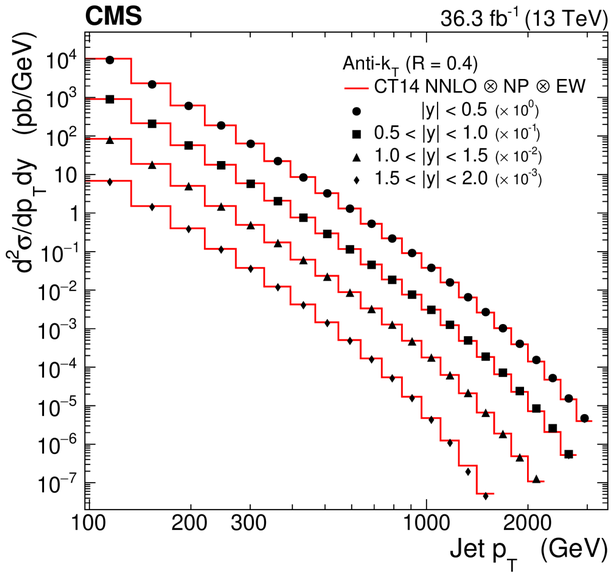
Data Analysis and Bayesian Methods Lecture 2

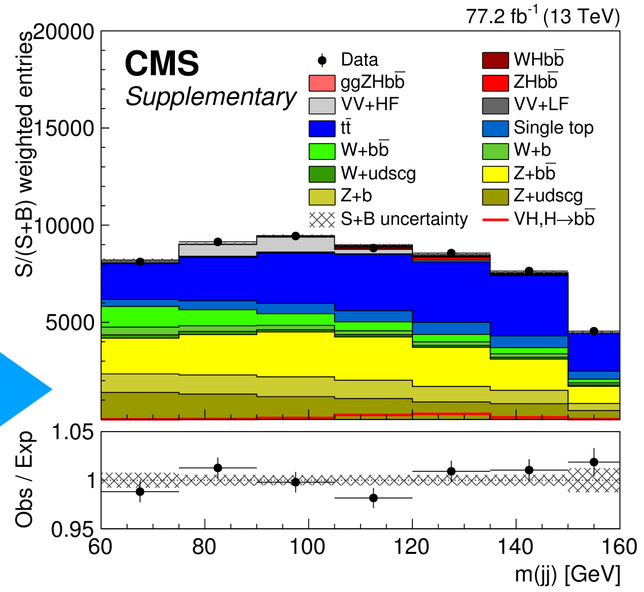
Maurizio Pierini

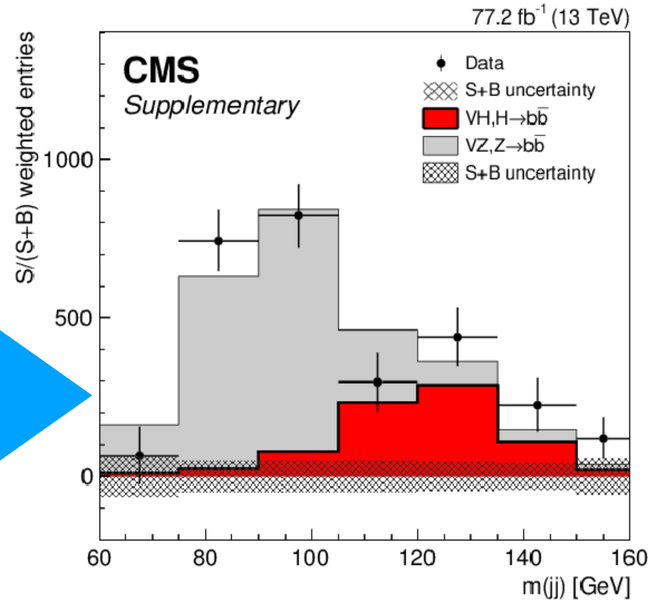
Lecture program

	Day1	Day2	Day3	Day4	Day5
Lecture	Introduction to probability and statistics	Data analysis in a nutshell	Bayesian inference	Bayesian inference	Data analysis beyond hypothesis test

A multiple-step process

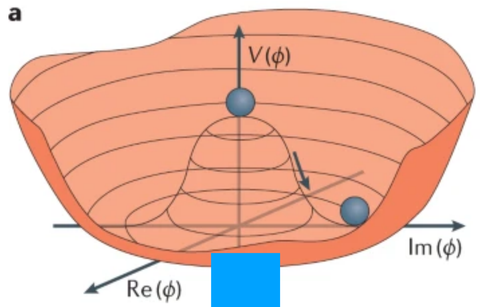


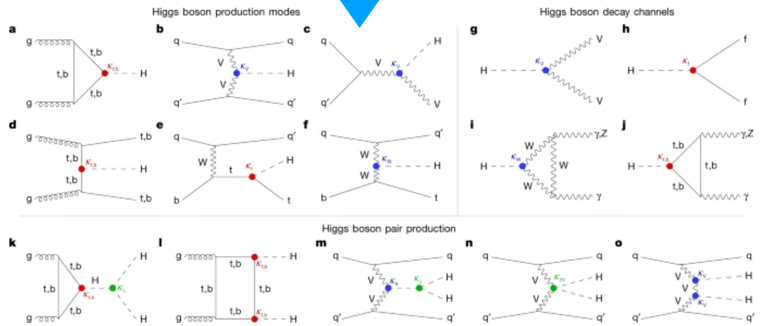


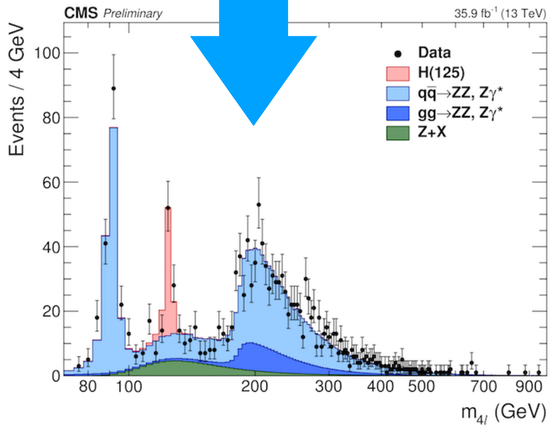


A supervised problem

- *Typically, HEP data analysis follows a top-down supervised problem*
  - *One starts with a specific process in mind*
  - *A theoretical framework allows to predict the experimental signature (qualitatively and quantitatively) through a Monte Carlo simulation*
  - *The data analysis is tailored on the process*
- *PROS: maximise sensitivity (e.g., can work on improving background rejection)*
- *CONS: poor generalization. Performance loss if the signal is different*
- *A supervised approach is ideal when you have a target in mind, e.g., Higgs@LHC, WIMP underground, a precision measurement of a Standard Model process*
- *For searches, one might need some additional tool with a different perspective (see Friday lecture)*







The program for these lectures

- *STEP 1: make sure that potentially interesting events enter your dataset (aka the trigger)*

- *STEP 2: define an event selection that selects a subset of your data, potentially enhanced with signal*
- *STEP 3: define a procedure to estimate the amount of residual background events in your selected sample*
- *STEP 4: extract the signal component (aka the measurement)*

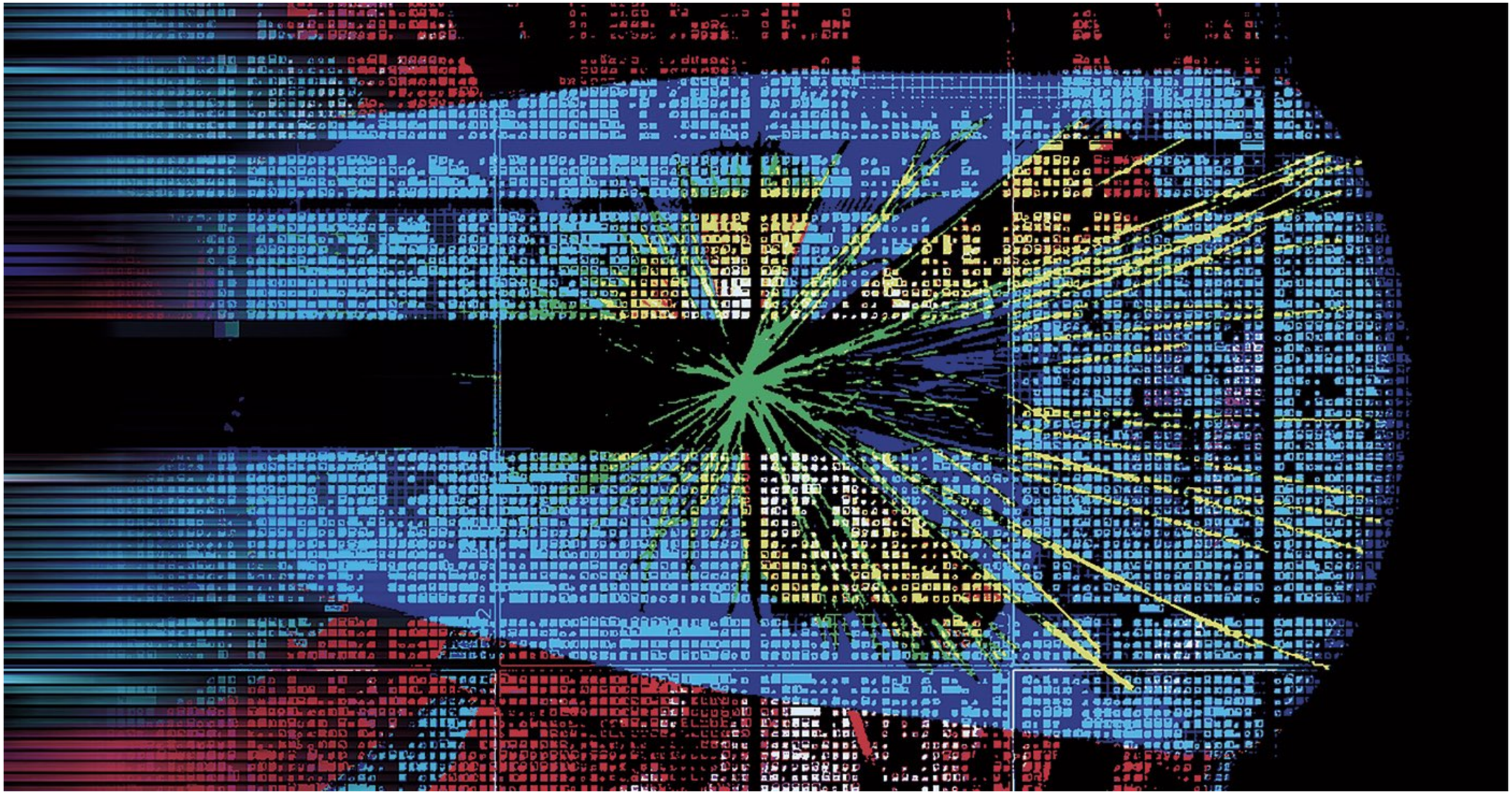
- *STEP 5: use the measurement to learn something about nature (aka phenomenology)*

Monday

Tuesday

Wednesday and Thursday

Friday: a degression on signal "agnostic" analyses



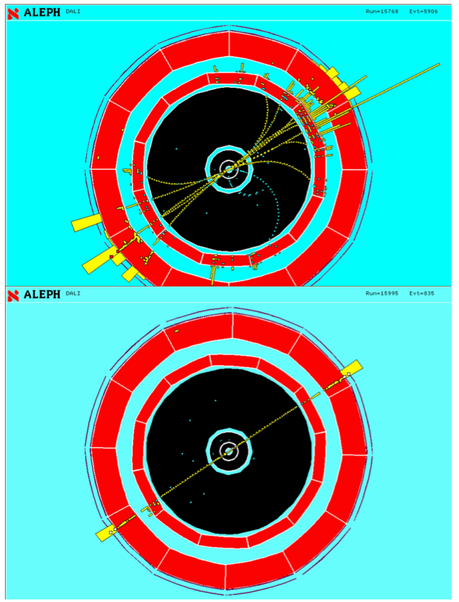
STEP 1: trigger selection

The need of a trigger

- *Ideally, one would like to be able to store and analyse each individual events*
- *In practice this is never done typically because of limited resources (storage, bandwidth, etc)*
- *Sometimes this is harmless*
  - *In a clean environment, one just needs to reject the obvious background (e+e- colliders, underground experiments, …)*
- *Sometimes it's a challenge*
  - *At the LHC, one cannot store everything*
  - *Difficult choices have to be made very early*

