $tW$  ( $H \rightarrow \ell\nu + jets$ ) ( $\Gamma/m = 0.3$ , LH)
- (q)X**
  - $tW$  ( $H \rightarrow \ell\nu + jets$ ) ( $\Gamma/m = 0.2$ , LH)
  - $tW$  ( $H \rightarrow \ell\nu + jets$ ) ( $\Gamma/m = 0.1$ , LH)
  - $Y_{-3/2} \rightarrow Y_{-3/2} + jets$  ( $\Gamma/m = 0.1$ , LH)
  - $Y_{-3/2} \rightarrow Y_{-3/2} + jets$  ( $\Gamma/m = 0.1$ , LH)
- Pair prod.**
  - $RR \rightarrow b\bar{q}q$  ( $B(RZ) = 1$ )
  - $RR \rightarrow b\bar{q}q$  ( $B(t\bar{t}) = 1$ )
  - $RR \rightarrow b\bar{q}q$  (Singlet)
  - $RR \rightarrow \ell\nu + jets$  (Doublet)
  - $RR \rightarrow \ell\nu + jets$  (Singlet)
  - $TT \rightarrow \ell\nu + jets$  (Singlet and Doublet)
  - $BB \rightarrow \ell\nu + jets$  ( $B(H) = 1$ )
  - $BB \rightarrow \ell\nu + jets$  ( $B(tZ) = 1$ )
  - $BB \rightarrow \ell\nu + jets$  (Doublet)
  - $BB \rightarrow \ell\nu + jets$  (Singlet)

## Overview of CMS B2G Results



2 TeV

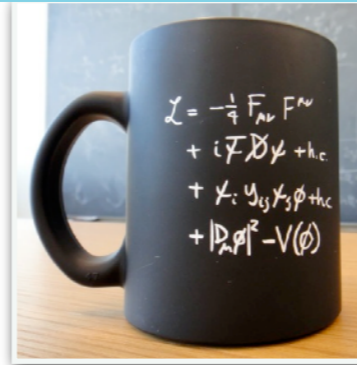


# Reducing model space

Direct searches  
New resonances

Indirect searches  
Beyond our energy reach

BSM  
model space



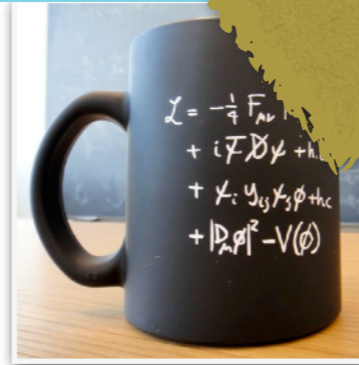
Precision measurements  
Validating the Standard Model

# Reducing model space

Direct searches  
New resonances

Indirect searches  
Beyond our energy reach

Anomaly &  
Unconventional  
B  
model

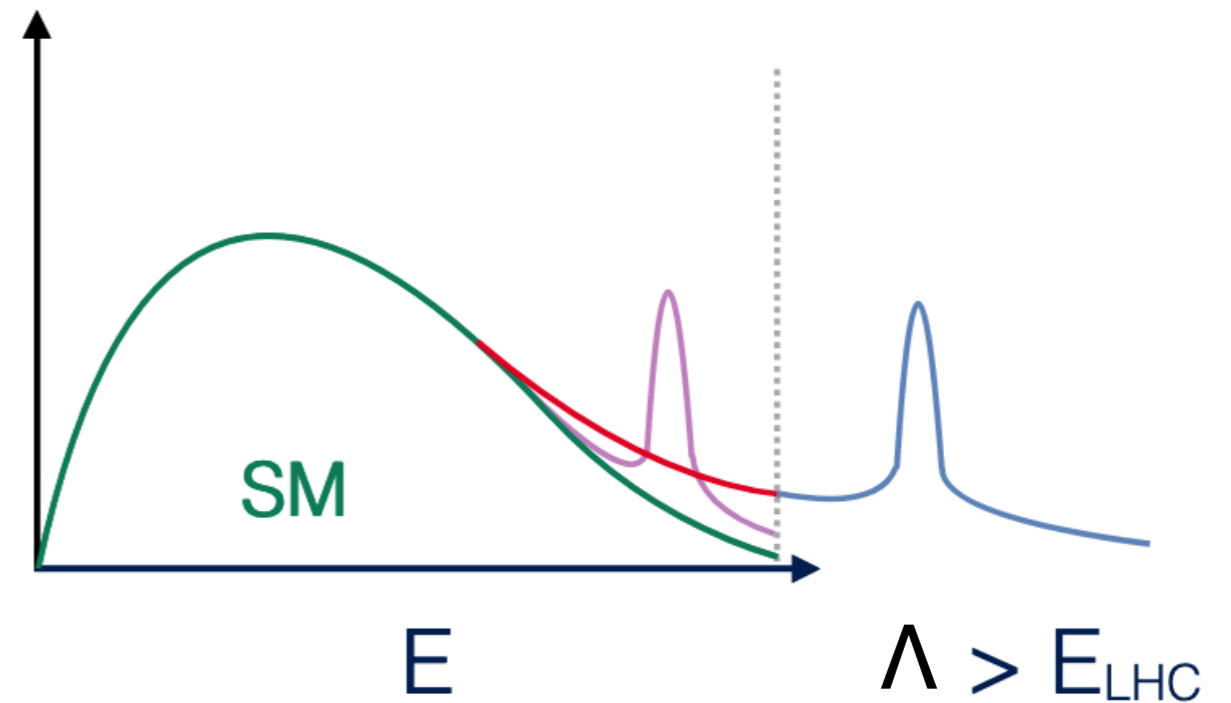


Precision measurements  
Validating the Standard Model



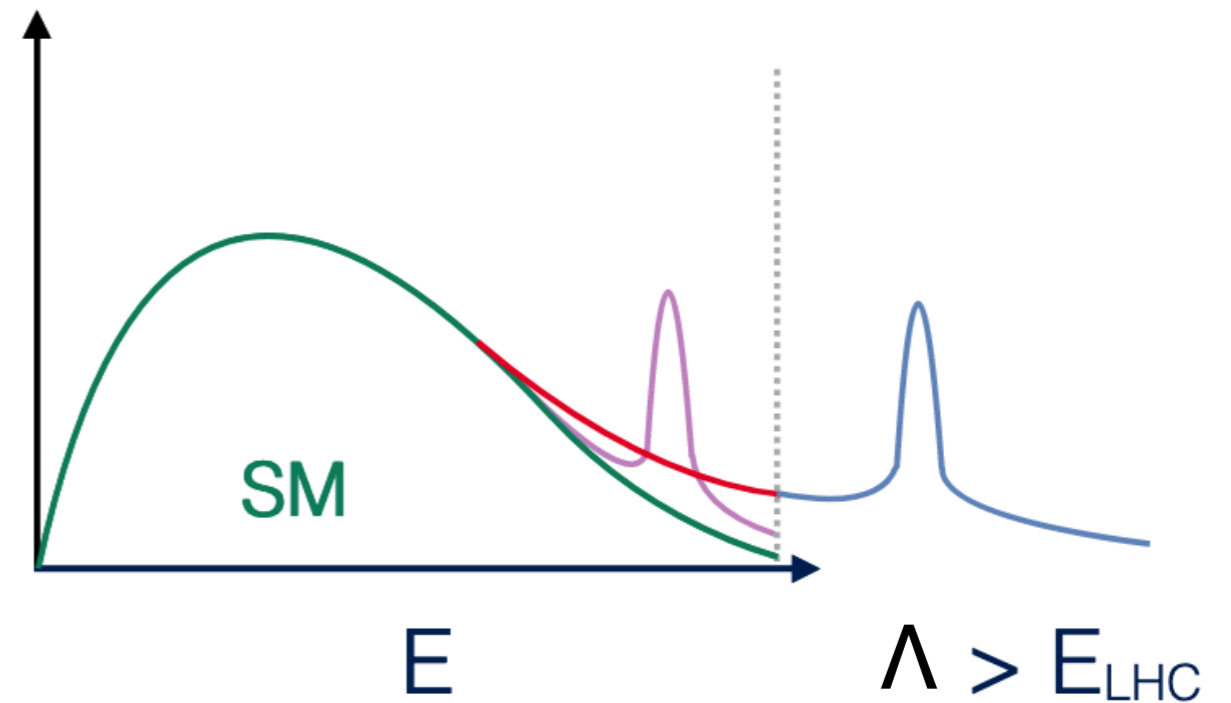
# Effective Field Theory (EFT) Intro

- New physics at scale  $\Lambda \gg E_{LHC}$  ?



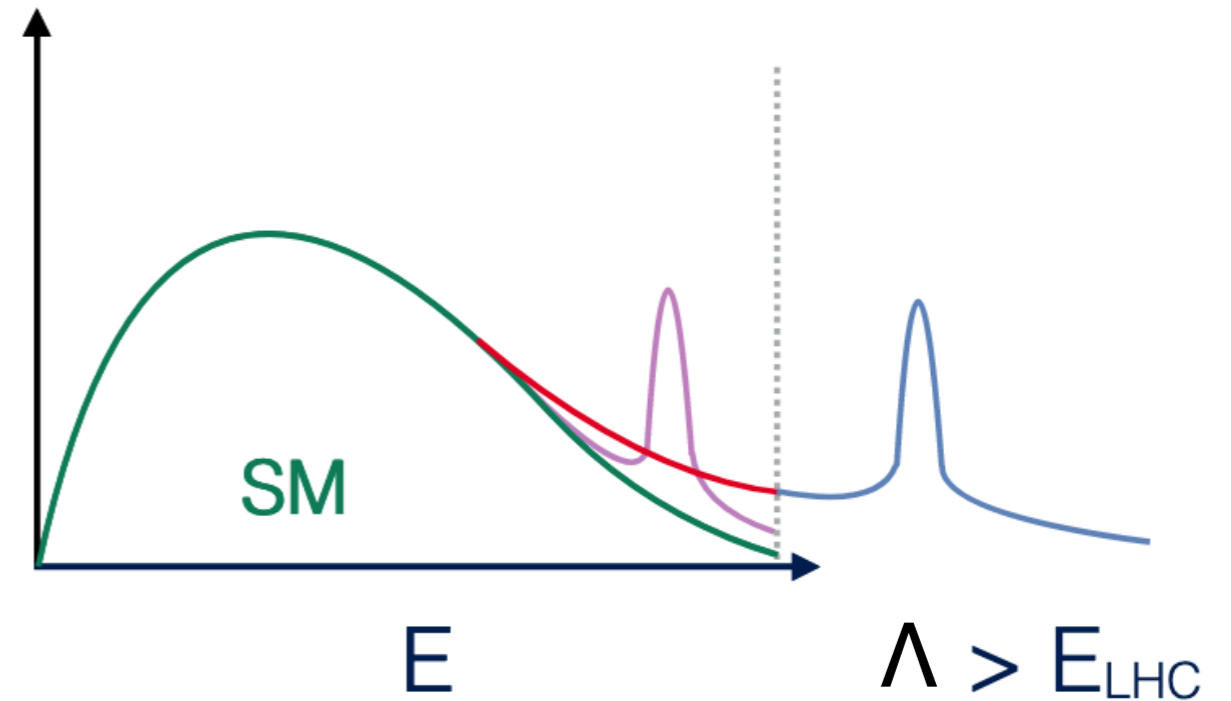
# Effective Field Theory (EFT) Intro

- New physics at scale  $\Lambda \gg E_{LHC}$  ?
  - Indirect effects may appear at LHC



# Effective Field Theory (EFT) Intro

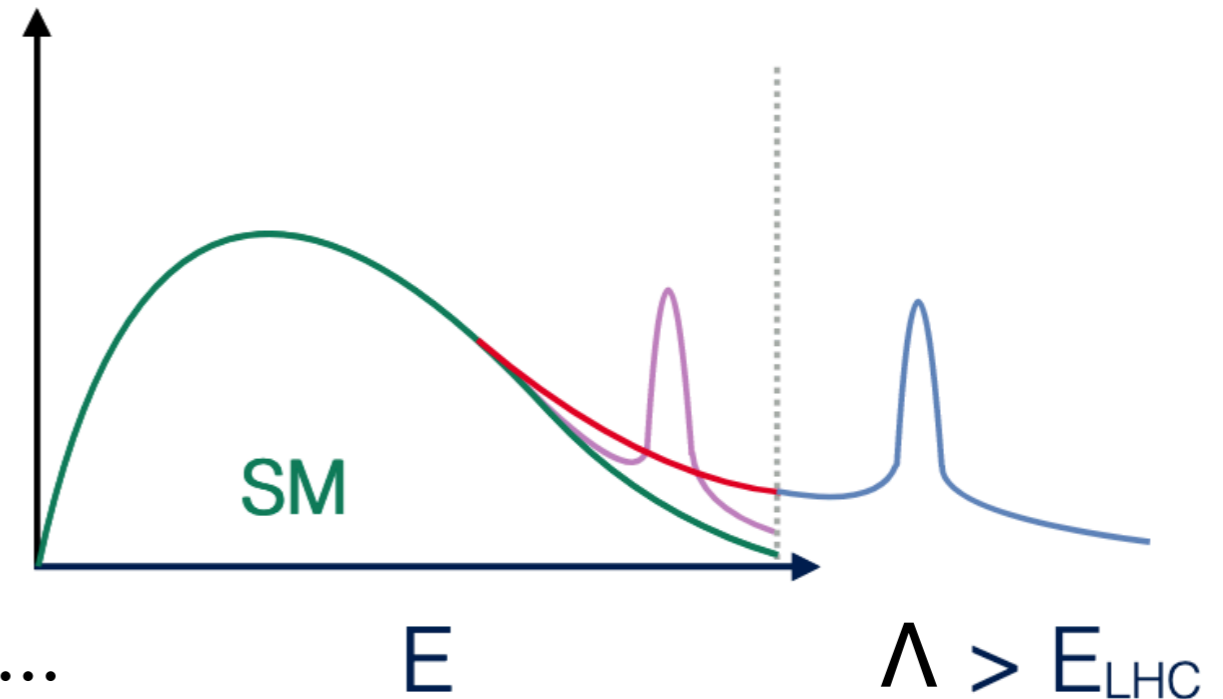
- New physics at scale  $\Lambda \gg E_{LHC}$  ?
  - Indirect effects may appear at LHC
  - Perform series expansion in  $1/\Lambda$



# Effective Field Theory (EFT) Intro

- New physics at scale  $\Lambda \gg E_{LHC}$ ?
  - Indirect effects may appear at LHC
  - Perform series expansion in  $1/\Lambda$
  - Enumerate all SM symmetry-preserving higher-dimension operators: “SMEFT”

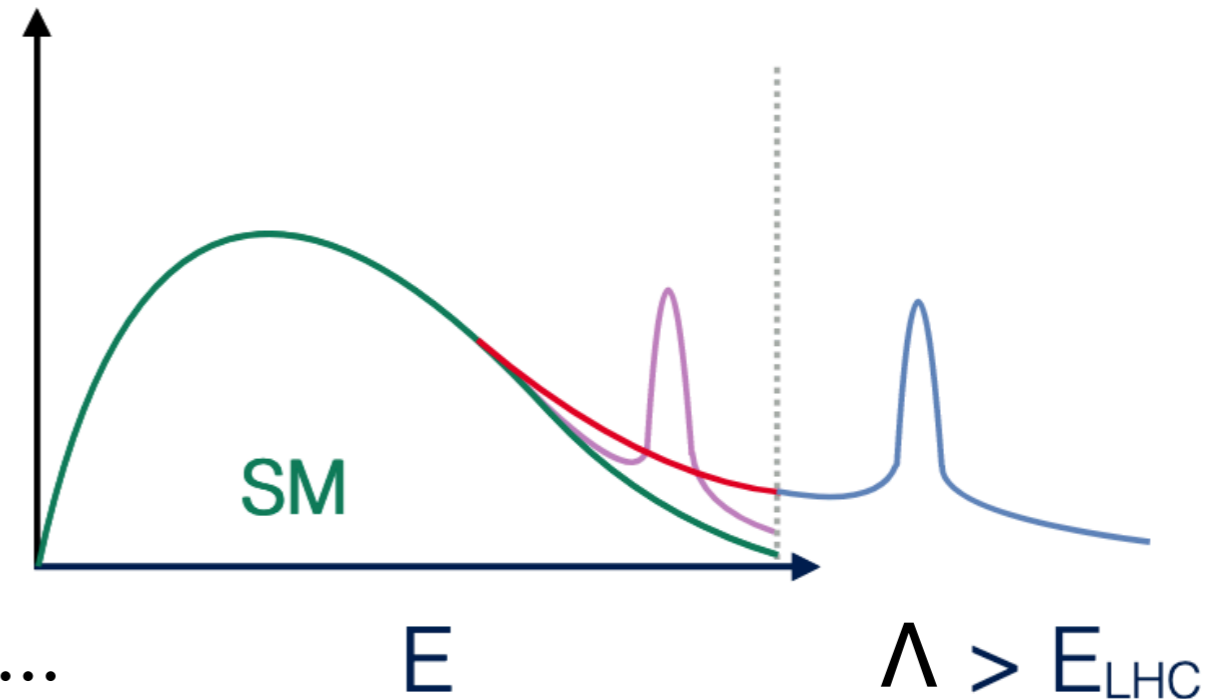
$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i \frac{c_i}{\Lambda} \mathcal{O}_i^{(5)} + \sum_i \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \dots$$



# Effective Field Theory (EFT) Intro

- New physics at scale  $\Lambda \gg E_{LHC}$  ?
  - Indirect effects may appear at LHC
  - Perform series expansion in  $1/\Lambda$
  - Enumerate all SM symmetry-preserving higher-dimension operators: “SMEFT”

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i \frac{c_i}{\Lambda} \mathcal{O}_i^{(5)} + \sum_i \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \dots$$

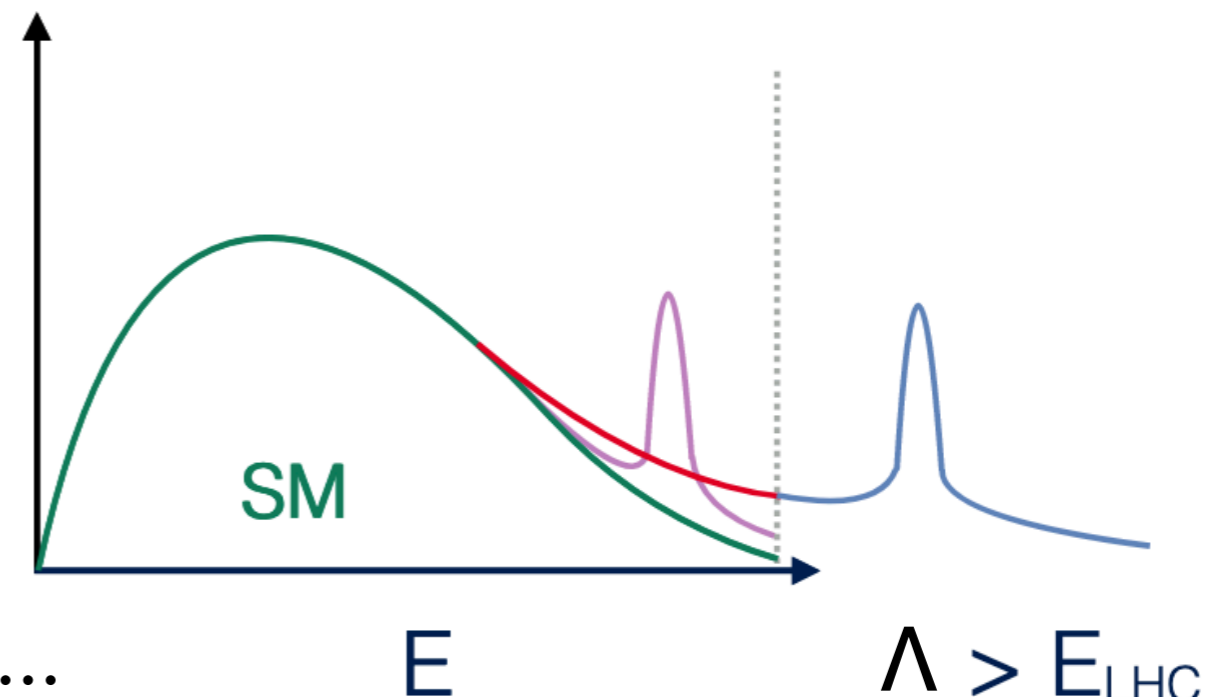


- Field redefinition allows many equivalent expansions
  - Dim-6 consensus: Warsaw basis [1008.4884](https://arxiv.org/abs/1008.4884)

# Effective Field Theory (EFT) Intro

- New physics at scale  $\Lambda \gg E_{LHC}$ ?
  - Indirect effects may appear at LHC
  - Perform series expansion in  $1/\Lambda$
  - Enumerate all SM symmetry-preserving higher-dimension operators: “SMEFT”

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i \frac{c_i}{\Lambda} \mathcal{O}_i^{(5)} + \sum_i \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \dots$$

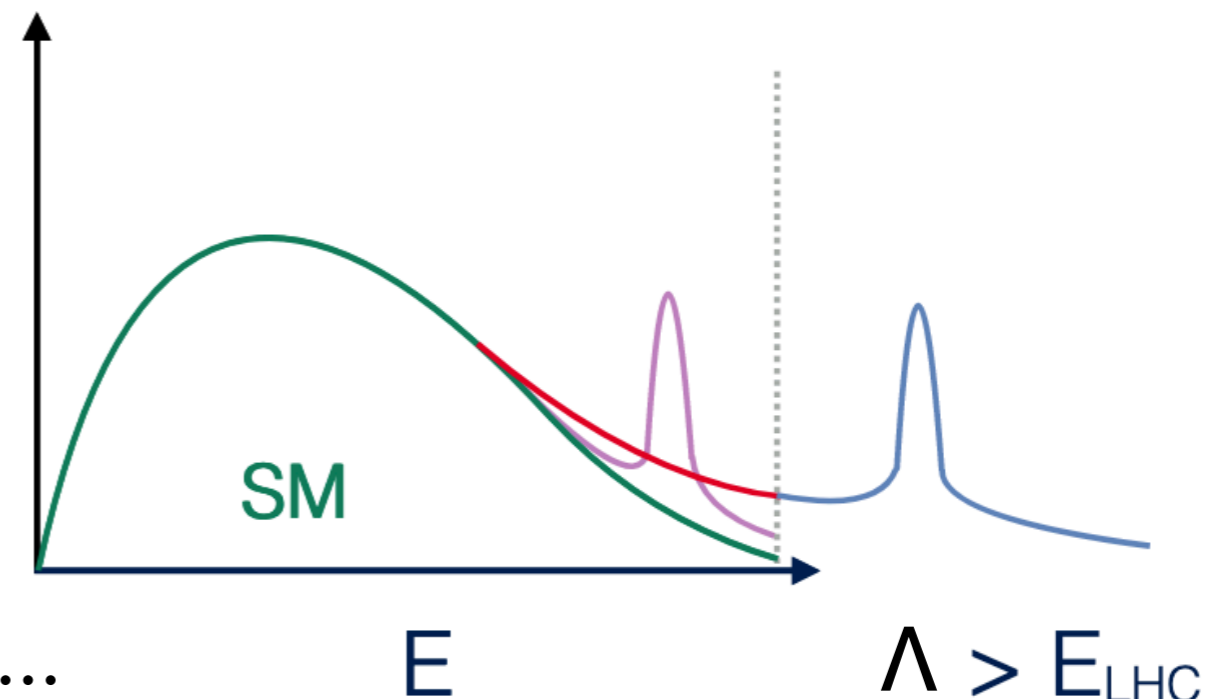


- Field redefinition allows many equivalent expansions
  - Dim-6 consensus: Warsaw basis [1008.4884](#)
  - Looking forward: field space geometry!
    - Scalars ([1511.00724](#) [2108.03240](#))
    - Vectors (DeWitt, 1980)
    - Spinors ([2006.05831](#))
    - With derivatives: jet bundles ([EFT WG meeting](#))

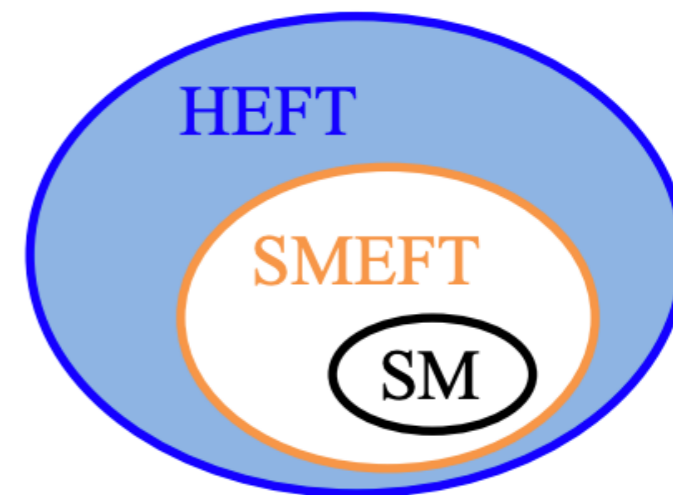
# Effective Field Theory (EFT) Intro

- New physics at scale  $\Lambda \gg E_{LHC}$  ?
  - Indirect effects may appear at LHC
  - Perform series expansion in  $1/\Lambda$ 
    - Enumerate all SM symmetry-preserving higher-dimension operators: “SMEFT”

$$\mathcal{L} = \mathcal{L}_{SM} + \sum_i \frac{c_i}{\Lambda} \mathcal{O}_i^{(5)} + \sum_i \frac{c_i}{\Lambda^2} \mathcal{O}_i^{(6)} + \dots$$

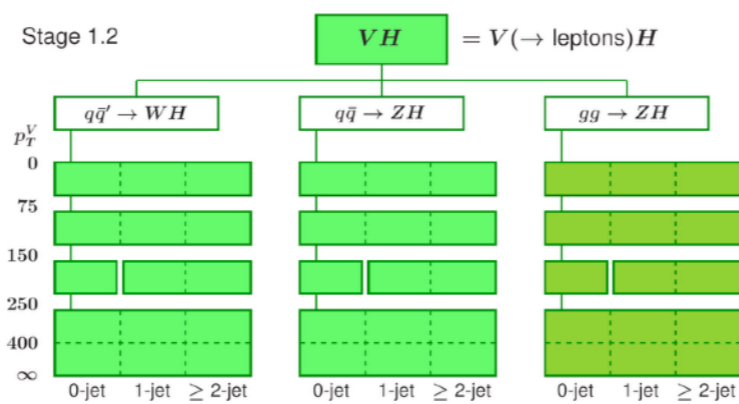
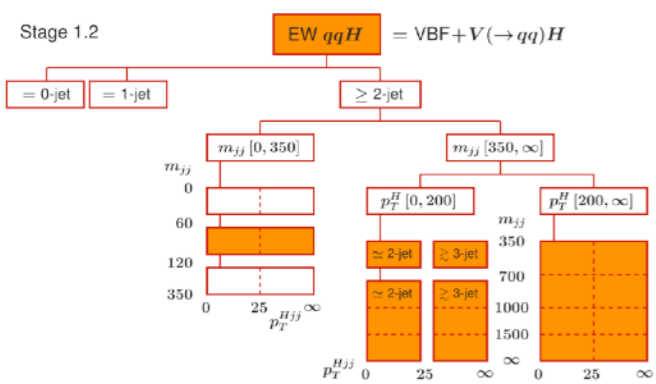
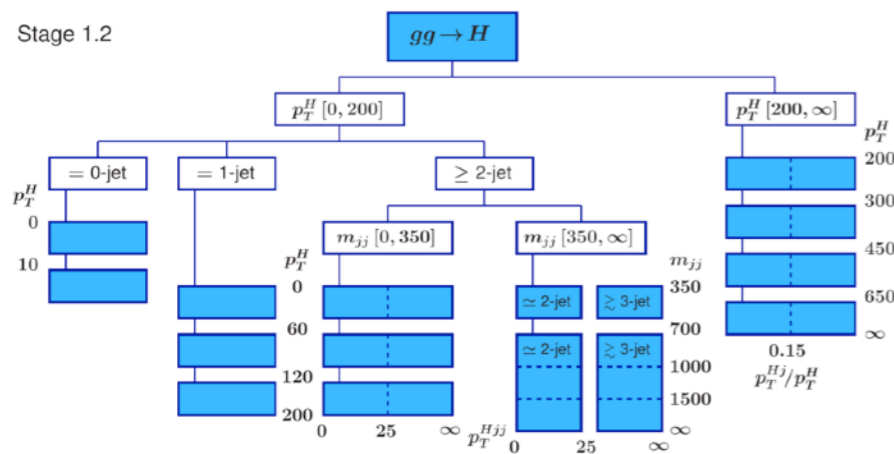


- Field redefinition allows many equivalent expansions
  - Dim-6 consensus: Warsaw basis [1008.4884](#)
  - Looking forward: field space geometry!
    - Scalars ([1511.00724](#) [2108.03240](#))
    - Vectors (DeWitt, 1980)
    - Spinors ([2006.05831](#))
    - With derivatives: jet bundles ([EFT WG meeting](#))



# Towards a global EFT analysis

- Theorists combine STXS result into global picture
  - Including also  $W/Z/t$  unfolded results

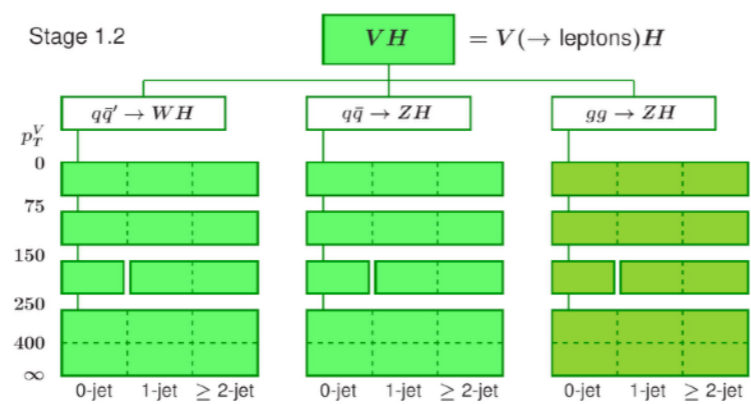
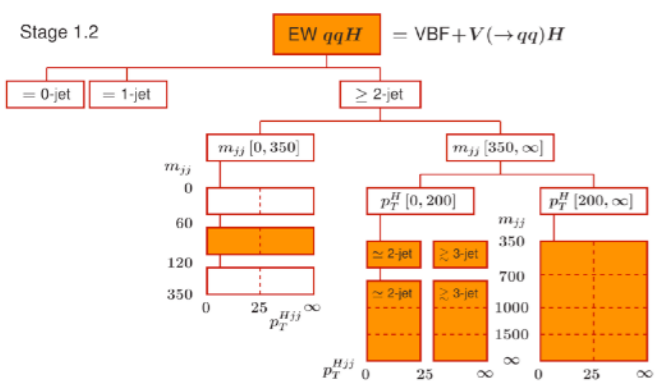
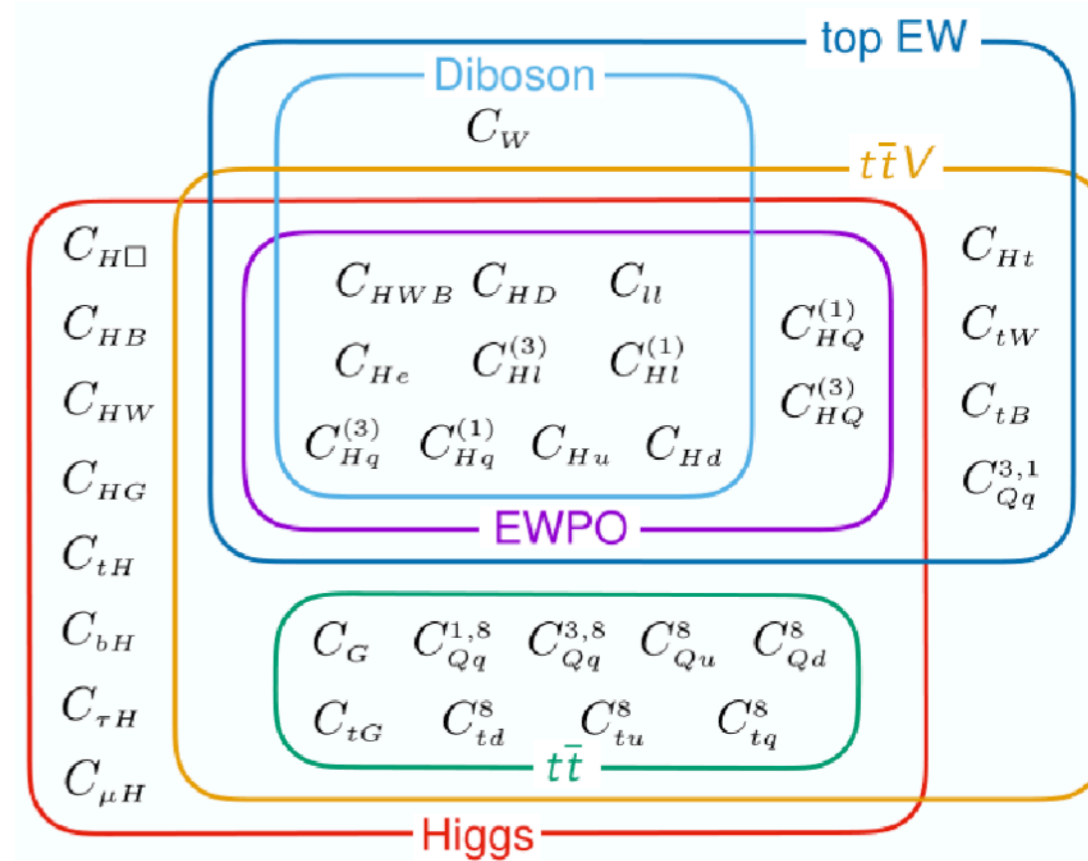
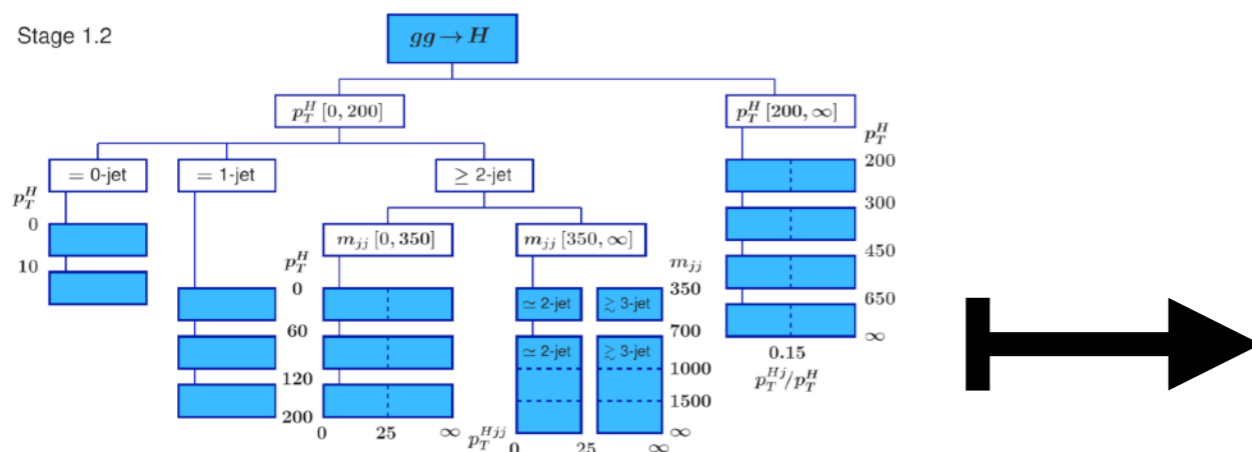




# Towards a global EFT analysis

- Theorists combine STXS result into global picture
  - Including also  $W/Z/t$  unfolded results
- 2499 dim-6 operators  $\rightarrow$  34 with flavor symmetry restrictions

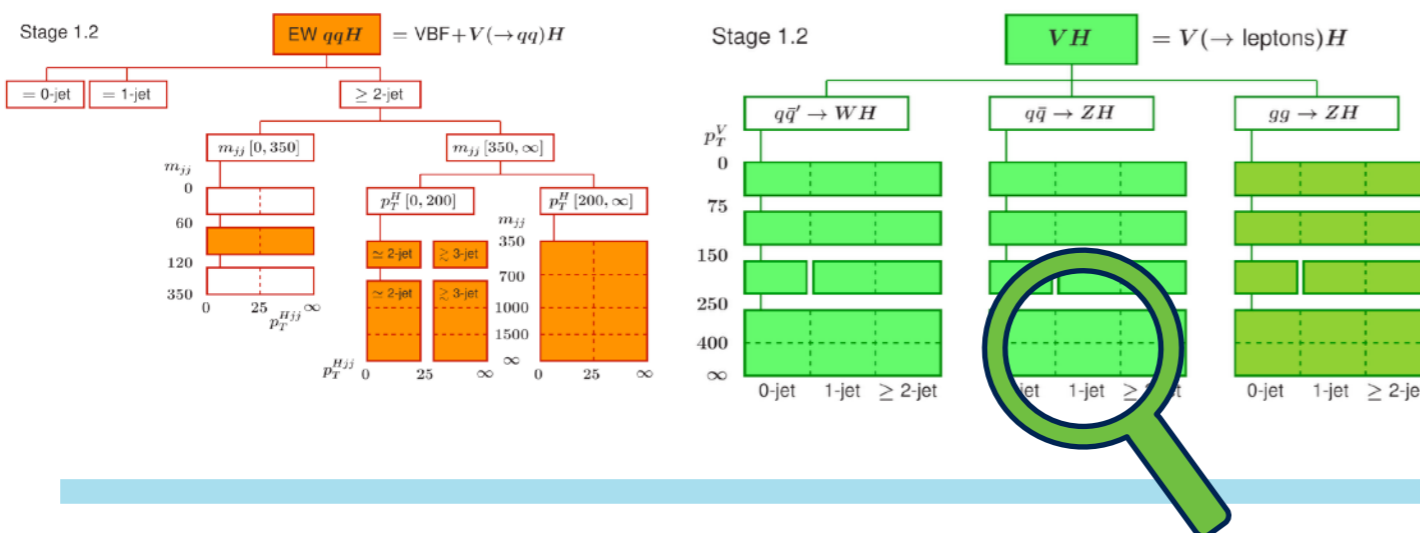
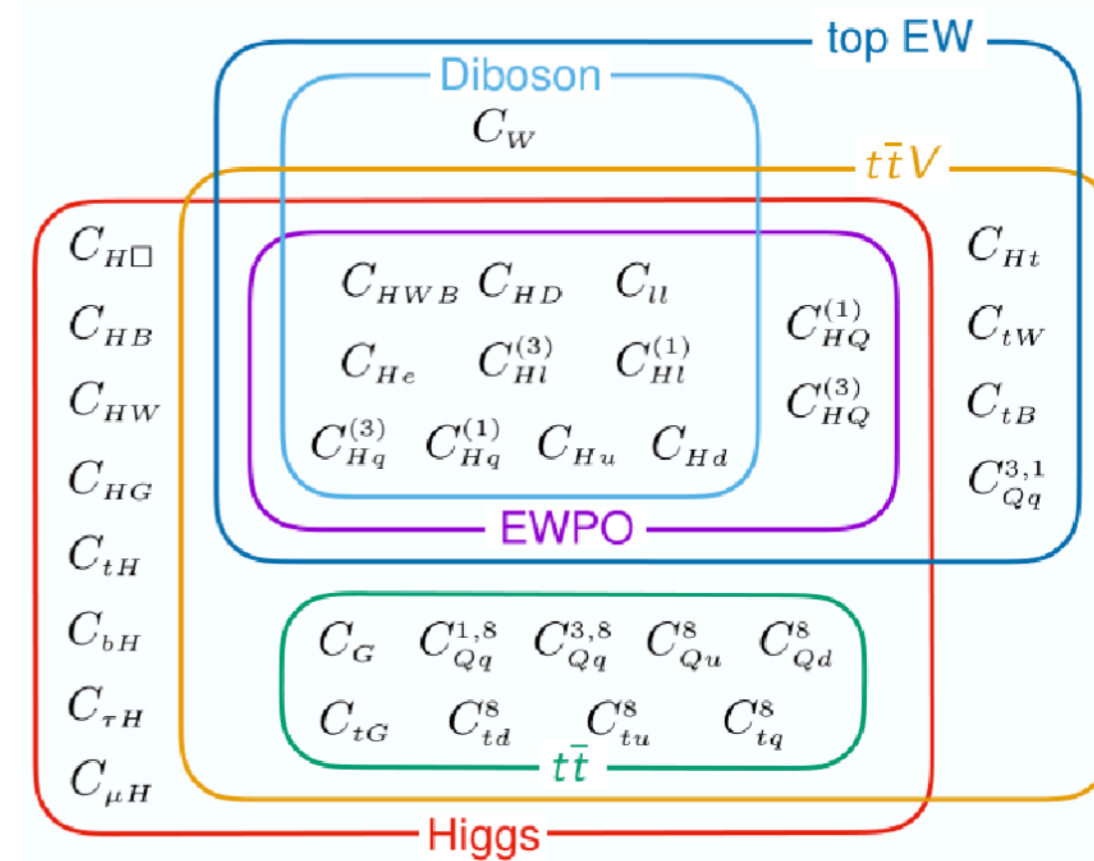
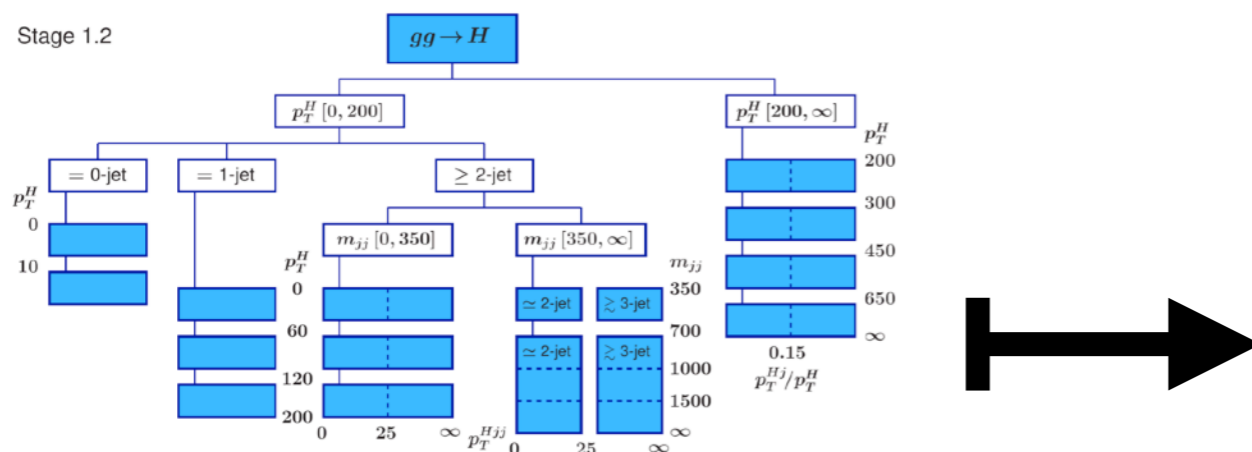
$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{i=1}^{2499} \frac{C_i}{\Lambda^2} \mathcal{O}_i$$