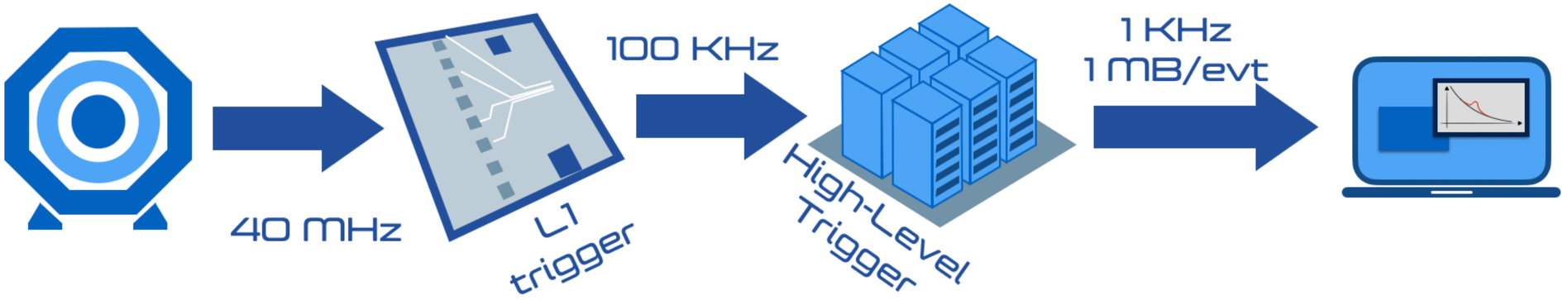
R. Tenchini "Trigger @LEP"

- All systems, except TPC, delivered signals suitable for L1 trigger purpose (i.e much faster than 11 μs)
- L1 criterion was to require the presence of at least one single particle candidate, charged or neutral, from one or more system

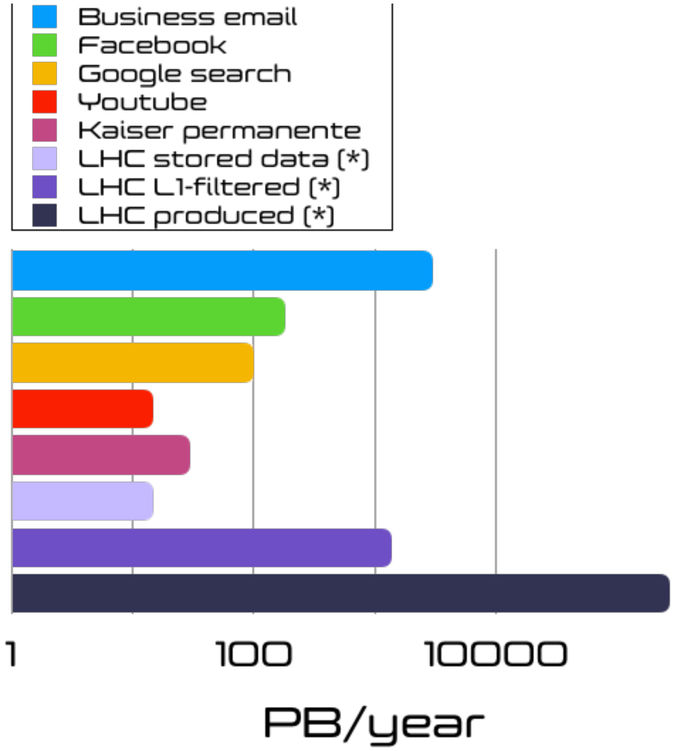


At LEP, the trigger consisted in a set of algorithms requiring some activity, to reject beam-gas interactions and to discard/tag obvious noise

How big is big?



- *The amount of produced data is too much to be stored*
  - *1,000 times the data generated by google searches+youtube+facebook back in 2013*
  - *Reduced to 5x(google searches+youtube+facebook) after first filtering*
  - *Can only store 5% of those*



(*) Only two big experiments (ATLAS and CMS), only RAW data

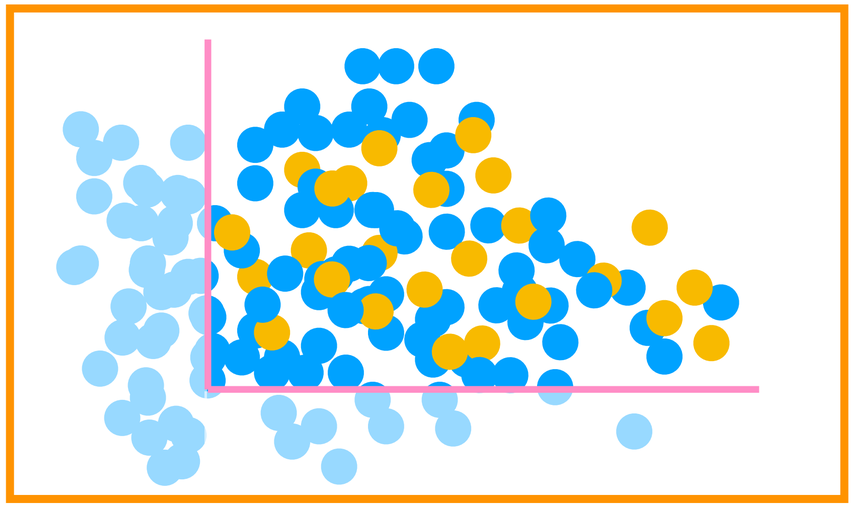
Online vs Offline

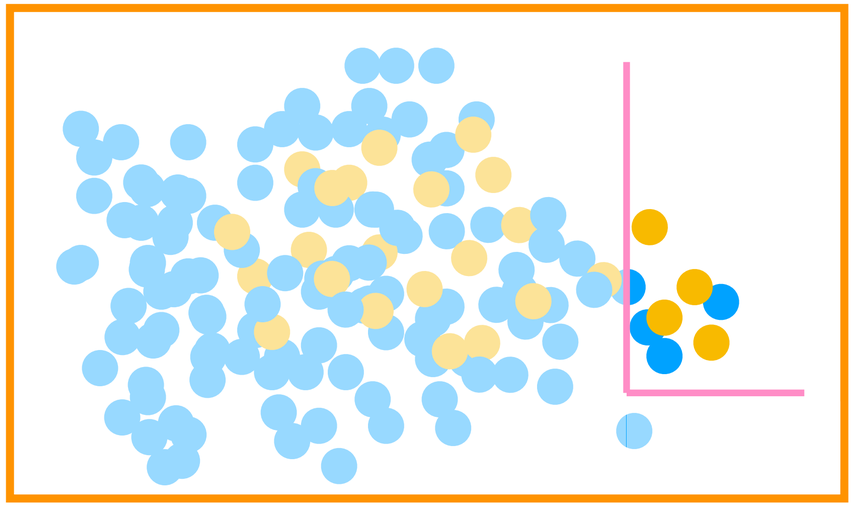
- *STEP 1: make sure that potentially interesting events enter your dataset (aka the trigger)*
- *STEP 2: define an event selection that selects a subset of your data, potentially enhanced with signal*

- *These are actually two aspects of the same problem*
  - *Both consists in applying a set of requirements to select a subset of the events*
- *They differ in scope and for practical reasons*
  - *Efficiency vs Purity*
  - *Accuracy vs Speed*

Efficiency vs Purity

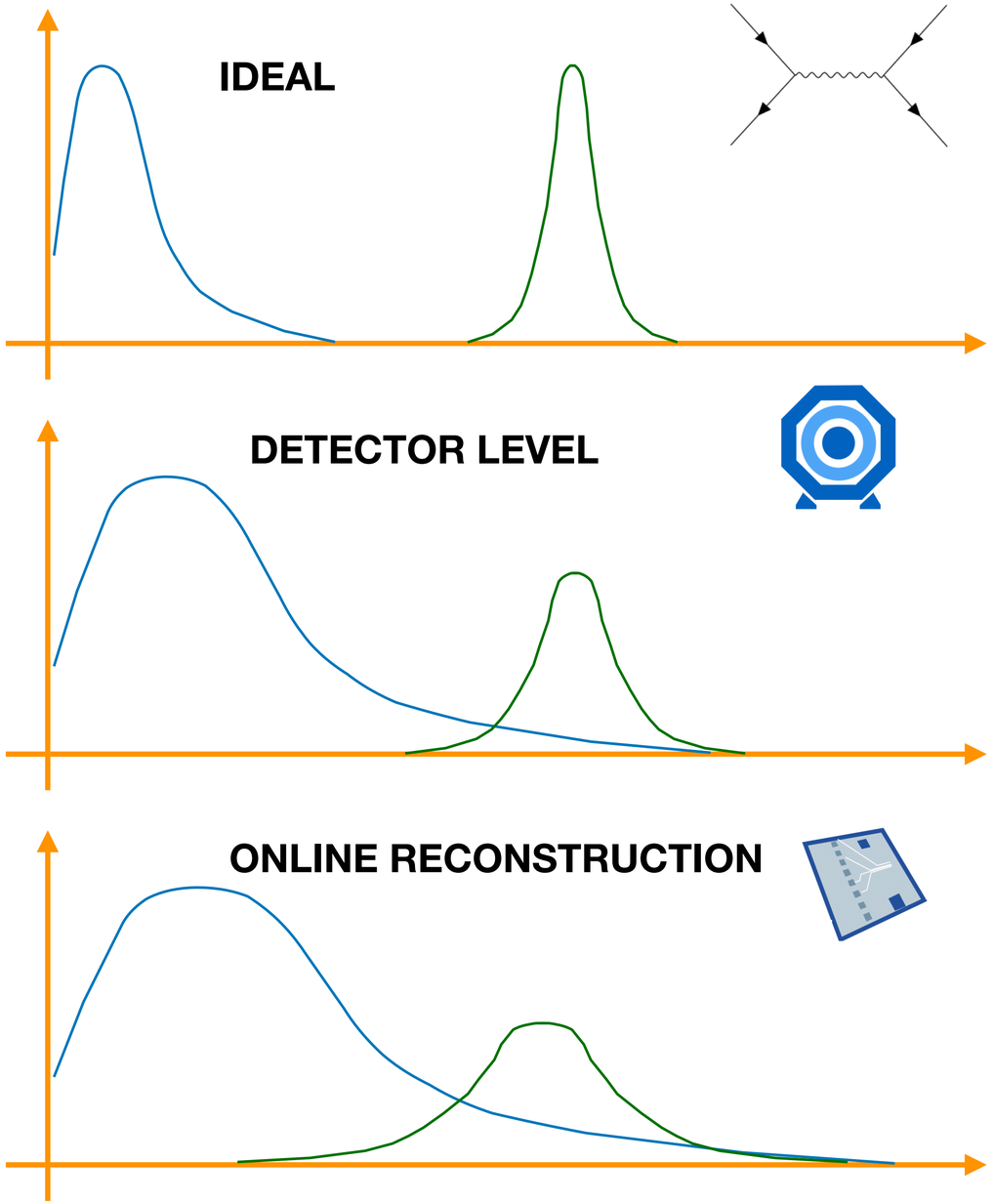
- *Efficiency is the fraction of signal events that would pass your selection*
  - *aka True positive rate, recall, etc.*
- *Purity is a measure of how large is the fraction of signal events in the selected dataset*
  - *Measured as $S/B, S/\sqrt{B}, S/\sqrt{S+B}$, depending on the context*
- *Maximizing efficiency (what one would do in a trigger) implies a loose selection*
- *Maximizing purity (what one would do in a data analysis offline) implies a tight selection*