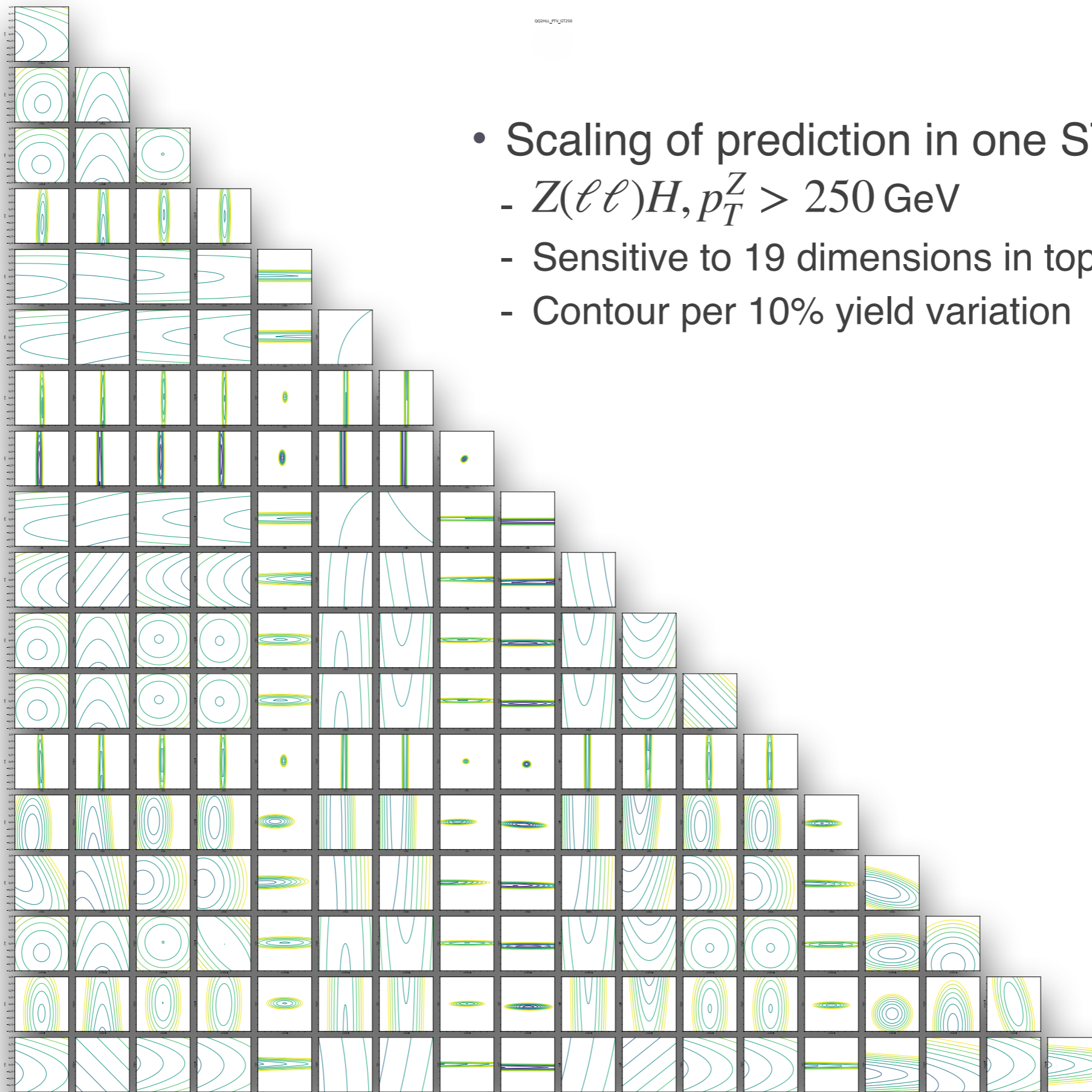


# Towards a global EFT analysis

- Theorists combine STXS result into global picture
  - Including also  $W/Z/t$  unfolded results
- 2499 dim-6 operators  $\rightarrow$  34 with flavor symmetry restrictions

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{i=1}^{2499} \frac{C_i}{\Lambda^2} \mathcal{O}_i$$



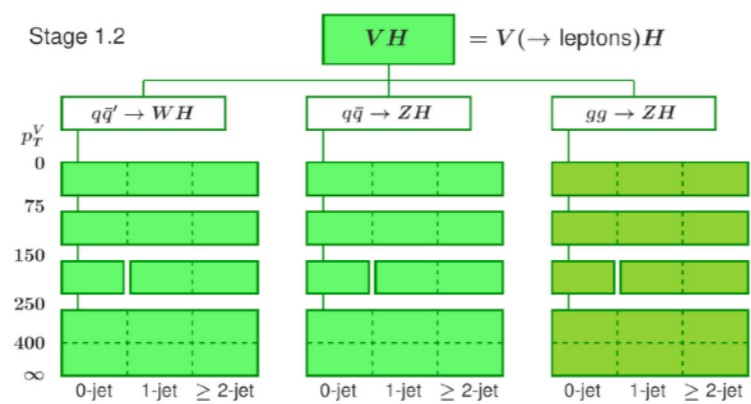
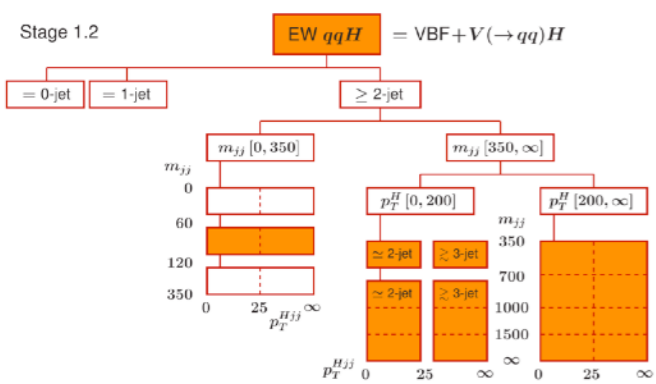
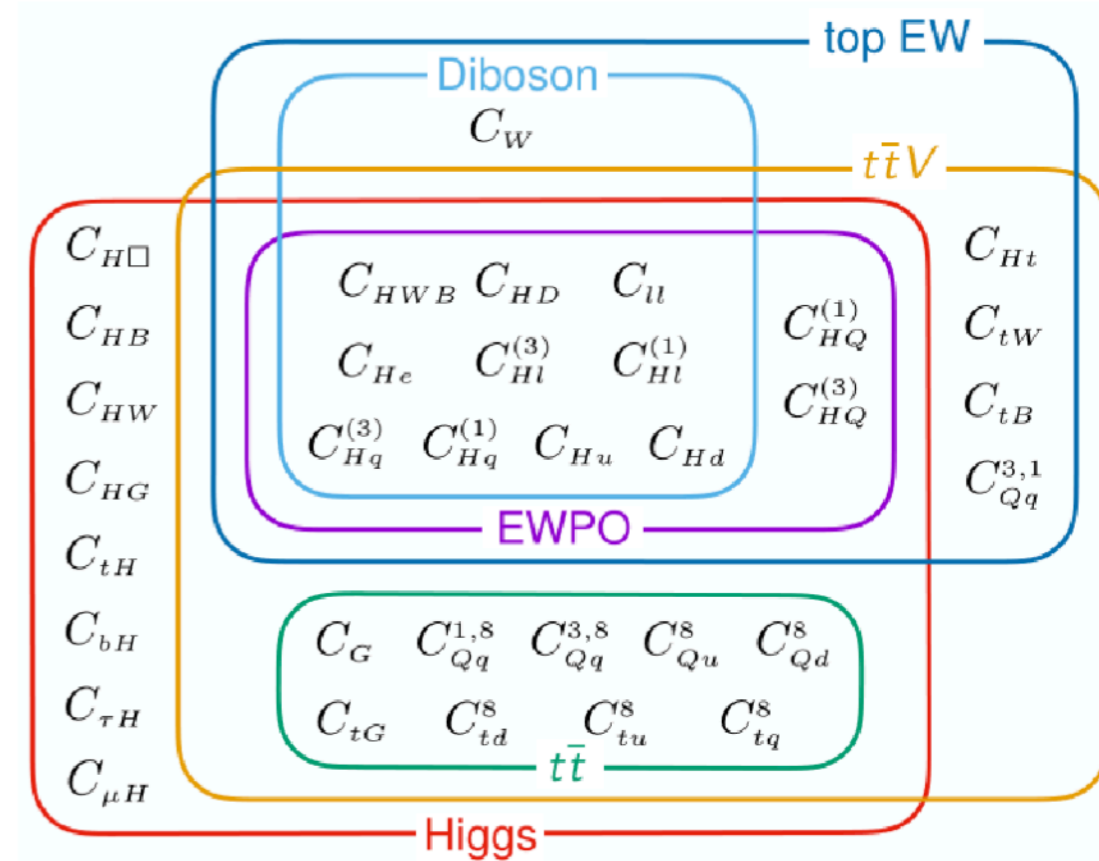
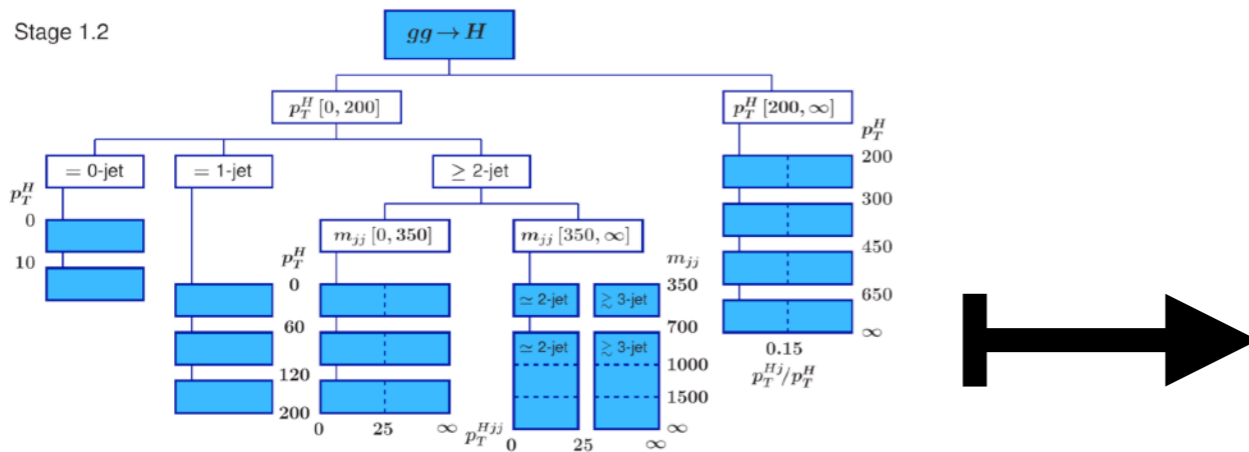


- Scaling of prediction in one STXS bin
  - $Z(\ell\ell)H, p_T^Z > 250 \text{ GeV}$
  - Sensitive to 19 dimensions in topU3l model
  - Contour per 10% yield variation

# Towards a global EFT analysis

- Theorists combine STXS result into global picture
- 2499 dim-6 operators  $\rightarrow$  34 with flavor symmetry restrictions

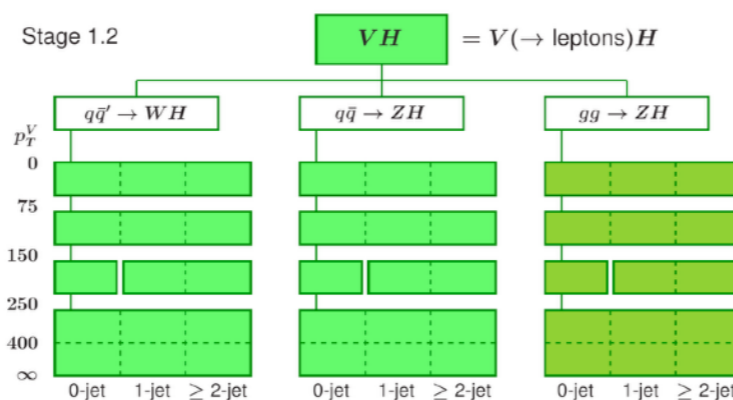
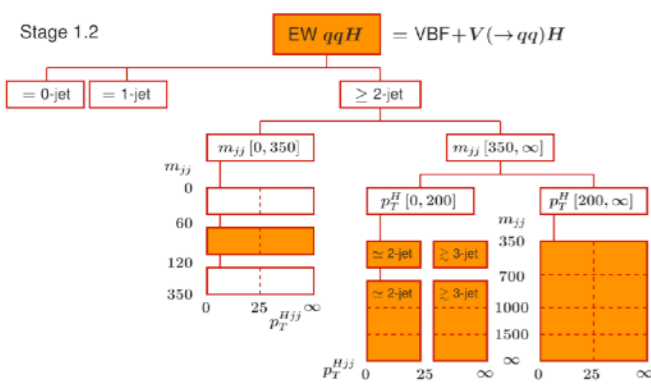
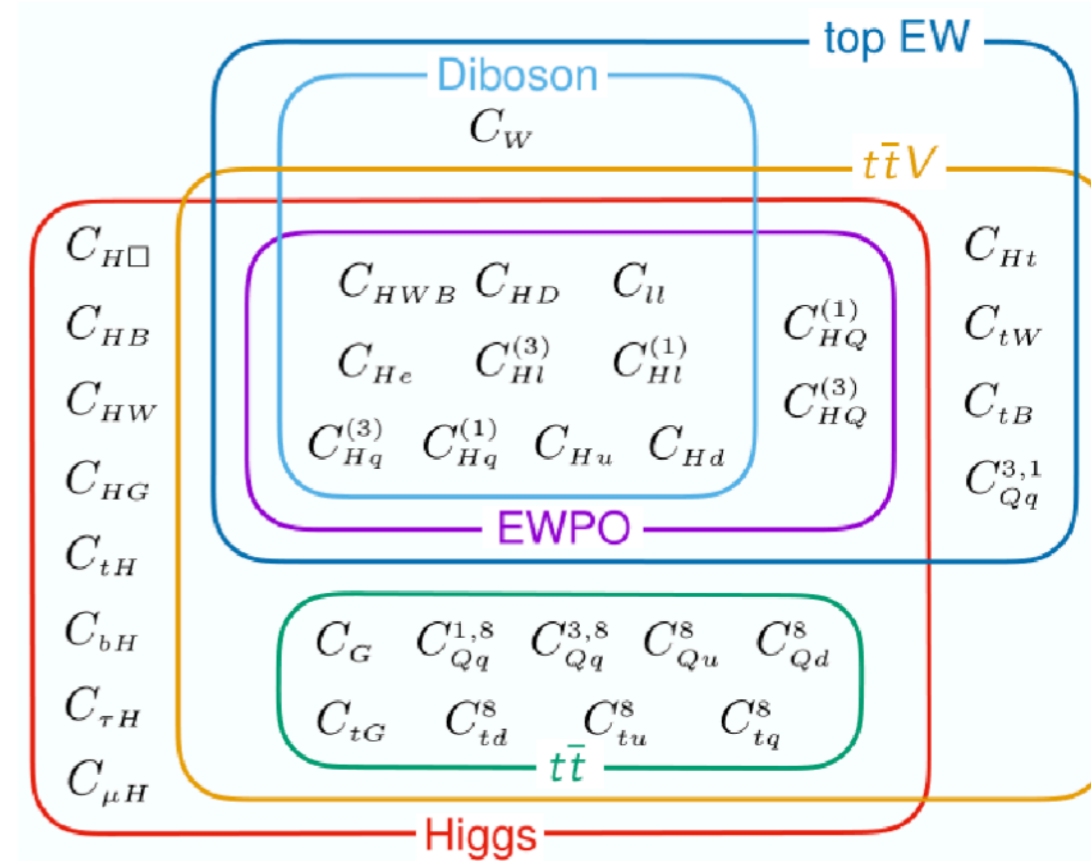
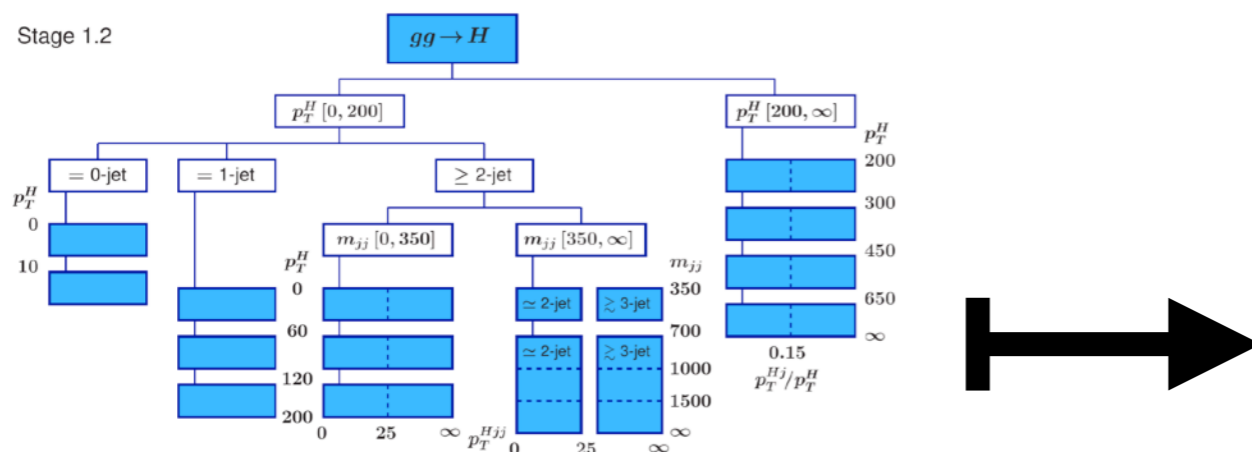
$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{i=1}^{2499} \frac{C_i}{\Lambda^2} \mathcal{O}_i$$



# Towards a global EFT analysis

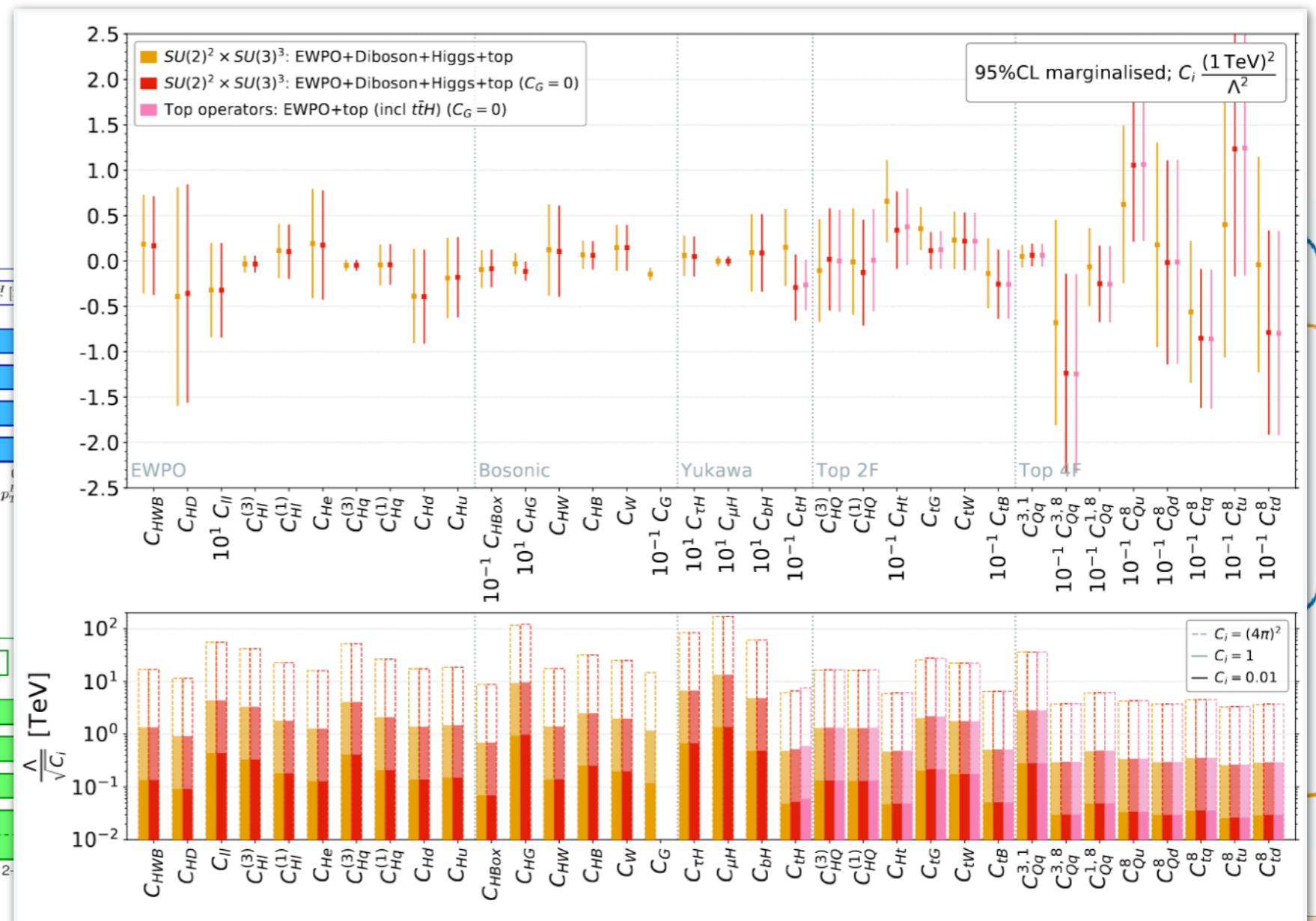
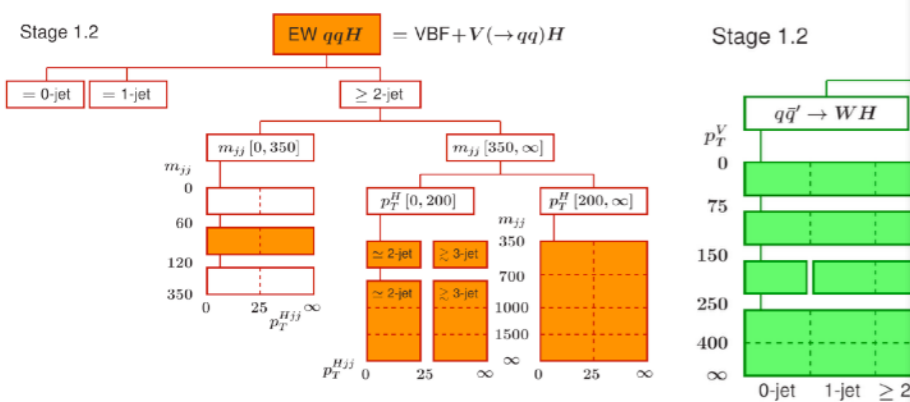
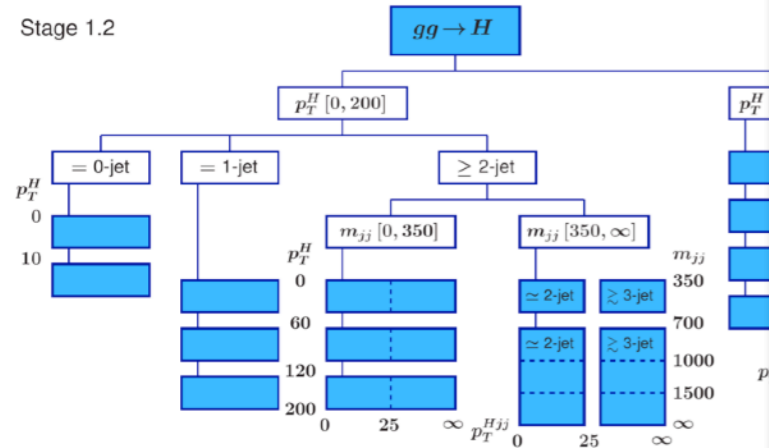
- Theorists combine STXS result into global picture
- 2499 dim-6 operators  $\rightarrow$  34 with flavor symmetry restrictions
- Result for either fixed  $\Lambda$  or fixed  $C_i$

$$\mathcal{L}_{\text{SMEFT}} = \mathcal{L}_{\text{SM}} + \sum_{i=1}^{2499} \frac{C_i}{\Lambda^2} \mathcal{O}_i$$



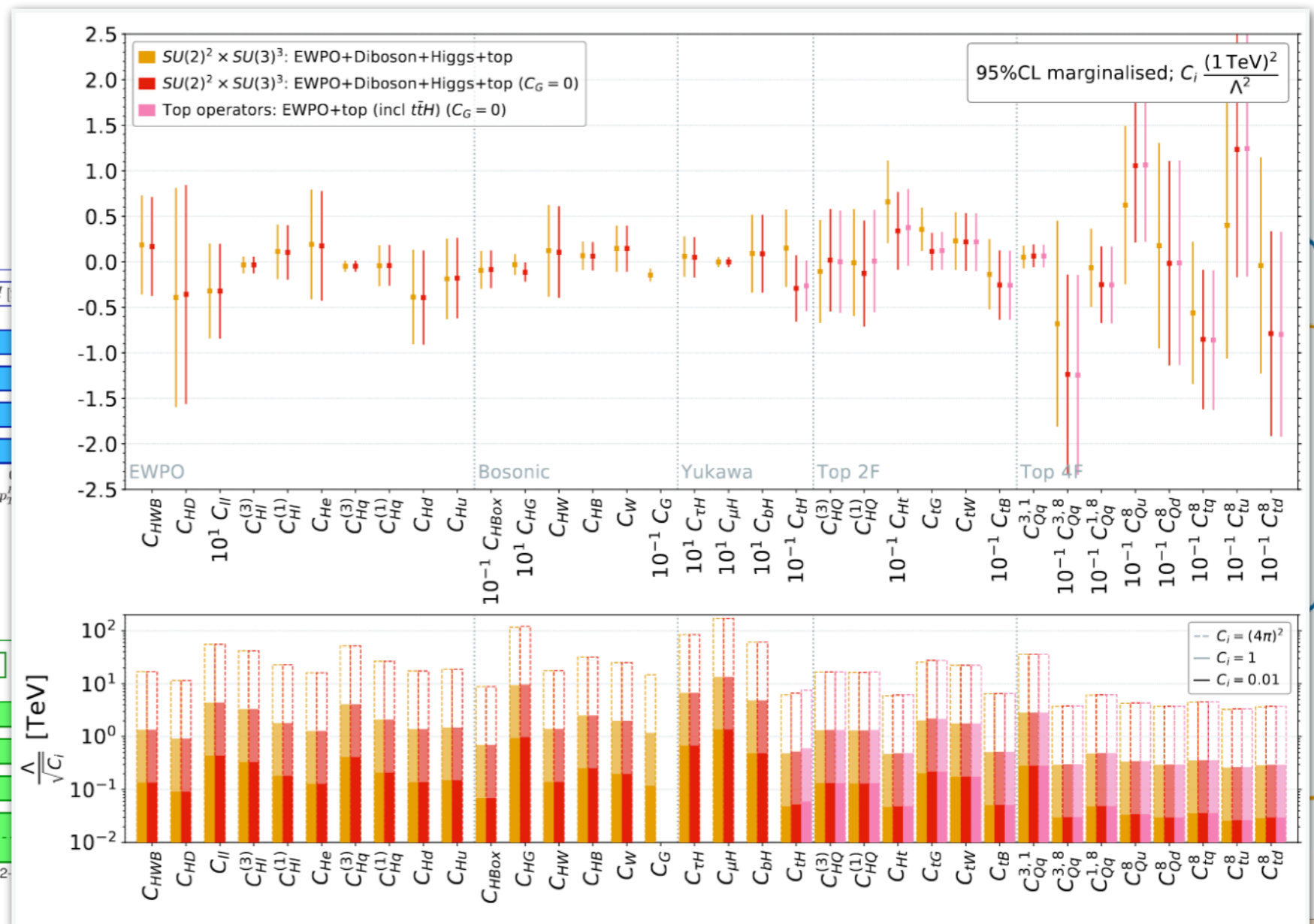
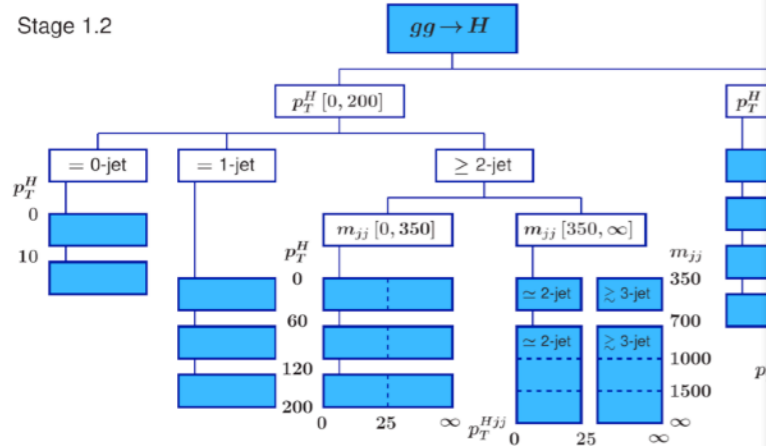
# Towards a global EFT analysis

- Theorists combine STXS result into global picture
- 2499 dim-6 operators  $\rightarrow$  34 with flavor symmetry restrictions
- Result for either fixed  $\Lambda$  or fixed  $C_i$



# Towards a global EFT analysis

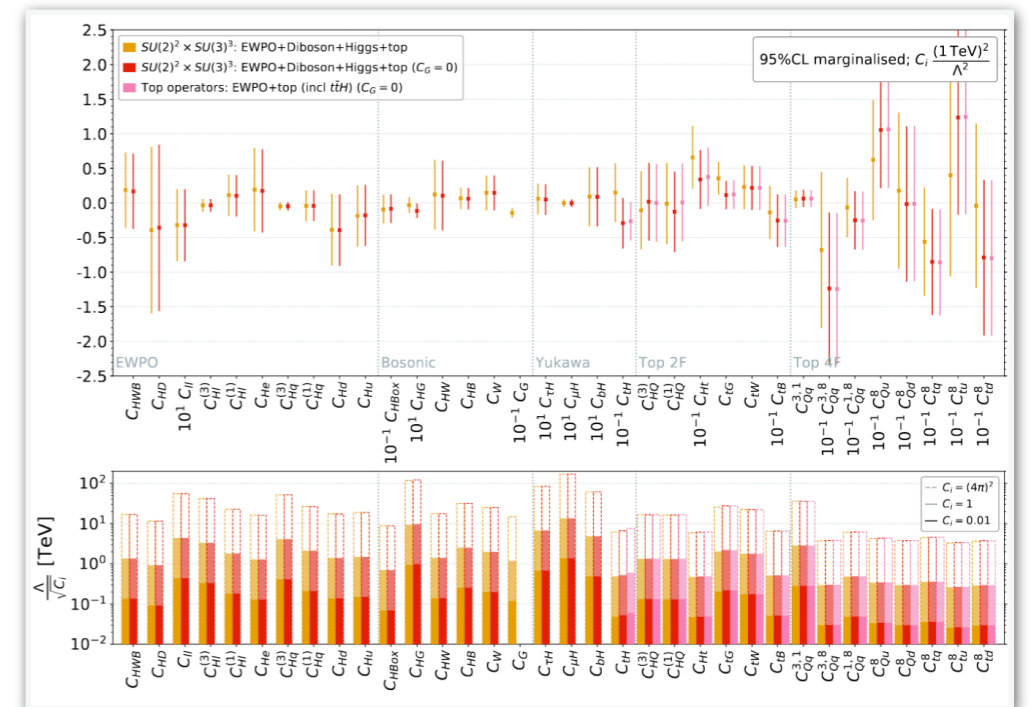
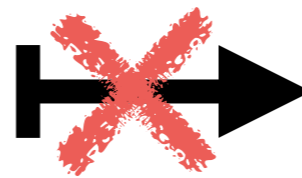
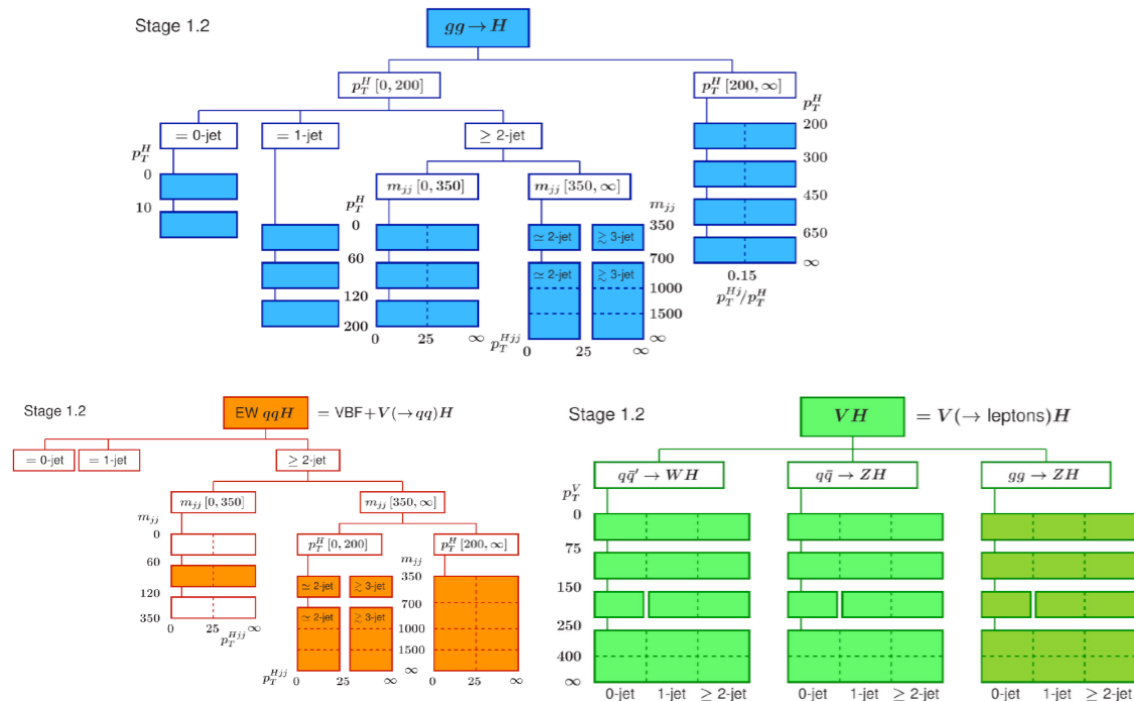
- Theorists combine STXS result into global picture
- 2499 dim-6 operators  $\rightarrow$  34 with flavor symmetry restrictions
- Result for either fixed  $\Lambda$  or fixed  $C_i$



Some directions  $> 13 \text{ TeV!}$

# Issues with delegating global analysis

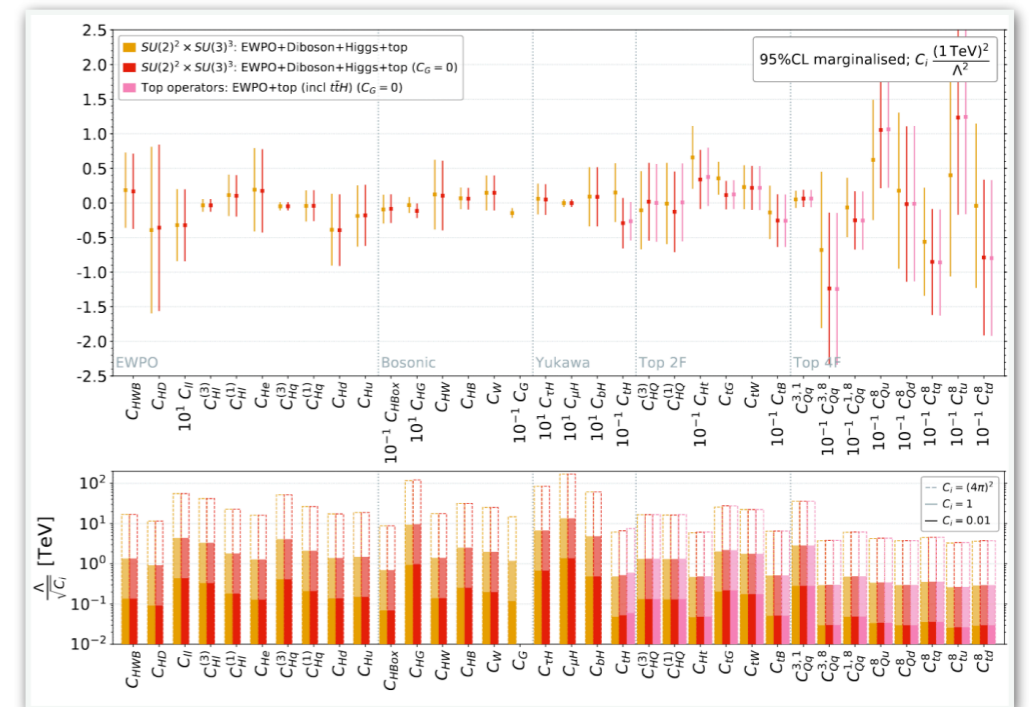
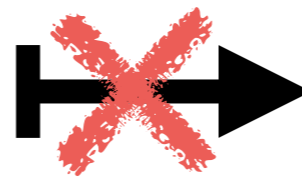
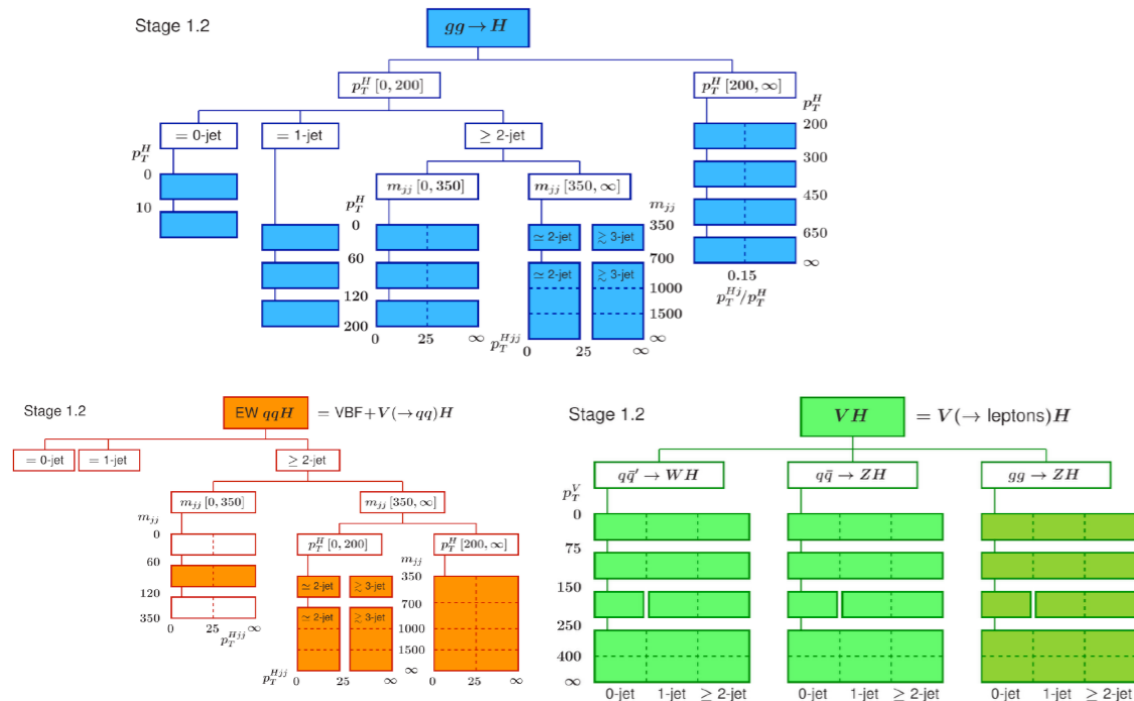
- Limited correlation information in CMS/ATLAS results
  - Could fix with public likelihood ([arxiv:2109.04981](https://arxiv.org/abs/2109.04981)), but workflow challenges remain





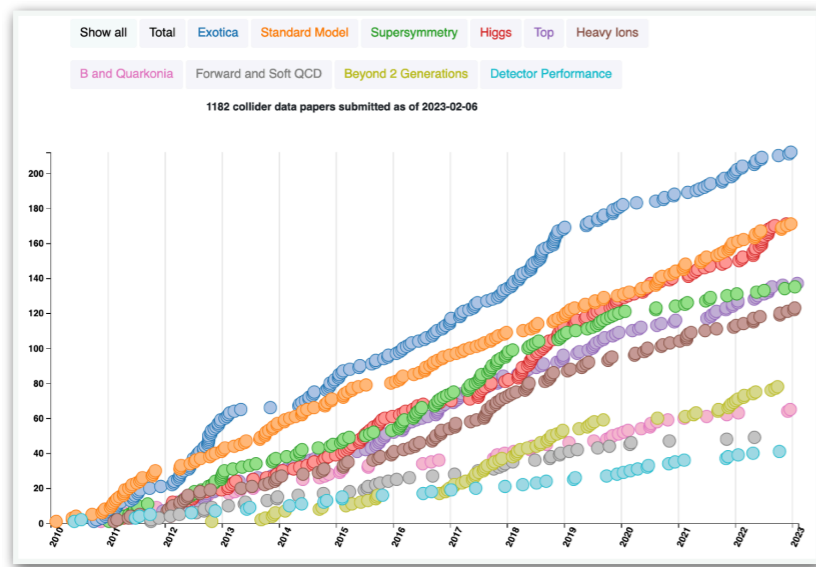
# Issues with delegating global analysis

- Limited correlation information in CMS/ATLAS results
  - Could fix with public likelihood ([arxiv:2109.04981](https://arxiv.org/abs/2109.04981)), but workflow challenges remain
- Experiments choose the observables
  - Higgs STXS: some coefficient dependence integrated out
  - W/Z/top: “rare SM process search” reinterpretation
    - Add boosted hadronic channels!
  - Often ignore overlap and background EFT scaling

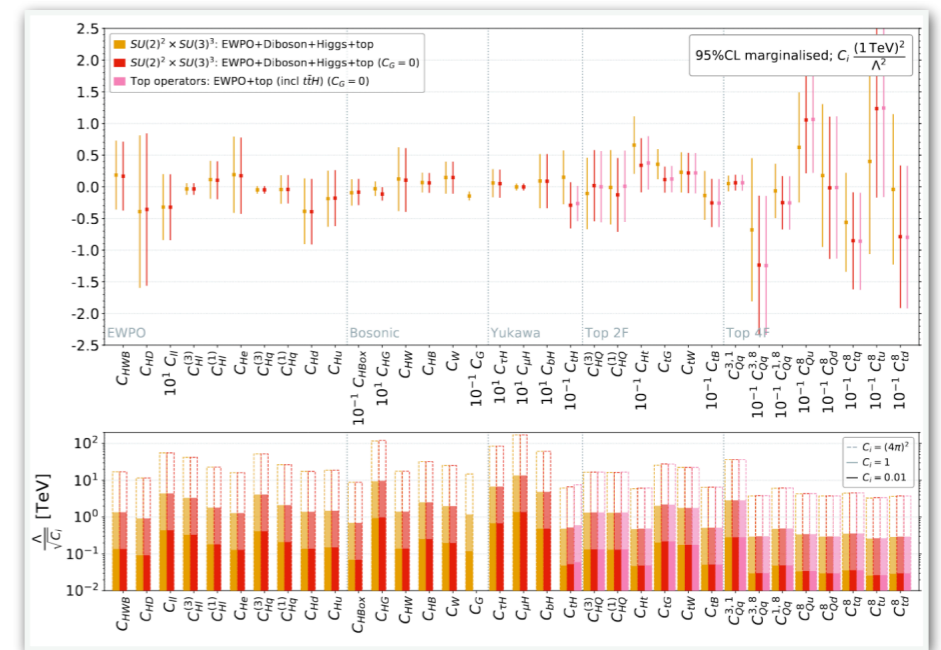
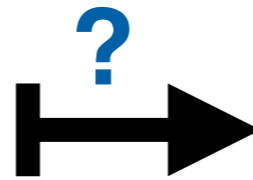




# More generally...



CMS publication count



Indirect searches for new physics

# The power of combination

- A common theme

## Joining forces: the NOvA-T2K joint fit

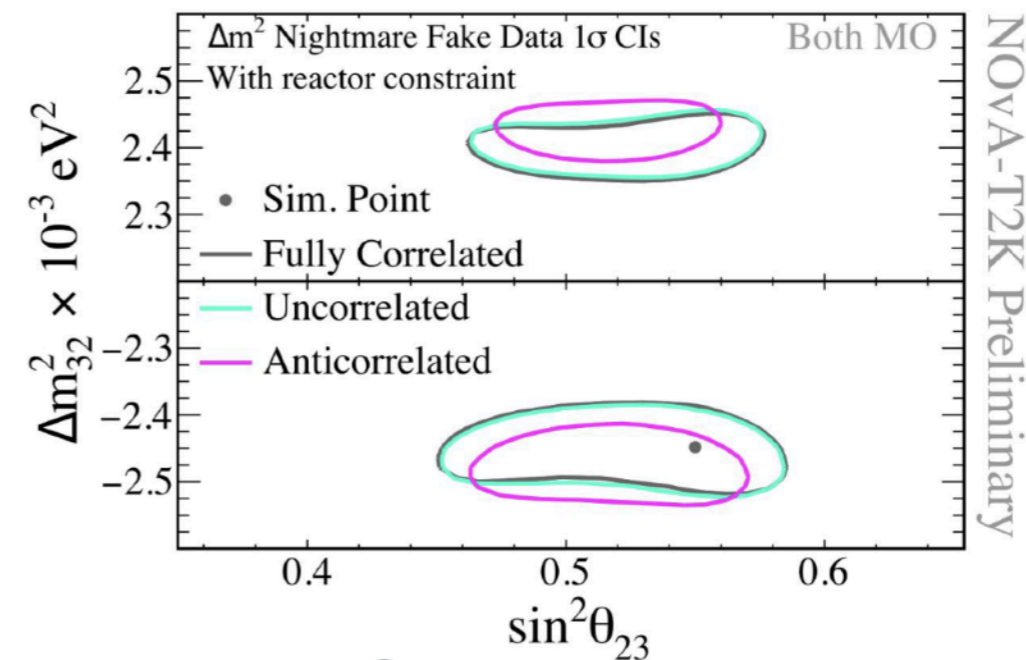
Scope: resolving degeneracies

This is a preliminary result

- The experiments have different analysis approaches driven by contrasting detector designs  
→ combination is far from trivial
- Challenge: When? What? How? to correlate common physics parameters between the two experiments.
  - Flux & detector models → uncorrelated
  - XS model → there will be dragons:  
even if energy is different and model used are different, the underlying processes are the same.