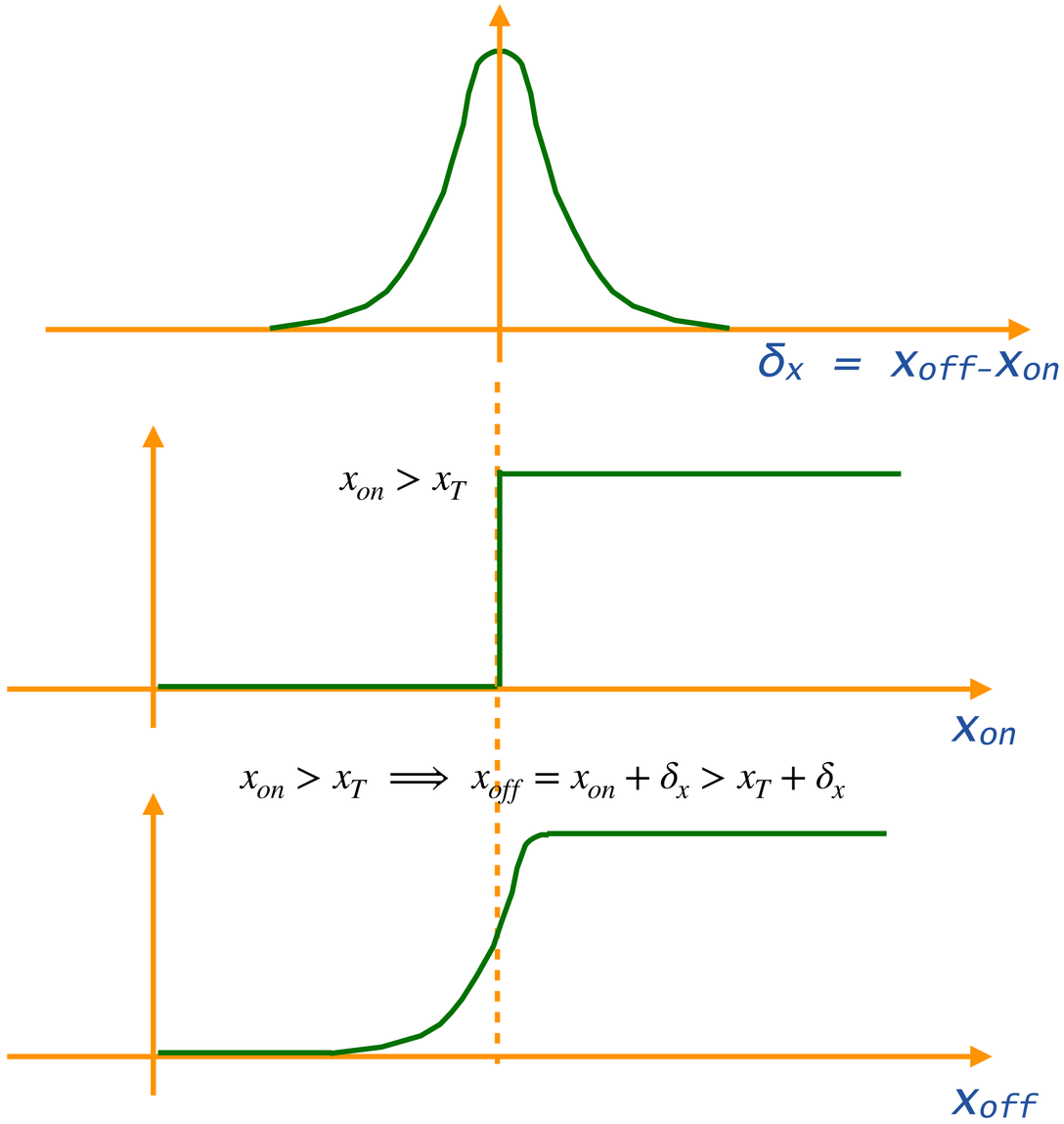
Accuracy vs Speed

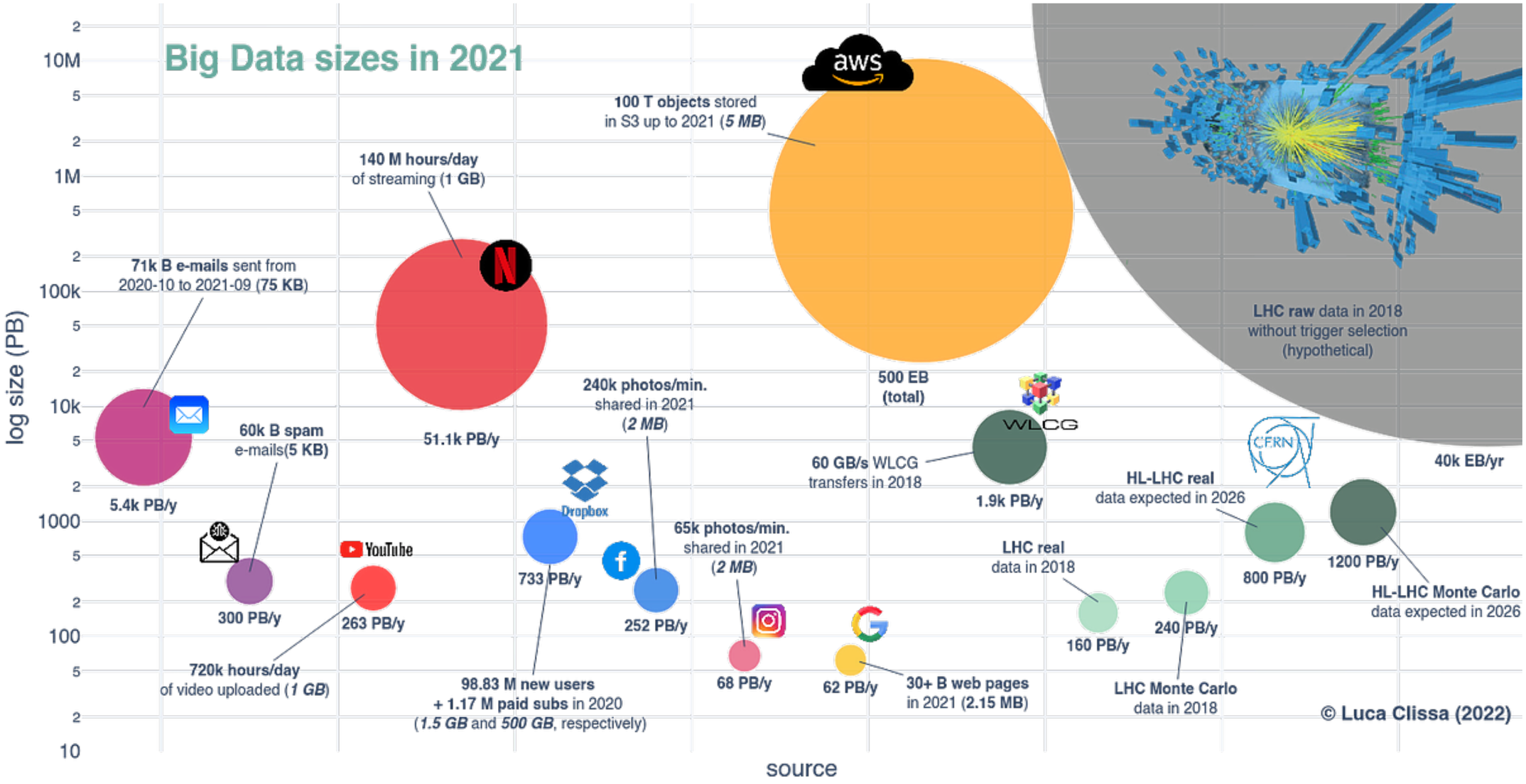
- *When applying a selection, one pays for the limited detector resolution*
  - *The amount of background leaking in the selected region is inflated by poor resolution*
- *When working in real time one has limited resources*
  - *limited computing power*
  - *little time to run complex calculations*
- *Detector performance are not exploited at best*
  - *coarser algorithms, limited input information (e.g., no particle tracking)*



Turn-on curve

- You have a discriminating quantity x and two estimates of it
  - x_{off}: an accurate measurements of x
  - x_{on}: a coarser measurement of x
- A turn-on curve models how the distribution of x_{off} is affected by a cut on x_{on}
- A typical analysis would work beyond the turn-on
  - constant efficiency loss with measurable uncertainty
  - some analysis could try to model efficiency (ε) along the curve and reweigh events by 1/ε

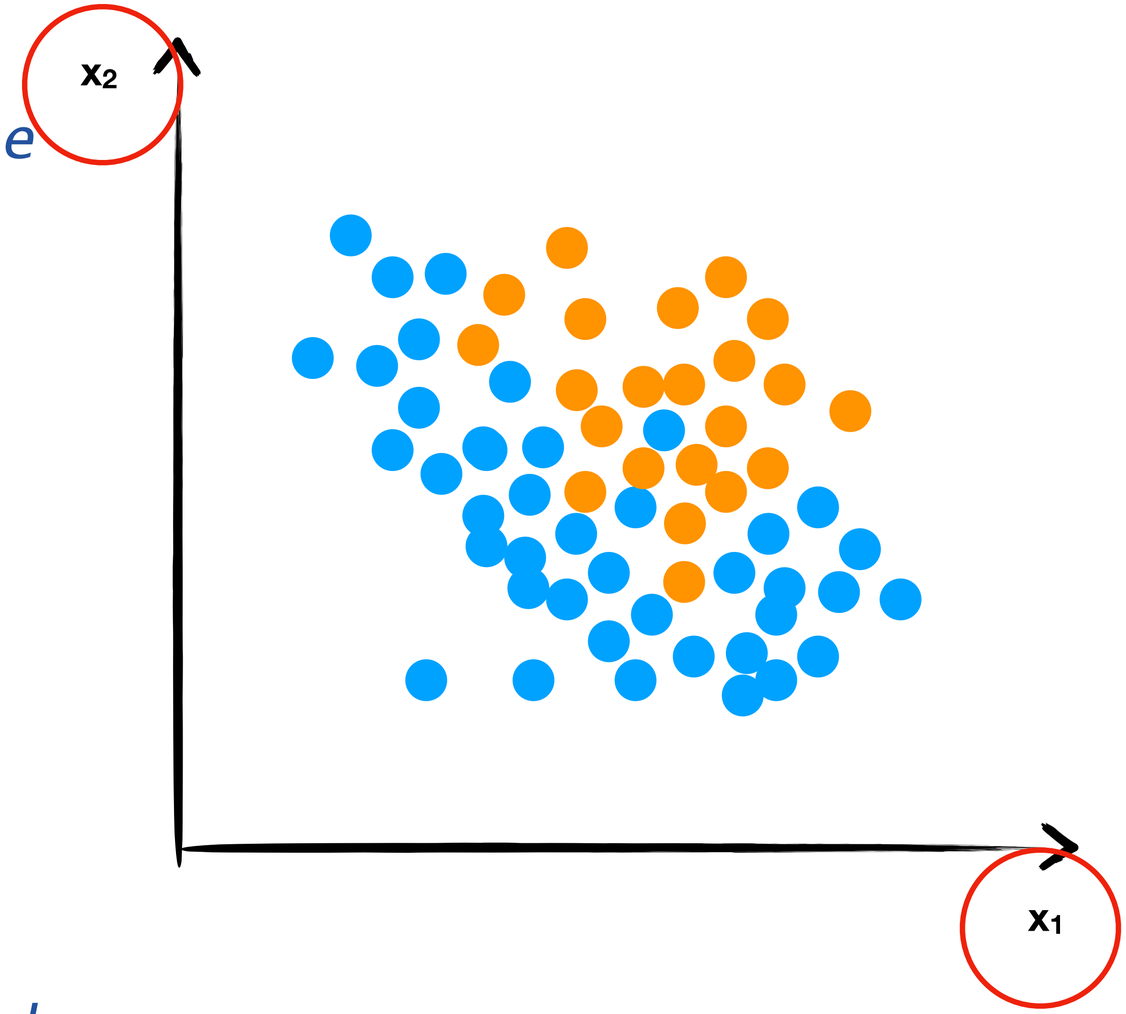




STEP 2: data selection

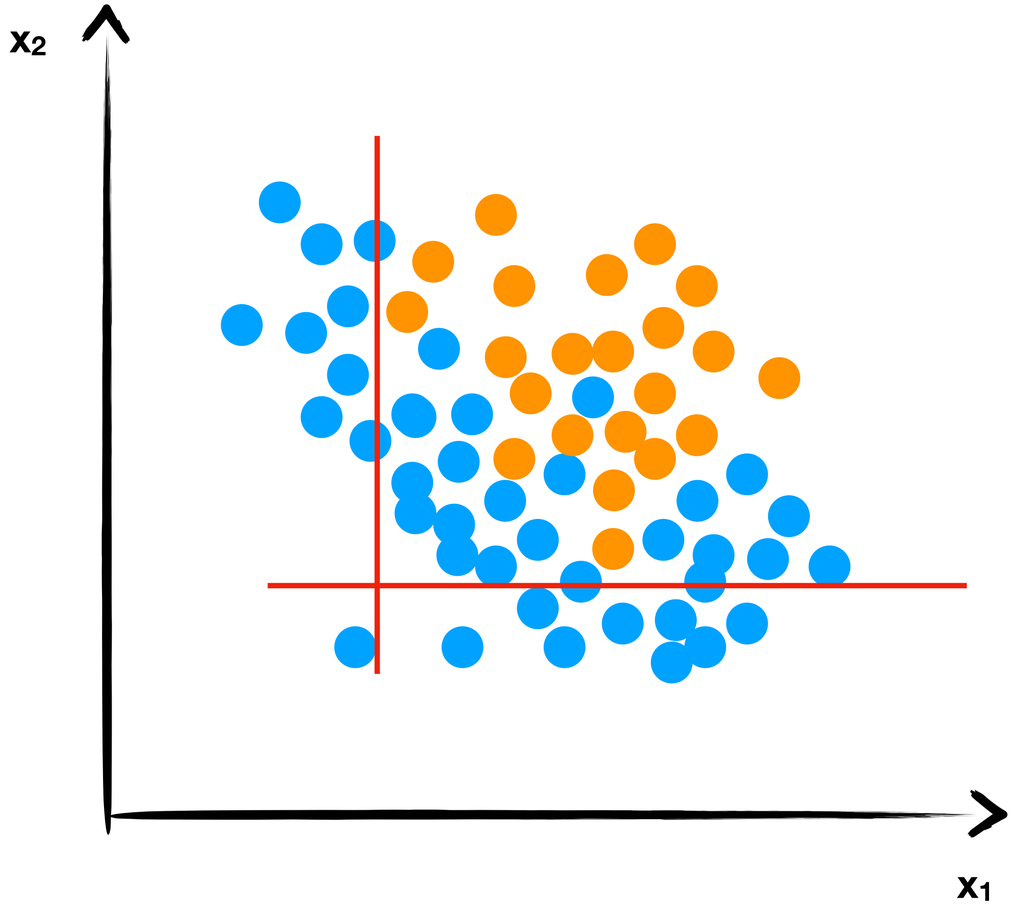
What you need to decide

- **What to cut on**
  - S vs B separation depends on the discriminating quantity you use
  - Depending on the quantity on axis, selection can be more or less efficient
- How to cut
  - linear vs non-linear cuts
- Where to cut
  - trade-off between efficiency and purity



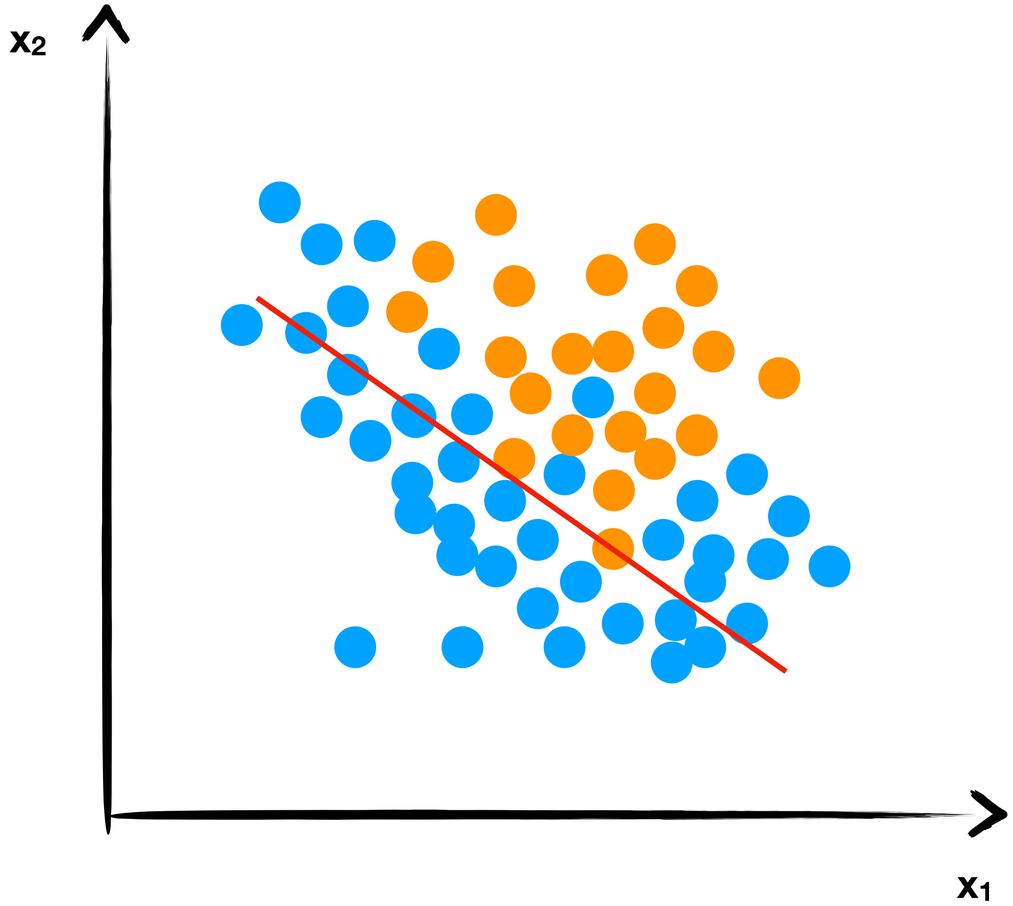
What you need to decide

- *What to cut on*
  - *S vs B separation depends on the discriminating quantity you use*
  - *Depending on the quantity on axis, selection can be more or less efficient*
- ***How to cut***
  - *linear vs non-linear cuts*
- *Where to cut*
  - *trade-off between efficiency and purity*



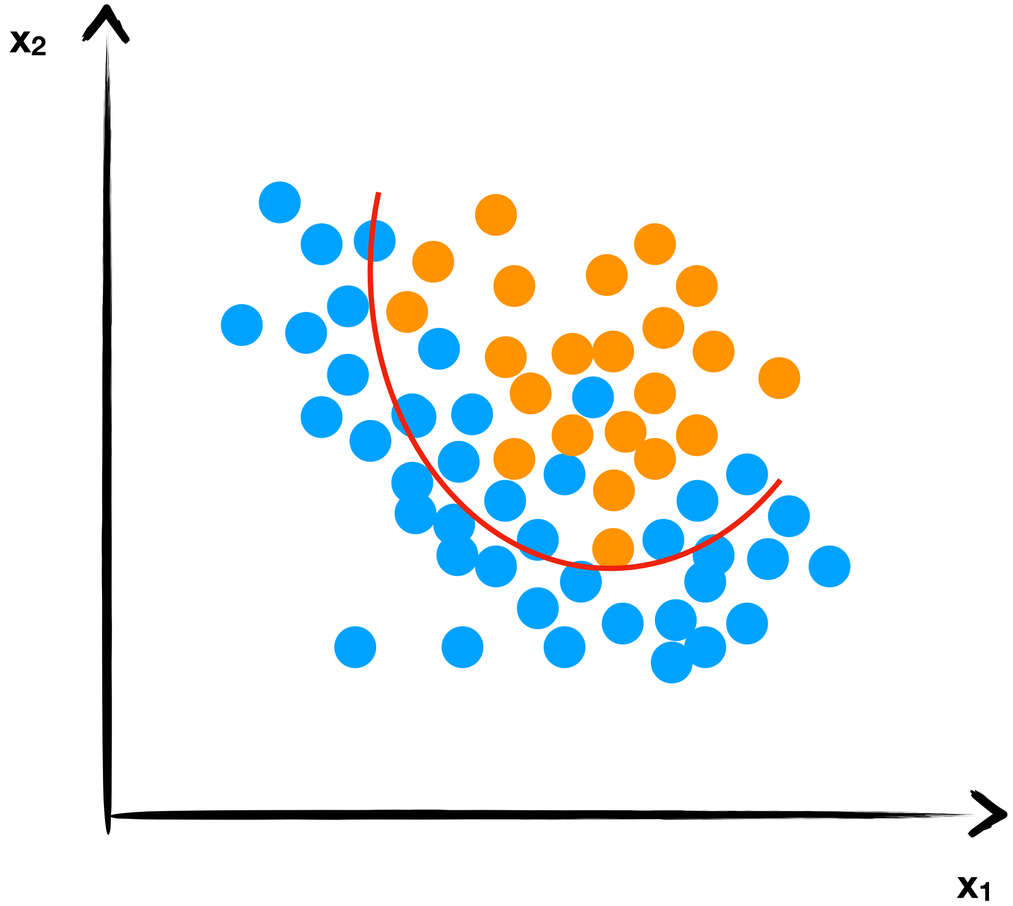
What you need to decide

- *What to cut on*
  - *S vs B separation depends on the discriminating quantity you use*
  - *Depending on the quantity on axis, selection can be more or less efficient*
- ***How to cut***
  - *linear vs non-linear cuts*
- *Where to cut*
  - *trade-off between efficiency and purity*



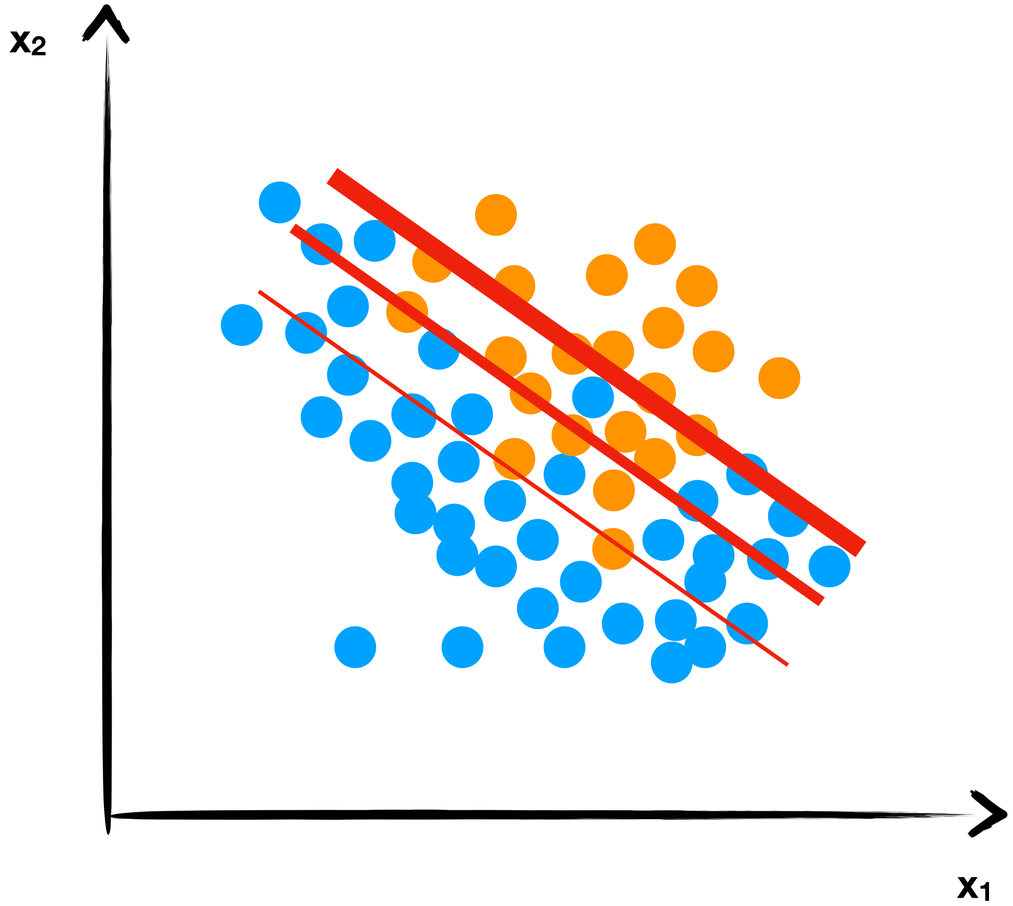
What you need to decide

- *What to cut on*
  - *S vs B separation depends on the discriminating quantity you use*
  - *Depending on the quantity on axis, selection can be more or less efficient*
- ***How to cut***
  - *linear vs non-linear cuts*
- *Where to cut*
  - *trade-off between efficiency and purity*



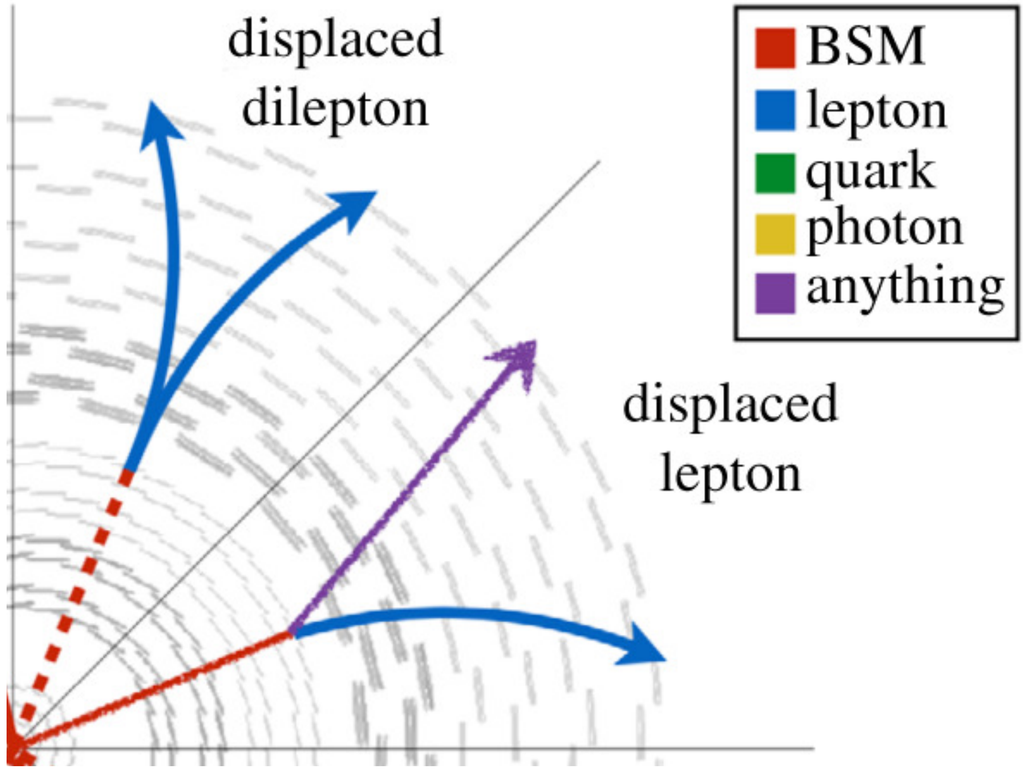
What you need to decide

- *What to cut on*
  - *S vs B separation depends on the discriminating quantity you use*
  - *Depending on the quantity on axis, selection can be more or less efficient*
- *How to cut*
  - *linear vs non-linear cuts*
- ***Where to cut***
  - *trade-off between efficiency and purity*