Post-hoc strategy: we can learn from this!

Paper in preparation, material from [FNAL W&C](#).



**Lessons Learned:**  
plan ahead on dataset  
combination!!!

# The power of combination

- A common theme

([arxiv:1802.08296](https://arxiv.org/abs/1802.08296), [arxiv:1902.06765](https://arxiv.org/abs/1902.06765))

## Joining forces: the NOvA-T2K joint fit

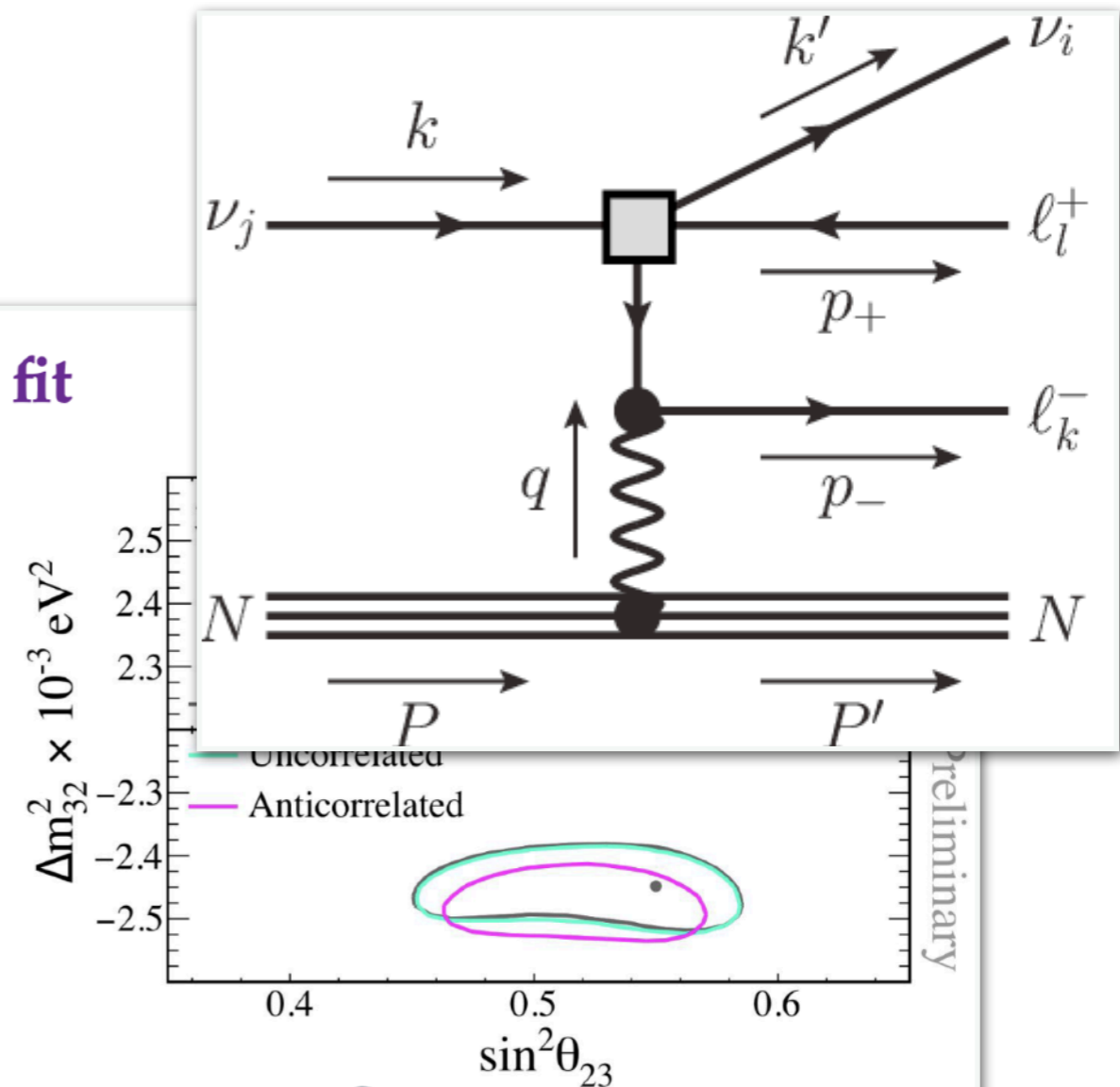
Scope: resolving degeneracies

This is a preliminary result

- The experiments have different analysis approaches driven by contrasting detector designs → combination is far from trivial
- Challenge: When? What? How? to correlate common physics parameters between the two experiments.
  - Flux & detector models → uncorrelated
  - XS model → there will be dragons: even if energy is different and model used are different, the underlying processes are the same.

Post-hoc strategy: we can learn from this!

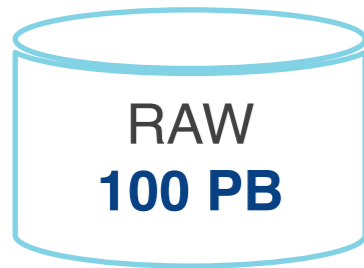
Paper in preparation, material from [FNAL W&C](#).



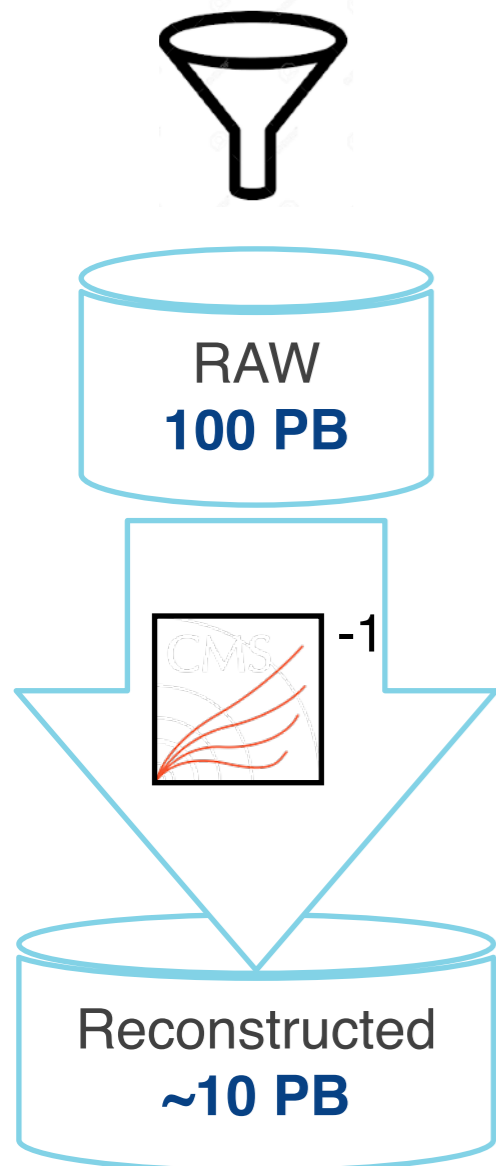
**Lessons Learned:**  
plan ahead on dataset combination!!!



# Reducing data space



# Reducing data space



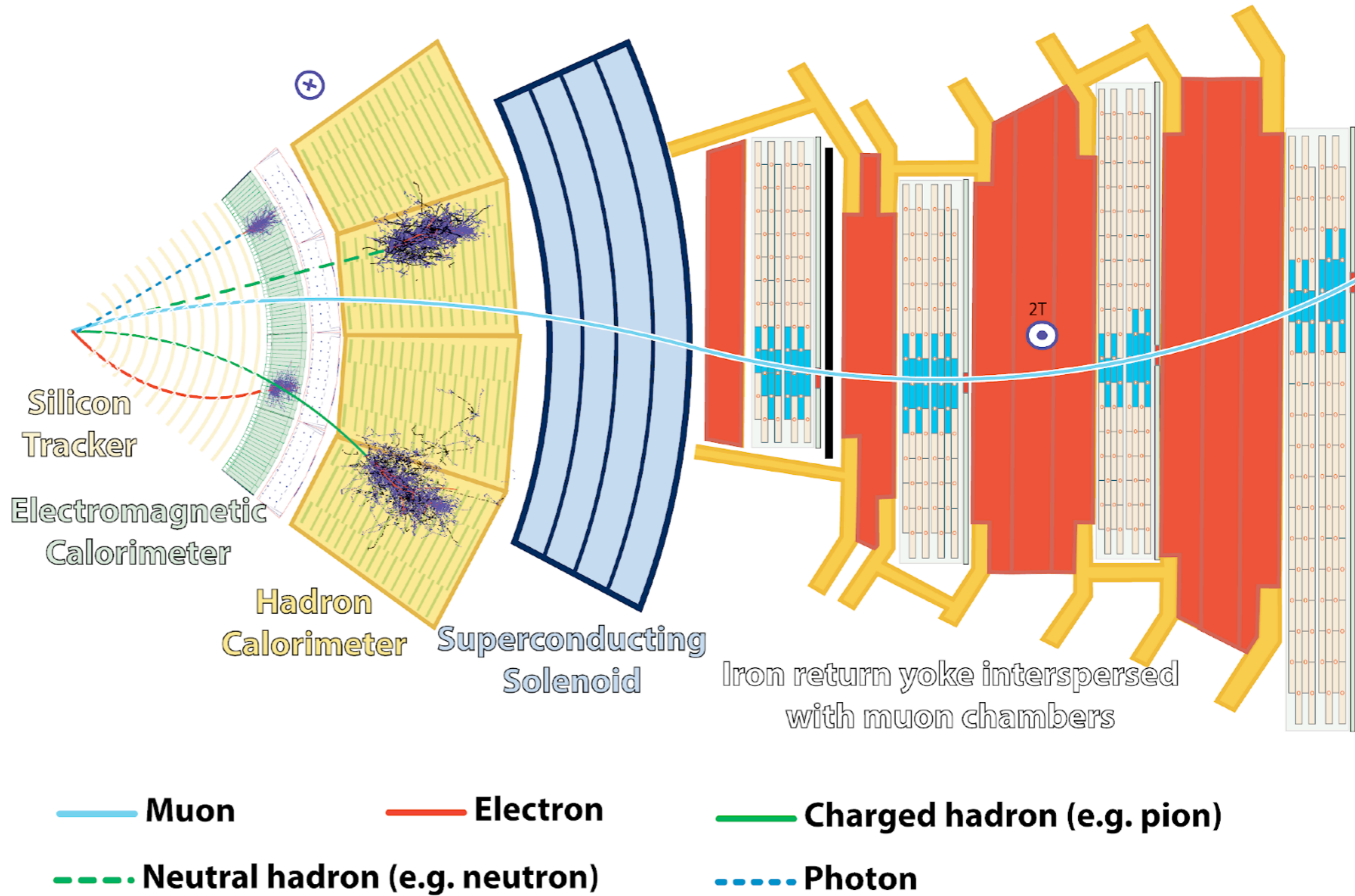
# Reducing data space



RAW  
**100 PB**



Reconstructed  
**~10 PB**

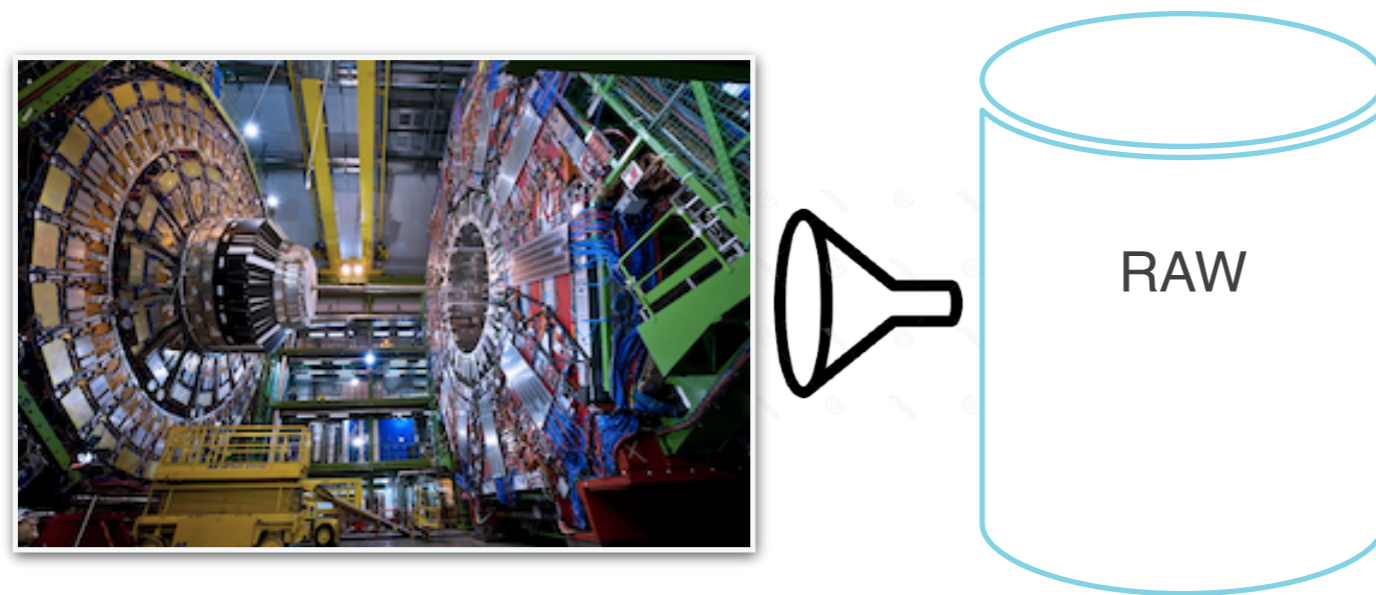


- Muon
- Electron
- Charged hadron (e.g. pion)
- - - Neutral hadron (e.g. neutron)
- - - Photon

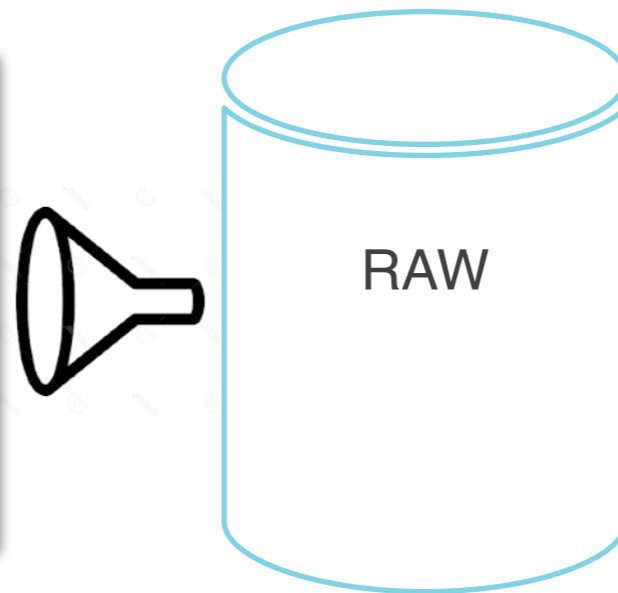
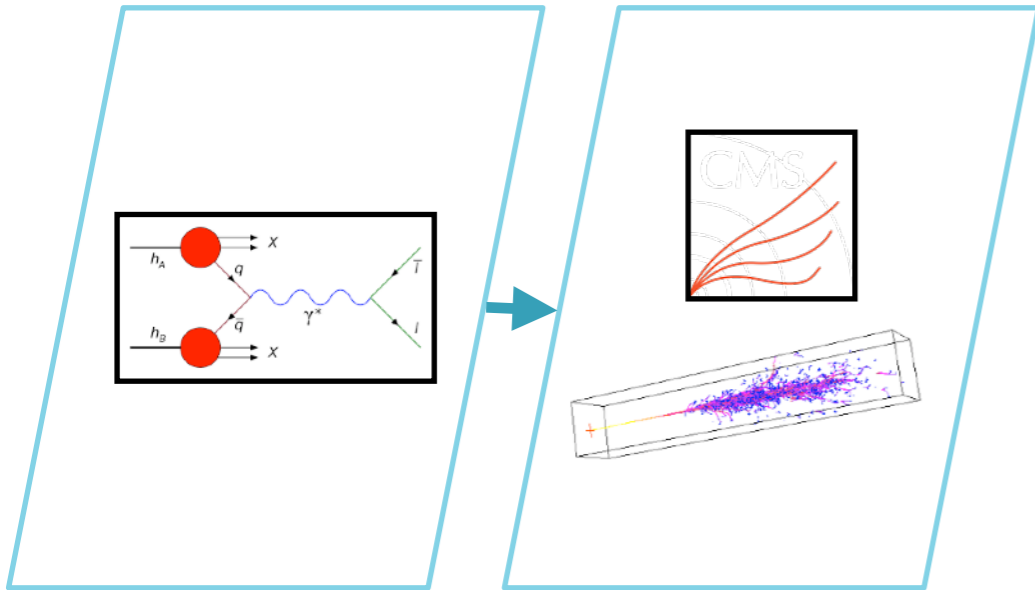