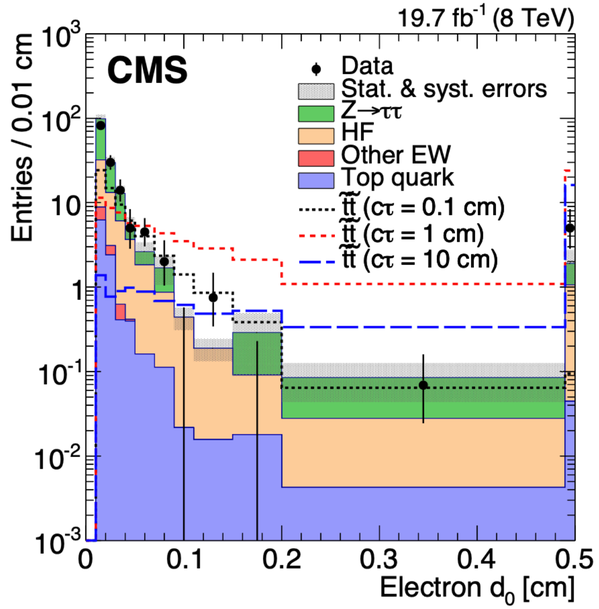
What to cut on

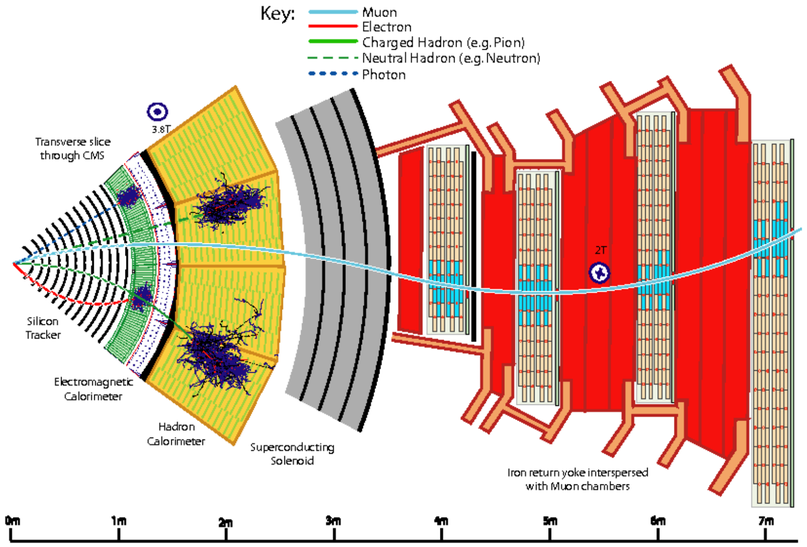
- *An event selection is a multi-dimensional problem*
  - *Several quantities can discriminate S vs B for different reasons*
- ***Example:***
  - *Signal: long-lived particles at LHC decaying to electrons*
  - *Signature: electrons displaced from collision point*
  - *Backgrounds*
    - *Events with real electrons from SM processes*
    - *Fake events (random association of a track to a calorimeter deposit)*

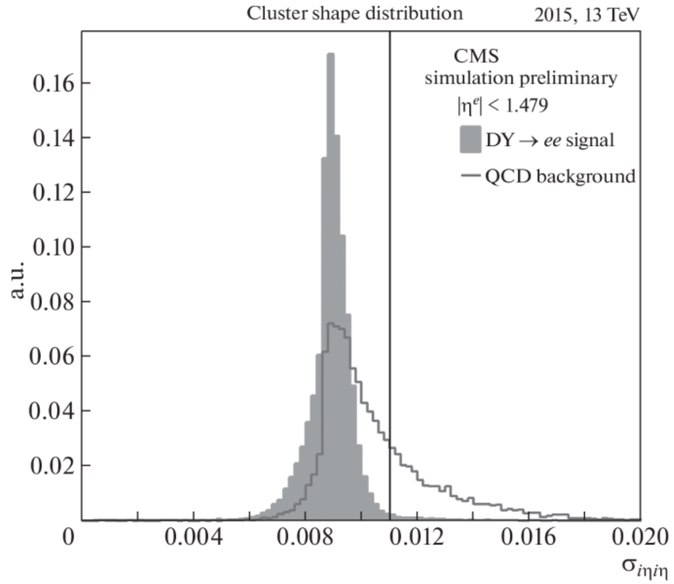


What to cut on

- *This is the easiest case*
  - *one can define two quantities which are (to a large extent) independent*
    - *Track displacement from primary vertex typically relies on the tracker*
    - *An electron ID score uses the calorimeter information*

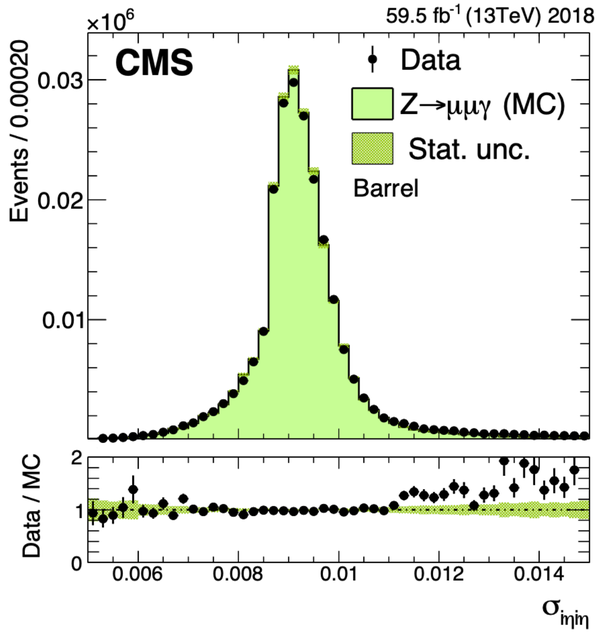


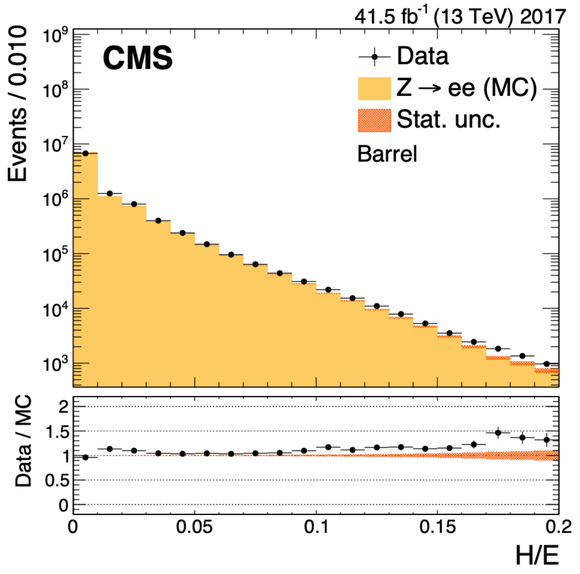




1-Dim vs N-Dim cut

- *Real problem and state-of-art solutions are typically more complicated*
  - *There is no magic quantity that gives optimal separation (but you can try to build one)*
  - *Several quantities can be defined, based on same inputs and correlated*
  - *Just using the quantity with best discrimination might not be optimal*
  - *Using more quantities can improve separation*

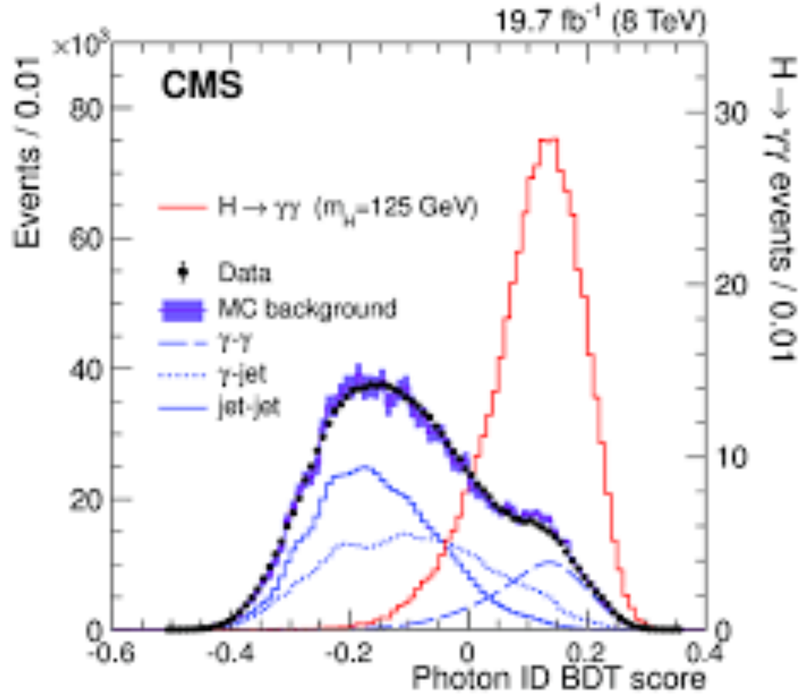




Different approaches

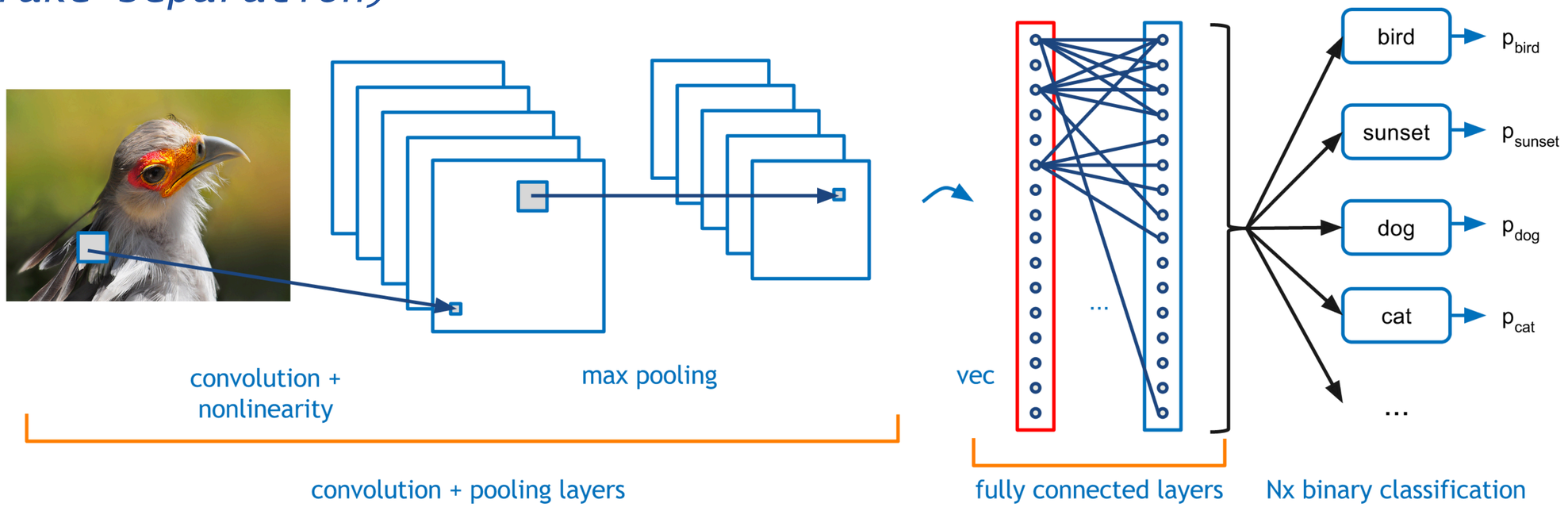
- *Cut-based selection*
  - *Select a portion of the N-dim space, through a set of cuts on each quantity*
- *MVA-based selection*
  - *Combine quantities in a single discriminator*
    - *N-dim likelihood (when correlations are known)*
    - *Machine learning e.g., BDT, NN, etc (when they are not)*

Variable	Barrel (tight WP)	Endcaps (tight WP)
$\sigma_{i\eta i\eta}$	<0.010	<0.035
$\lvert\Delta\eta_{\text{in}}^{\text{seed}}\rvert$	<0.0025	<0.005
$\lvert\Delta\phi_{\text{in}}\rvert$	<0.022 rad	<0.024 rad
H/E	$<0.026 + 1.15\,\text{GeV}/E_{\text{SC}}$ $+0.032\rho/E_{\text{SC}}$	$<0.019 + 2.06\,\text{GeV}/E_{\text{SC}}$ $+0.183\rho/E_{\text{SC}}$
$I_{\text{combined}}/E_{\text{T}}$	$<0.029 + 0.51\,\text{GeV}/E_{\text{T}}$	$<0.0445 + 0.963\,\text{GeV}/E_{\text{T}}$
$\lvert 1/E - 1/p\rvert$	$<0.16\,\text{GeV}^{-1}$	$<0.0197\,\text{GeV}^{-1}$
Number of missing hits	≤ 1	≤ 1
Pass conversion veto	Yes	Yes



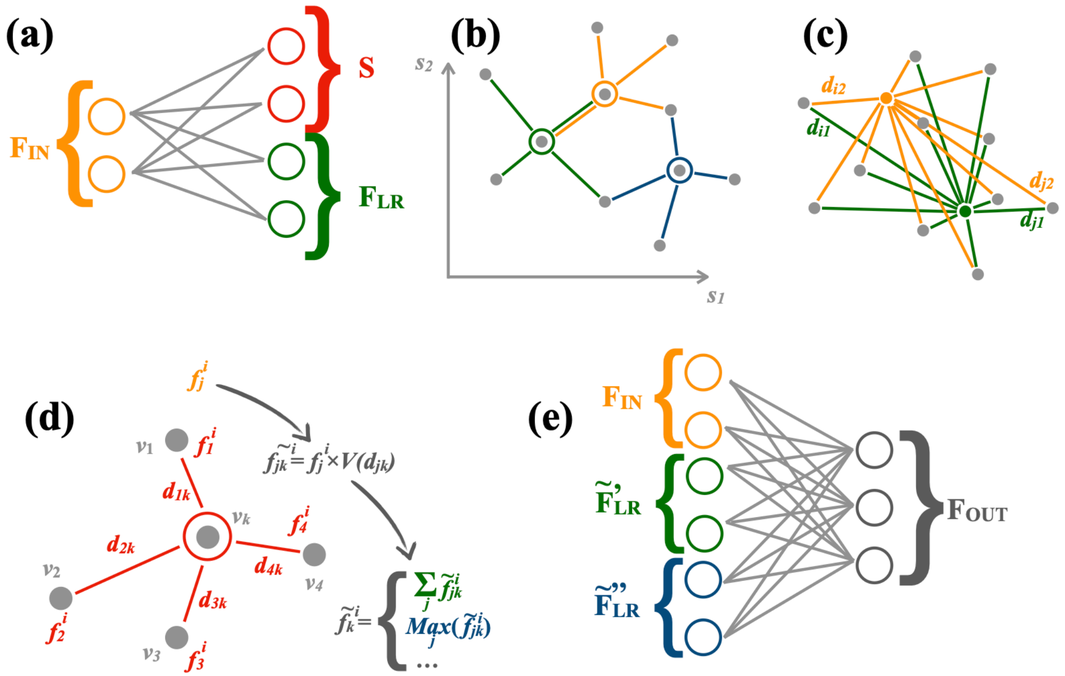
A new approach: end-to-end

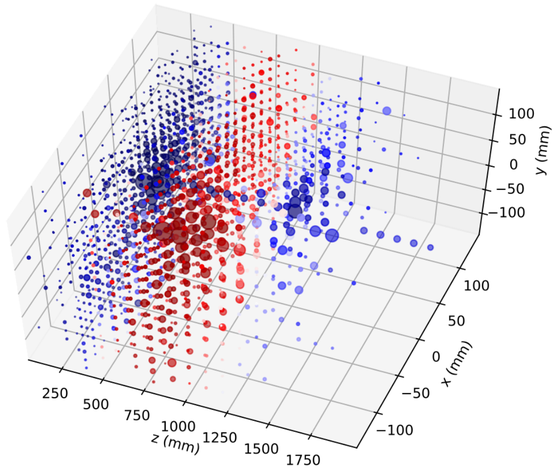
- *Traditional approaches exploits high-level features (hlf) built based on physics intuition*
- *With Deep learning, it is now possible to start from raw data and engineer high-level quantities*
- *The hlf definition is optimized together with the task (electron vs fake separation)*



A new approach: end-to-end

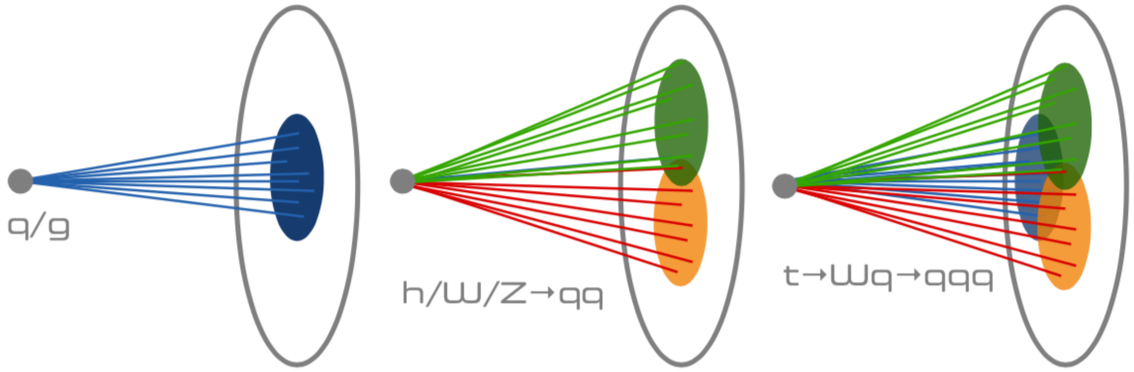
- *Multi-task problem with single train*
  - *cluster energy deposits into photon/electron candidates*
  - *build hlf quantities from the cluster*
  - *maximise the separation between signal and background*

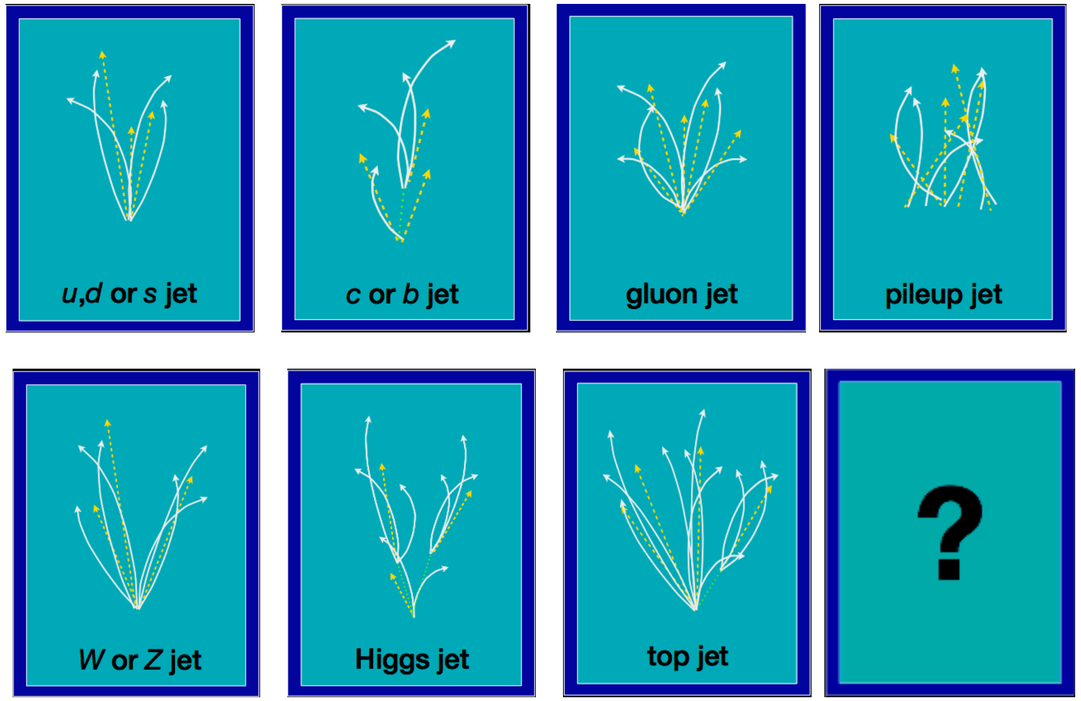




How to cut

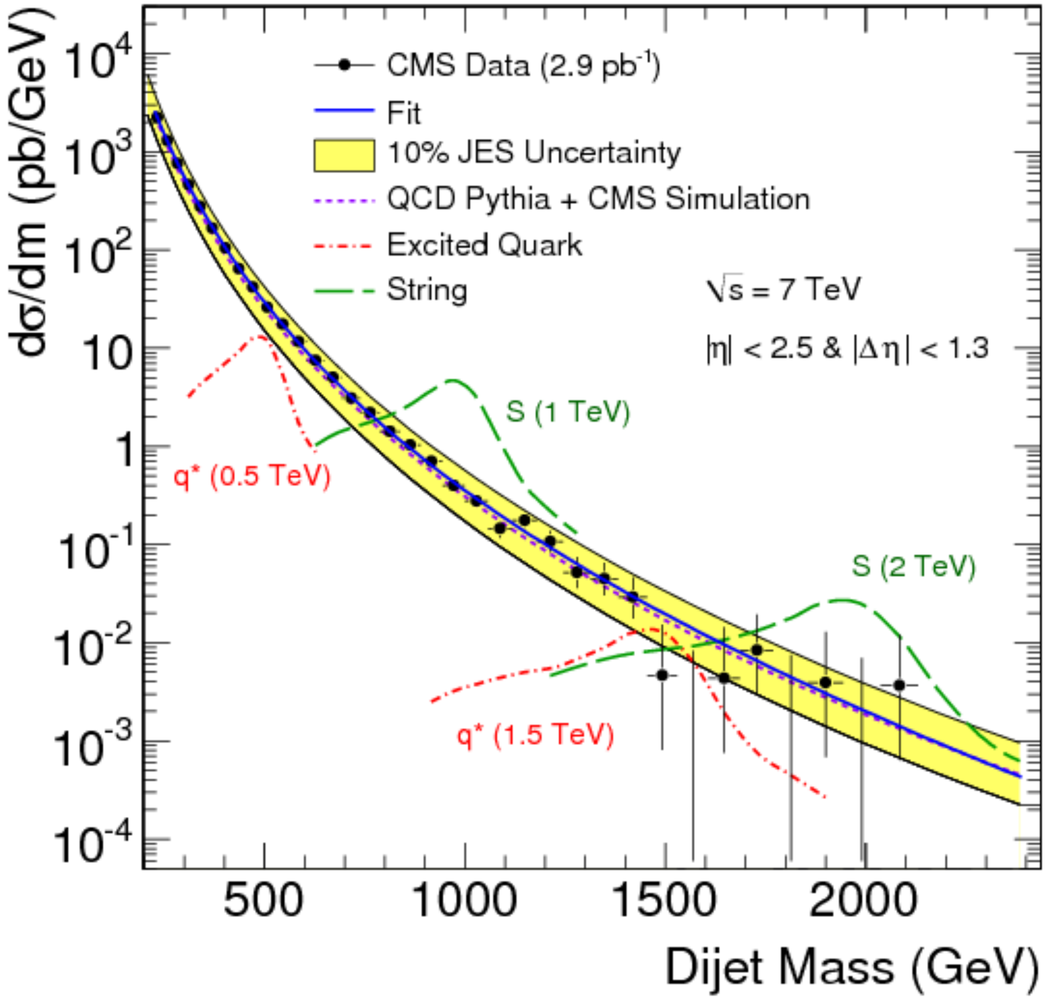
- *Typically, one puts a "flat cut" on each quantity*
  - *This quantity could be a physics-inspired function*
  - *Or an MVA score*
- *When doing so, one has to take care of the impact this has on other quantities*
- *Example:*
  - *jet tagging: identify which kind of particle started a jet*
  - *Jet mass can do it, but the jet substructure provides extra information*
  - *Usually one b builds an MVA, e.g., a Neural Network*





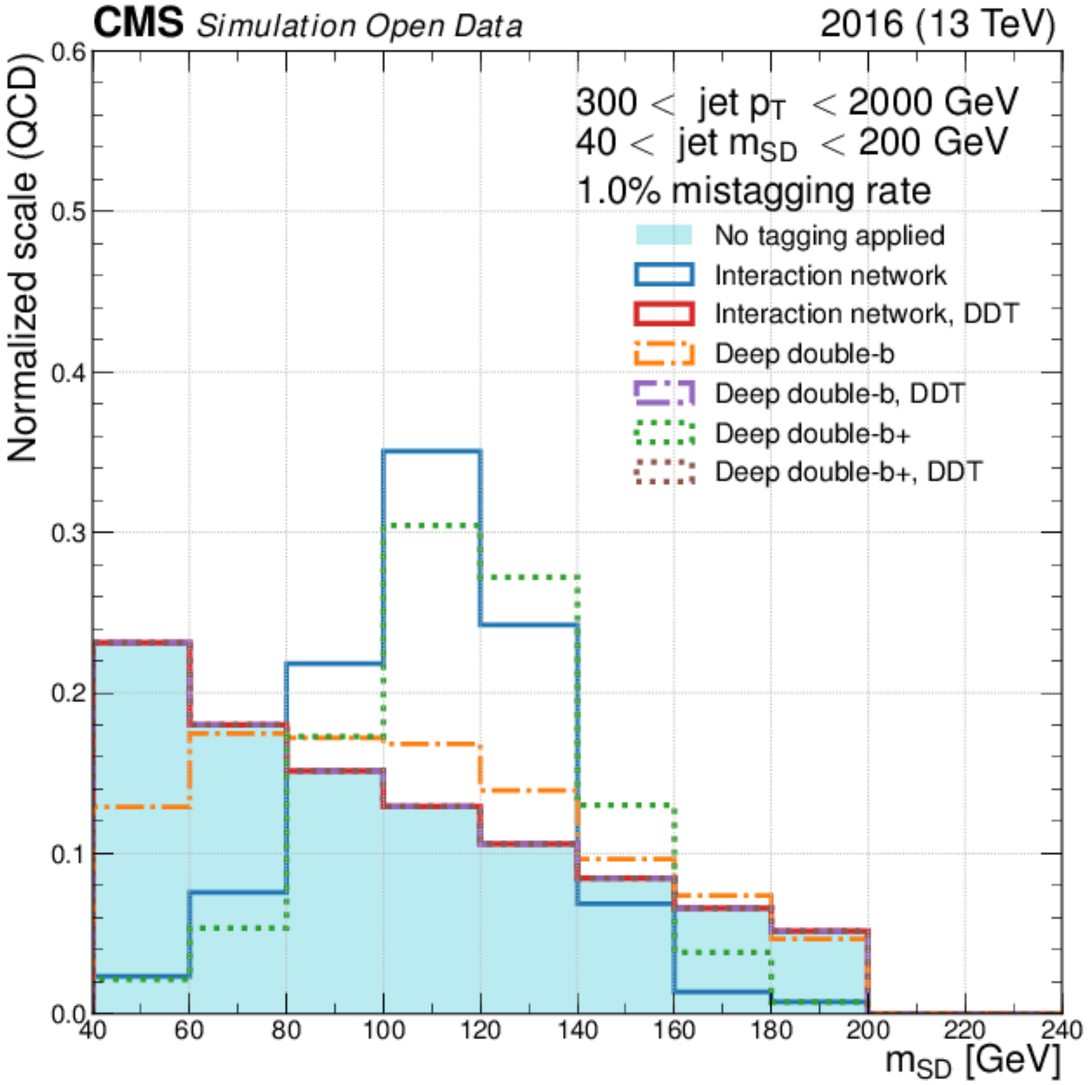
A dijet resonance search

- *Typically, one puts a "flat cut" on each quantity*
  - *This quantity could be a physics-inspired function*
  - *Or an MVA score*
- *When doing so, one has to take care of the impact this has on other quantities*
- *Example:*
  - *jet tagging: identify which kind of particle started a jet*
  - *Jet mass can do it, but the jet substructure provides extra information*
  - *Usually one b builds an MVA, e.g., a Neural Network*