# Inference: the reality

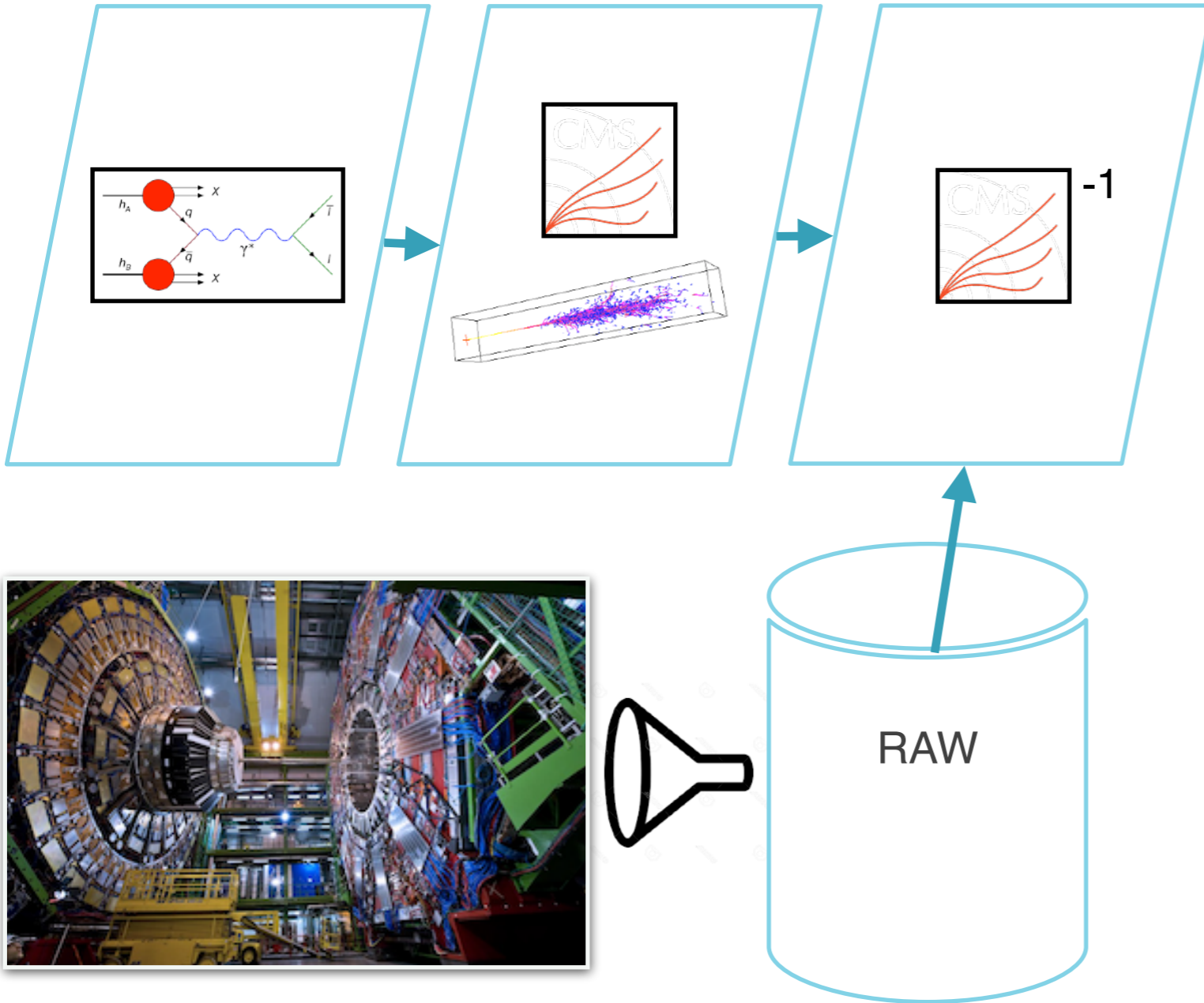
# Inference: the reality



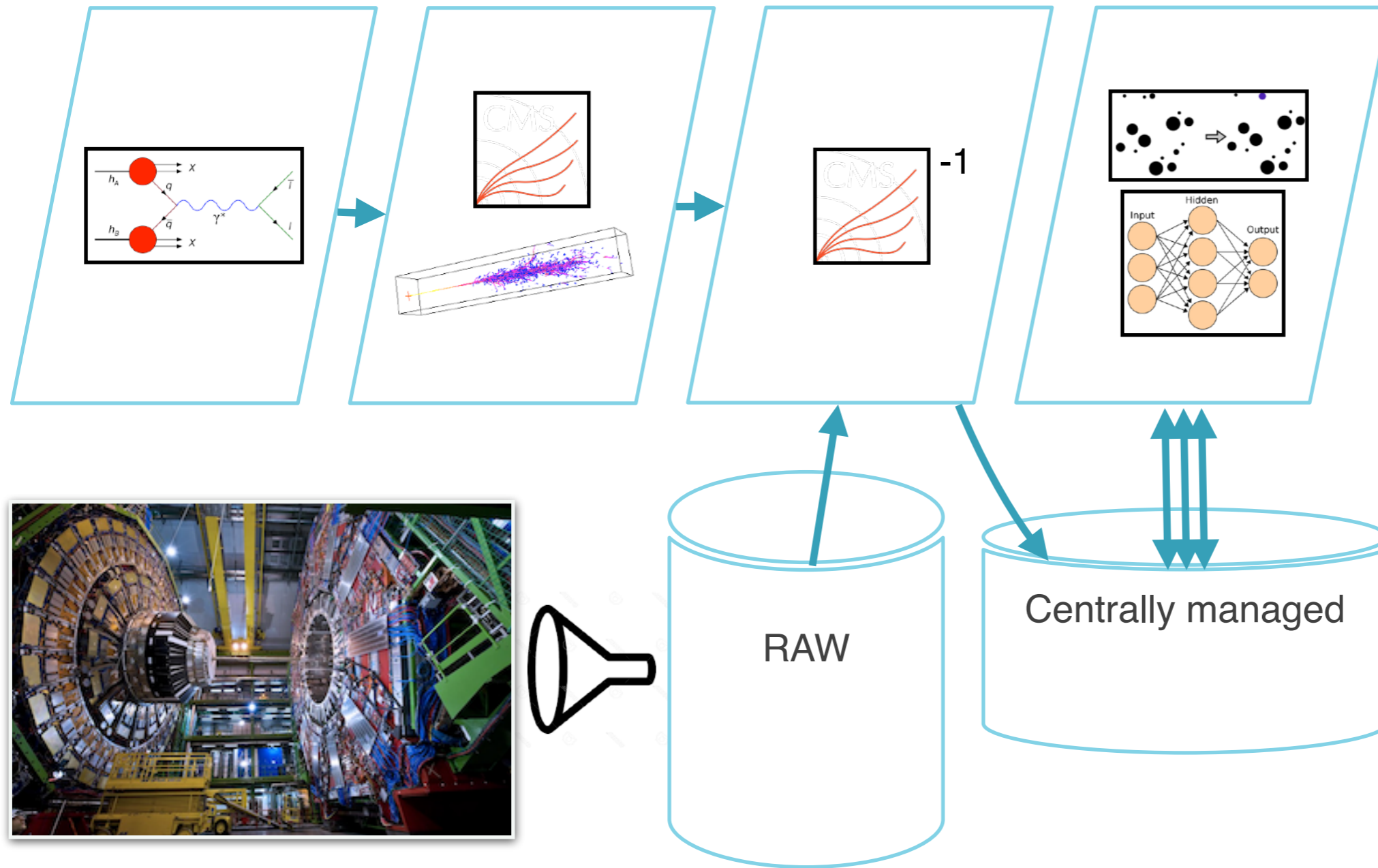
# Inference: the reality



# Inference: the reality

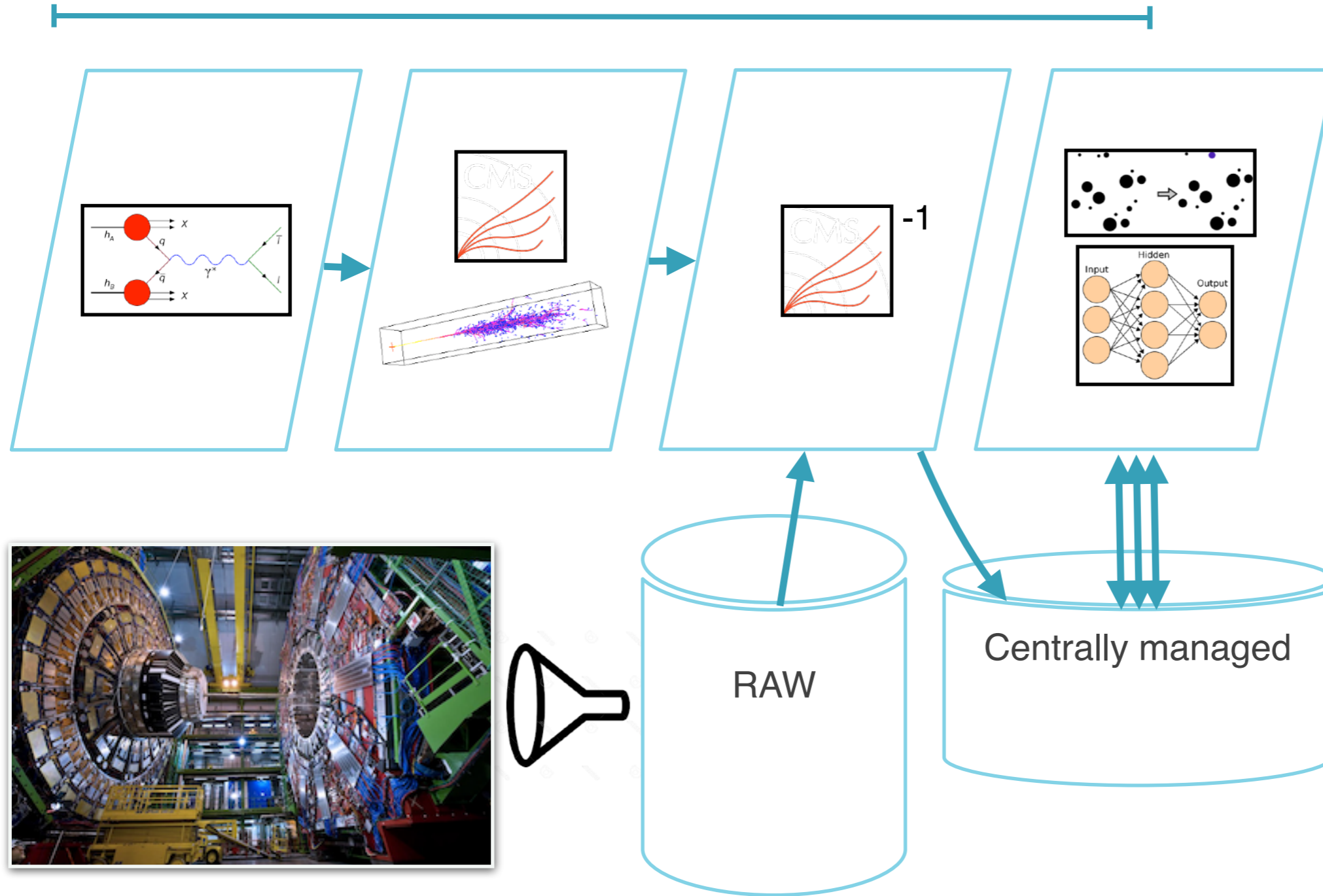


# Inference: the reality



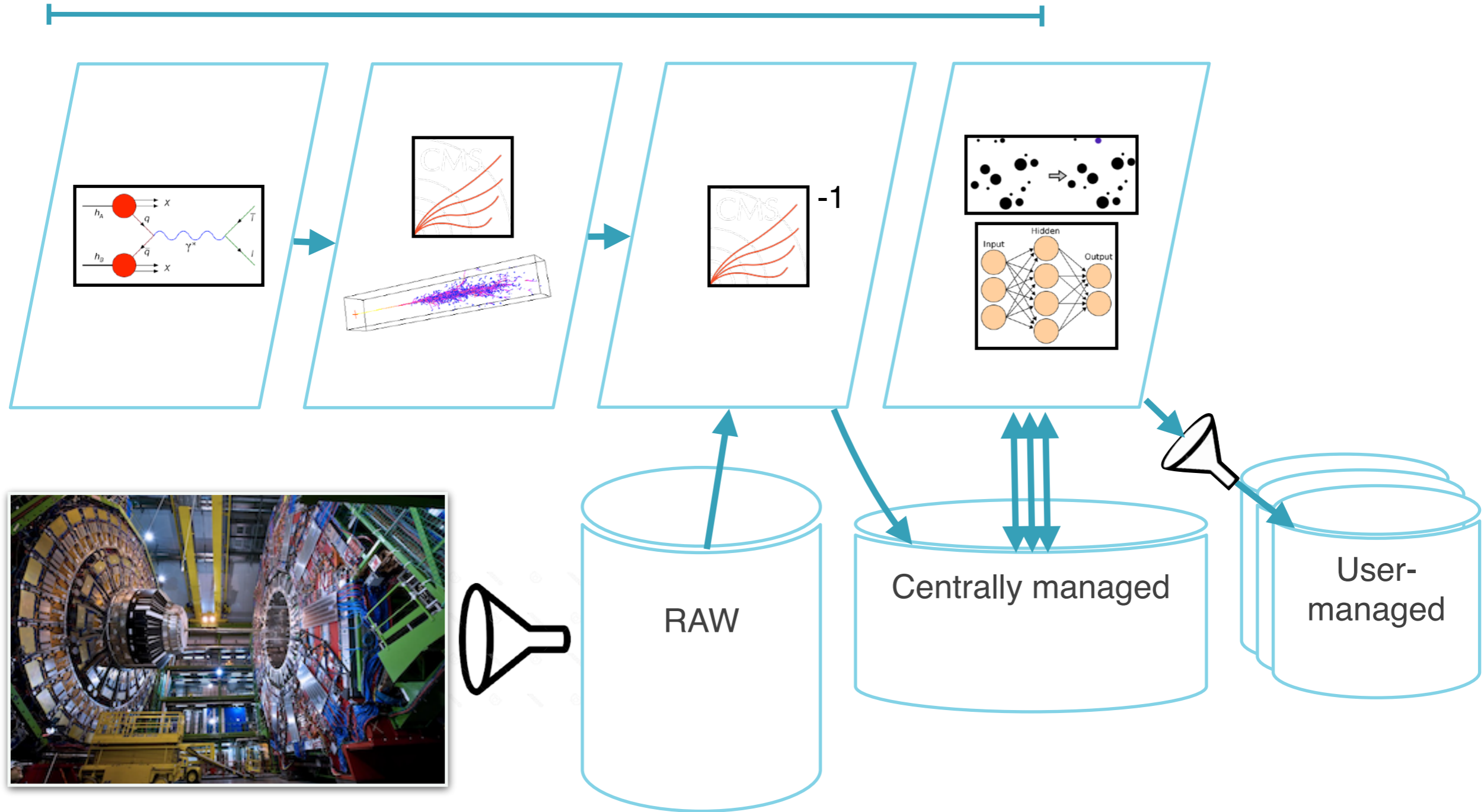
# Inference: the reality

Centrally planned, executed



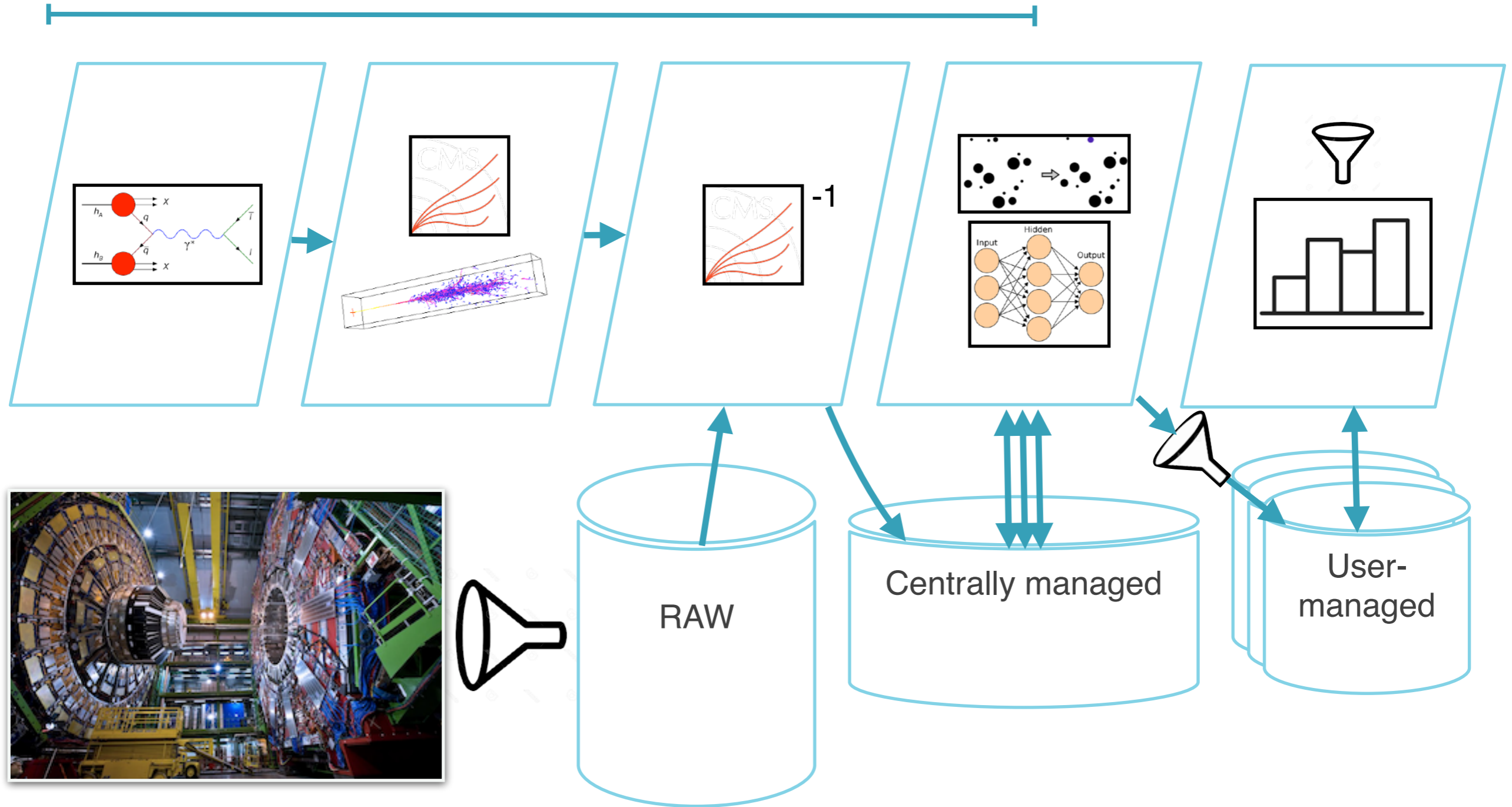
# Inference: the reality

Centrally planned, executed



# Inference: the reality

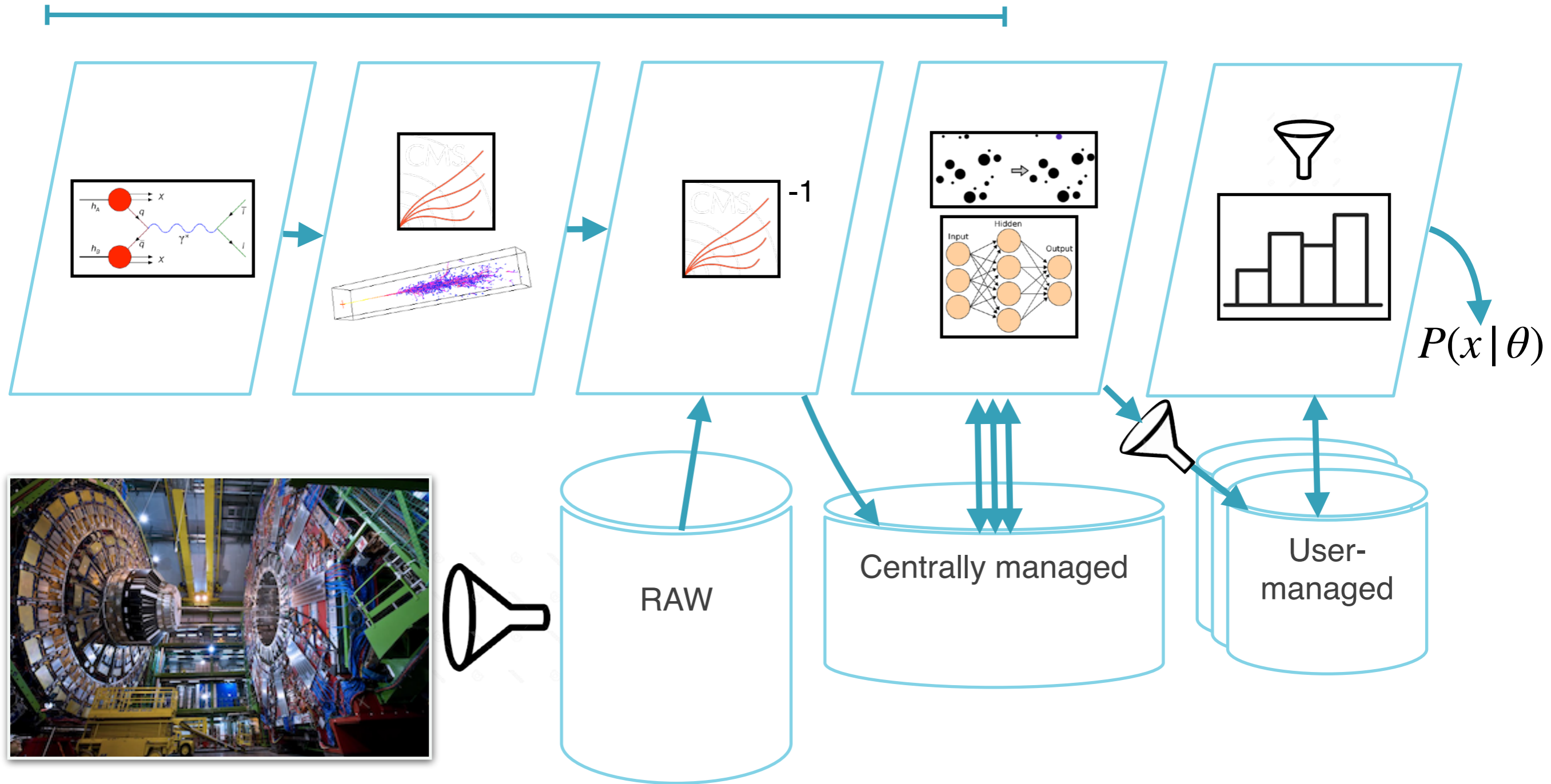
Centrally planned, executed





# Inference: the reality

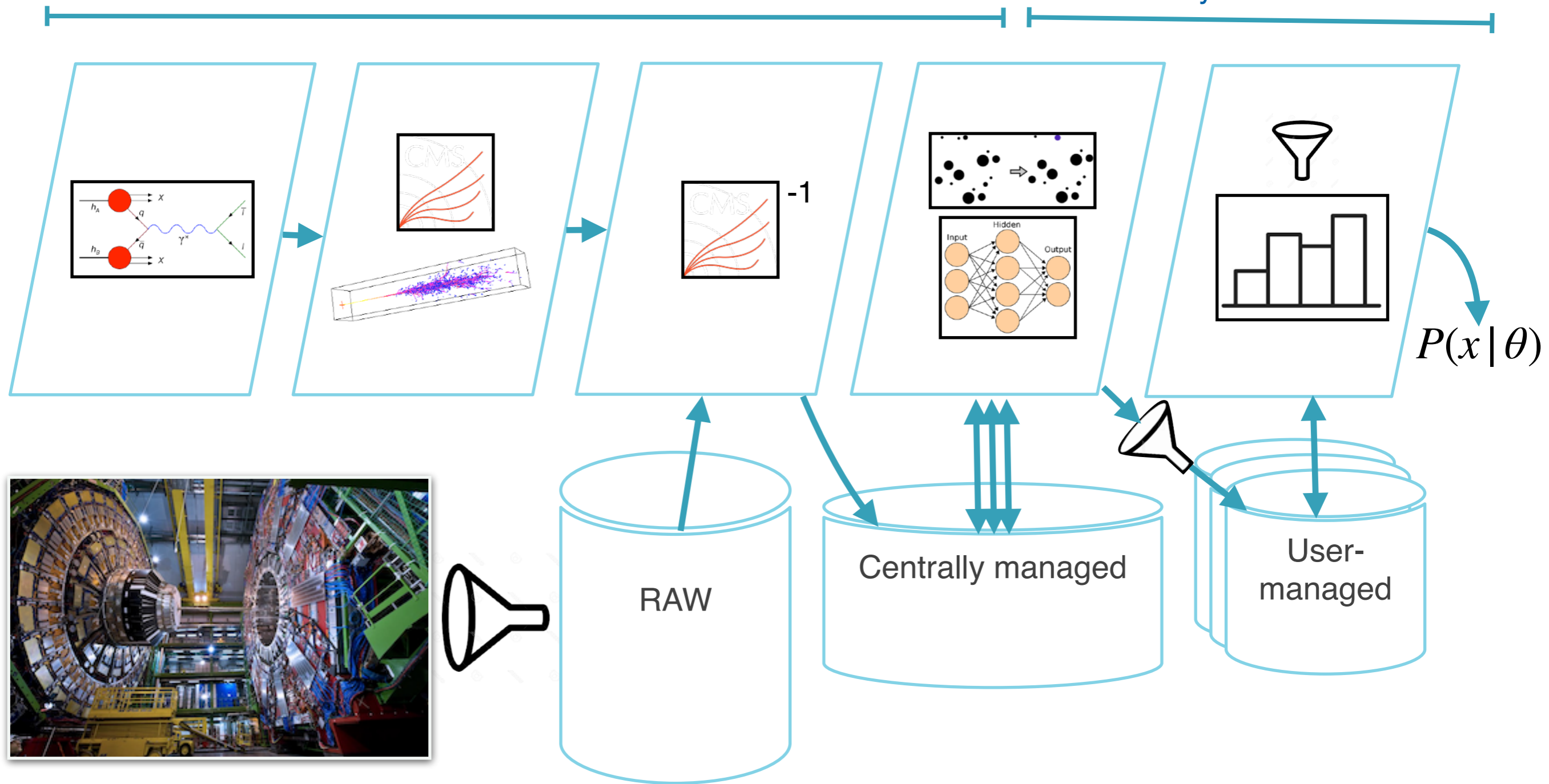
Centrally planned, executed



# Inference: the reality

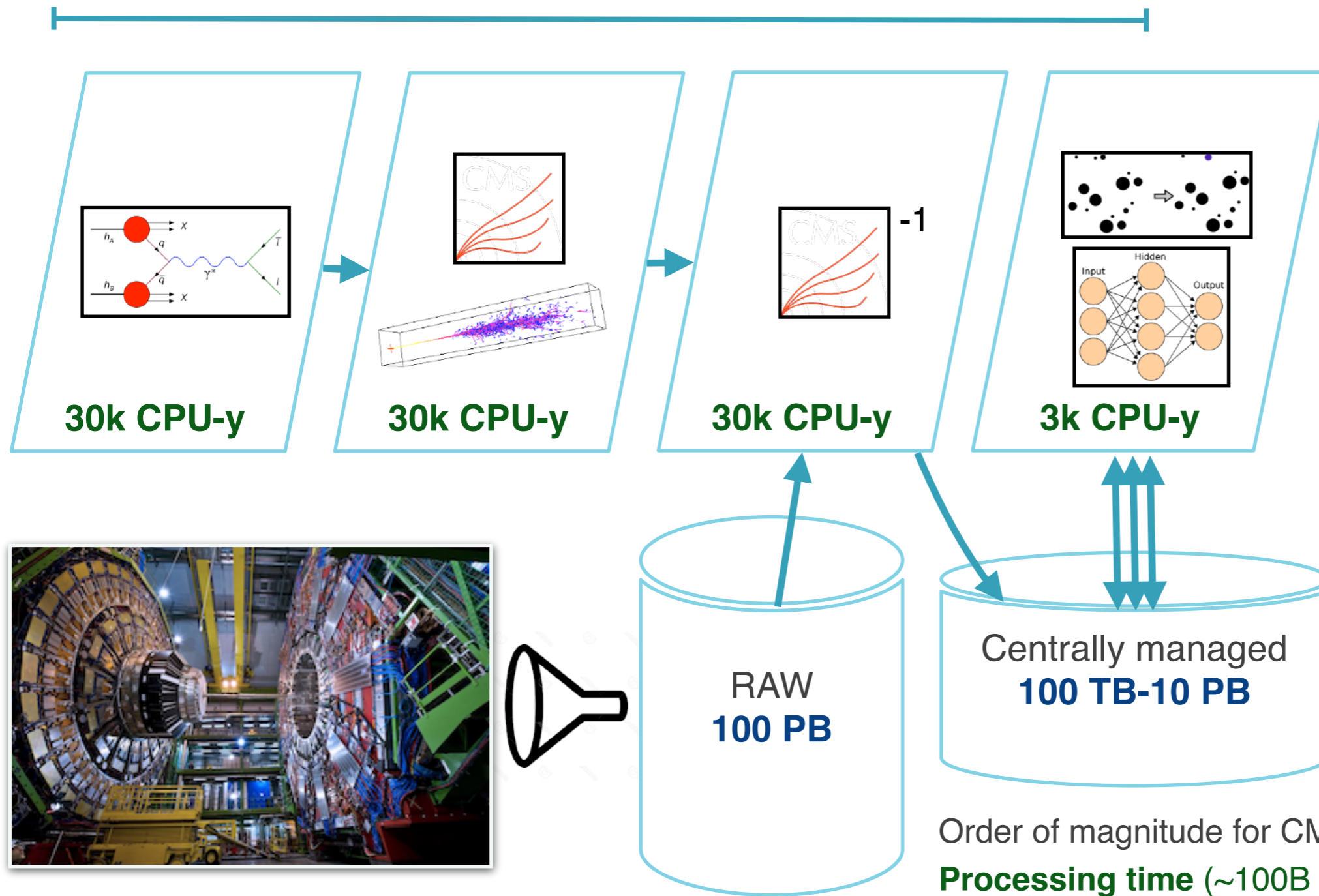
Centrally planned, executed

Analyst / Scientist



# Inference: the reality

Centrally planned, executed



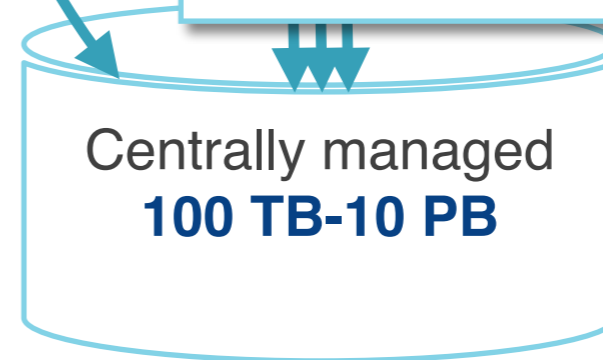
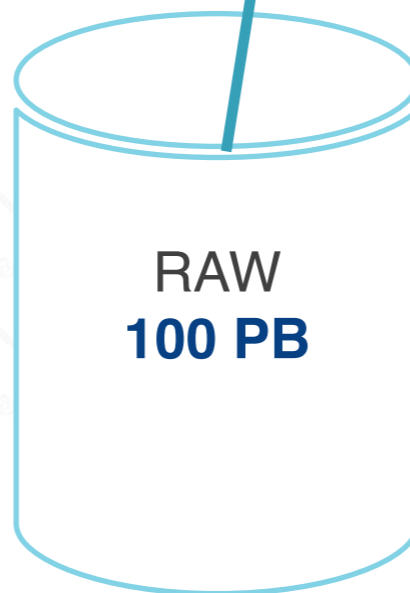
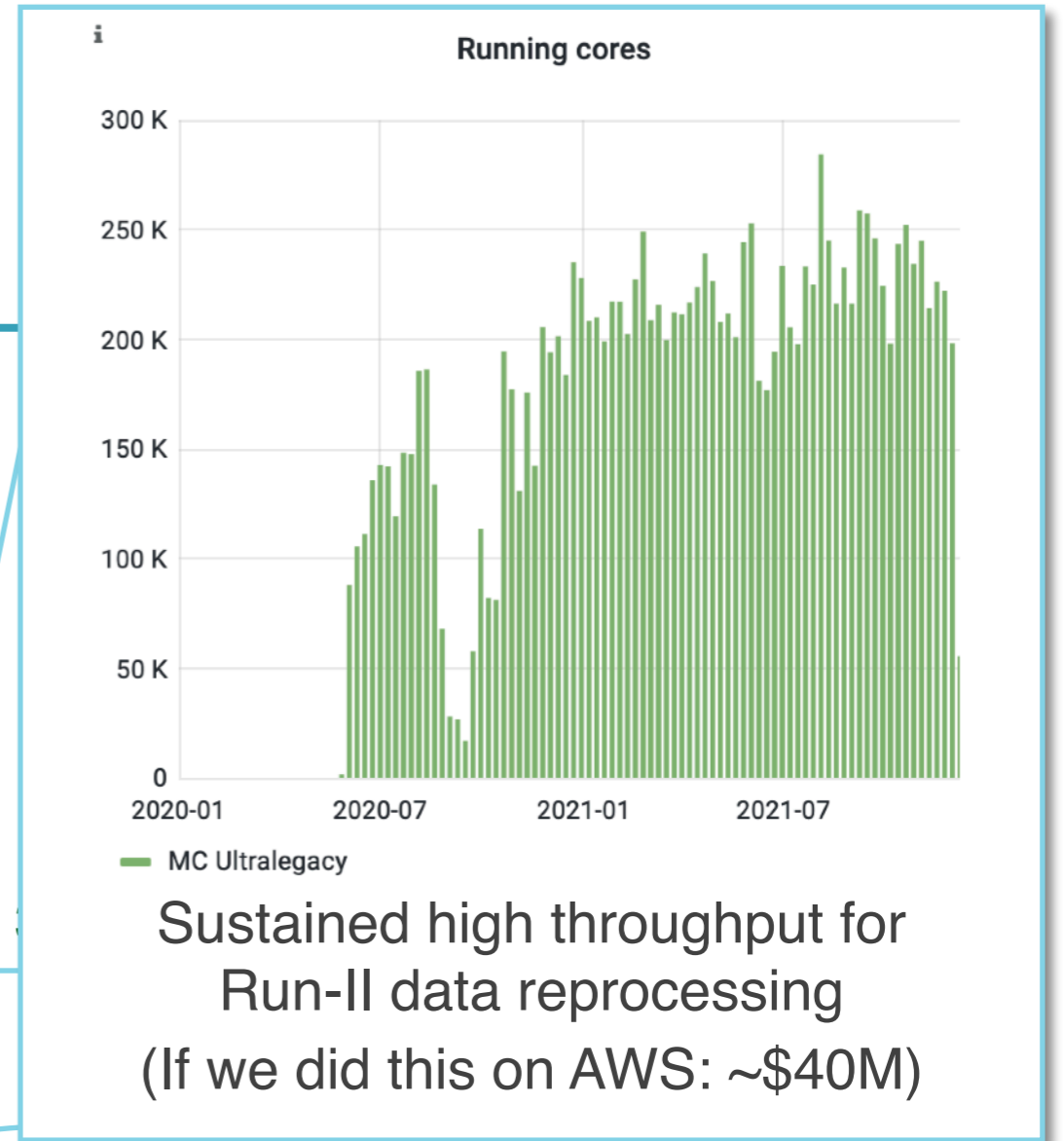
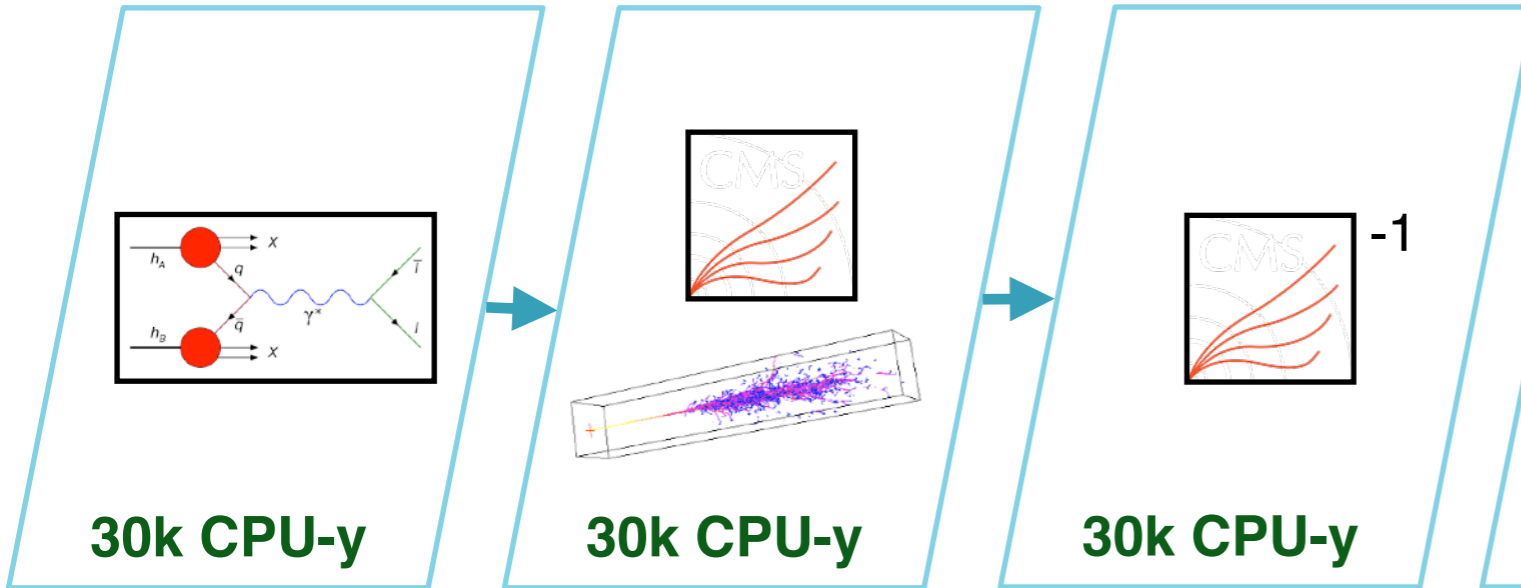
Order of magnitude for CMS Run-II (2016-18)

**Processing time** (~100B events)

**Data volume on disk**

# Inference: the reality

Centrally planned, executed



Order of magnitude for CMS Run-II (2016-18)

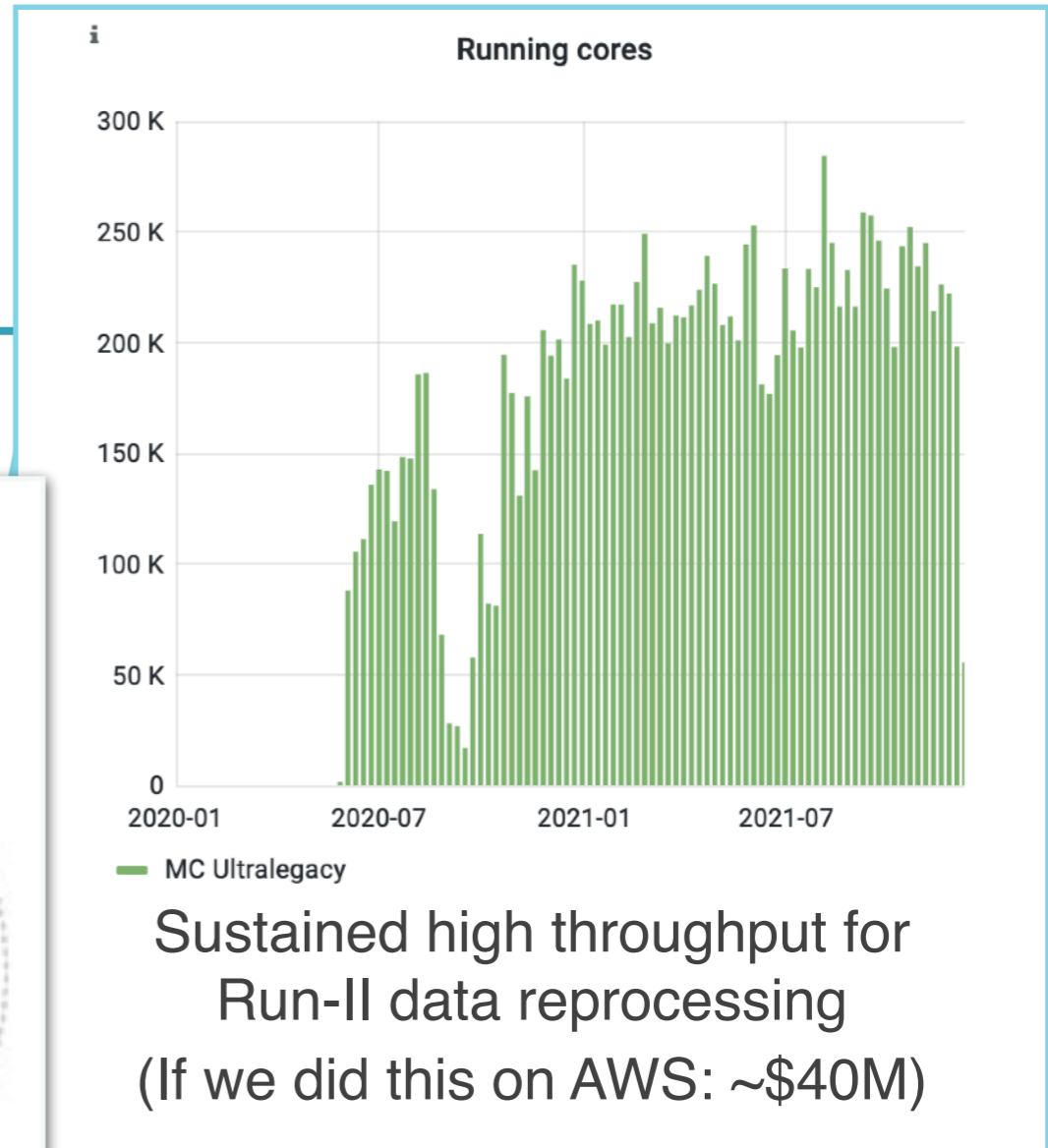
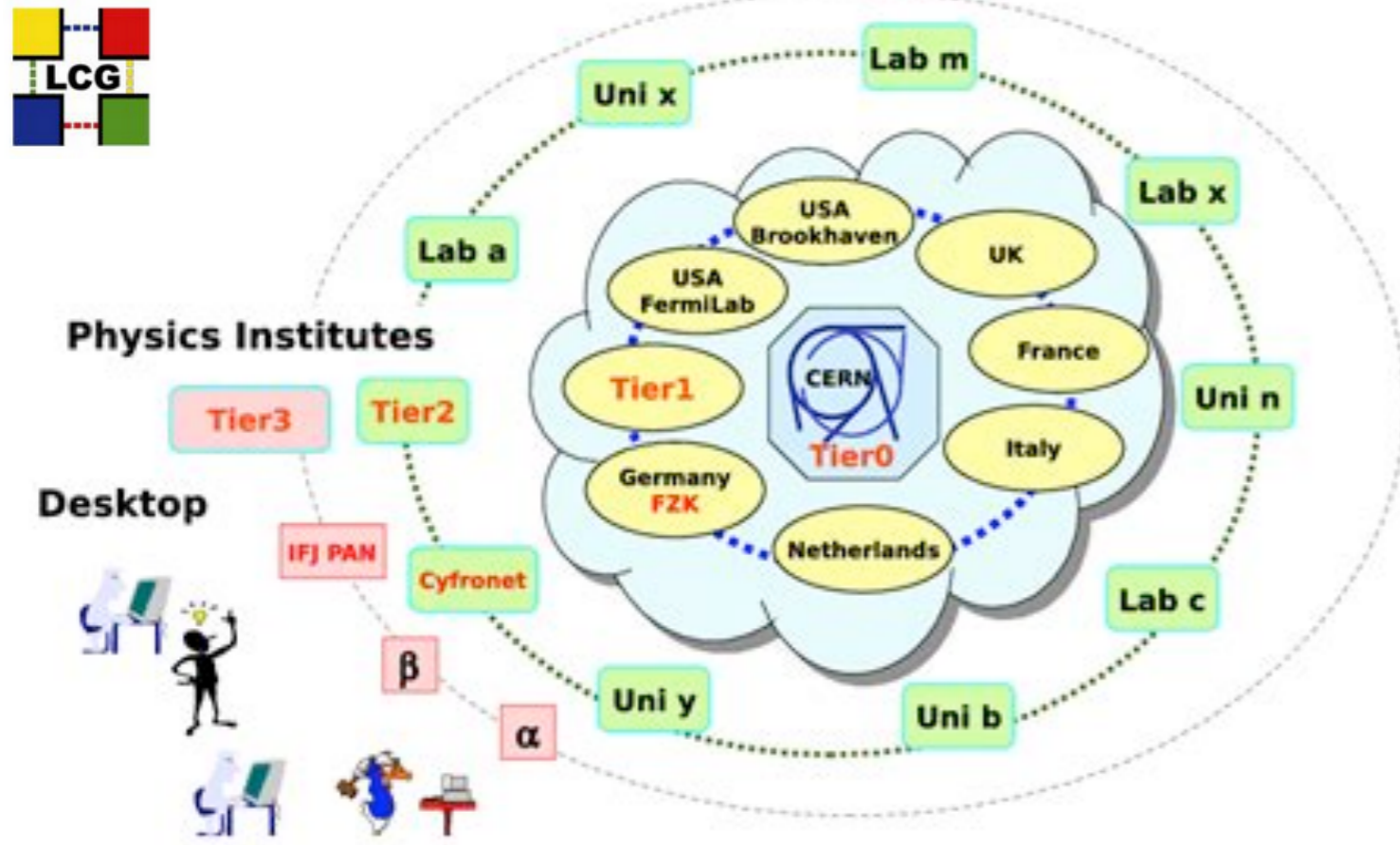
**Processing time** (~100B events)

**Data volume on disk**

# Inference: the reality

Centrally planned, executed

WLCG: cloud of the 2000s



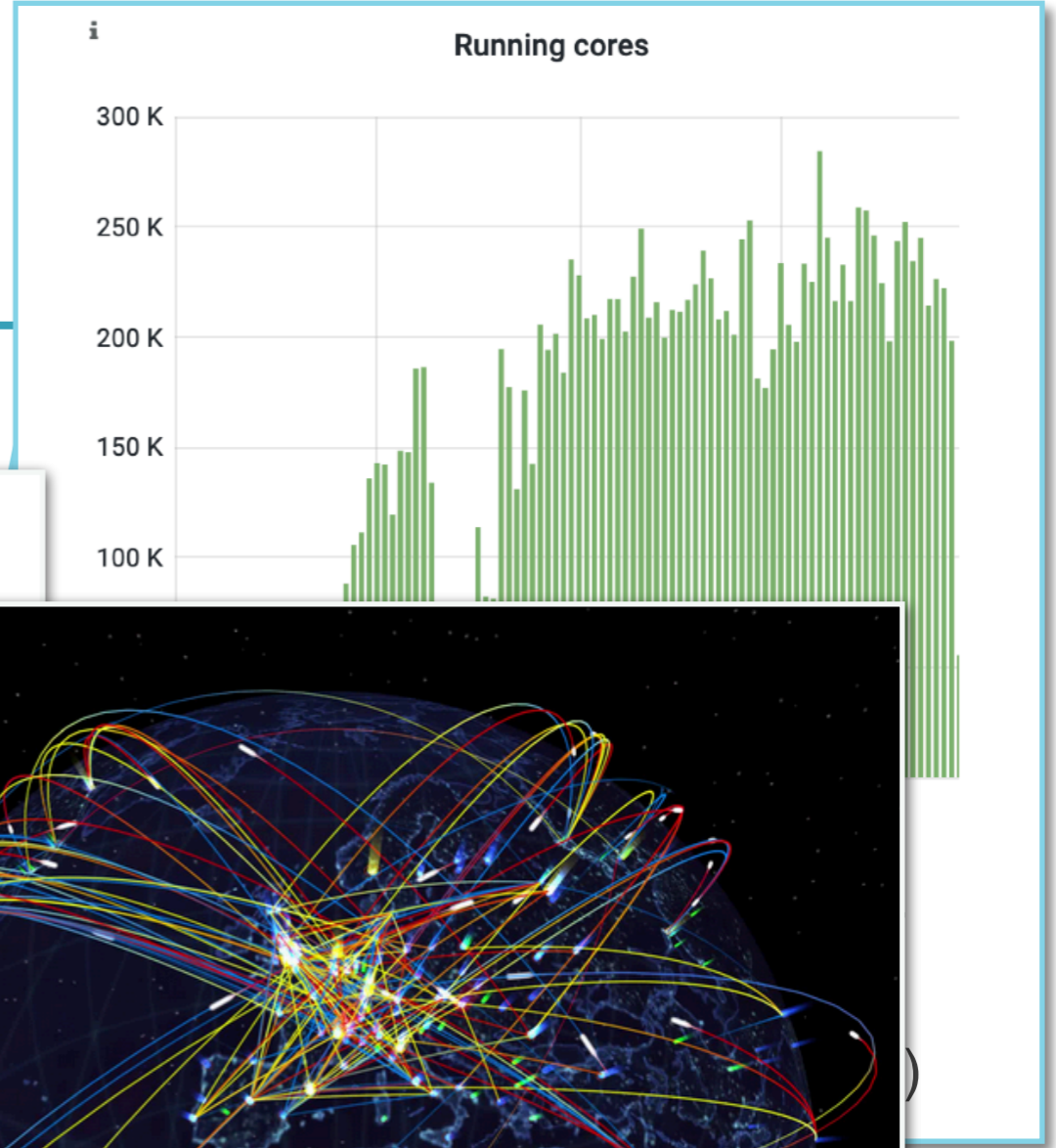
Centrally managed  
10 TB-10 PB

Order of magnitude for CMS Run-II (2016-18)  
**Processing time** (~100B events)  
**Data volume on disk**

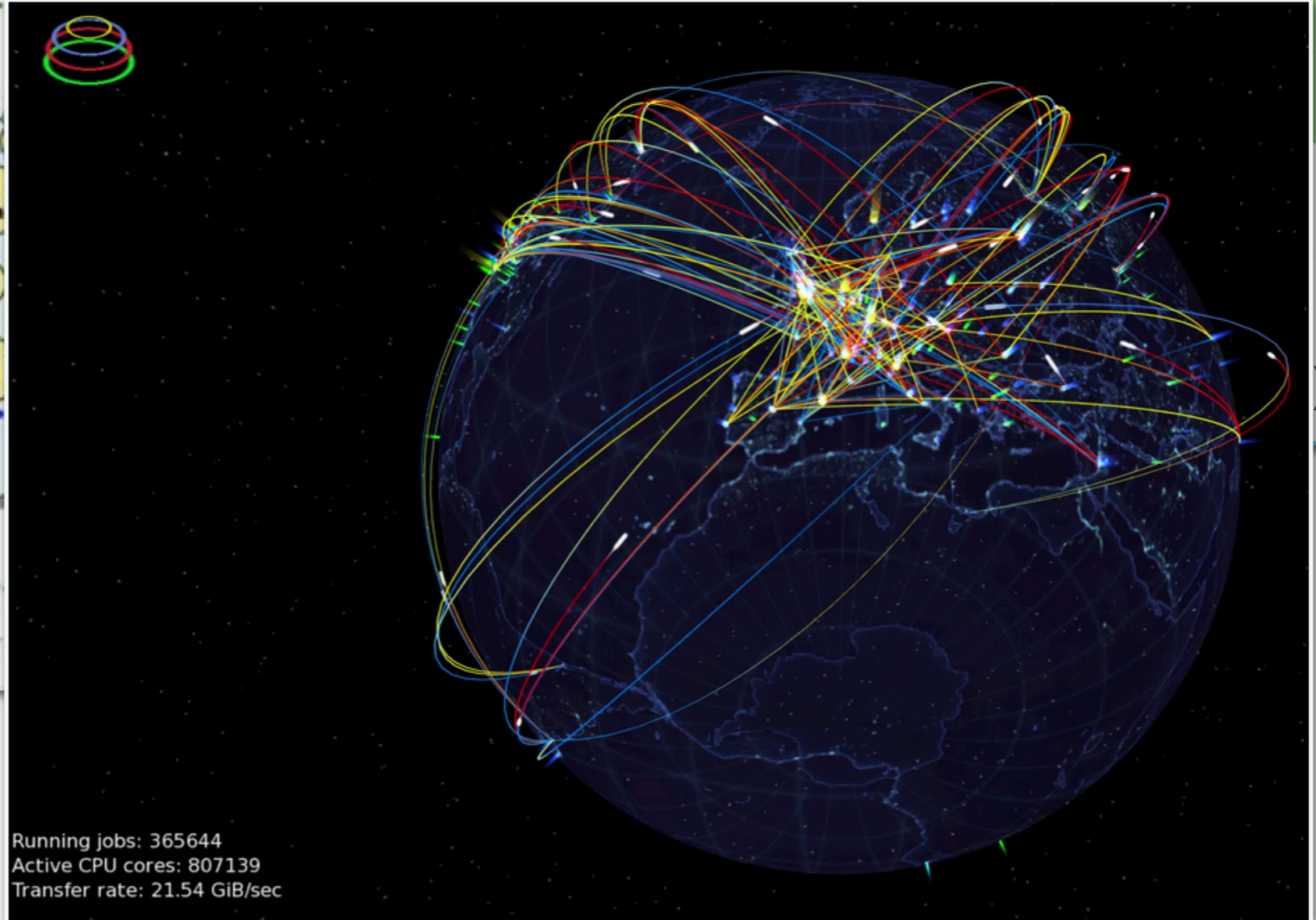
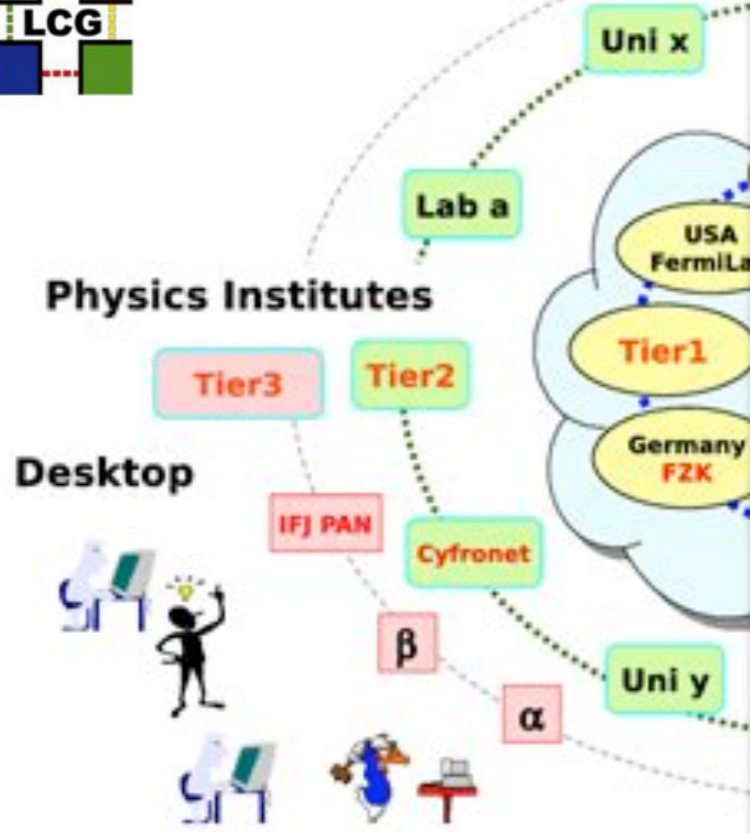