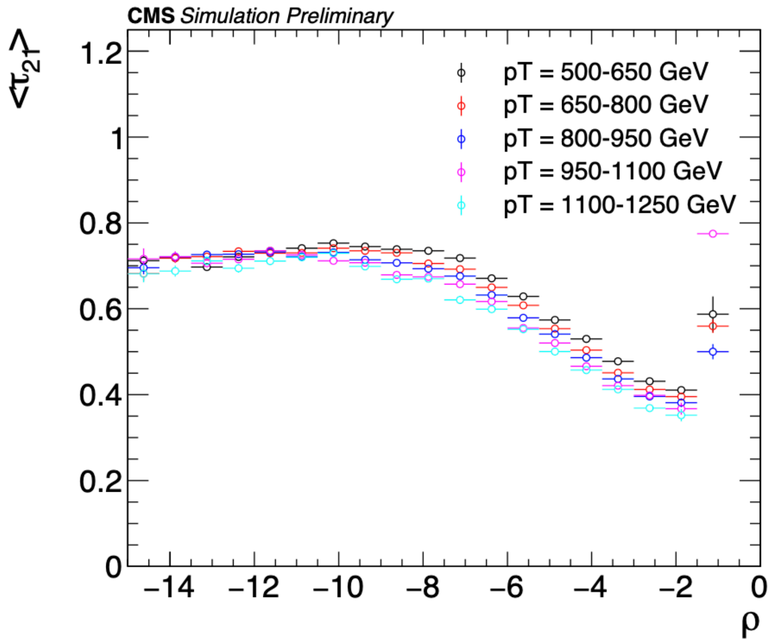
Mass sculpting

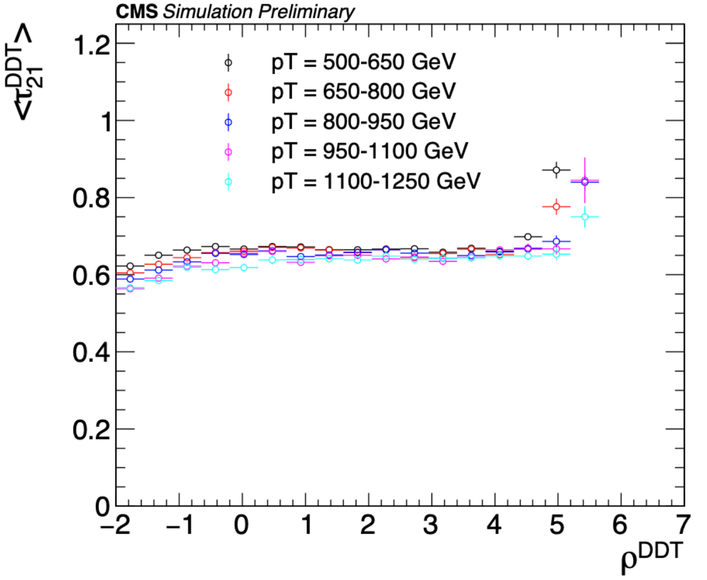
- *Very often, a NN learns information that is exploited in physics inspired quantities*
  - *A cut on the MVA score can learn kinematic and alert it*
  - *This can affect background distribution. For instance, a jet ID cut can change the distribution of dijet mass, creating a bump in the background*
  - *This can make a search for a new resonance more complicated*



Designed decorrelated taggers

- *The problem is visible even w/o machine learning*
  - *Jet substructure before NNs was done with physics-inspired jet substructure*
  - *Jet substructure shows a correlation with* $\rho = \log(m^2/p_T^2)$, *used in the search to extract the background*
  - *The correlation is critical in the regime of a typical search* $(\rho > -8)$
- *A two step approach*
  - *Most of the dependence is linear and can be corrected by hand, trading* ρ *for* $\rho^{DDT} = \log\left(\frac{m}{p_T \mu}\right)$ $(\mu = 1\ GeV)$
  - *The residual dependent is removed trading* τ_{21} *for* $\tau_{21}^{DDT} = \tau_{21} - k \times \rho$
- *At that point, a flat cut on* τ_{21}^{DDT} *is applied*



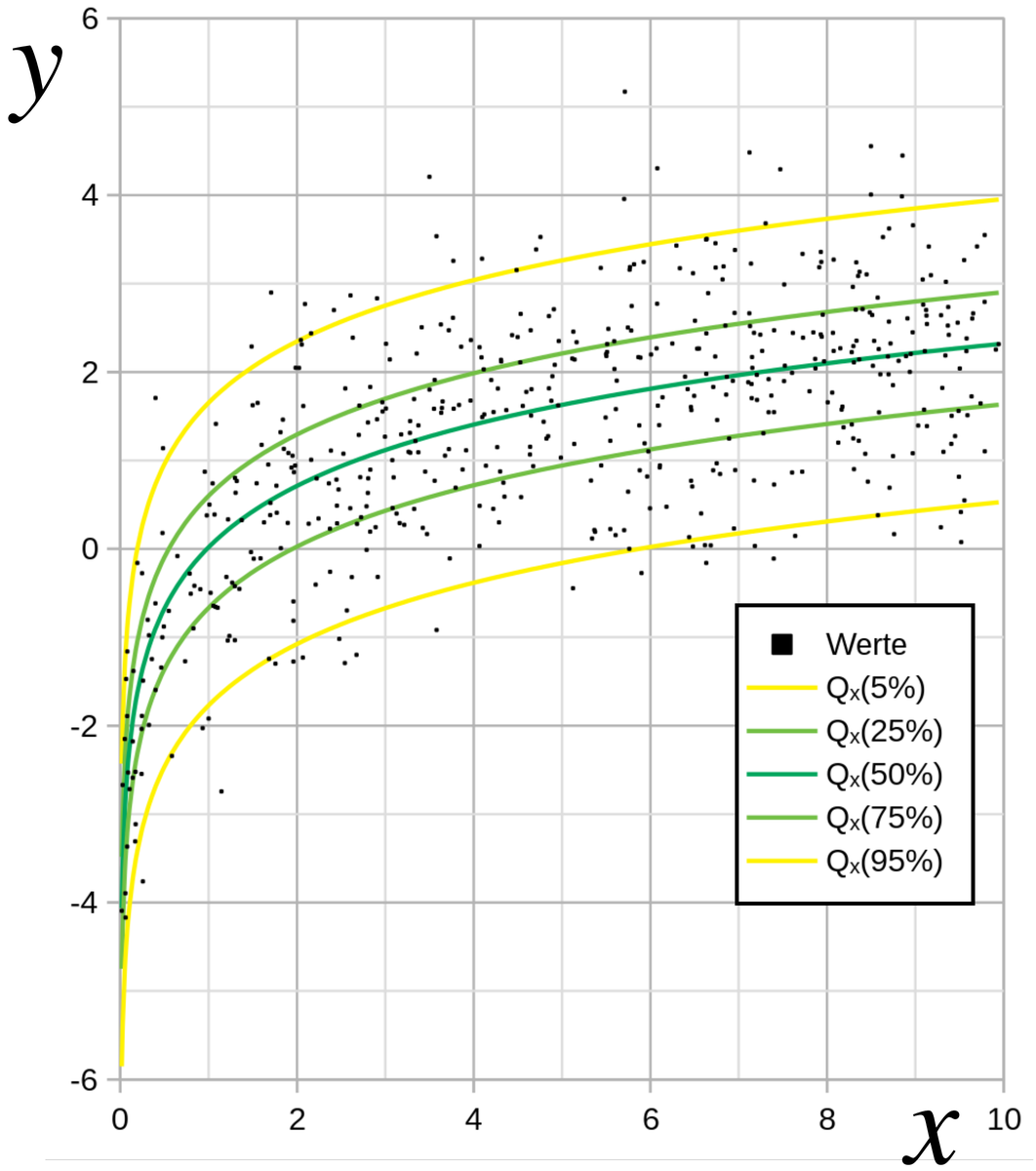


Quantile Regression

- *DDT approaches can be generalized to non-linear dependencies using a learnable non-linear function*
  - *One can use machine learning to find the contour defined by* $y > y_T(x)$
  - $y_T(x)$ *is learned through a neural network taking* x *as input*
  - *The training is performed using a specific loss function*

$$L(y_i^p, y_i) = max[q(y_i - y_i^p),\ (q-1)(y_i - y_i^p)]$$

- *N-dim generalization: The big advantage of this approach (as usual with neural networks) is that* x *can actually be a vector of input quantities*

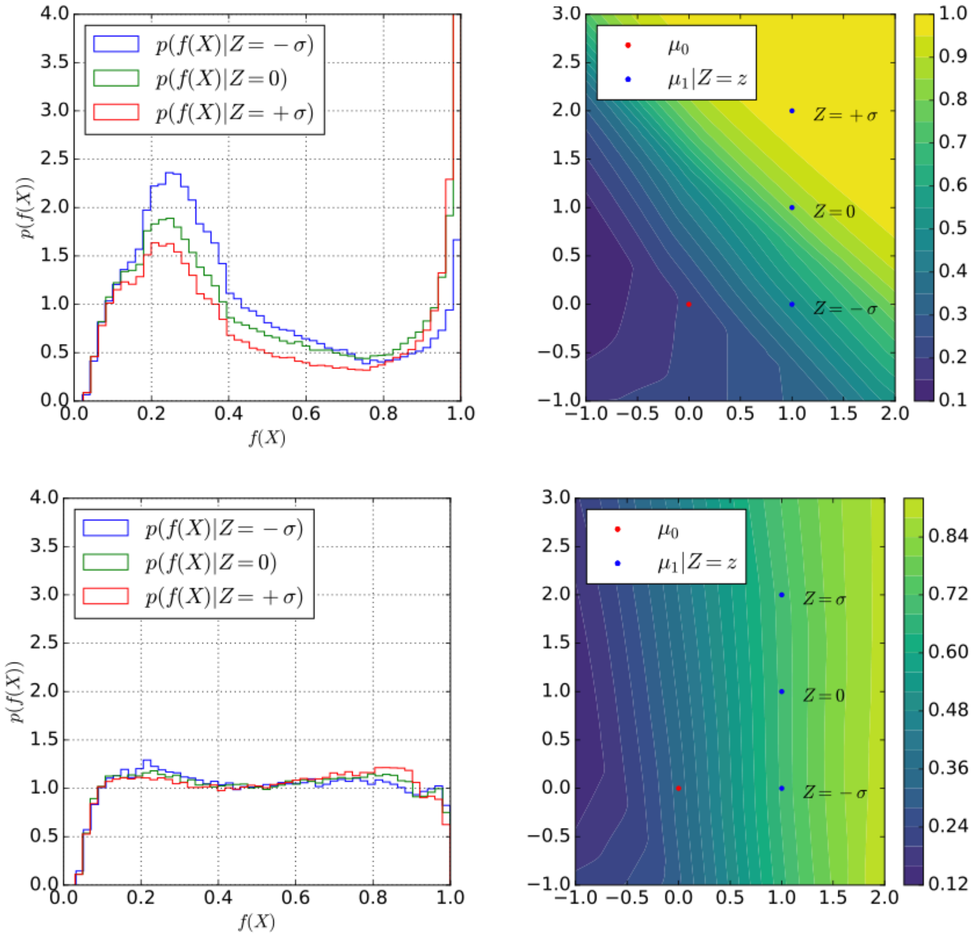


Adversarial learning

- *When your quantity y is a trained algorithm you have a further option wrt quantile regression, DDT etc.*
  - *You can prevent y from learning x through an adversarial learning at training time, minimizing the correlation between x and y while constructing y*
- *BE CAREFUL:*
  - *this might alleviate the problem but not remove it*
  - *Adversarial trainings can be tricky: two terms of the likelihood fighting against each other could introduce a training instability*

$$E(\theta_f, \theta_r) = \mathcal{L}_f(\theta_f) - \mathcal{L}_r(\theta_f, \theta_r)$$