# Inference: the reality

Centrally planned, executed



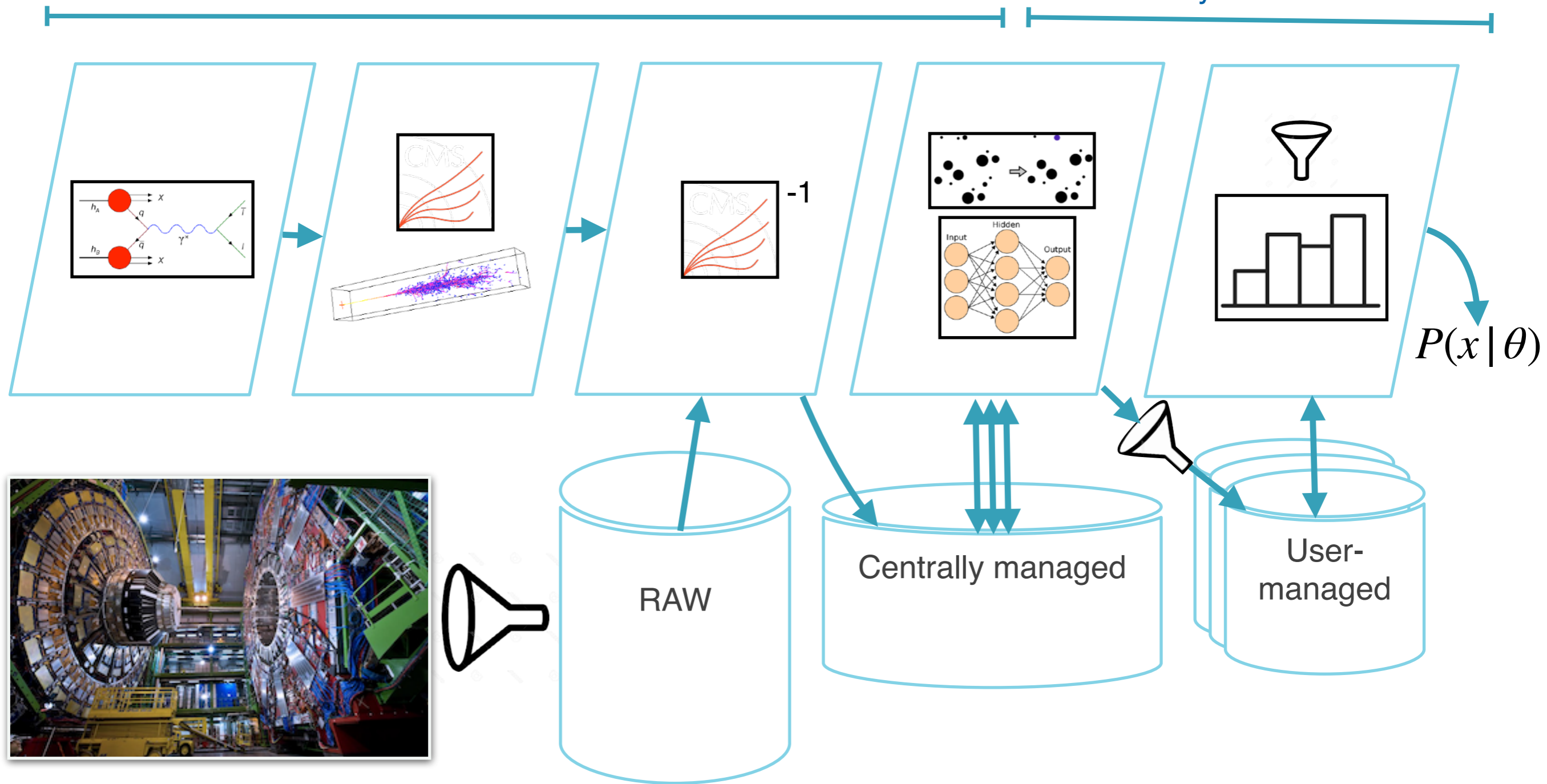
WLCG: cloud of the 2000s



# Inference: the reality

Centrally planned, executed

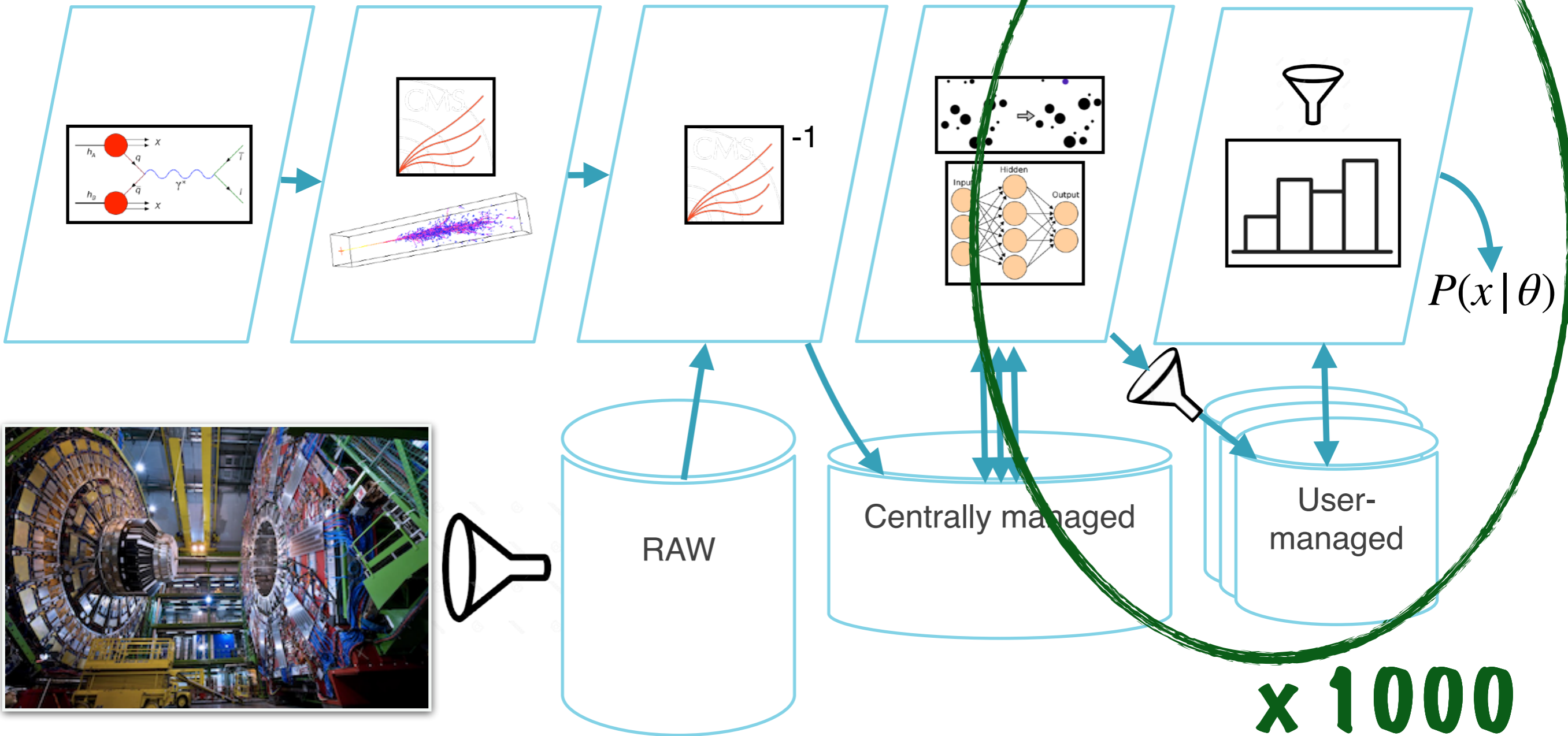
Analyst / Scientist



# Inference: the reality

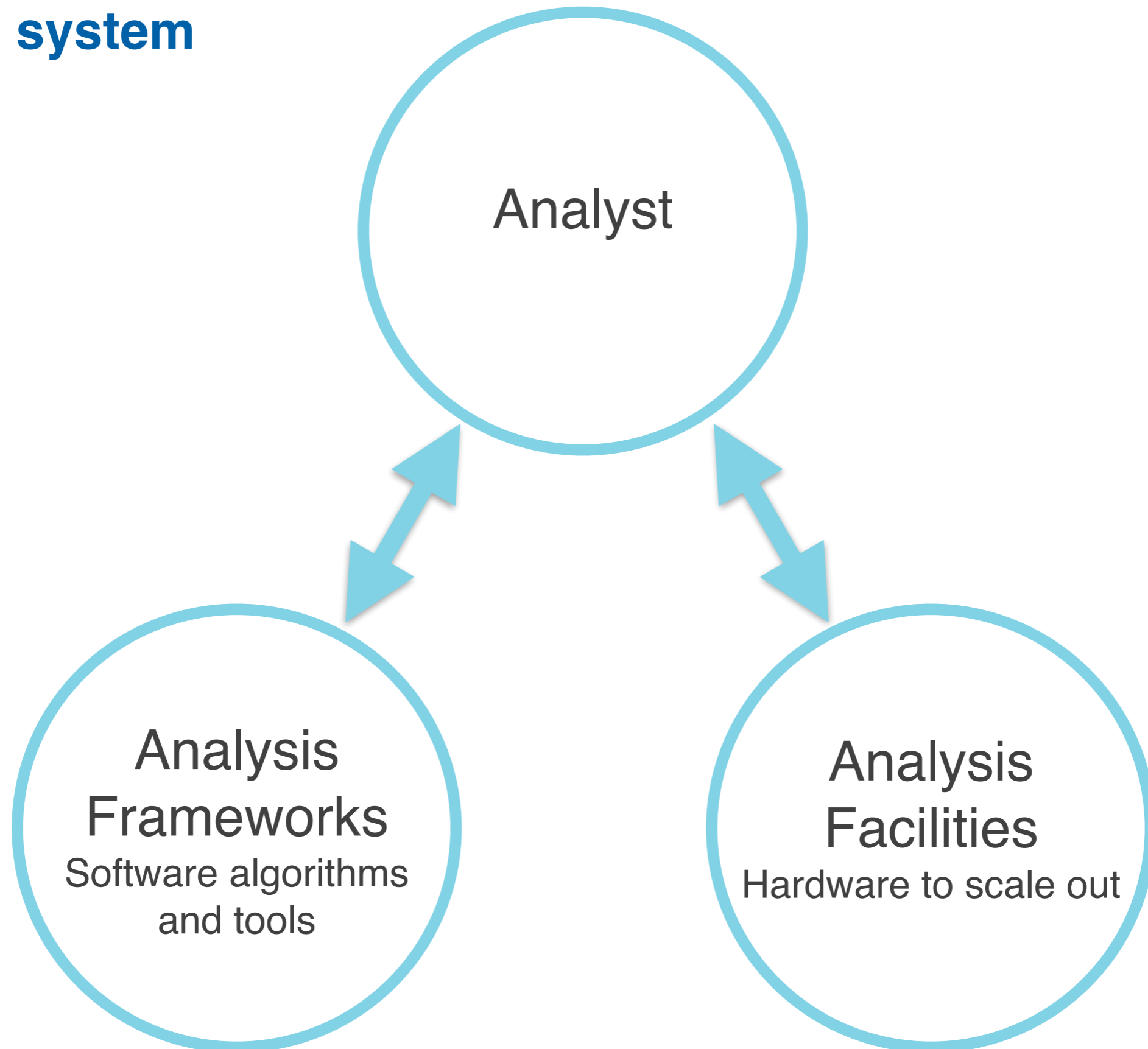
Centrally planned, executed

Analyst / Scientist

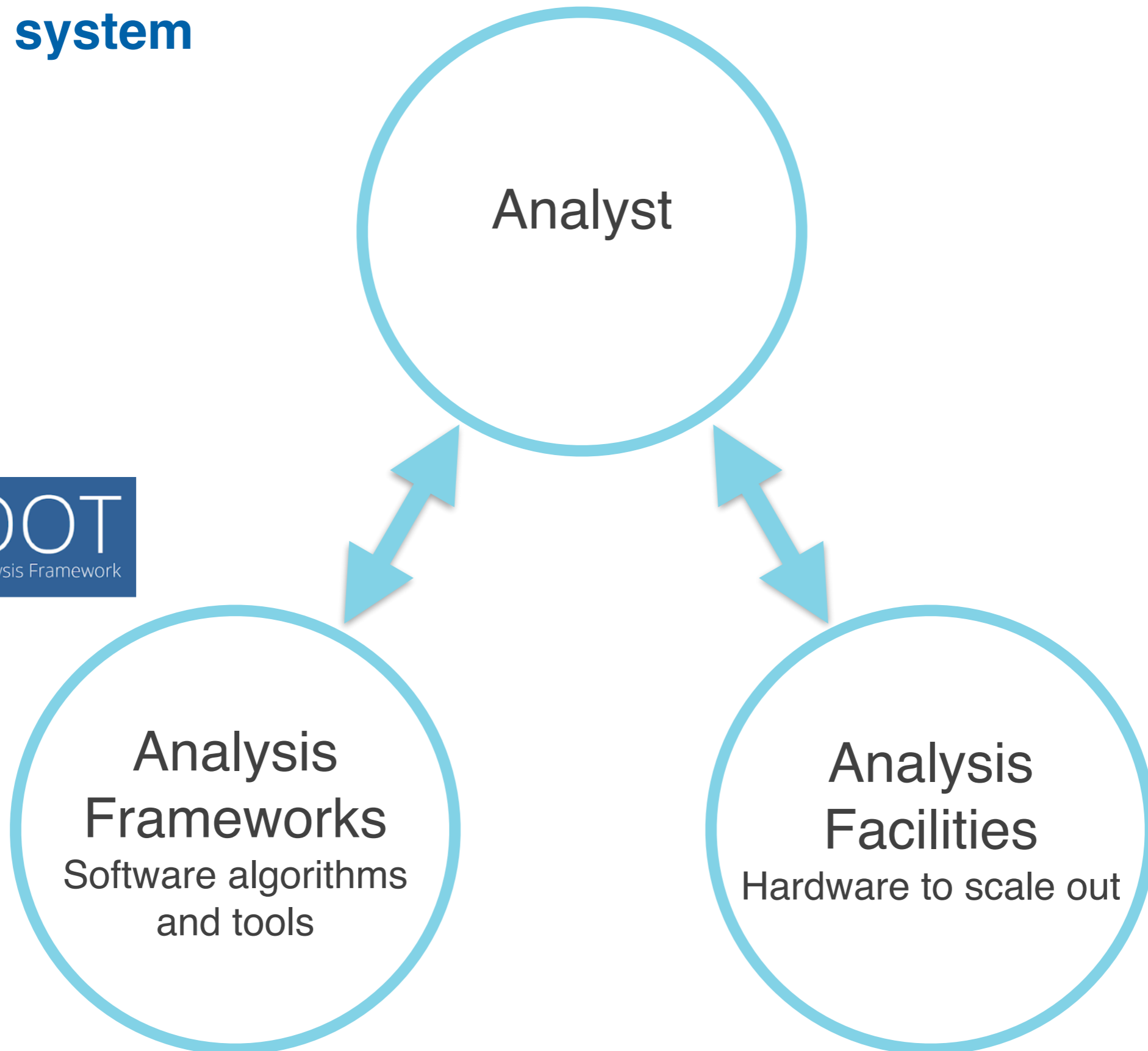




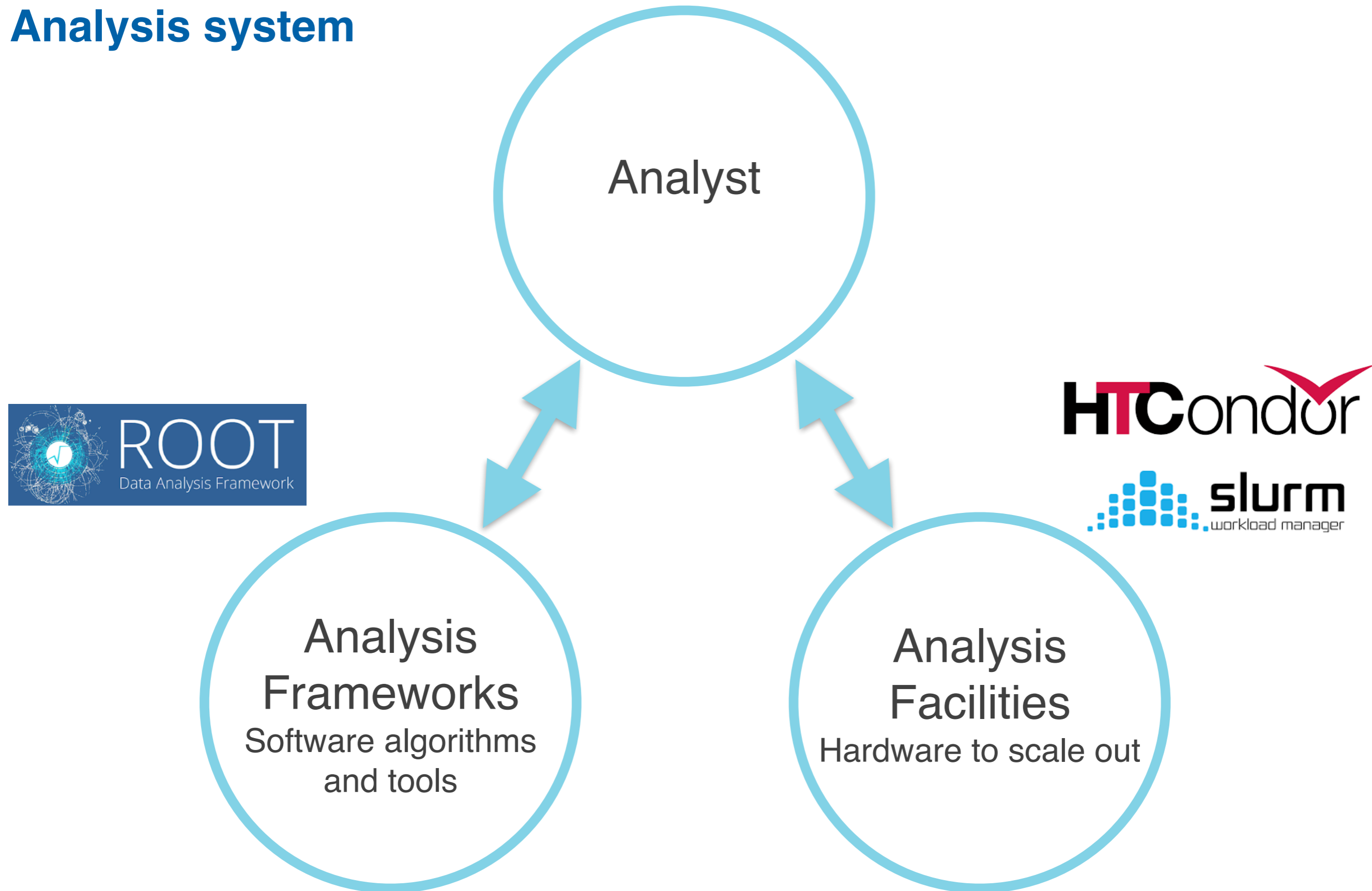
# Analysis system



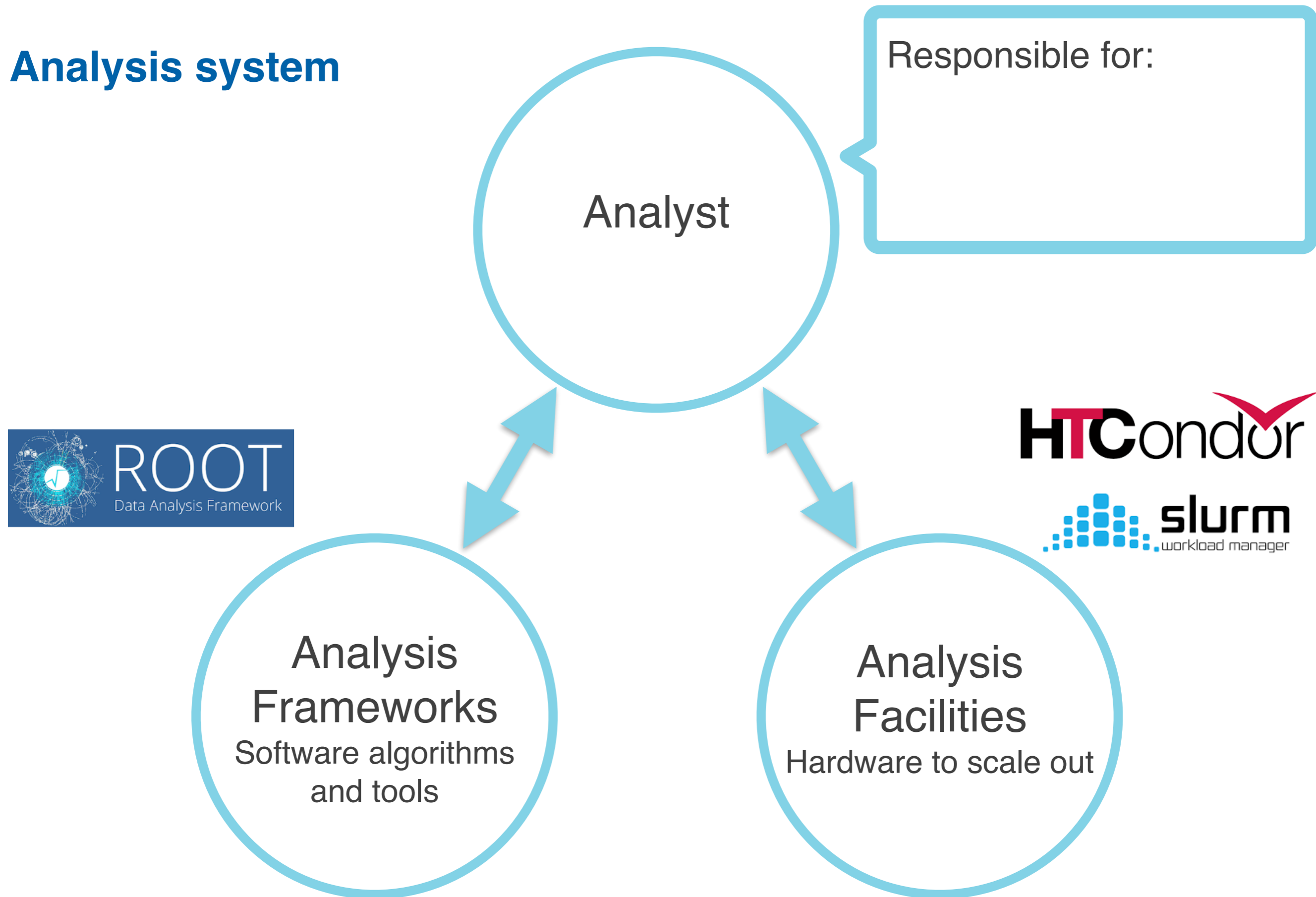
# Analysis system



# Analysis system



# Analysis system

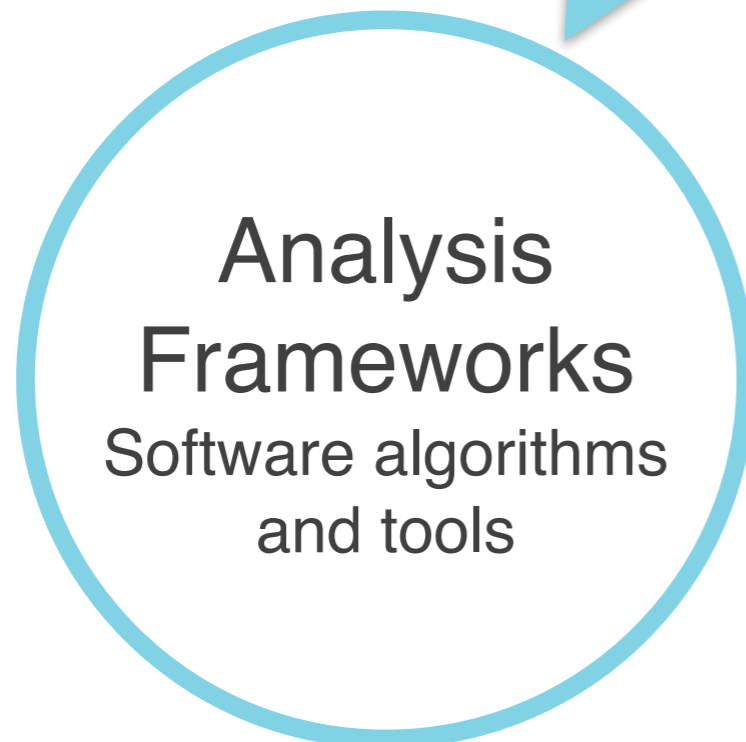


# Analysis system



Responsible for:

- Data access



# Analysis system



Analyst

- Responsible for:
- Data access
  - Query planning



Analysis Frameworks  
Software algorithms and tools

Analysis Facilities  
Hardware to scale out

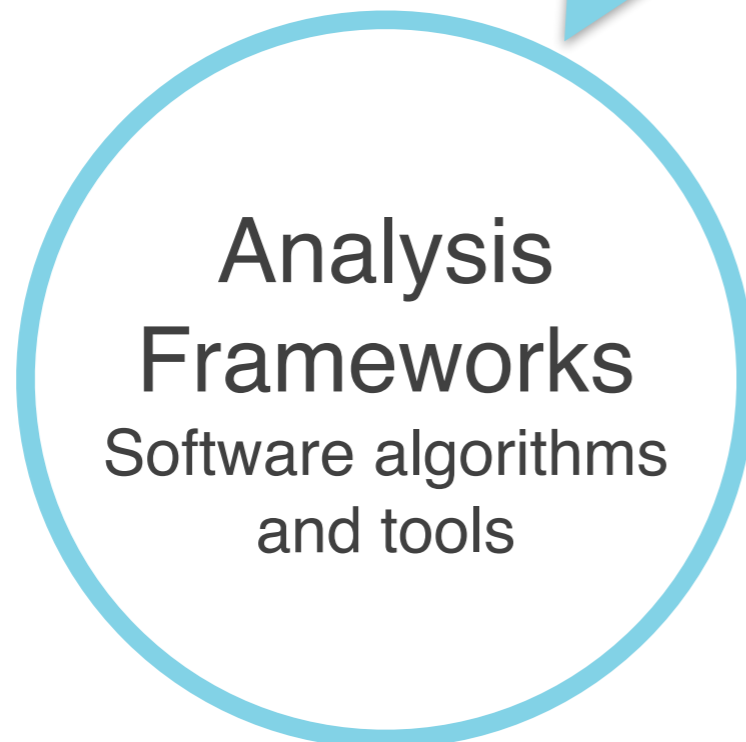


# Analysis system



Responsible for:

- Data access
- Query planning
- Workflow scale-out



# Analysis system



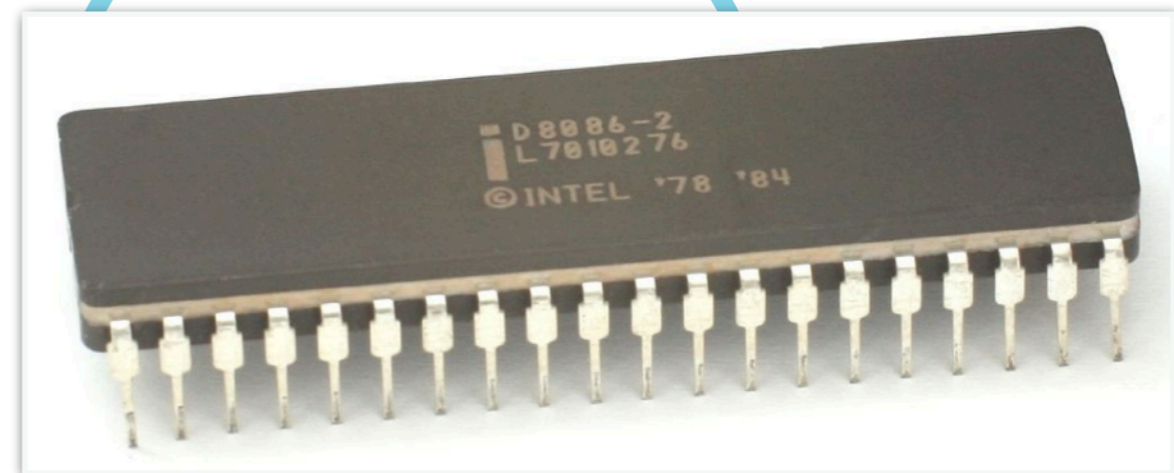
Analyst

Responsible for:

- Data access
- Query planning
- Workflow scale-out

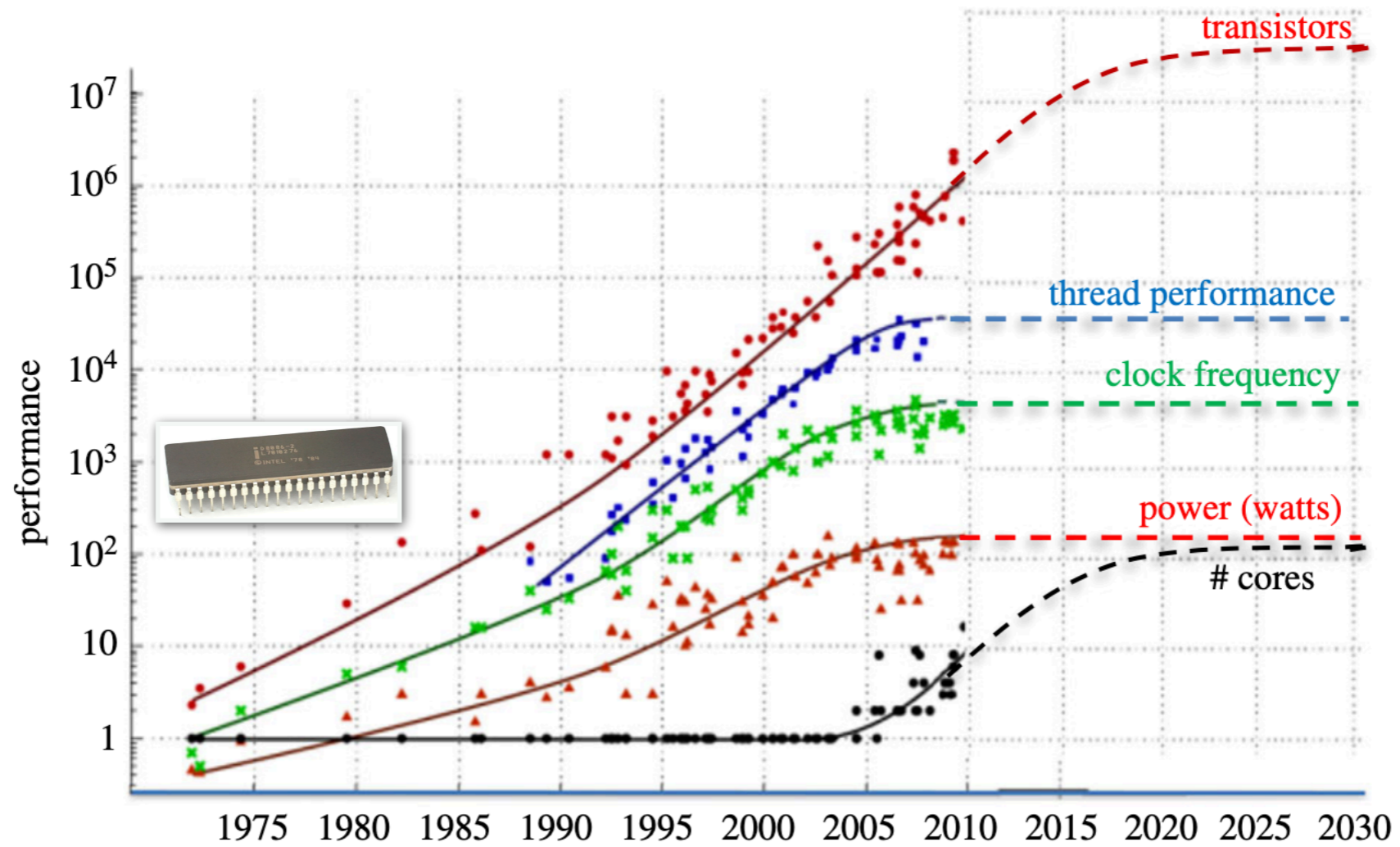


Analysis  
Frameworks  
Software algorithms  
and tools

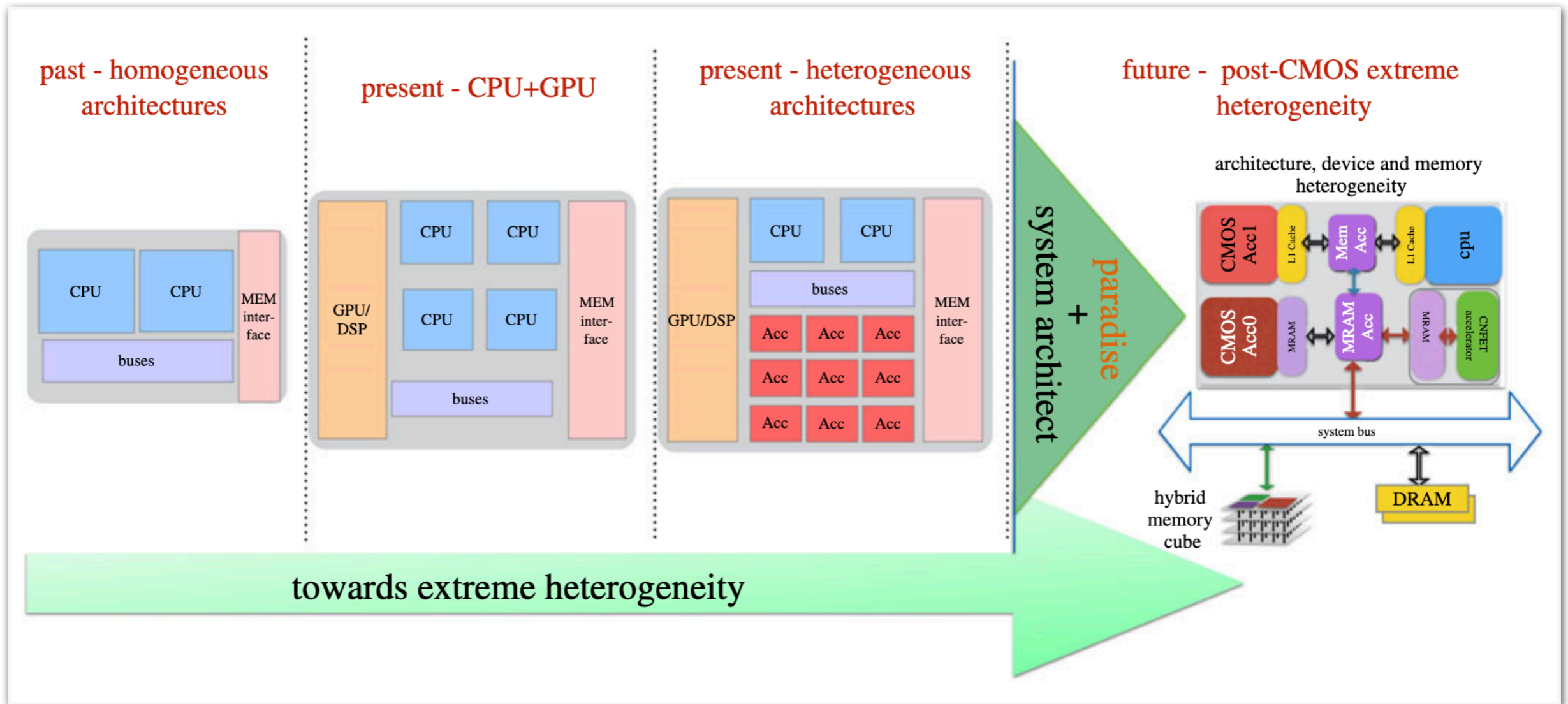




# Aside: RIP x86?



# Aside: RIP x86?



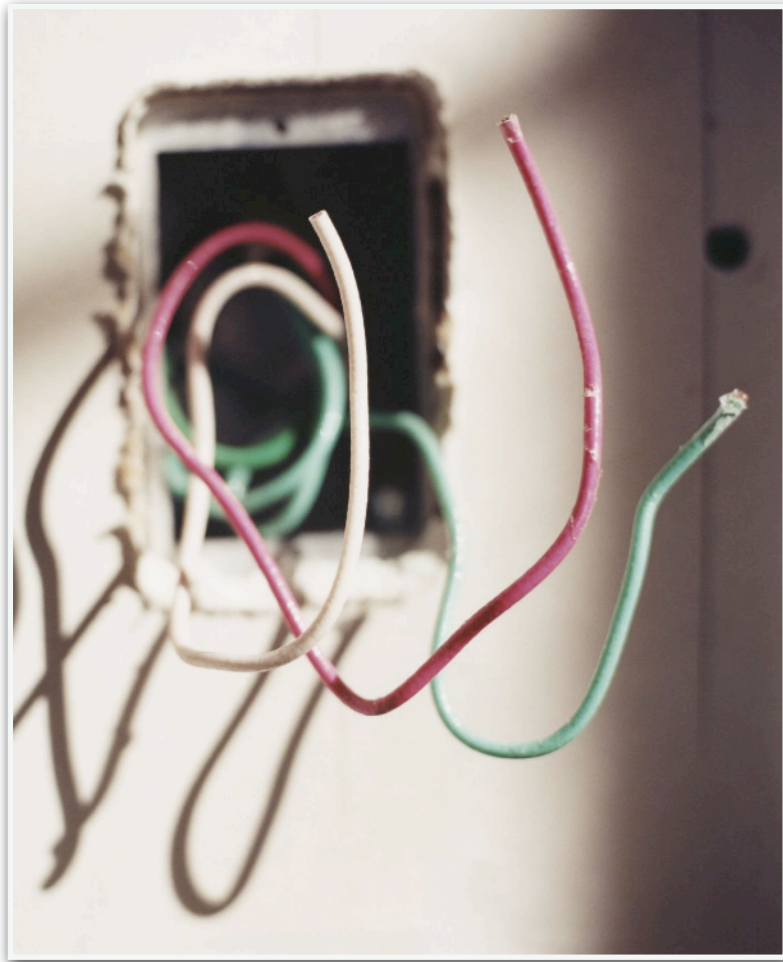
# Interfaces



# Interfaces

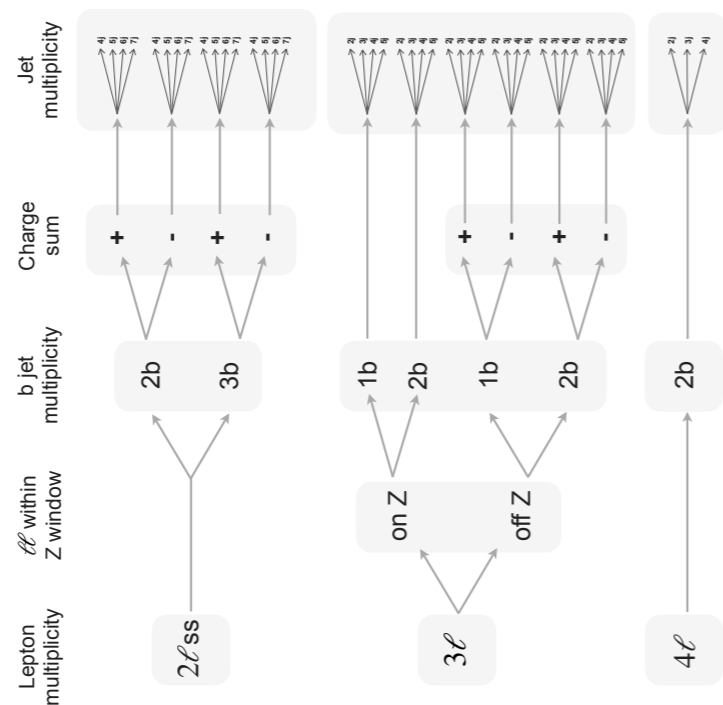


# Interfaces



# Scaling challenges

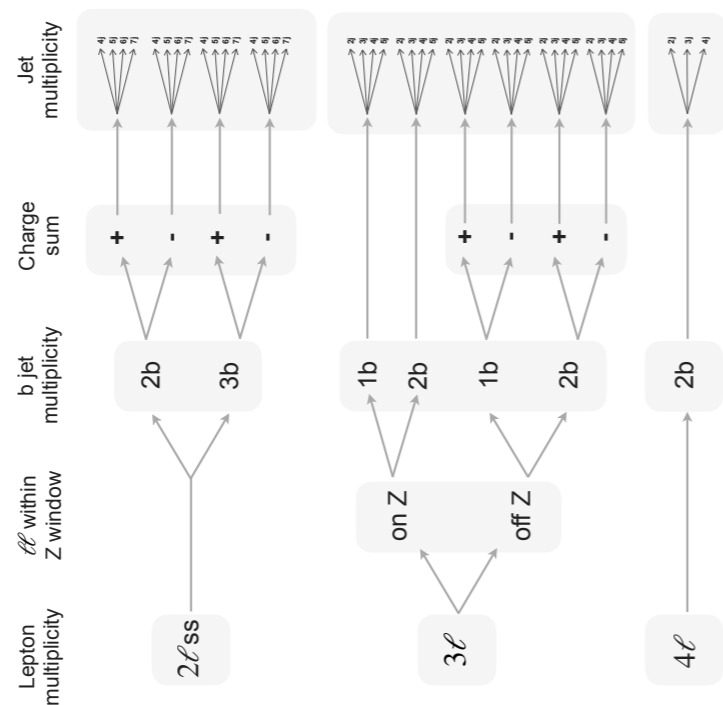
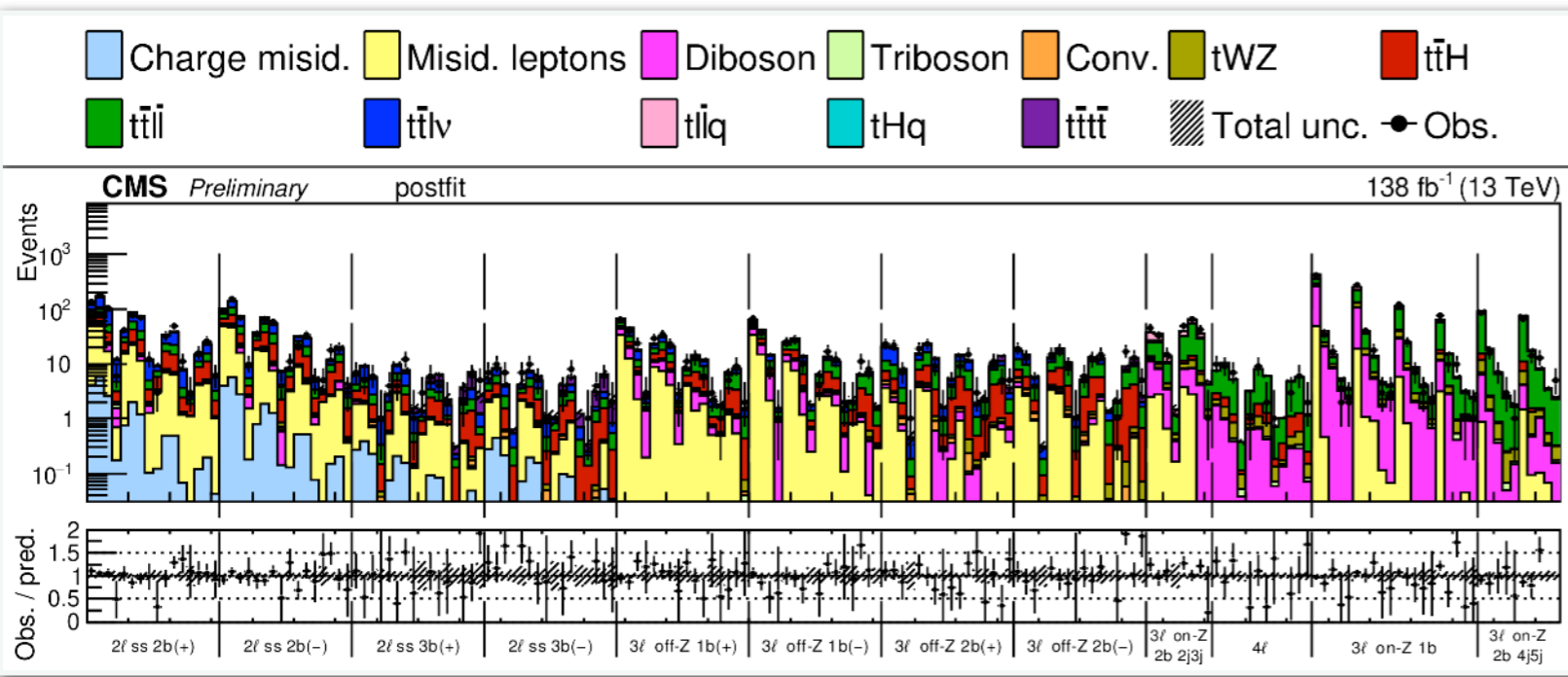
- EFT search: many observables to constrain high-dim. parameter space



[TOP-22-006](#)

# Scaling challenges

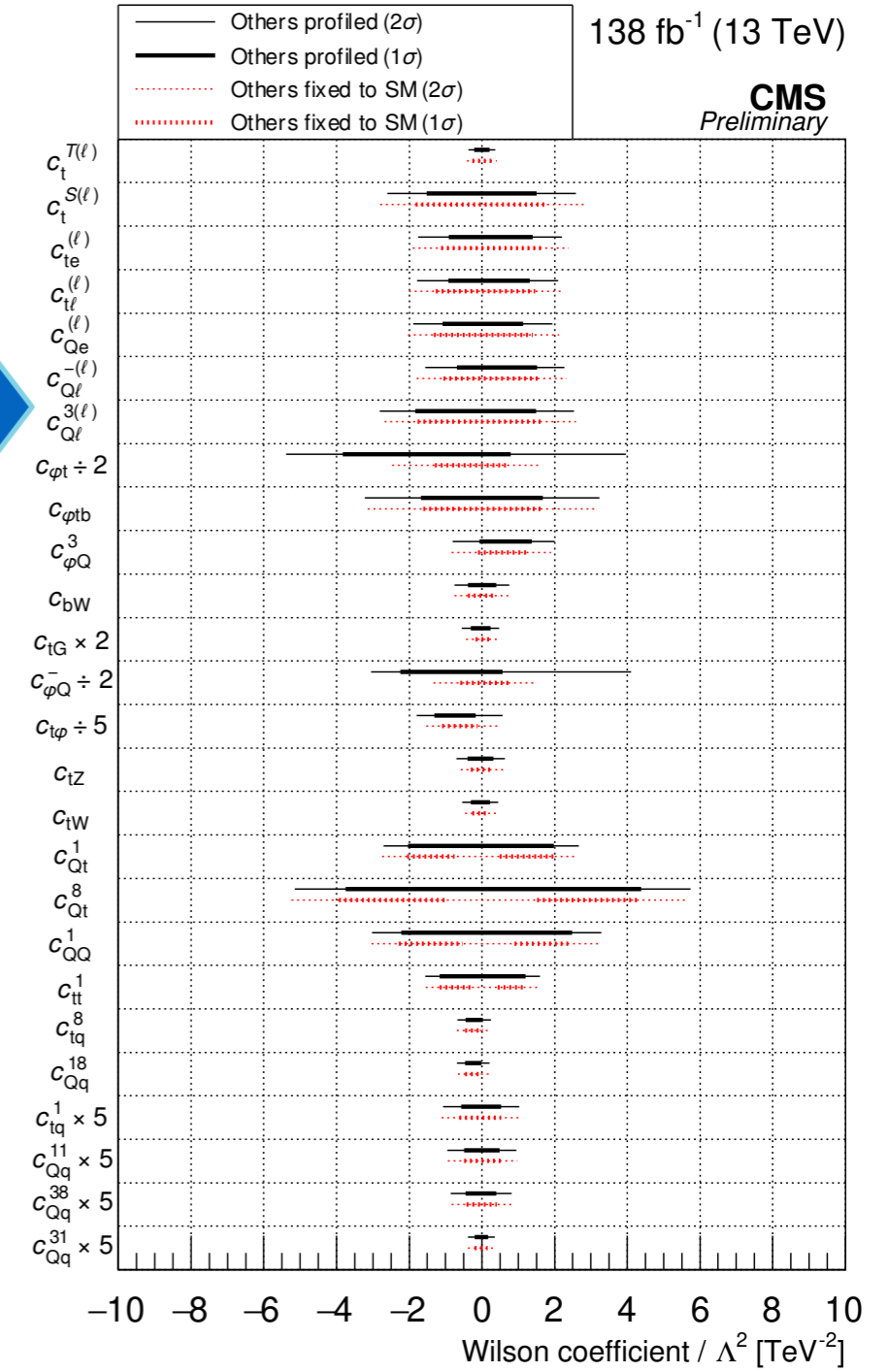
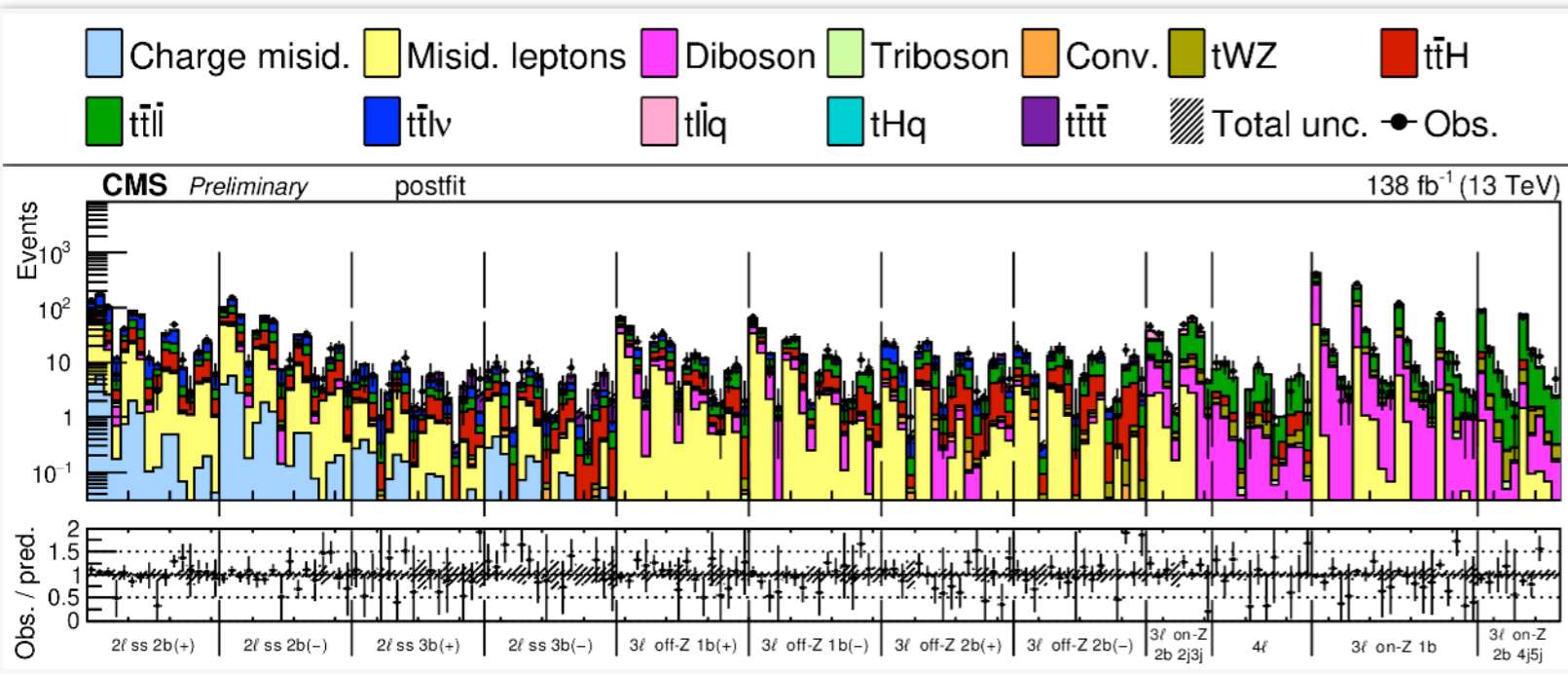
- EFT search: many observables to constrain high-dim. parameter space



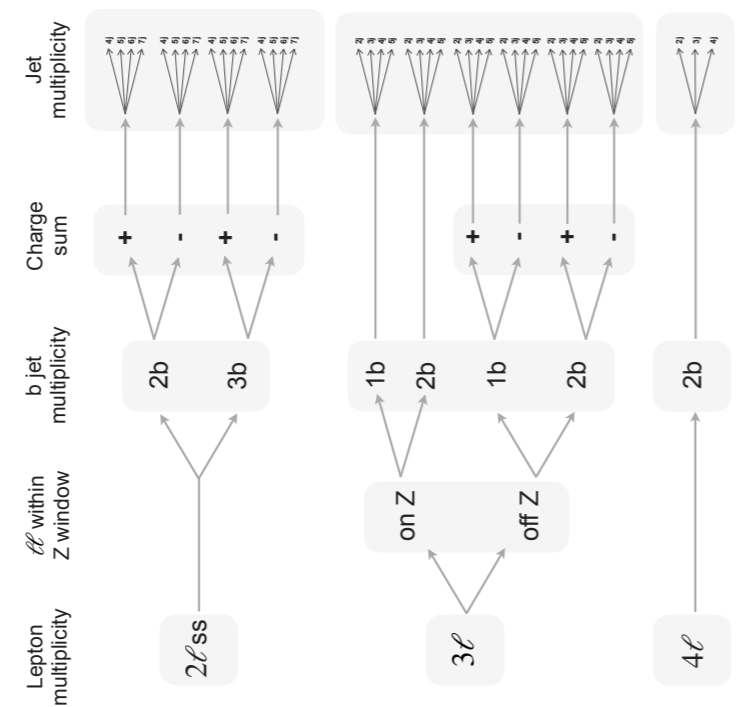
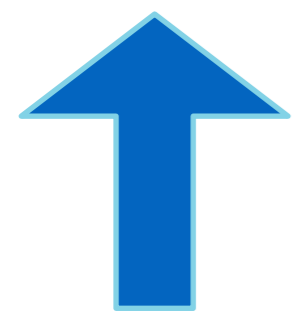
[TOP-22-006](#)

# Scaling challenges

- EFT search: many observables to constrain high-dim. parameter space



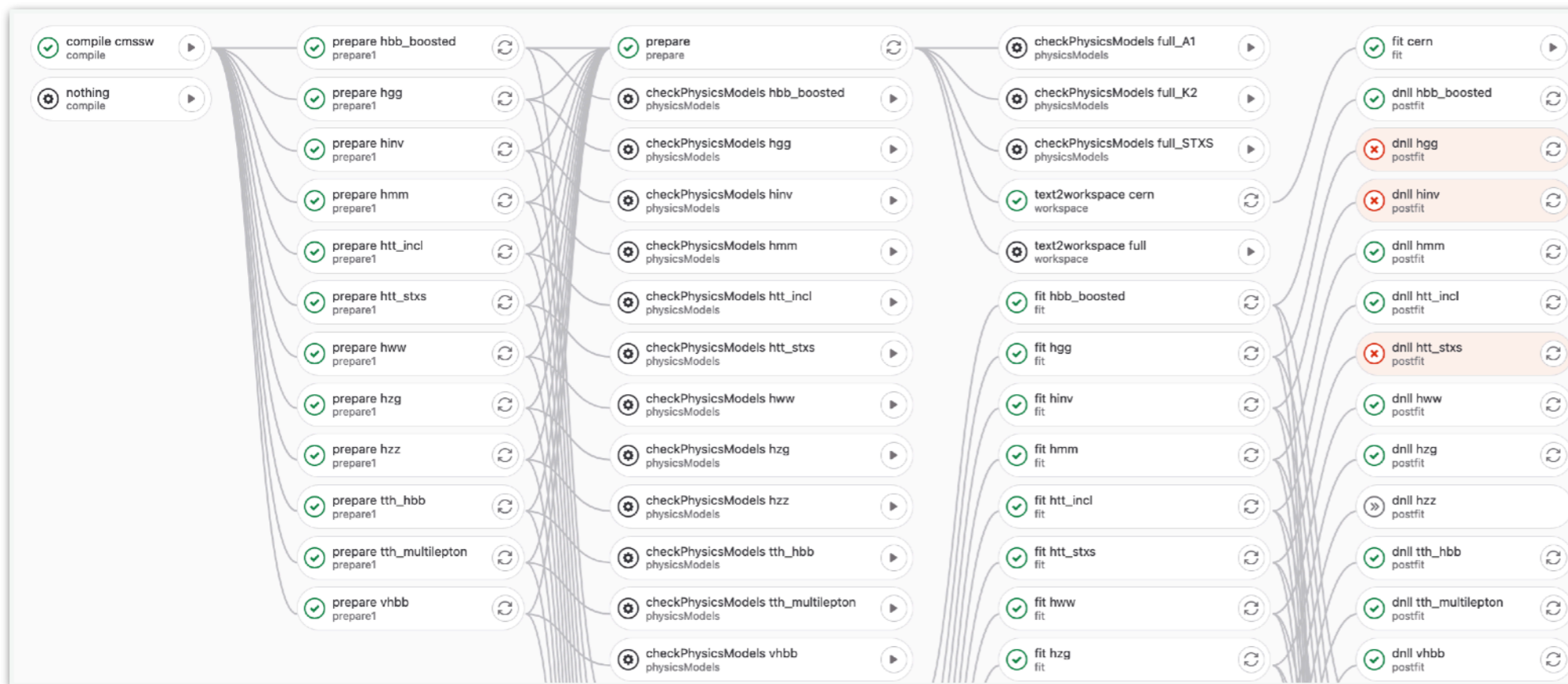
TOP-22-006





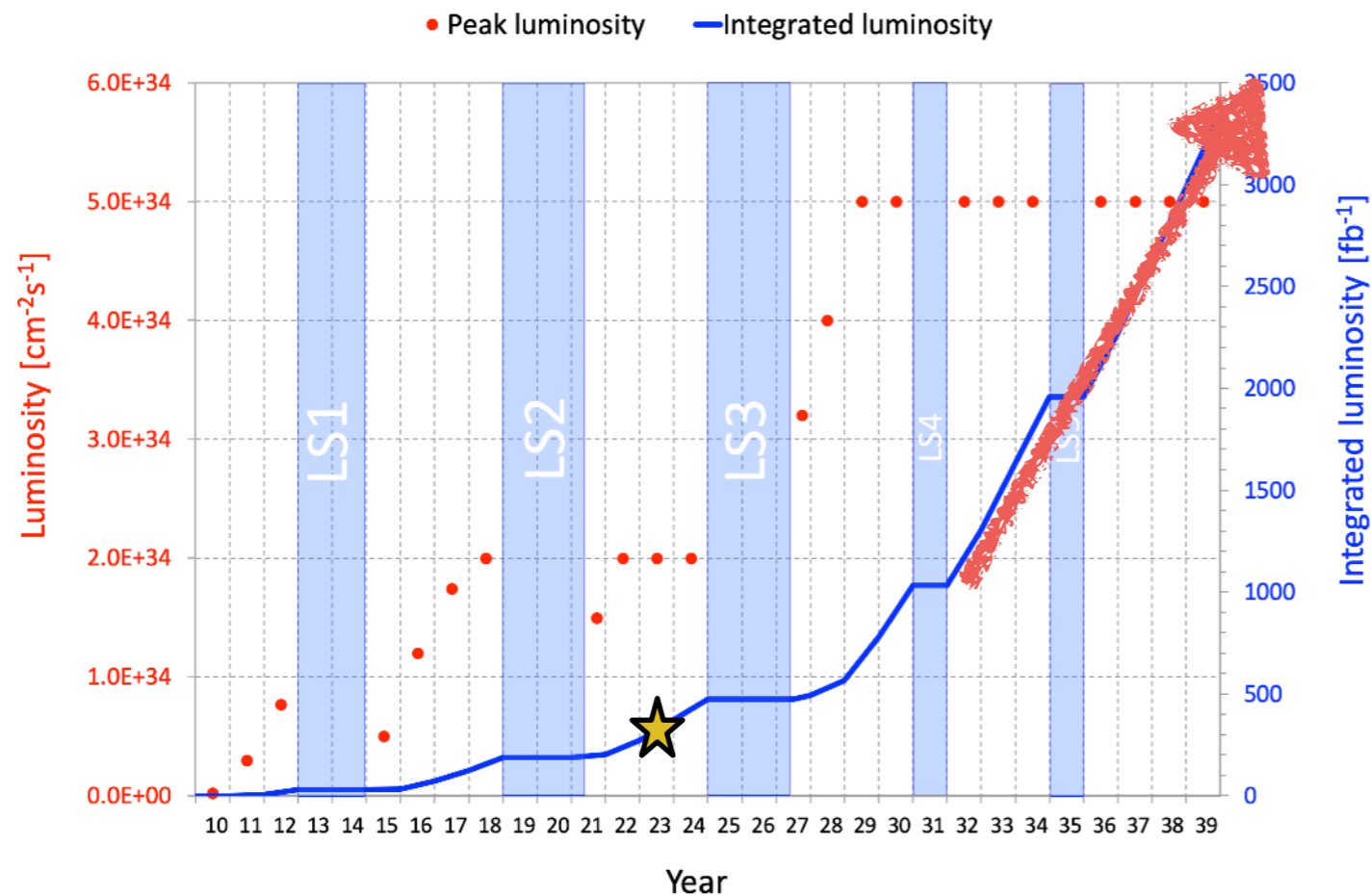
# Scaling challenges

- Higgs combination: bringing a dozen complex analysis descriptions together
  - Maximum likelihood fit: 10k parameters, 30h minimizations
  - Workflow management tools become a requirement



# Looking forward

- High-Luminosity LHC: 20x current dataset
  - One view:  $\sqrt{20} \approx 4.5x$  statistical precision
  - But also, 20x observables at same precision: constrain more dimensions!

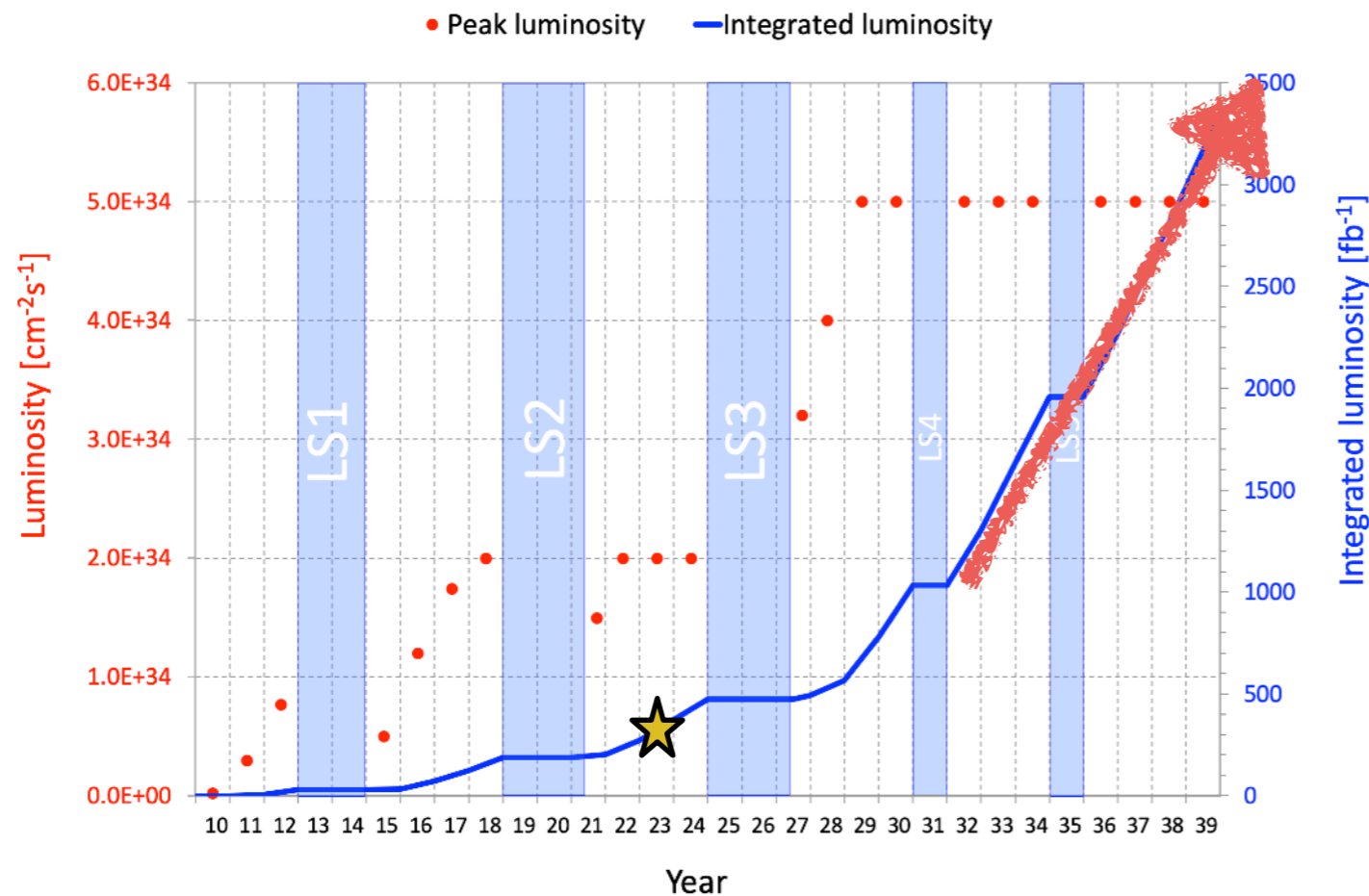


HL-LHC (pre-covid) nominal schedule

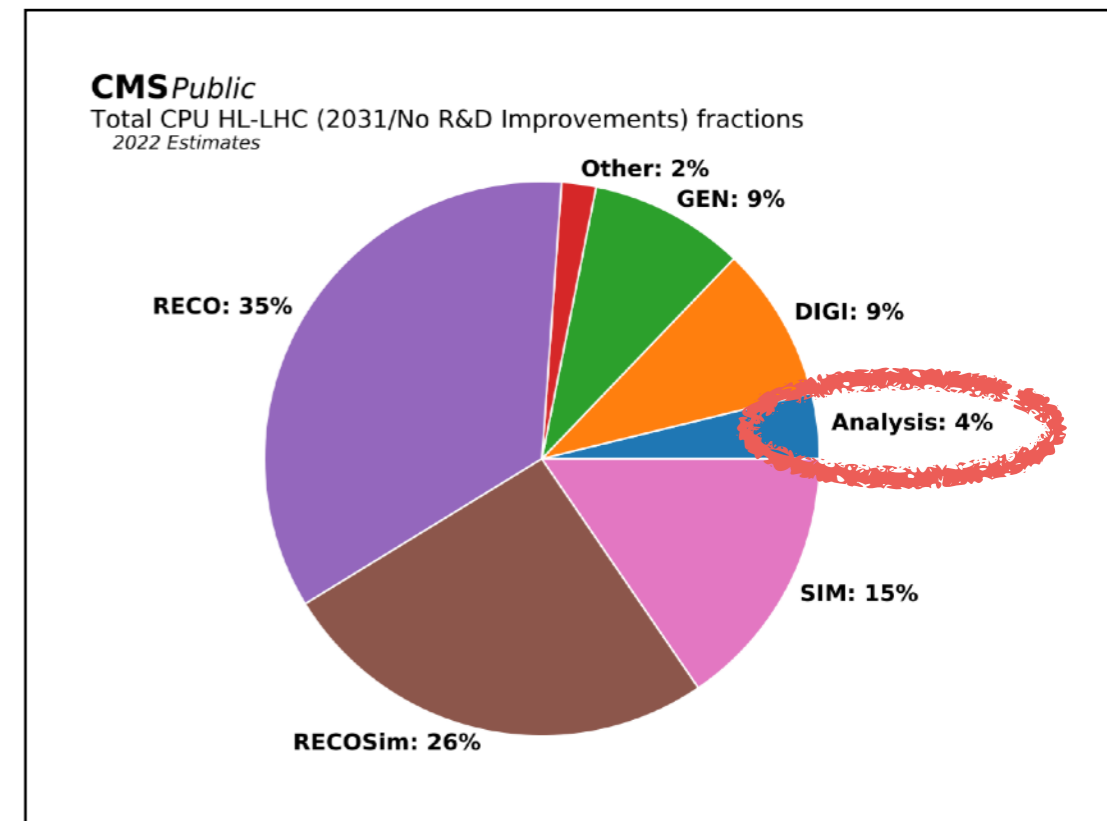


# Looking forward

- High-Luminosity LHC: 20x current dataset
  - One view:  $\sqrt{20} \approx 4.5x$  statistical precision
  - But also, 20x observables at same precision: constrain more dimensions!
- Analysis not directly driving compute capacity projections
  - Lever arm



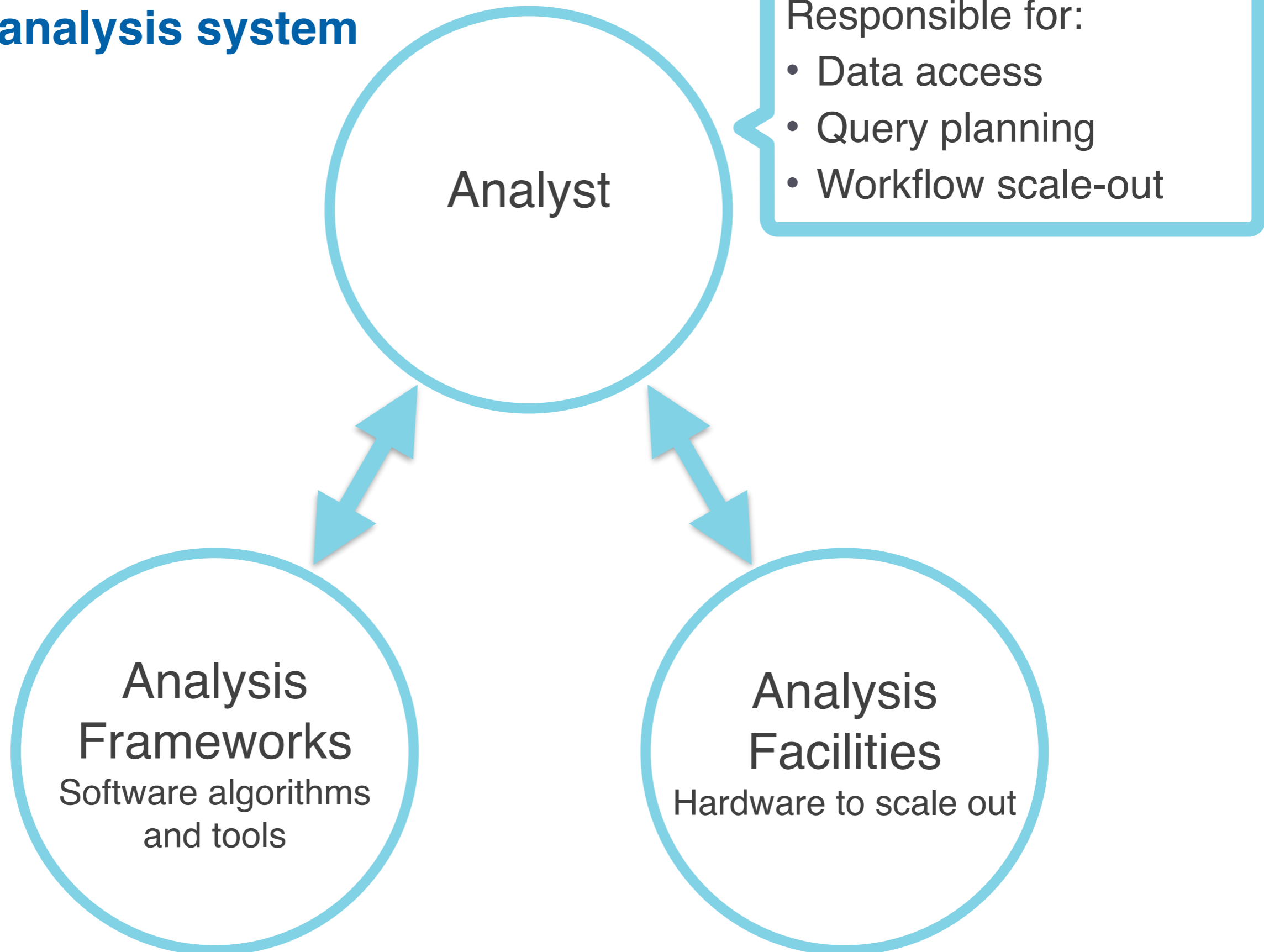
HL-LHC (pre-covid) nominal schedule



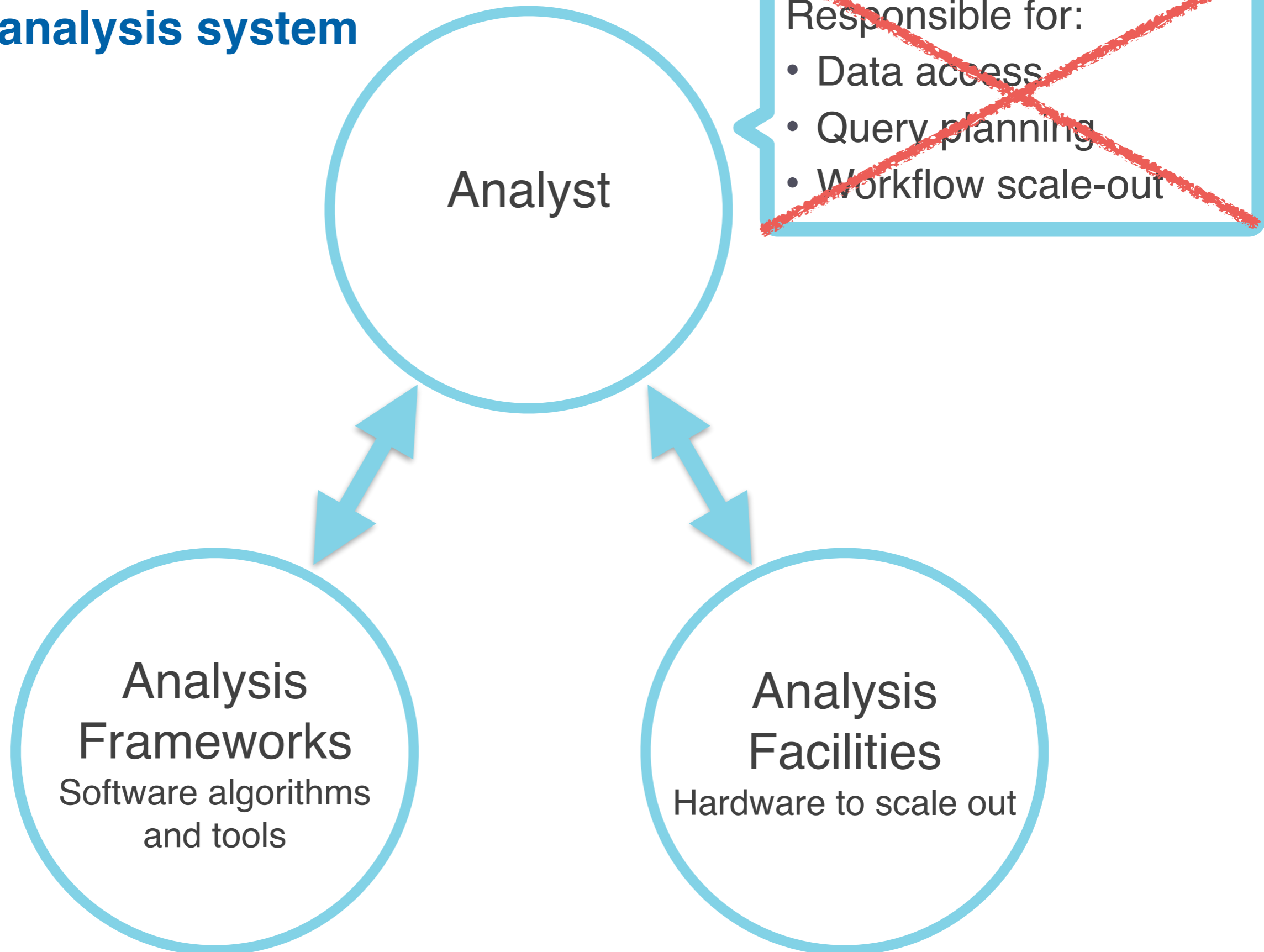
CMS Computing Projections



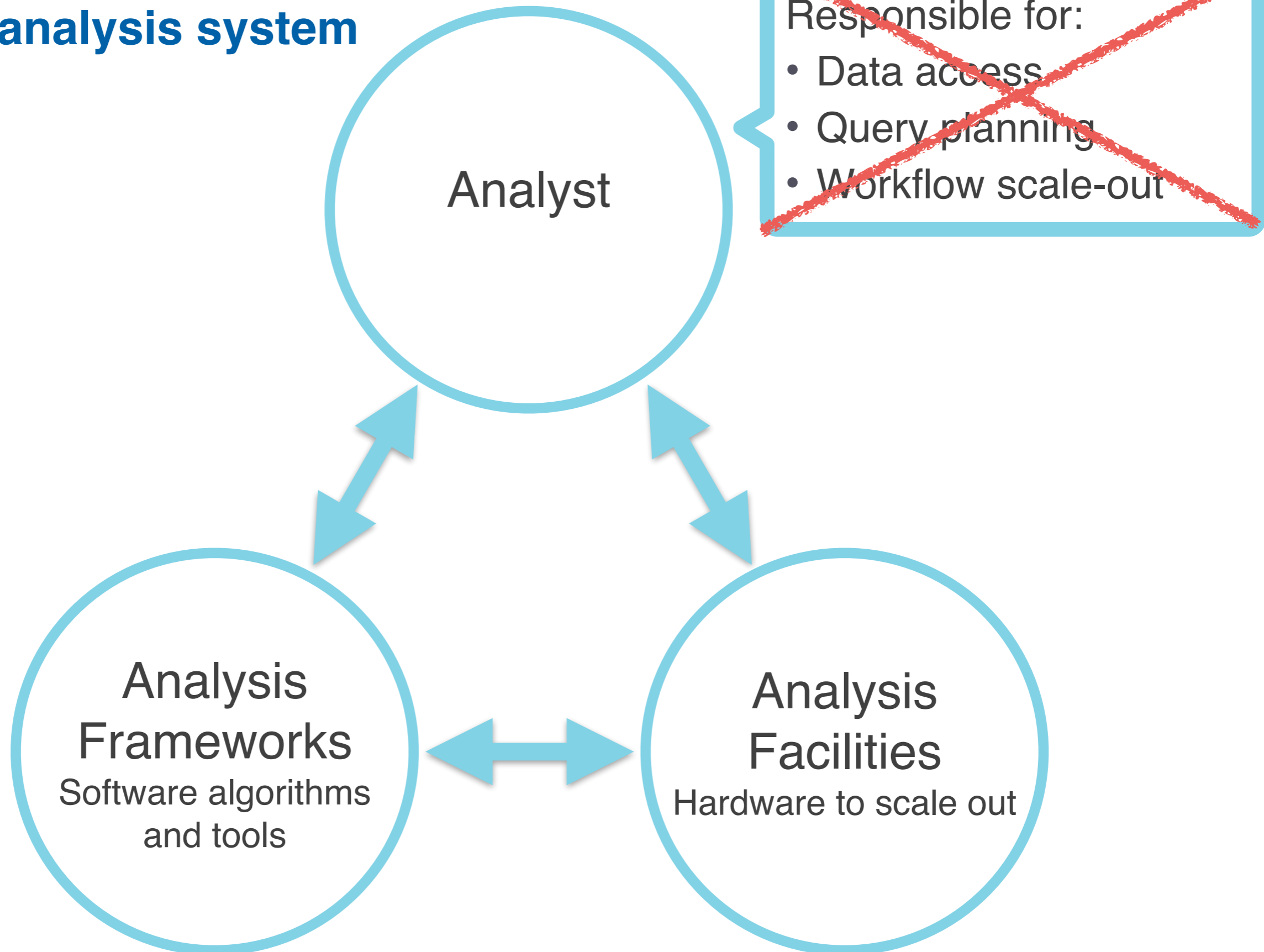
# A better analysis system



# A better analysis system

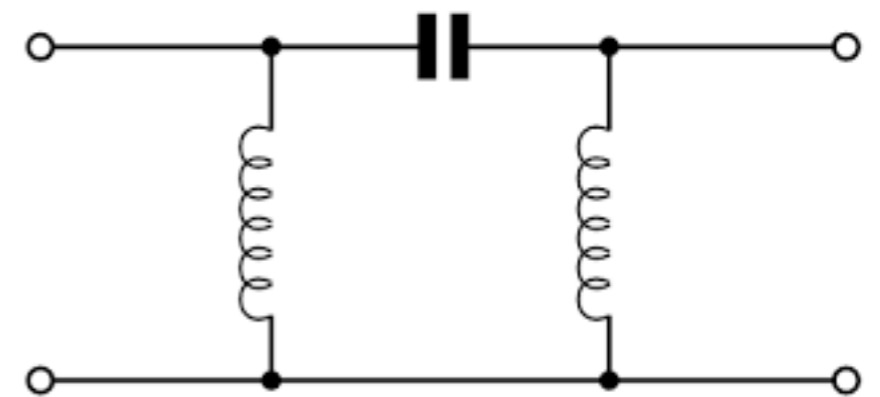


# A better analysis system



# Impedance Mismatches

- ROOT File Format ↔ Machine Learning
- Big data ↔ Python
- HEP Physicist ↔ Industry



# Big Data

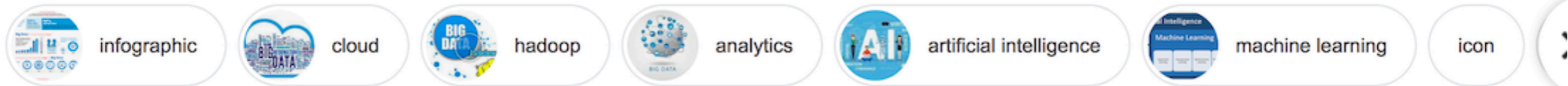


big data



All News Images Videos Books More Settings Tools  
Size Color Usage rights Type Time Clear

Collections SafeSearch



1024 × 682

10 Parameters for Big Data Assessment ...  
analyticsinsight.net



800 × 505

What is Big Data? Let's answer this ...  
towardsdatascience.com



1024 × 614

Big Data Analytics ...  
smartdatacollective.com



1280 × 720

What is Big Data? | Big Data Definition ...  
edureka.co



1366 × 768

DNS Infrastructure - Big Data Connector ...  
akamai.com



1838 × 1034

Data Analytics Overtakes Big Data ...  
flextrade.com



847 × 480

Importance of Big Data Analytics ...  
learntek.org



1800 × 1200

interesting ideas that harness big data ...  
bbvaopen4u.com





# Big Data

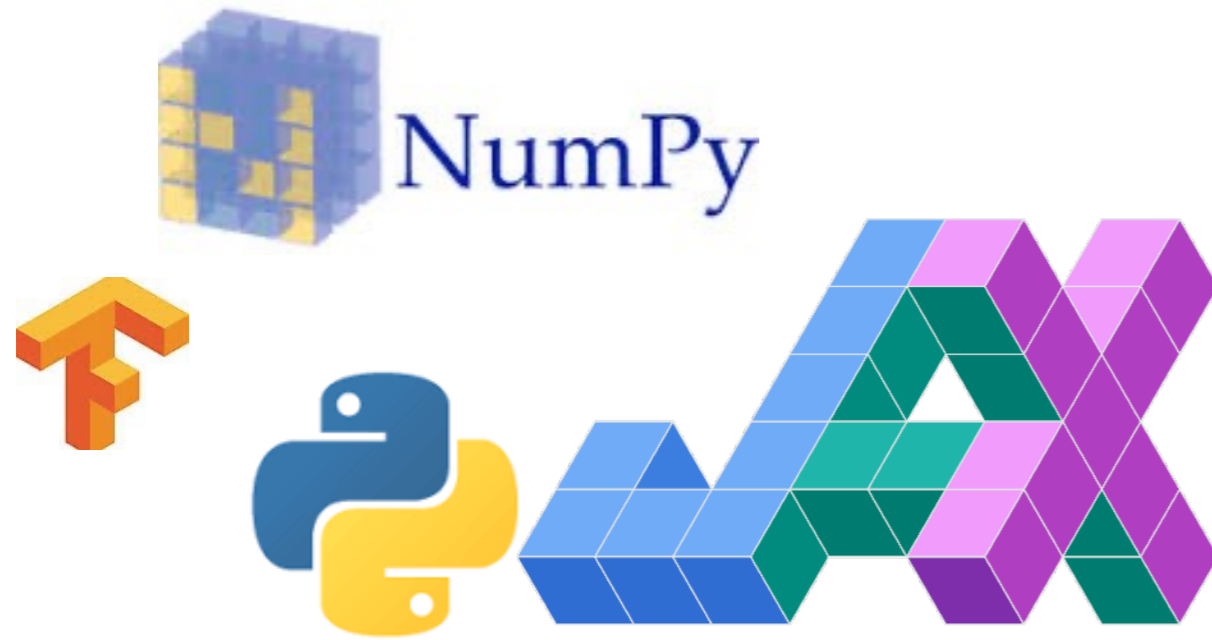


**ML / Quant / Science**  
**Array programming**



**Business Analytics**  
**SQL-like**

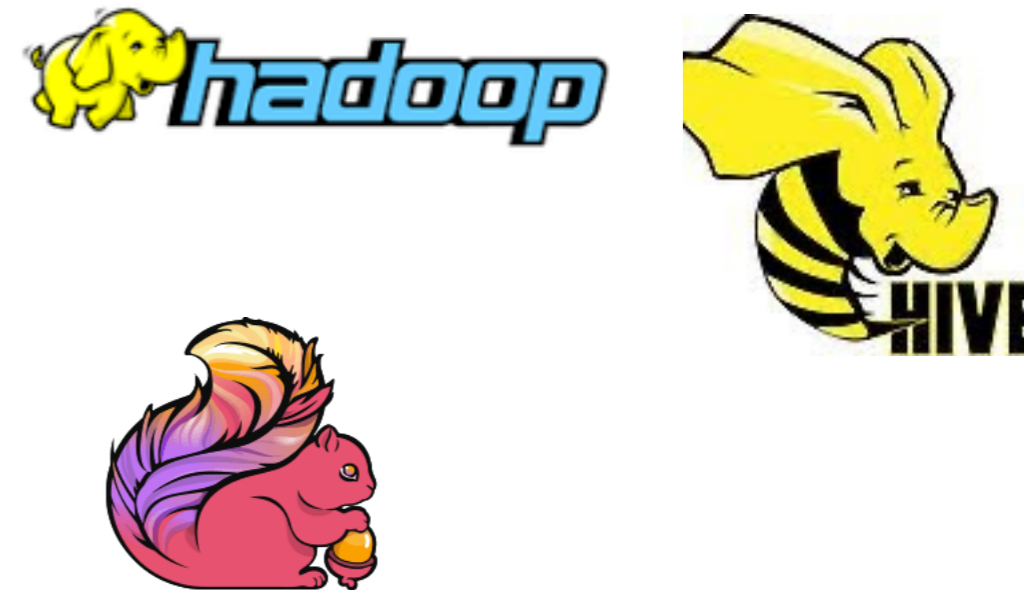
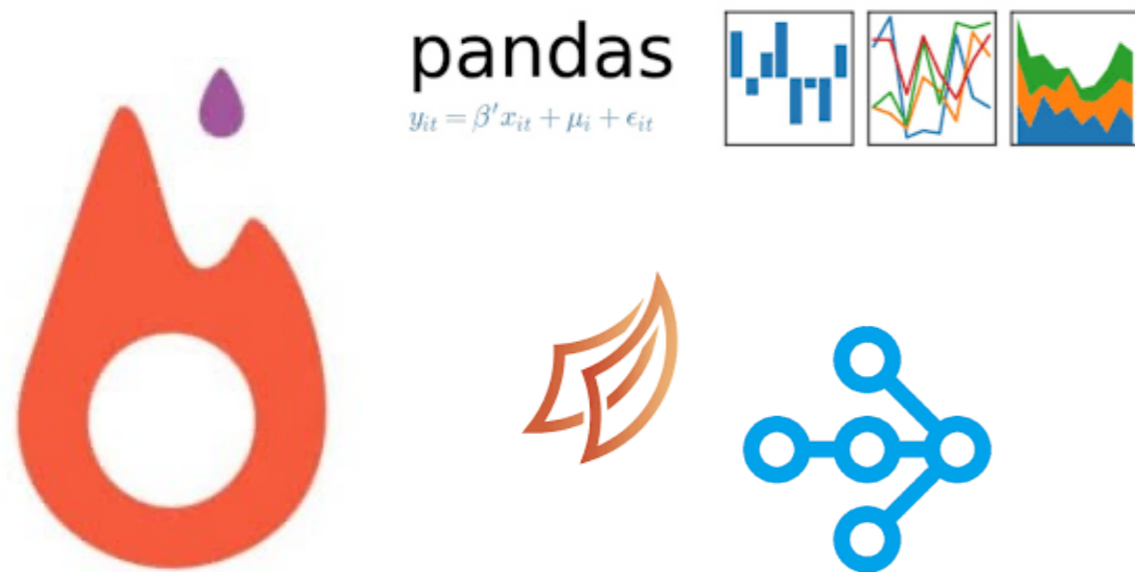
# Big Data



**ML / Quant / Science**  
**Array programming**



**Business Analytics**  
**SQL-like**



# Big Data



**ML / Quant / Science**  
**Array programming**

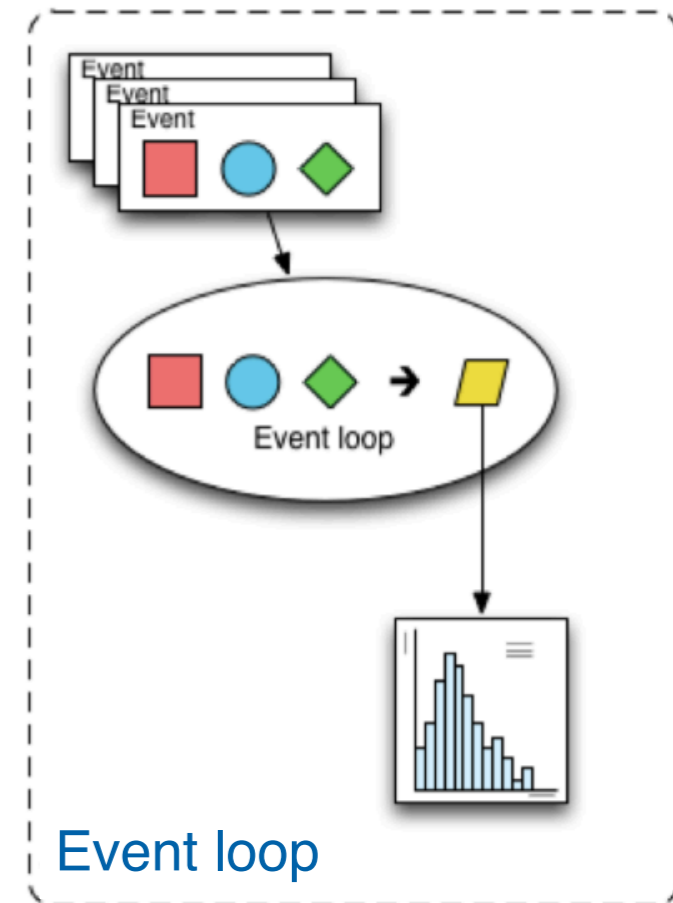


**Business Analytics**  
**SQL-like**



# The paradigm shift

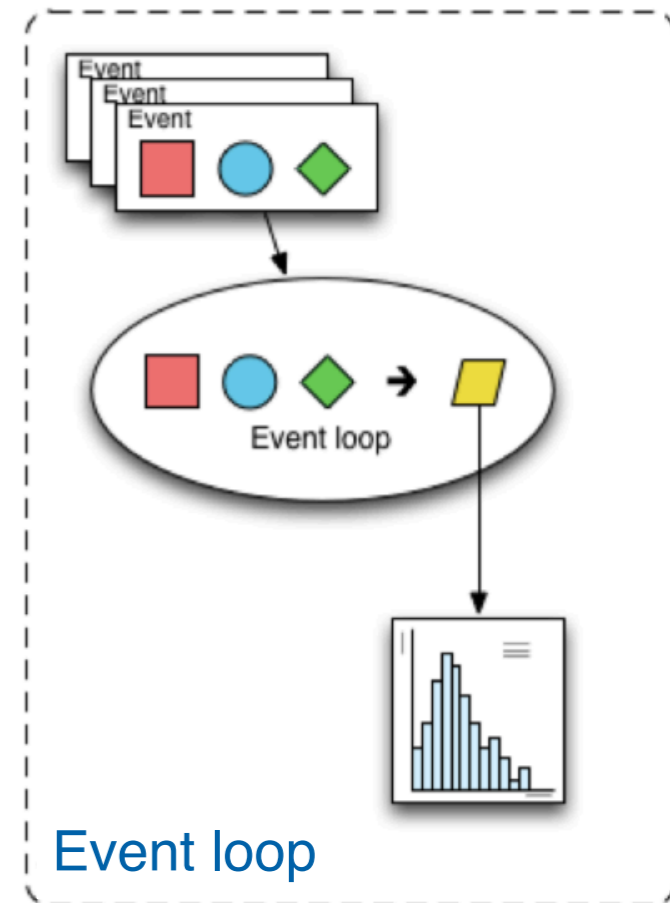
- Event loop analysis:
  - Load relevant values for a specific event into local variables
  - Evaluate several expressions
  - Store derived values
  - Repeat (explicit outer loop)



# The paradigm shift

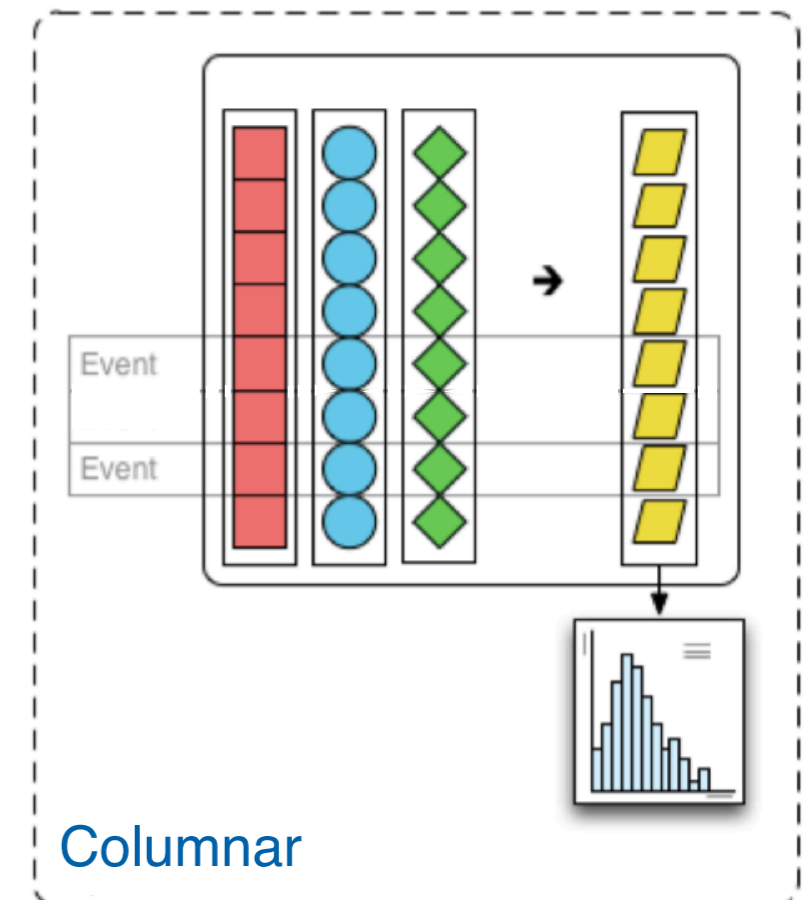
- Event loop analysis:

- Load relevant values for a specific event into local variables
- Evaluate several expressions
- Store derived values
- Repeat (explicit outer loop)



- Columnar analysis:

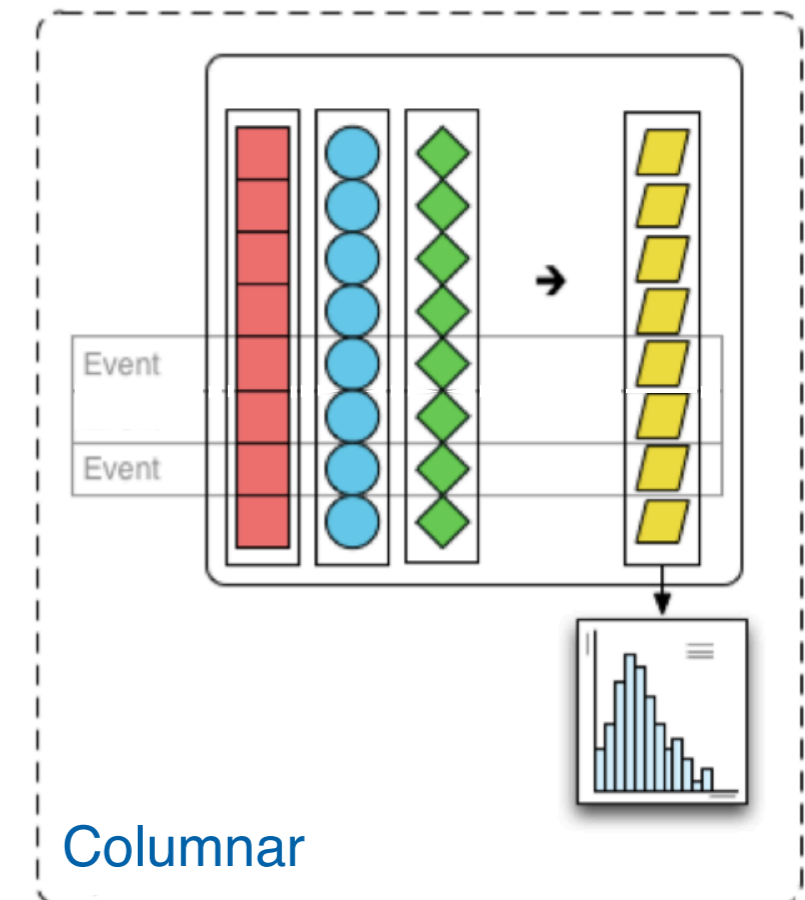
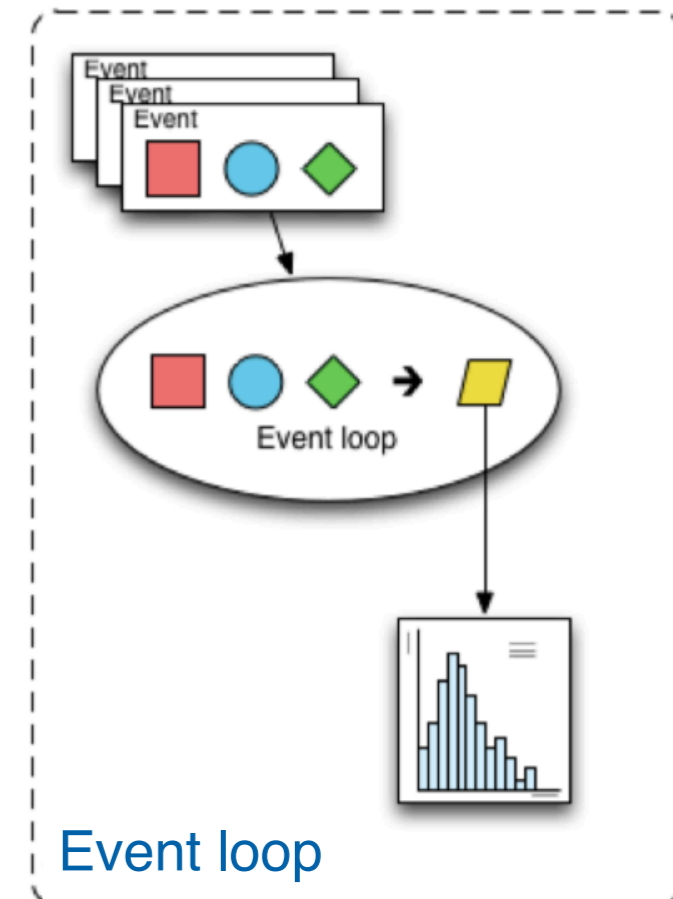
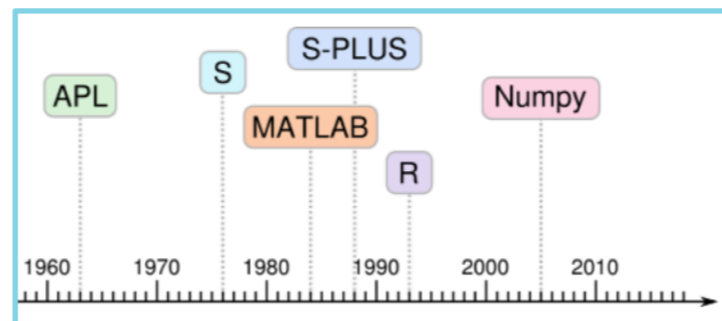
- Load relevant values for many events into contiguous arrays
- Evaluate several **array programming** expressions
  - Implicit *inner* loops
- Store derived values



# The paradigm shift

- Event loop analysis:
  - Load relevant values for a specific event into local variables
  - Evaluate several expressions
  - Store derived values
  - Repeat (explicit outer loop)
- Columnar analysis:
  - Load relevant values for many events into contiguous arrays
  - Evaluate several **array programming** expressions
    - Implicit *inner* loops
  - Store derived values

Array programming is not new!  
[APL demo on YouTube](#)



# Awkward Array: JSON-like data, NumPy-like idioms

```
array = ak.Array([
    [{"x": 1.1, "y": [1]}, {"x": 2.2, "y": [1, 2]}, {"x": 3.3, "y": [1, 2, 3]}],
    [],
    [{"x": 4.4, "y": [1, 2, 3, 4]}, {"x": 5.5, "y": [1, 2, 3, 4, 5]}]
])
```

```
output = []
for sublist in python_objects:
    tmp1 = []
    for record in sublist:
        tmp2 = []
        for number in record["y"][1:]:
            tmp2.append(np.square(number))
        tmp1.append(tmp2)
    output.append(tmp1)
```

**2.3 minutes to run**

(single-threaded on a 2.2 GHz processor with a dataset 10 million times larger than the one shown)

```
output = np.square(array["y", ..., 1:])
```

```
[
    [[], [4], [4, 9]],
    [],
    [[4, 9, 16], [4, 9, 16, 25]]
]
```

**4.6 seconds to run**

[SciPy2020 awkward presentation](#)

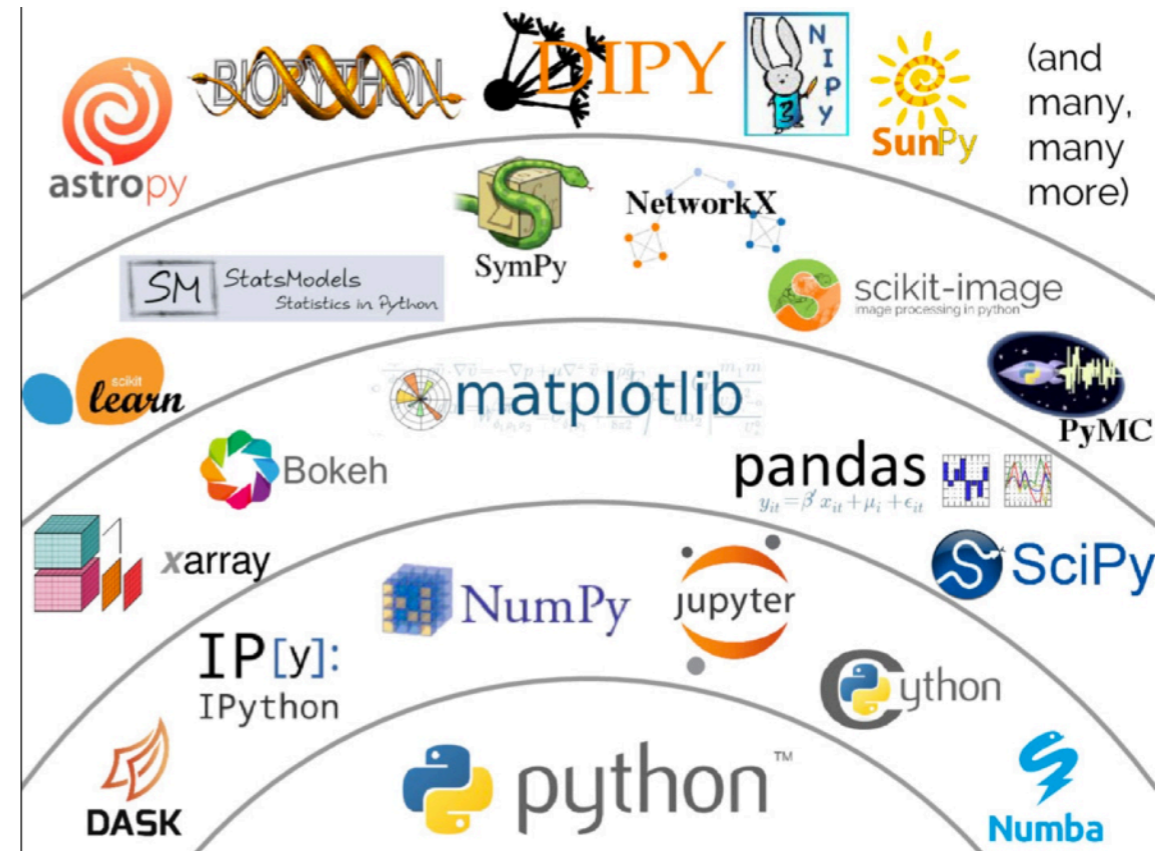
# The other paradigm shift

- From vertically-integrated solution to ecosystem

Experiment-specific and analysis frameworks



<https://root.cern.ch/>



Scientific Python Ecosystem





# Coffea project

- A user interface to *columnar analysis*
  - Optimized array programming kernels build an **expressive and performant** language
  - Seamless integration with ML tools due to shared interface



# Coffea project

- A user interface to *columnar analysis*
  - Optimized array programming kernels build an **expressive and performant** language
  - Seamless integration with ML tools due to shared interface
- An incubator for rapid prototyping
  - Fill in missing pieces of ecosystem
  - Good abstractions are factored out



# Coffea project

- A user interface to *columnar analysis*
  - Optimized array programming kernels build an **expressive and performant** language
  - Seamless integration with ML tools due to shared interface
- An incubator for rapid prototyping
  - Fill in missing pieces of ecosystem
  - Good abstractions are factored out
- A minimum viable product
  - Already used in several CMS publications
  - In use by ATLAS, ProtoDUNE collaborators
  - Early feedback builds ecosystem roadmap
    - Vibrant contributor community



# Coffea project

- A user interface to *columnar analysis*
  - Optimized array programming kernels build an **expressive and performant** language
  - Seamless integration with ML tools due to shared interface
- An incubator for rapid prototyping
  - Fill in missing pieces of ecosystem
  - Good abstractions are factored out
- A minimum viable product
  - Already used in several CMS publications
  - In use by ATLAS, ProtoDUNE collaborators
  - Early feedback builds ecosystem roadmap
    - Vibrant contributor community



**We might be in the business of putting ourselves out of business**

# Your project

- Talk about your community's analysis framework
  - e.g. at HSF Data Analysis Working Group



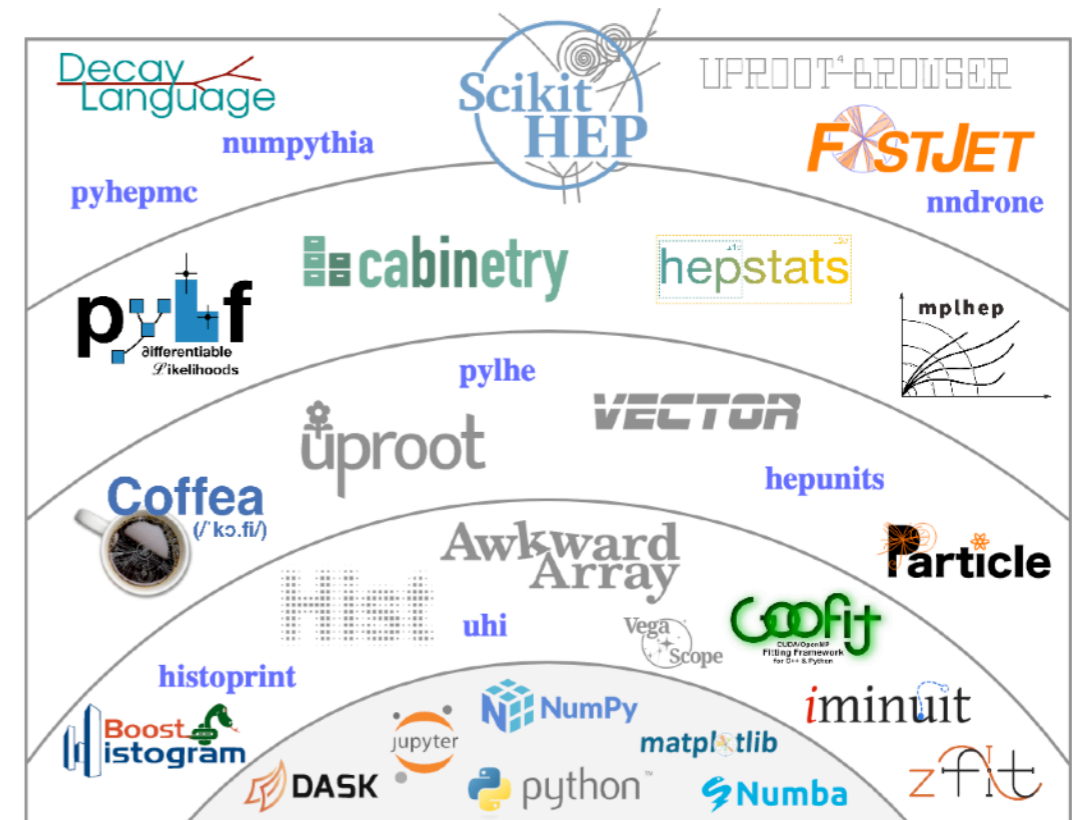
# Your project

- Talk about your community's analysis framework
  - e.g. at HSF Data Analysis Working Group
- Find shared abstractions
  - Generators?



# Your project

- Talk about your community's analysis framework
  - e.g. at HSF Data Analysis Working Group
- Find shared abstractions
  - Generators?
- Help build the ecosystem



# Scaling analyses

- Easy **transition** from local to distributed execution
  - Ideally no user code change



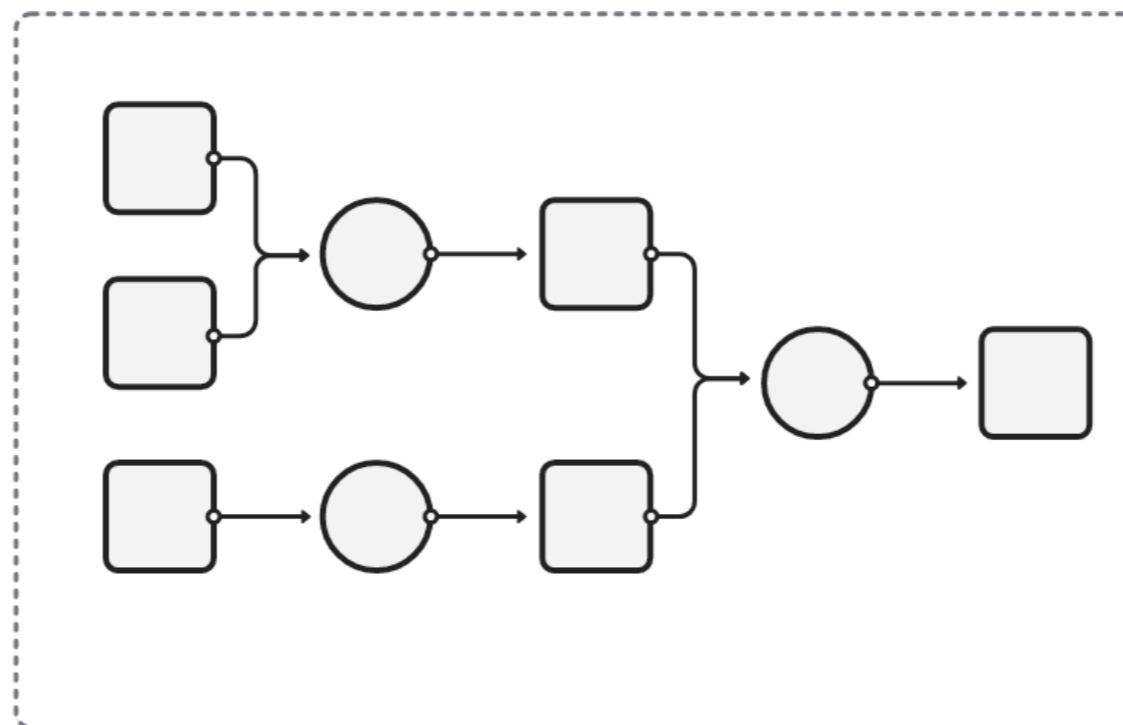
## Collections

(create task graphs)

- Dask Array
- Dask DataFrame
- Dask Bag
- Dask Delayed
- Futures



## Task Graph



## Schedulers

(execute task graphs)

- Single-machine (threads, processes, synchronous)
- Distributed





# Scaling analyses

- Easy **transition** from local to distributed execution
  - Ideally no user code change



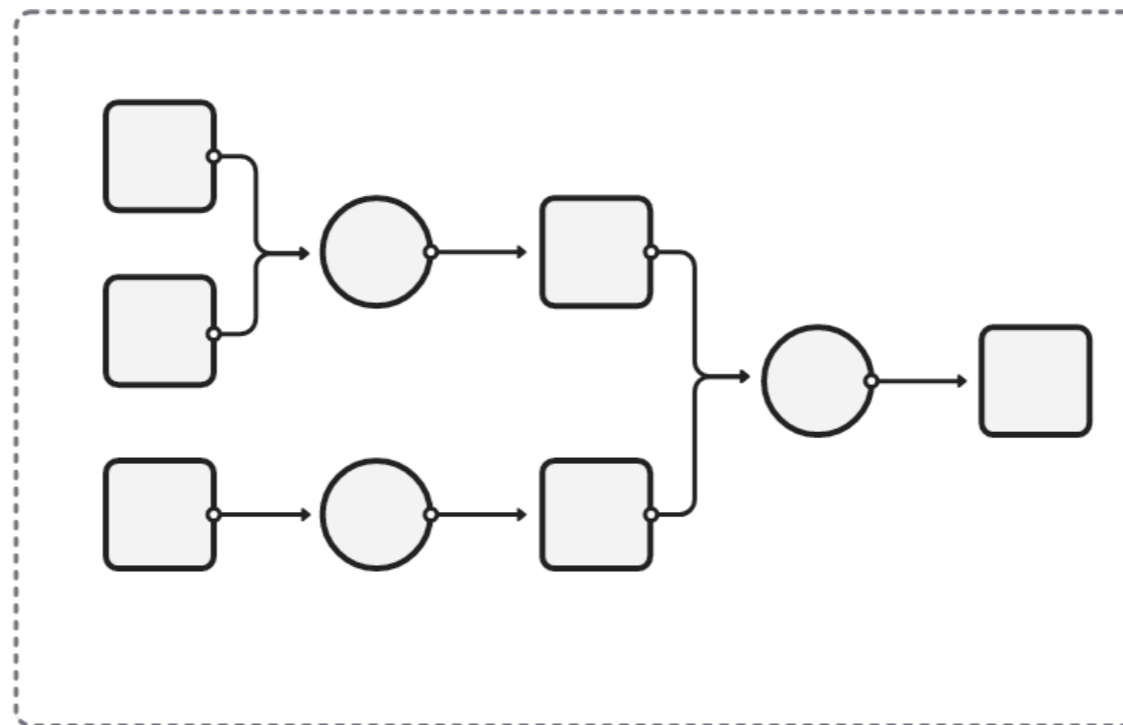
## Collections

(create task graphs)

- Dask Array
- Dask DataFrame
- Dask Bag
- Dask Delayed
- Futures
- Dask-awkward



## Task Graph



## Schedulers

(execute task graphs)

- Single-machine (threads, processes, synchronous)
- Distributed



# Scaling analyses

- Easy **transition** from local to distributed execution
  - Ideally no user code change



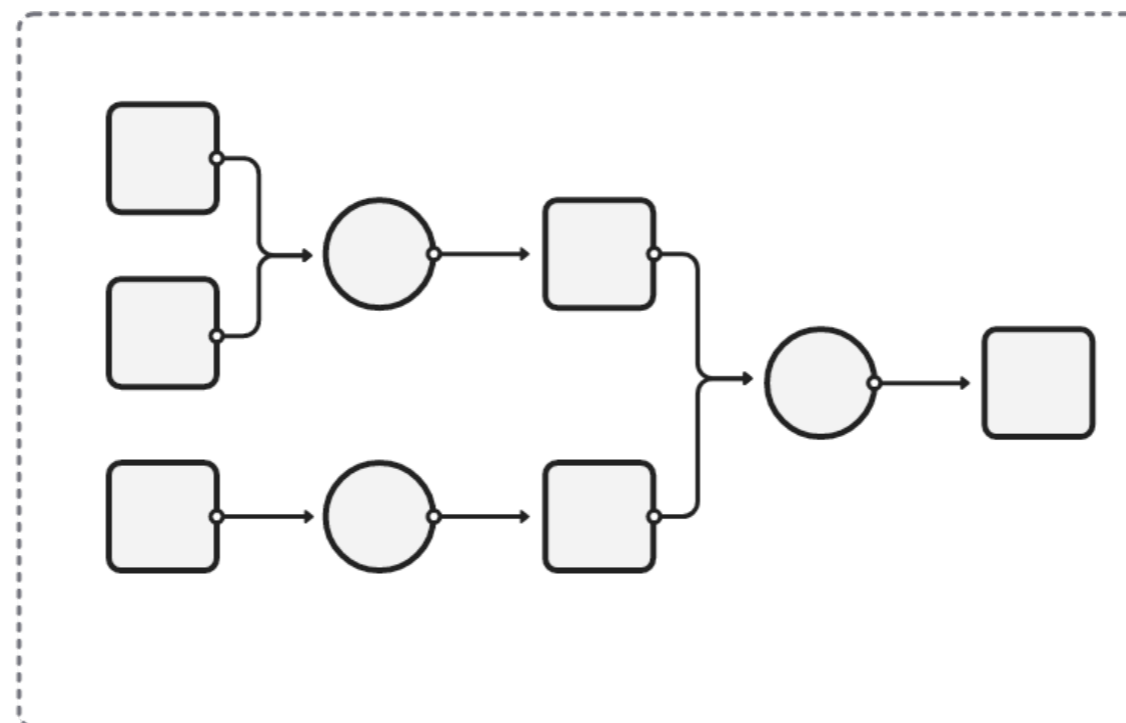
## Collections

(create task graphs)

- Dask Array
- Dask DataFrame
- Dask Bag
- Dask Delayed
- Futures
- Dask-awkward



## Task Graph



## Schedulers

(execute task graphs)

- Single-machine (threads, processes, synchronous)
- Distributed
- Ray
- WorkQueue



# Scaling analyses

- Easy **transition** from local to distributed execution
  - Ideally no user code change



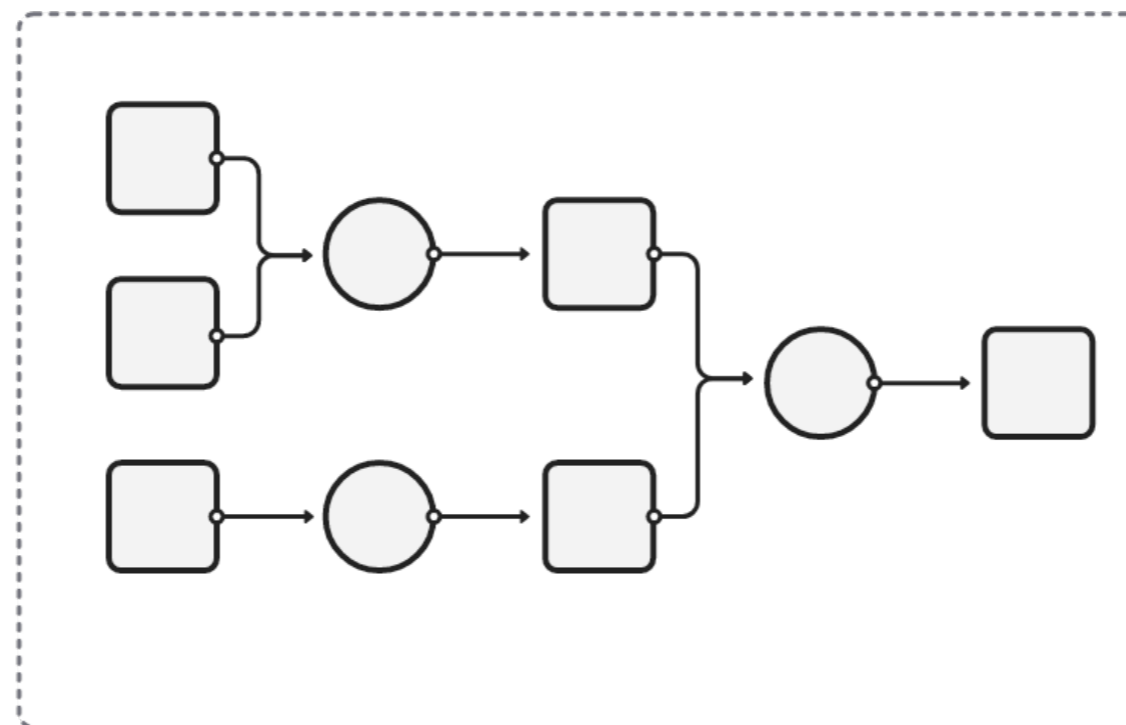
## Collections

(create task graphs)

- Dask Array
- Dask DataFrame
- Dask Bag
- Dask Delayed
- Futures
- Dask-awkward



## Task Graph



## Schedulers

(execute task graphs)

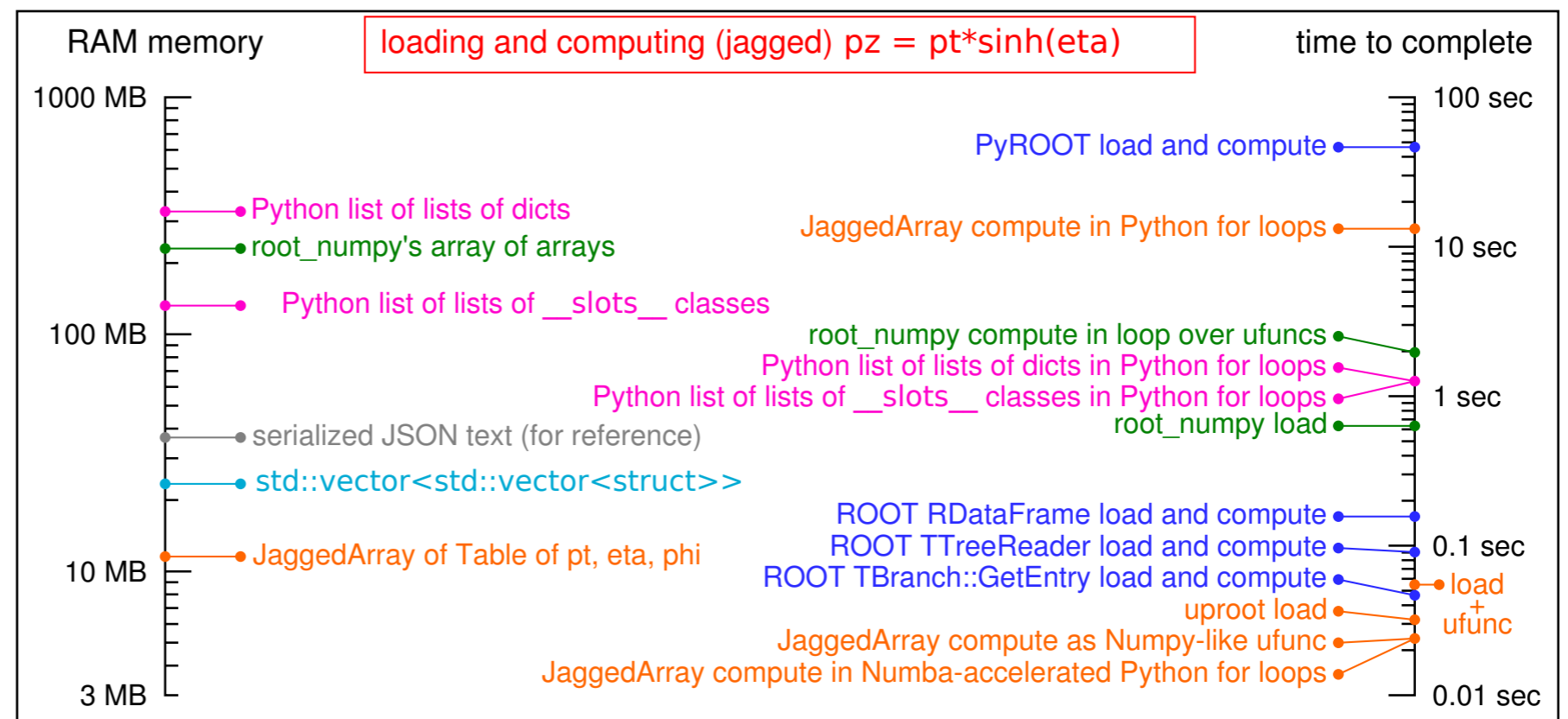
- Single-machine (threads, processes, synchronous)
- Distributed
- Ray
- WorkQueue

Batch queues (HTCondor, Slurm) are now resource provisioning



# Performance

- For library designers, important to know when we are fast *enough*
  - $\mu\text{s}$  to  $\text{ms}$  per event

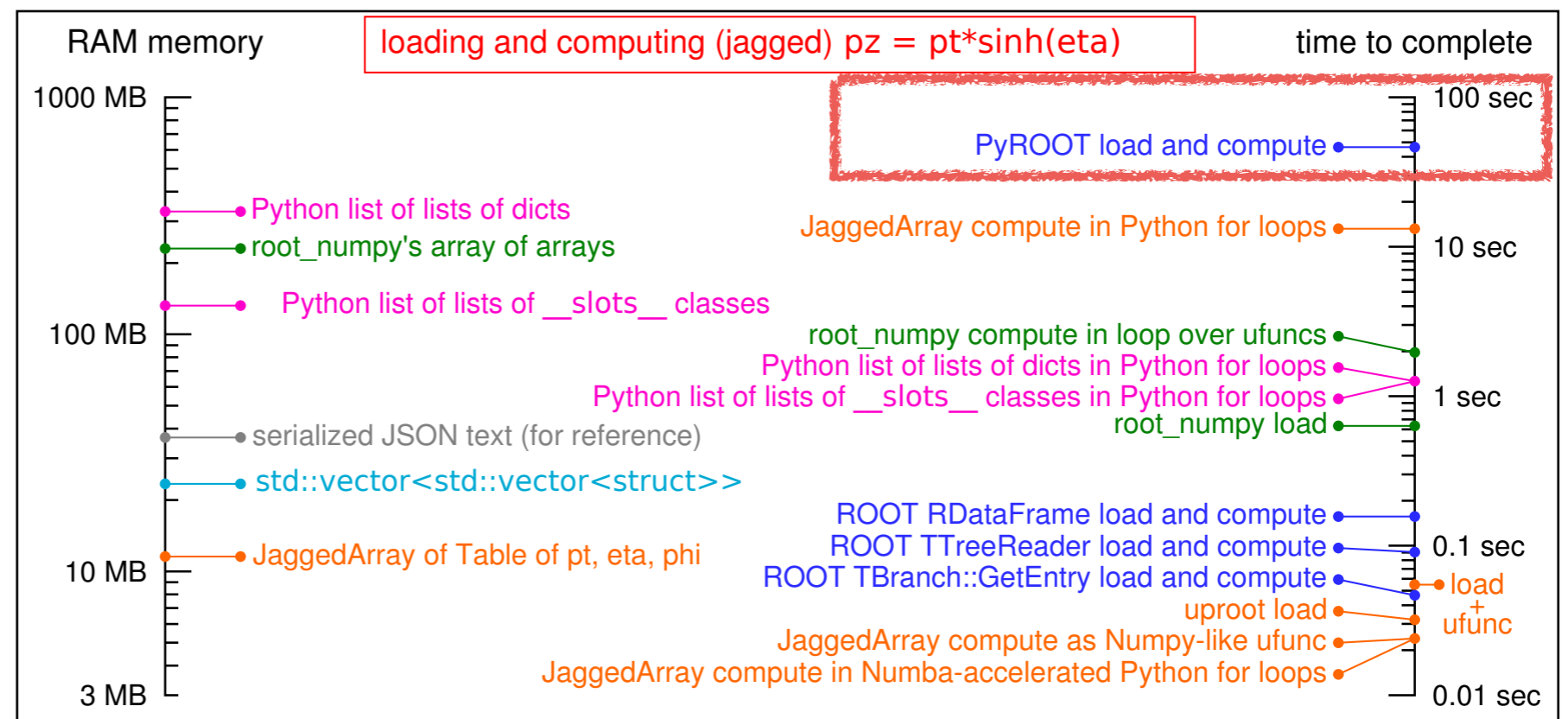


J. Pivarski



# Performance

- For library designers, important to know when we are fast *enough*
  - $\mu\text{s}$  to  $\text{ms}$  per event
- Users may have other plans

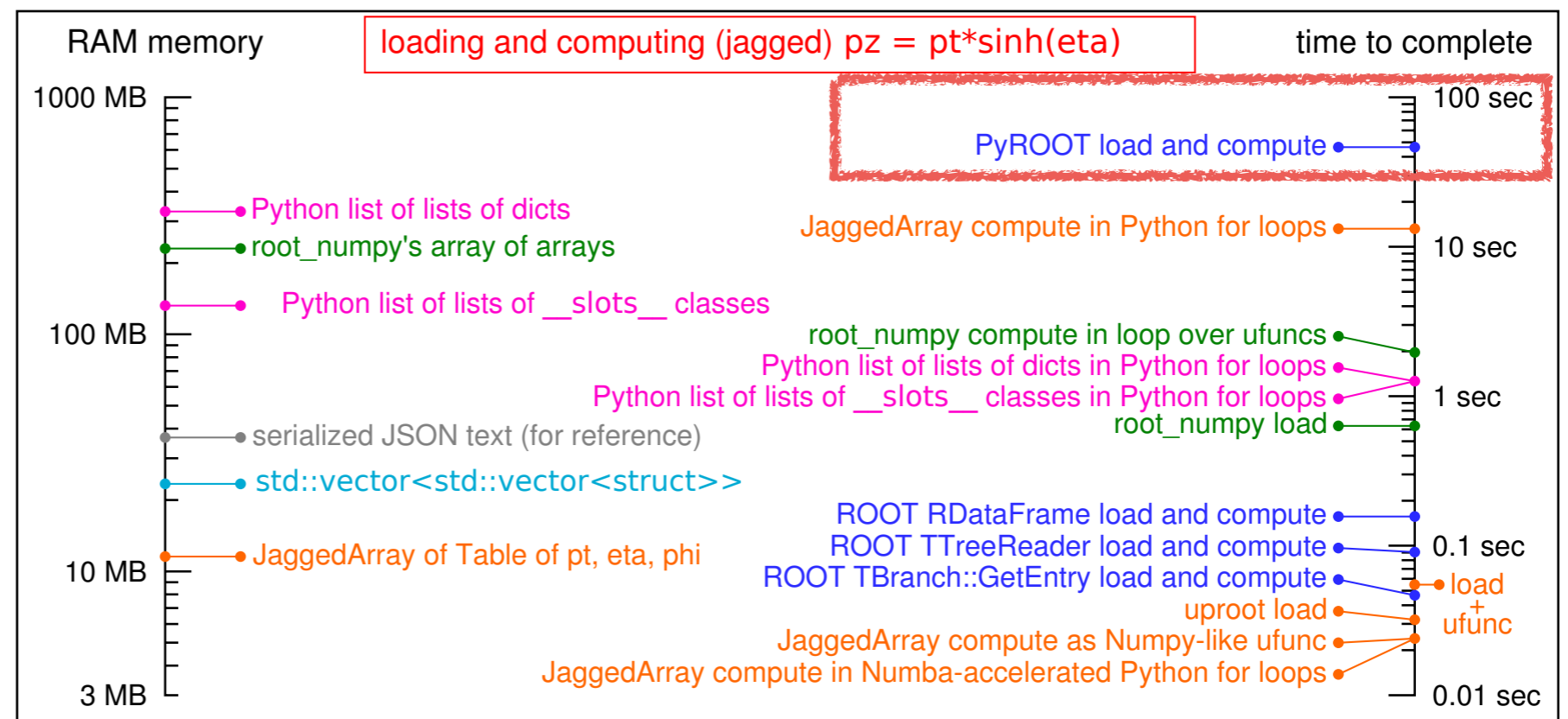


J. Pivarski



# Performance

- For library designers, important to know when we are fast *enough*
  - $\mu\text{s}$  to  $\text{ms}$  per event
- Users may have other plans



[N. Manganelli](#)

[J. Pivarski](#)

Benchmarking the code and coming out fastest is fantastic

- Factor 3x\* is small compared to the O(1000)-O(10000) improvement RDF/coffea have against TTree::Draw-based frameworks (I know of several)

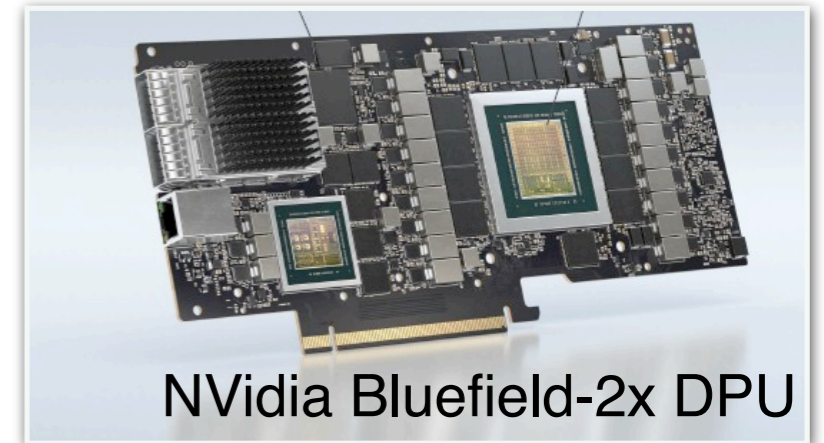


# Facility integration

- **Data delivery** is the bottleneck for columnar analysis at scale
  - True also for AI/ML workloads

# Facility integration

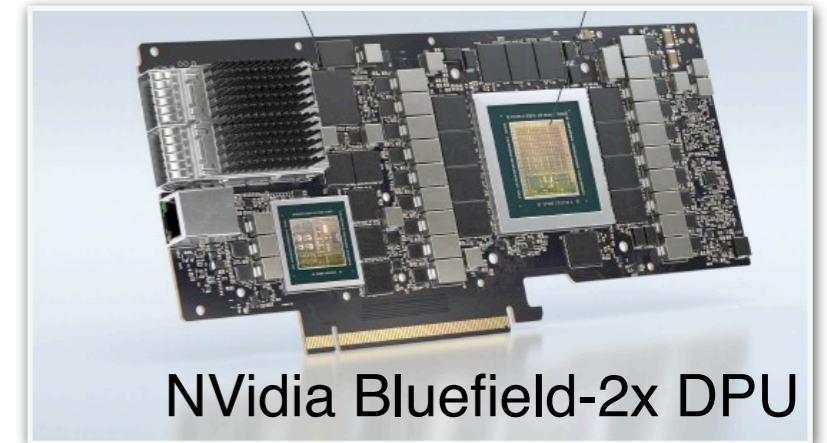
- **Data delivery** is the bottleneck for columnar analysis at scale
  - True also for AI/ML workloads
- Analysis facility integration would:
  - Reduce manual user data curation
  - Offload expensive algorithms to accelerators
  - Save compute and storage resources





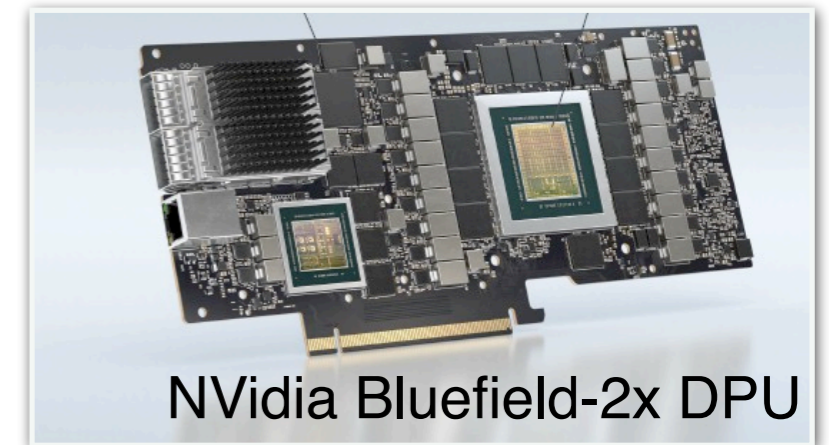
# Facility integration

- **Data delivery** is the bottleneck for columnar analysis at scale
  - True also for AI/ML workloads
- Analysis facility integration would:
  - Reduce manual user data curation
  - Offload expensive algorithms to accelerators
  - Save compute and storage resources
- ...to maintain fast *time to insight physics*

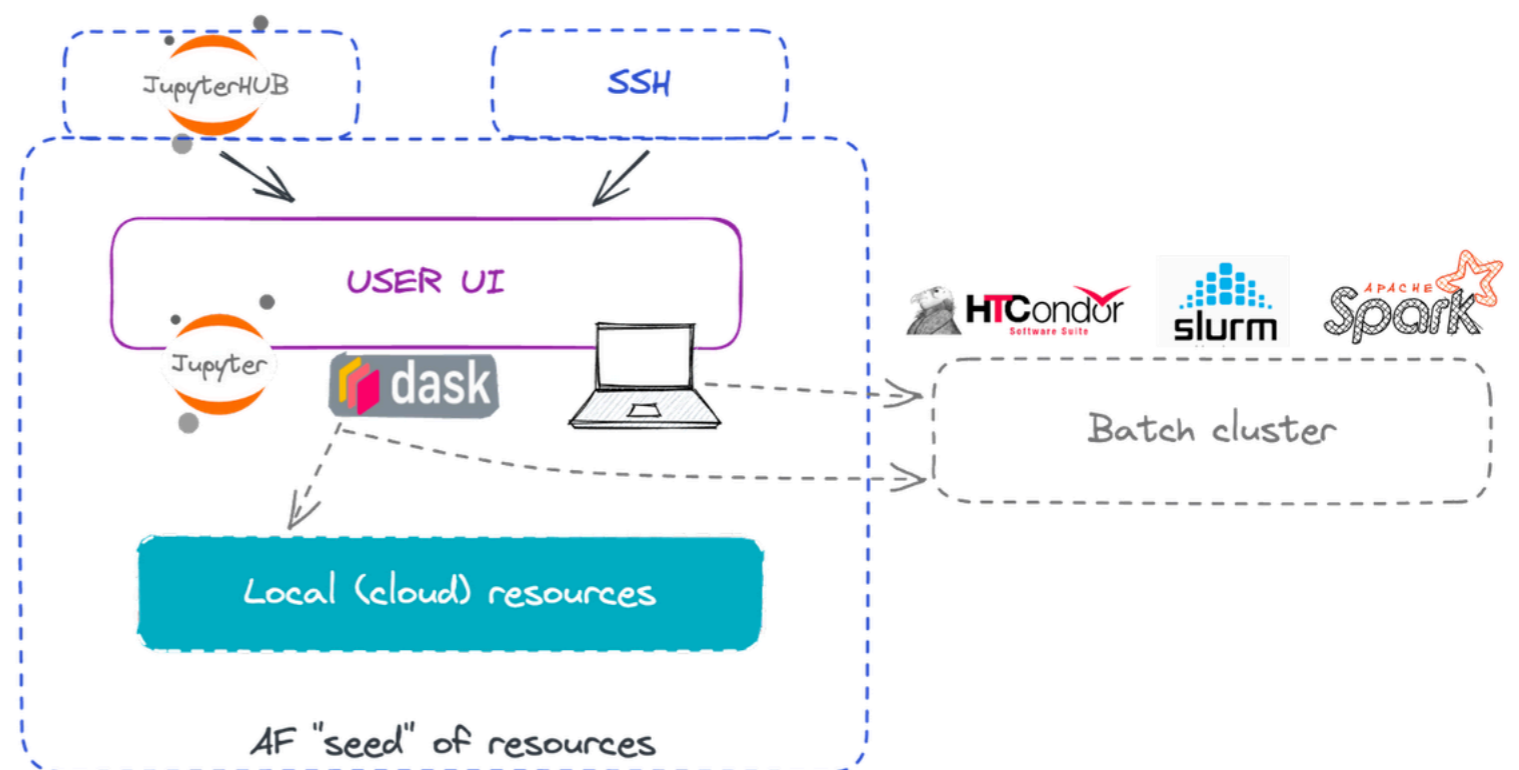


# Facility integration

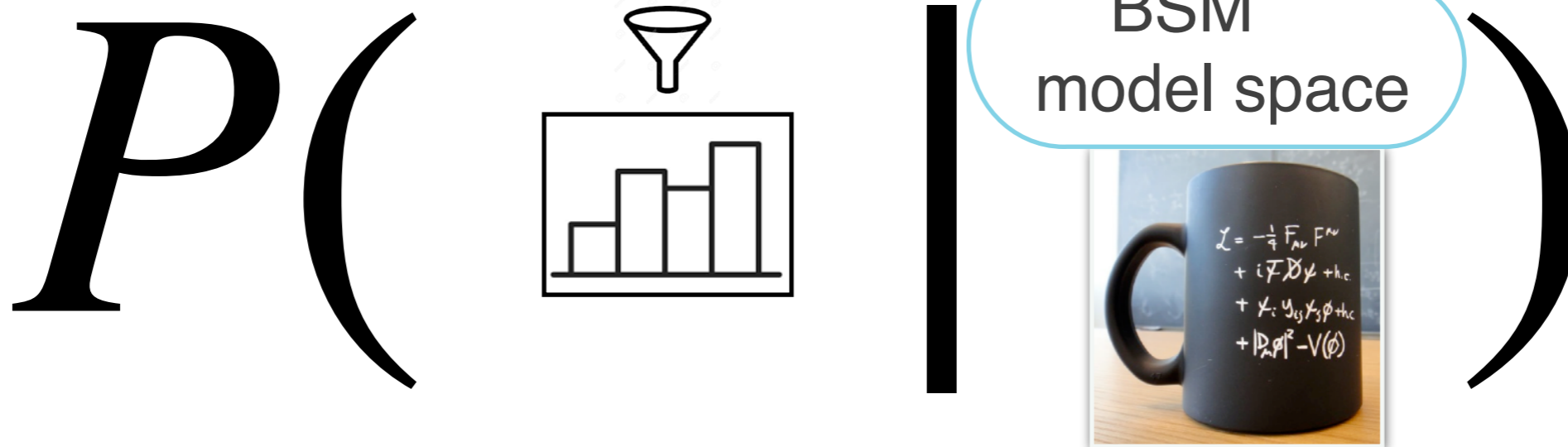
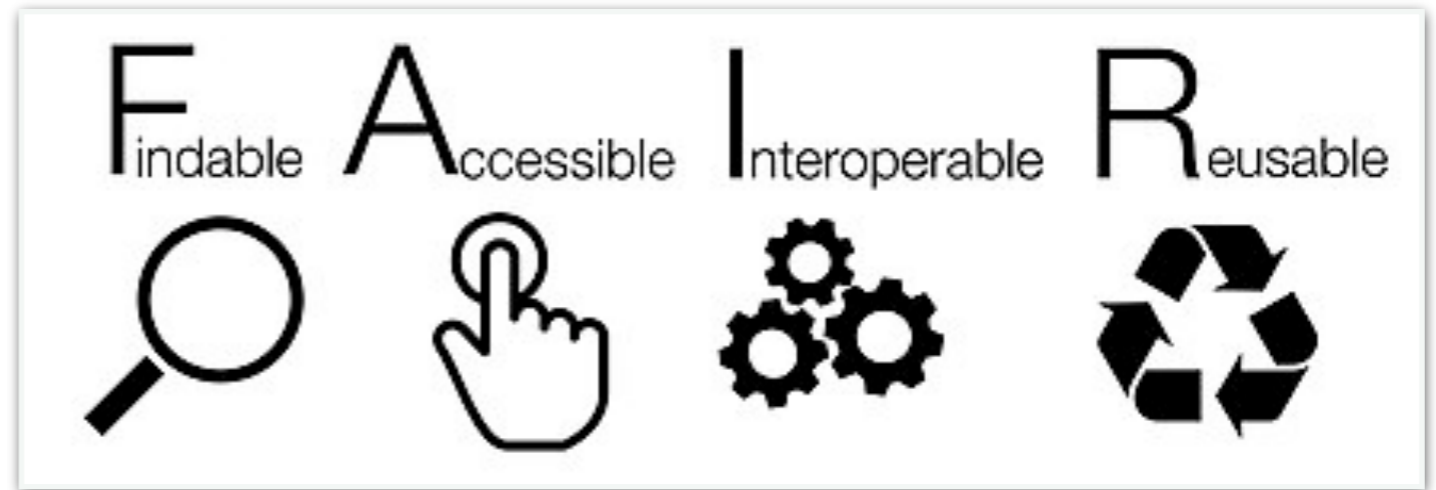
- **Data delivery** is the bottleneck for columnar analysis at scale
  - True also for AI/ML workloads
- Analysis facility integration would:
  - Reduce manual user data curation
  - Offload expensive algorithms to accelerators
  - Save compute and storage resources
- ...to maintain fast *time to insight physics*



- We have the playgrounds to realize this vision
  - Snowmass contrib: [arxiv:2203.10161](https://arxiv.org/abs/2203.10161)
  - [HSF AF white paper](#)



# FAIR Results



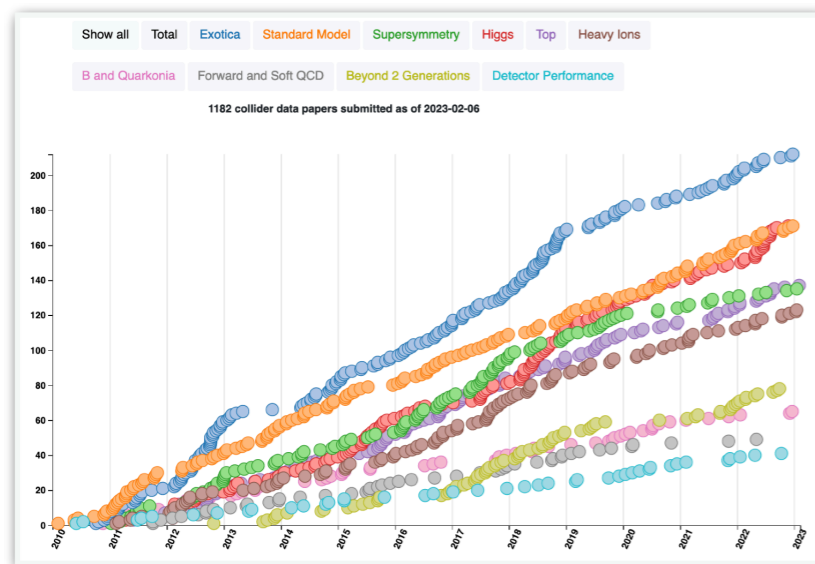
# Papers are a bottleneck

- When will papers fail to capture physics output?
  - EFT model space: interpretability challenge
    - Combinations of many observables may be key to find new physics
  - Data formats, software tools can meet this challenge
    - HEPData, HEP Statistics Serialization Standard

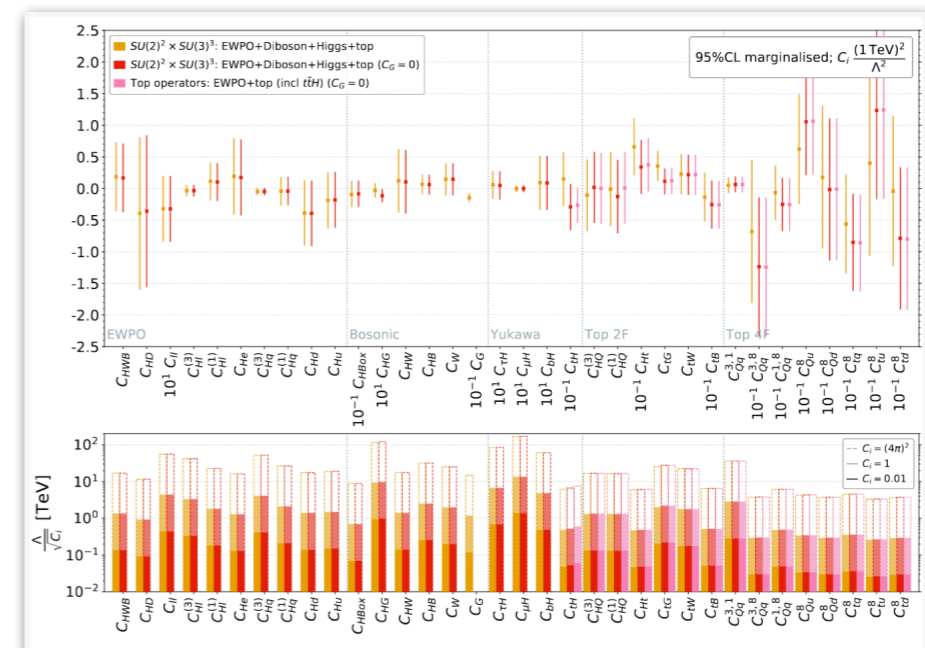
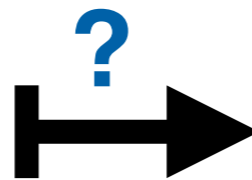


# Papers are a bottleneck

- When will papers fail to capture physics output?
  - EFT model space: interpretability challenge
    - Combinations of many observables may be key to find new physics
  - Data formats, software tools can meet this challenge
    - HEPData, HEP Statistics Serialization Standard



CMS publication count

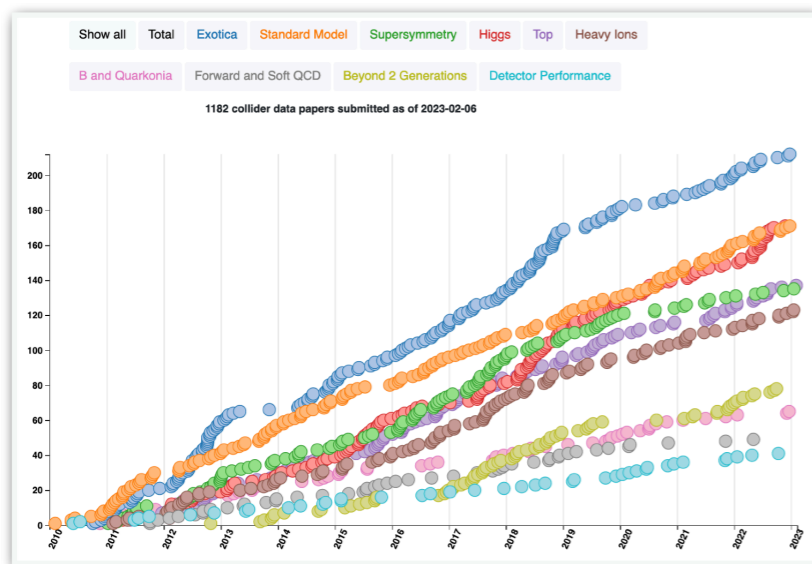
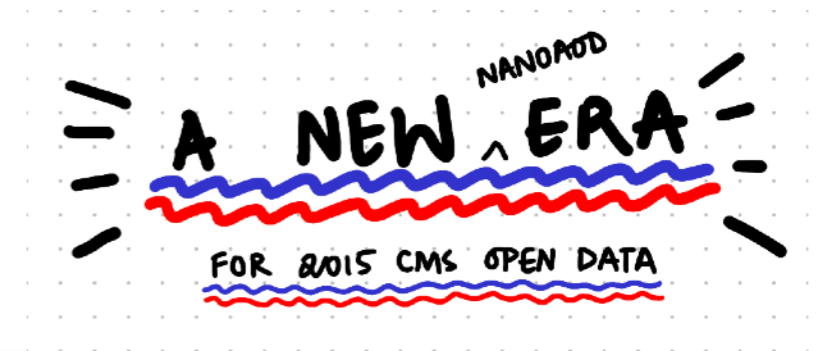


Indirect searches for new physics

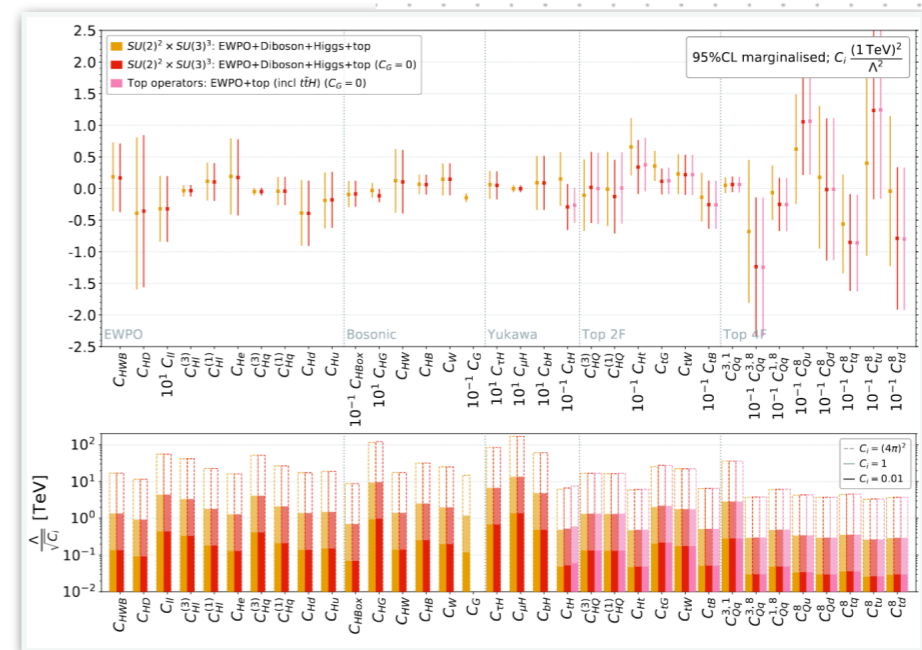
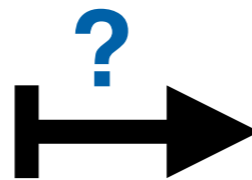


# Papers are a bottleneck

- When will papers fail to capture physics output?
  - EFT model space: interpretability challenge
    - Combinations of many observables may be key to find new physics
  - Data formats, software tools can meet this challenge
    - HEPData, HEP Statistics Serialization Standard
- Open data may enable new discovery
  - Why impose our priors? (Cross-checks expected)
    - Data, metadata, and tools need to be FAIR



CMS publication count

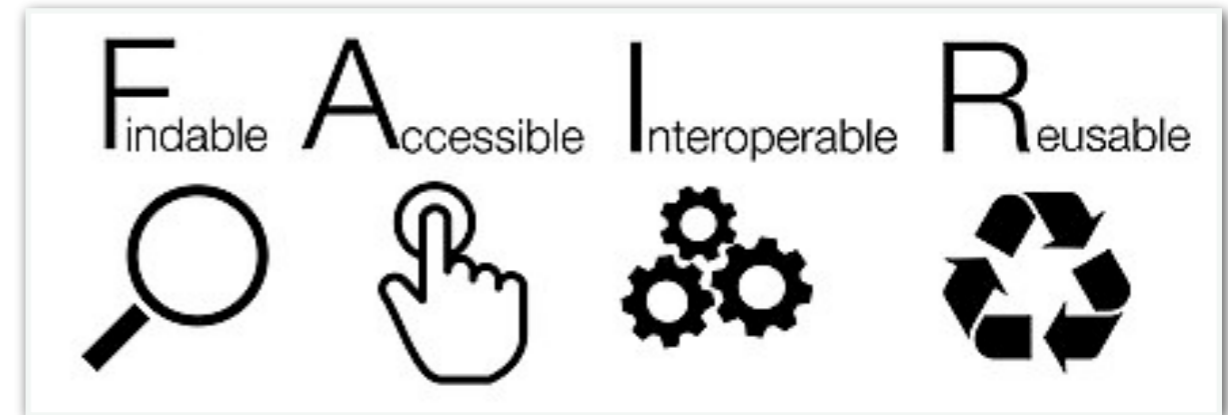


Indirect searches for new physics



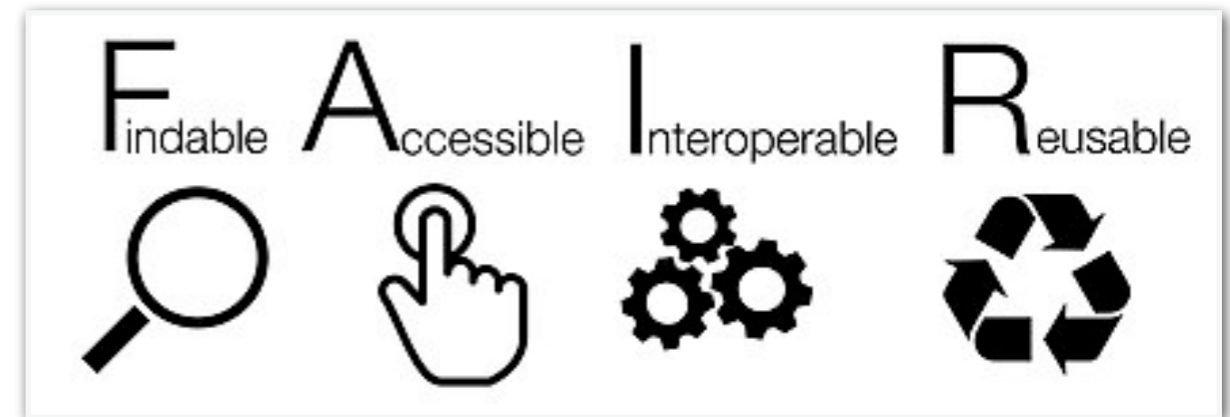
# Particle Physics Software Culture

- Analysis software is critical for HEP
  - But new collaborators can struggle
    - Chase down requirements / “recipes”
    - Join group with mature framework / toolset
  - Is our data and metadata FAIR?
    - Software sometimes viewed as competitive advantage



# Particle Physics Software Culture

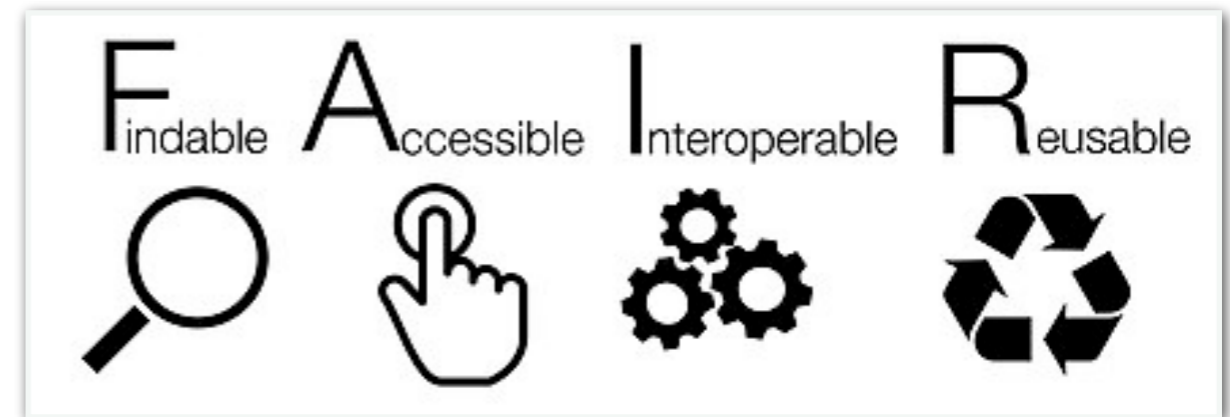
- Analysis software is critical for HEP
  - But new collaborators can struggle
    - Chase down requirements / “recipes”
    - Join group with mature framework / toolset
  - Is our data and metadata FAIR?
    - Software sometimes viewed as competitive advantage
- Need training & mentorship pipeline
  - In same sense as pixels, calorimeters, ...software detector?
  - Career paths: Traditional track, Research Software Engineer ([US-RSE](#)), ?



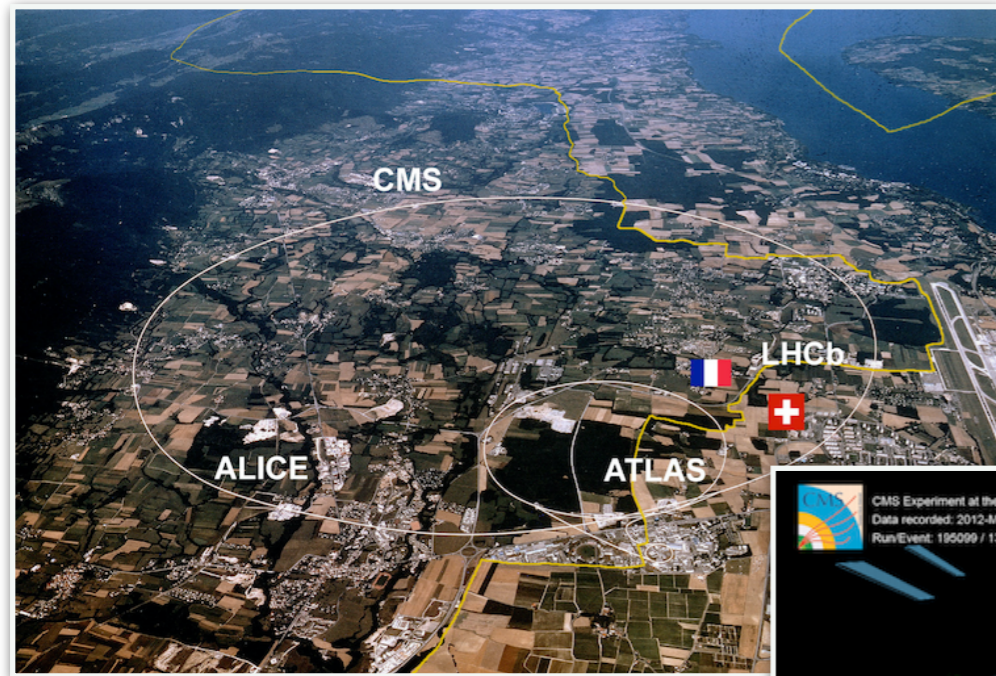


# Particle Physics Software Culture

- Analysis software is critical for HEP
  - But new collaborators can struggle
    - Chase down requirements / “recipes”
    - Join group with mature framework / toolset
  - Is our data and metadata FAIR?
    - Software sometimes viewed as competitive advantage
- Need training & mentorship pipeline
  - In same sense as pixels, calorimeters, ...software detector?
  - Career paths: Traditional track, Research Software Engineer ([US-RSE](#)), ?
- In the right direction: IRIS-HEP, HSF
  - Soon: DPF Coordination Panel for Software and Computing



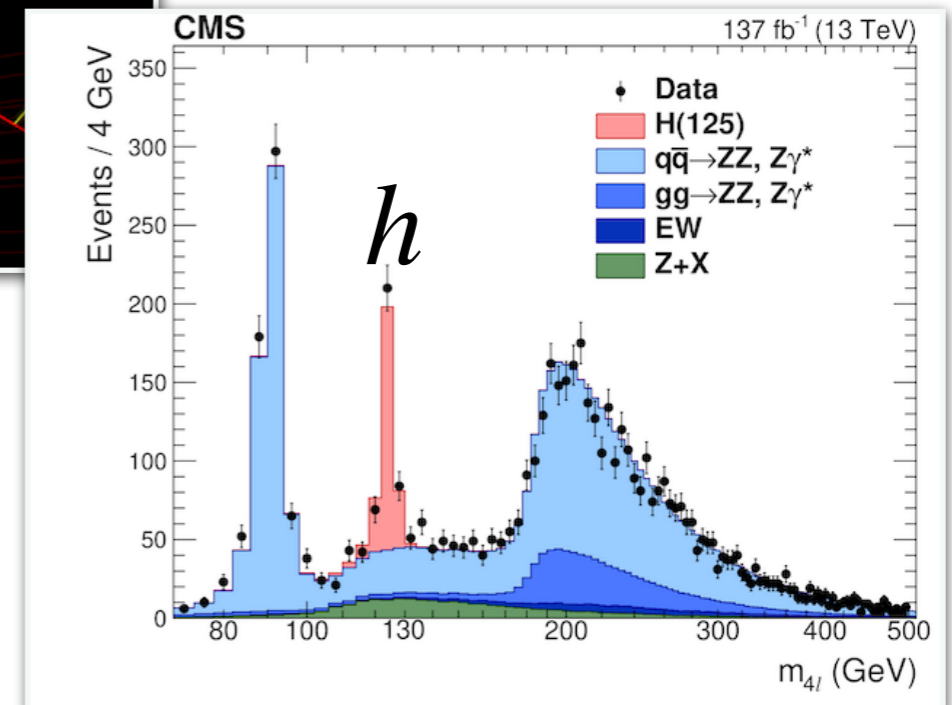
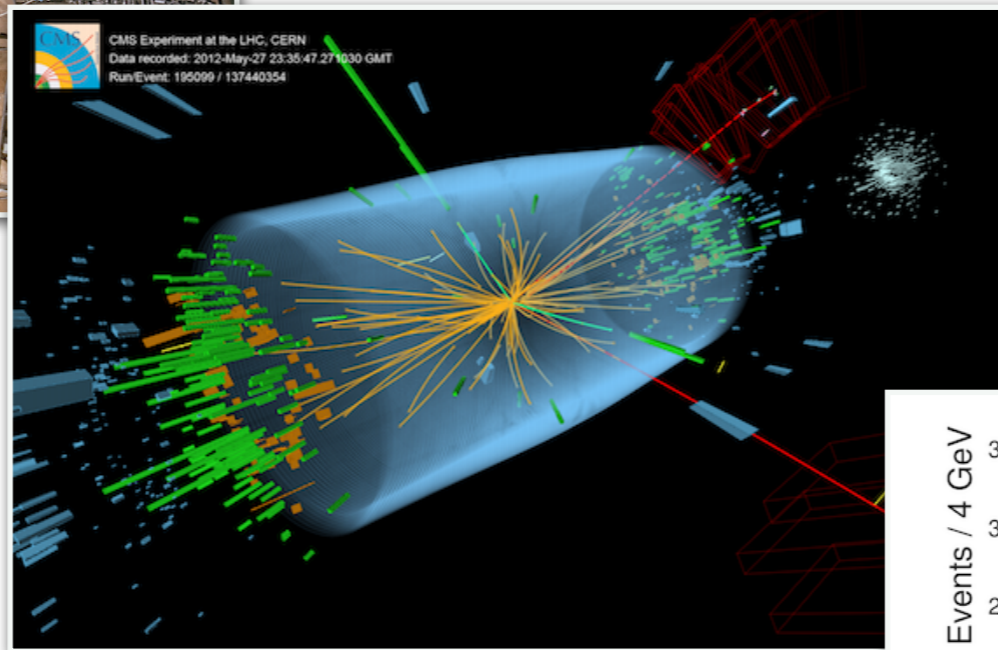
# HEP Experiment: three easy steps



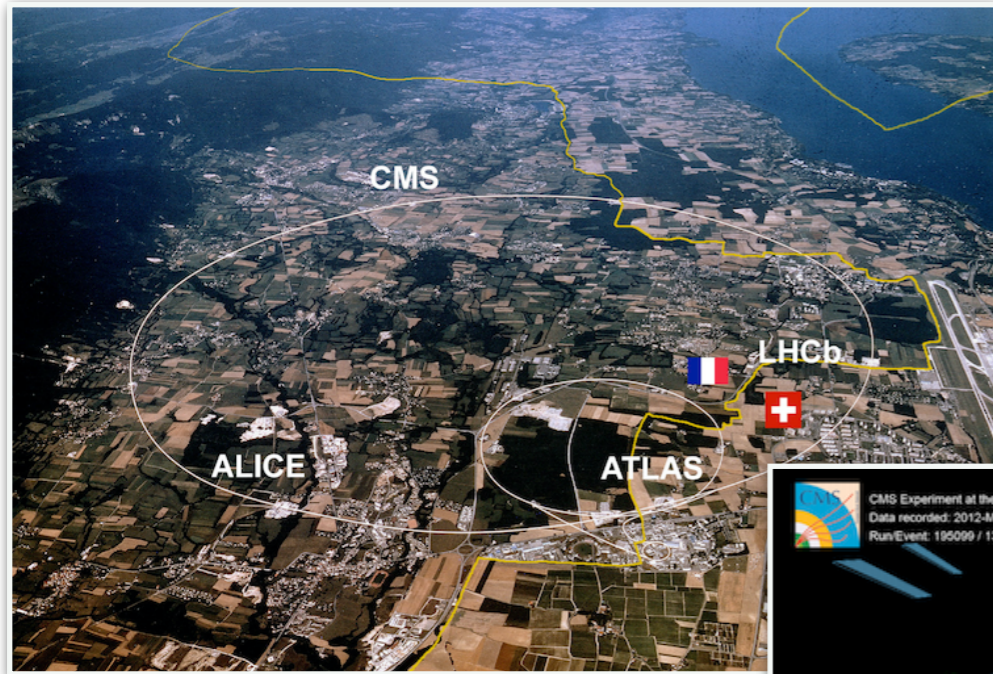
1. Collide particles

2. Take pictures

3. Infer parameters



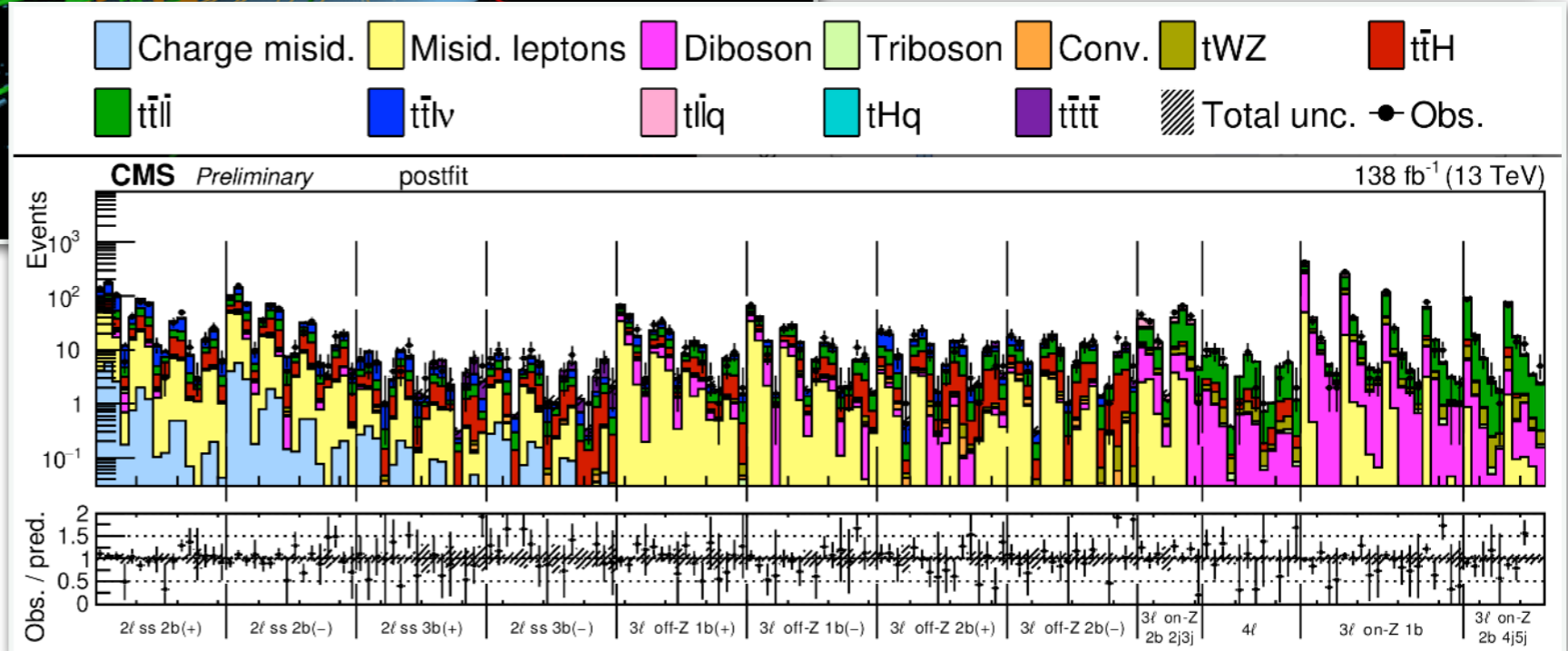
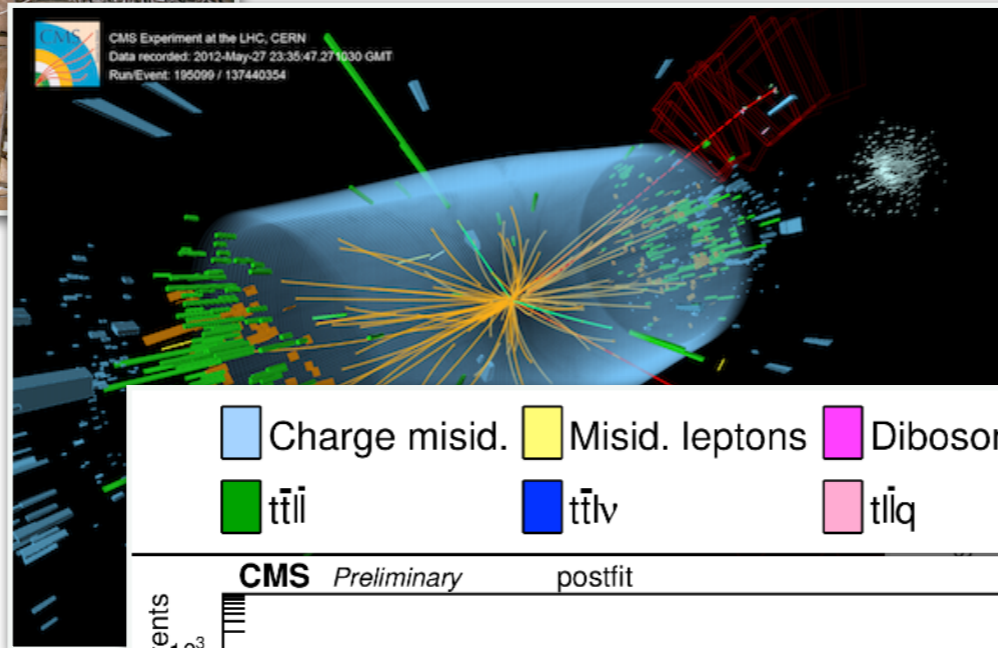
# HEP Experiment: three easy steps



1. Collide particles

2. Take pictures

3. Infer parameters



# Conclusion

# Conclusion

- A new kind of scaling challenge faces us
  - Complexity of data analysis

# Conclusion

- A new kind of scaling challenge faces us
  - Complexity of data analysis
- If we are to make the most of our data, we need innovations in *design*
  - Composable libraries and shared interfaces



# Conclusion

- A new kind of scaling challenge faces us
  - Complexity of data analysis
- If we are to make the most of our data, we need innovations in *design*
  - Composable libraries and shared interfaces
- To build a software detector, we need a software culture
  - Build an expert community, and keep it

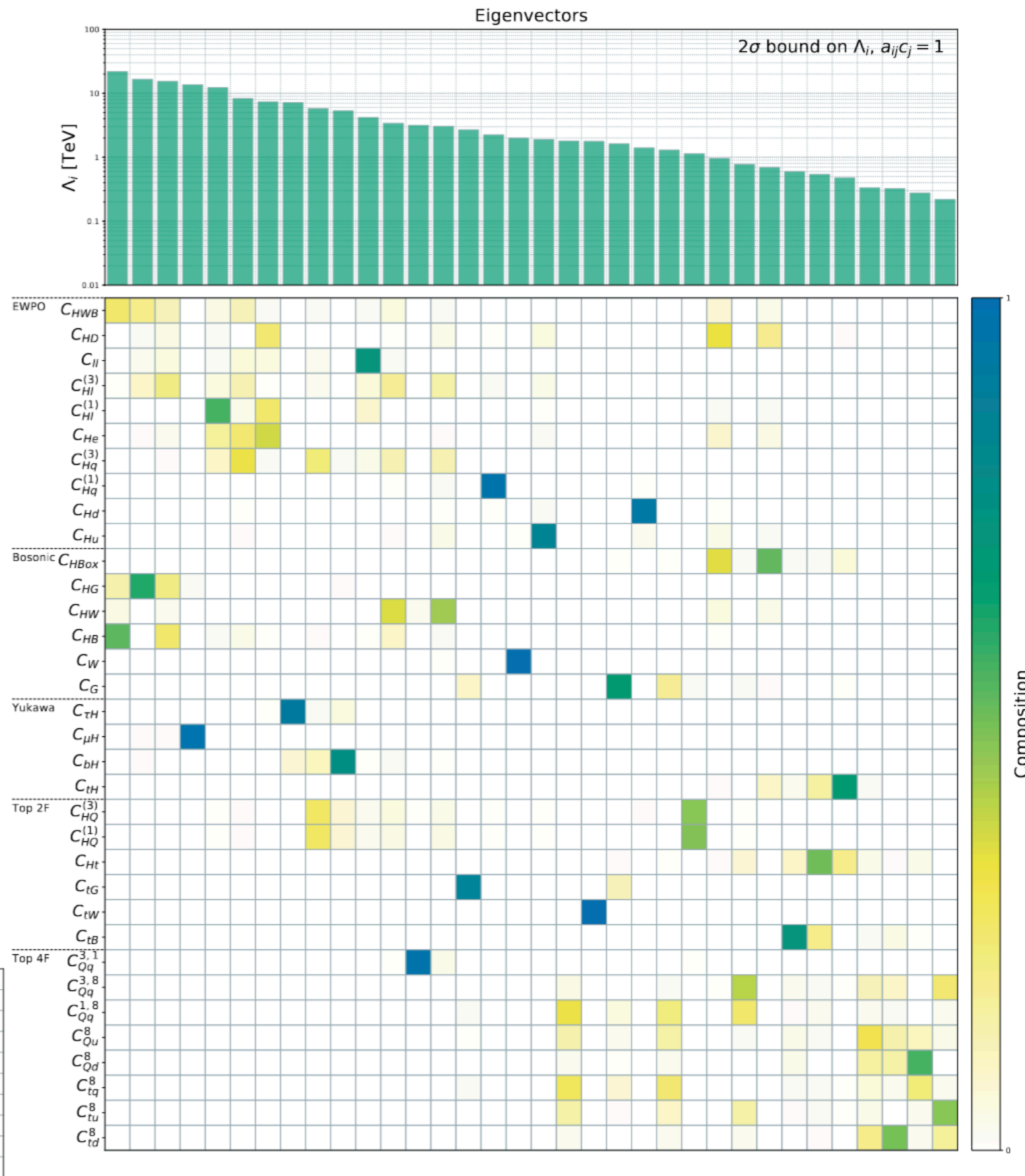


# Bonus slides



# SMEFT global fit

- From [arxiv:2012.02779](https://arxiv.org/abs/2012.02779)
- One such analysis



EWPO	6	35	59	-	99	95	99	-	74	18	73	7	-	7	-	94	-	65	-	-	90	-	-	2	-	4	-	-	1	-	-	-	-	-	
LEP WW	-	-	-	-	-	-	-	-	-	22	2	-	-	-	2	2	14	-	-	-	5	-	-	3	-	4	-	-	-	-	-	-	-	-	-
Run 1, $\mu_H$	15	13	7	5	-	-	21	2	8	-	5	-	2	-	-	2	-	-	-	-	-	-	27	1	14	1	3	8	-	-	-	-	-	-	
Run 2, $\mu_H$	29	28	16	94	-	-	45	4	19	-	53	3	39	10	-	4	-	-	13	-	5	-	40	2	33	3	10	27	-	-	-	-	-	-	
STXS	50	23	18	1	-	-	33	14	55	-	4	-	2	4	-	10	-	-	6	1	2	-	12	-	23	1	4	16	-	-	-	-	-	-	
LHC WV	-	-	-	-	3	-	6	-	3	29	-	46	-	2	11	4	-	-	2	-	-	-	6	-	7	-	-	2	-	-	-	-	-	-	
LHC Zjj	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	86	1	-	-	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	
$t\bar{t}$	-	-	-	-	-	-	-	-	-	-	-	-	-	-	57	-	-	99	84	48	-	79	3	4	15	12	4	28	32	58	90	61	95	-	-
single top	-	-	-	-	-	-	-	-	-	-	-	-	-	96	3	-	-	-	-	15	-	-	78	-	4	-	-	2	-	-	-	-	-	-	
$t\bar{t}V$	-	-	-	-	-	-	-	-	-	-	-	-	-	28	-	-	-	32	-	13	18	4	77	2	89	53	12	41	10	38	5	-	-	-	

Relative constraining power (%)