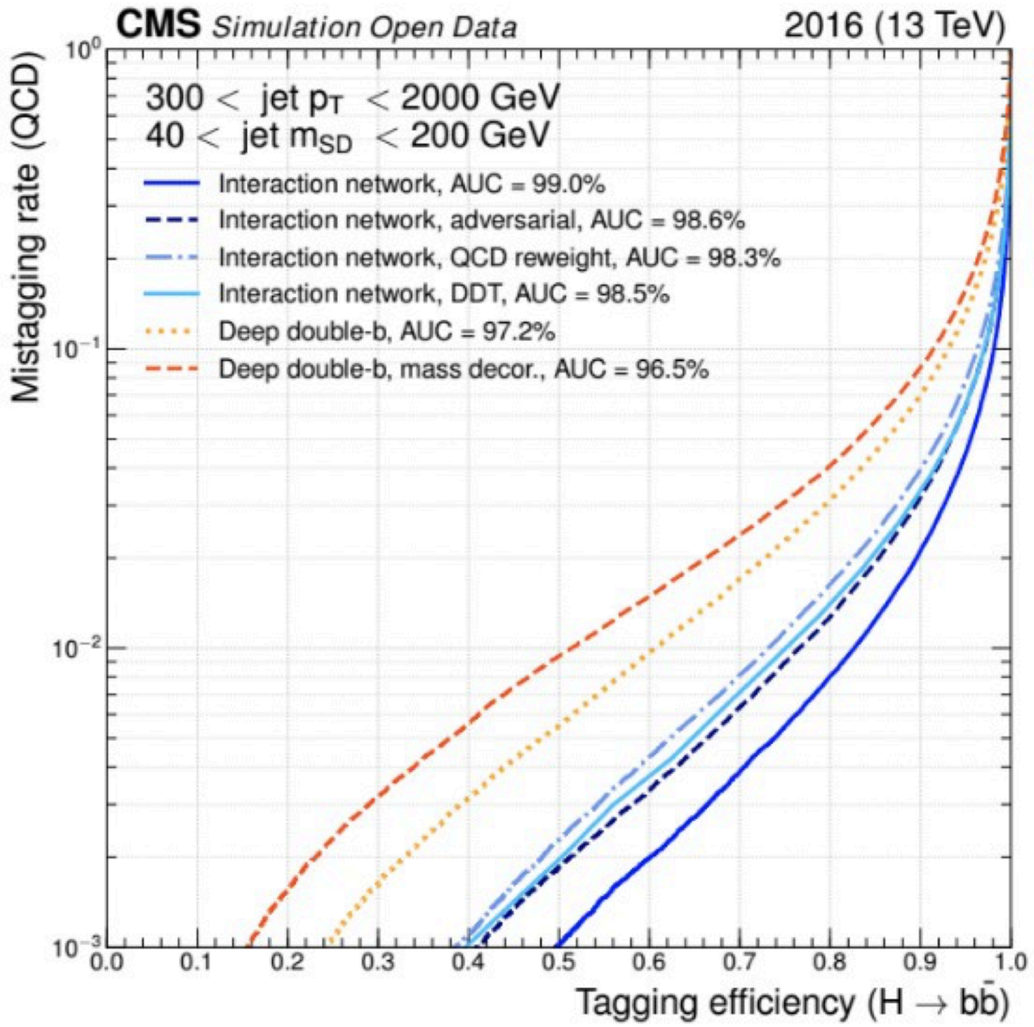
$$\hat{\theta}_f, \hat{\theta}_r = \arg\min_{\theta_f} \max_{\theta_r} E(\theta_f, \theta_r)$$



G. Loupe et al. https://arxiv.org/pdf/1611.01046.pdf

Adversarial learning

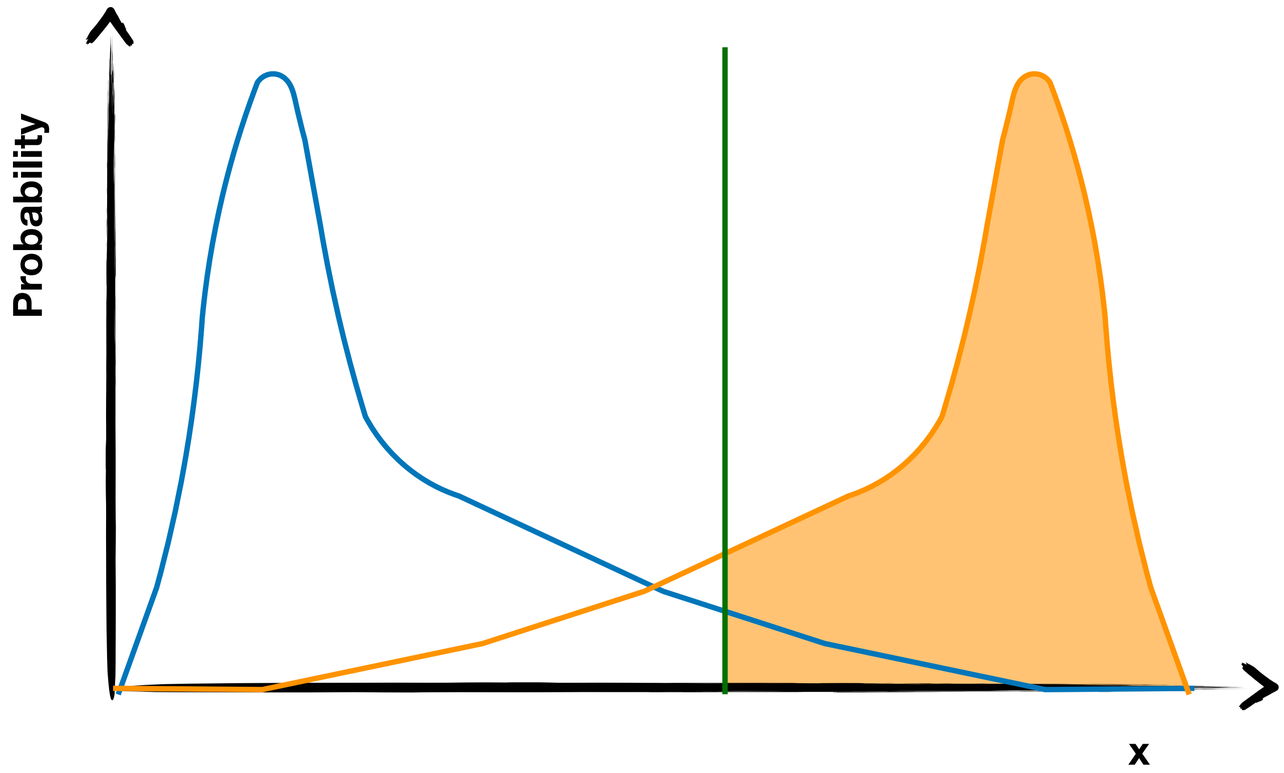
- *When your quantity y is a trained algorithm you have a further option wrt quantile regression, DDT etc.*
  - *You can prevent y from learning x through an adversarial learning at training time, minimizing the correlation between x and y while constructing y*
- *BE CAREFUL:*
  - *this might alleviate the problem but not remove it*
  - *Adversarial trainings can be tricky: two terms of the likelihood fighting against each other could introduce a training instability*



E. Moreno et al. https://arxiv.org/abs/1909.12285

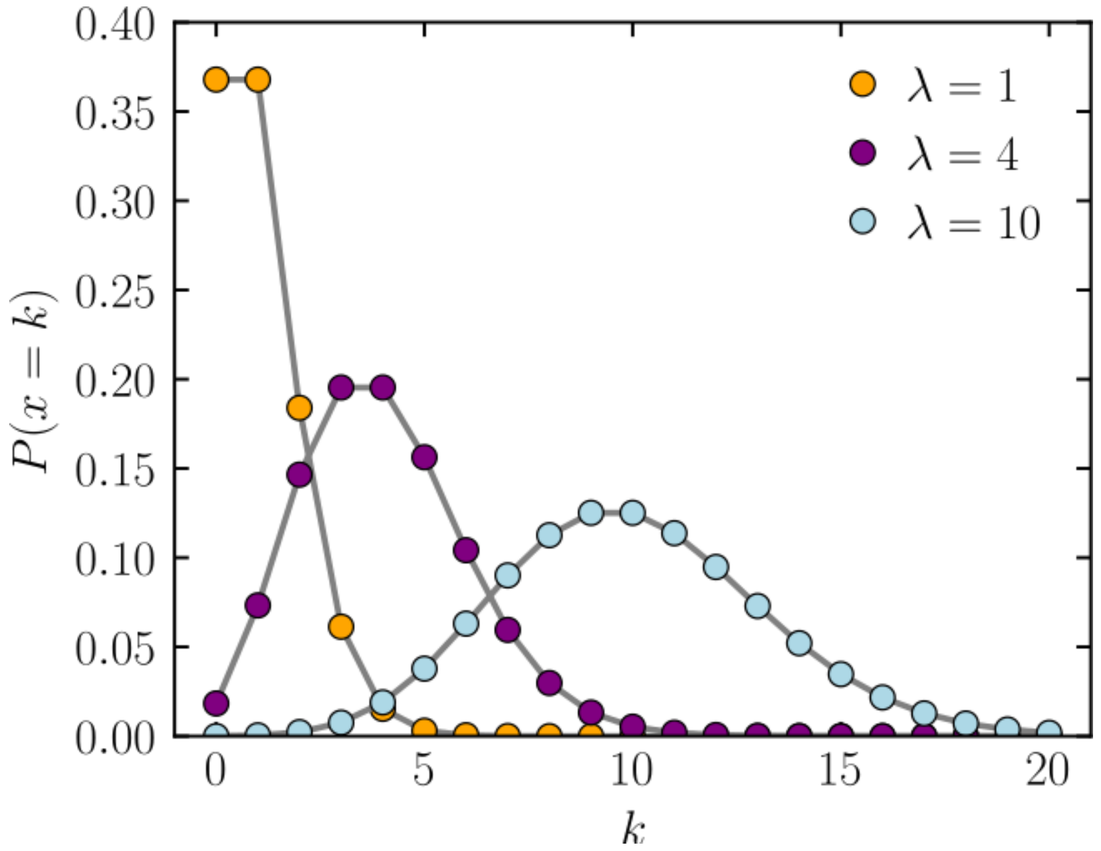
Where to cut

- *The cut threshold is chosen maximizing some figure of merit of purity.*
  - *Usually people set cuts that maximizes $S/\sqrt{B}$, when optimizing for a discovery*
- *Where is this coming from?*



Counting Experiment

- Given a population of N data entries (events)
- Apply a cut with efficiency ε, one expect εN events surviving
- The number of observed events is distributed according to a Poisson distribution
  - centred at $S=\varepsilon_S N_S$ for signal
  - centred at $B=\varepsilon_B N_B$ for background
- The total number of events is centred around S+B
- A Poisson converges quickly to a Gaussian with mean λ and RMS $\sqrt{\lambda}$



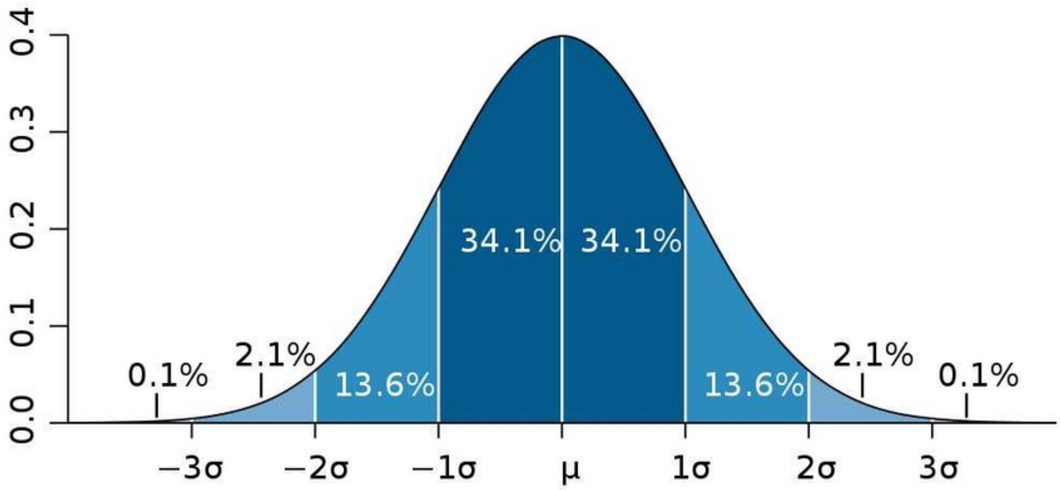
$$P(k|\lambda) = \frac{e^{-\lambda}\lambda^k}{k!}$$

$$E[k] = \lambda$$

$$Var[k] = \lambda$$

Number of sigmas

- When we talk about an observation, we quantify its strength in number of sigmas
  - The difference between the observed and expected yield, in units of the uncertainty
  - The uncertainty is that of a bkg-only distribution
  - One is minimizing the probability of a background-only distribution to mimic a signal
- Chances to discover are maximal when the signal would induced the largest possible excess wrt bkg-only distribution
  - Expected yield in presence of a signal: S+B
  - Expected yield in absence of a signal: B
  - $\sigma_B = \sqrt{Var[k_B]} = \sqrt{B}$ is the RMS in absence of a signal



$$\sharp\sigma = \frac{E[k_S + k_B] - E[k_B]}{\sqrt{Var[k_B]}} = \frac{S}{\sqrt{B}}$$

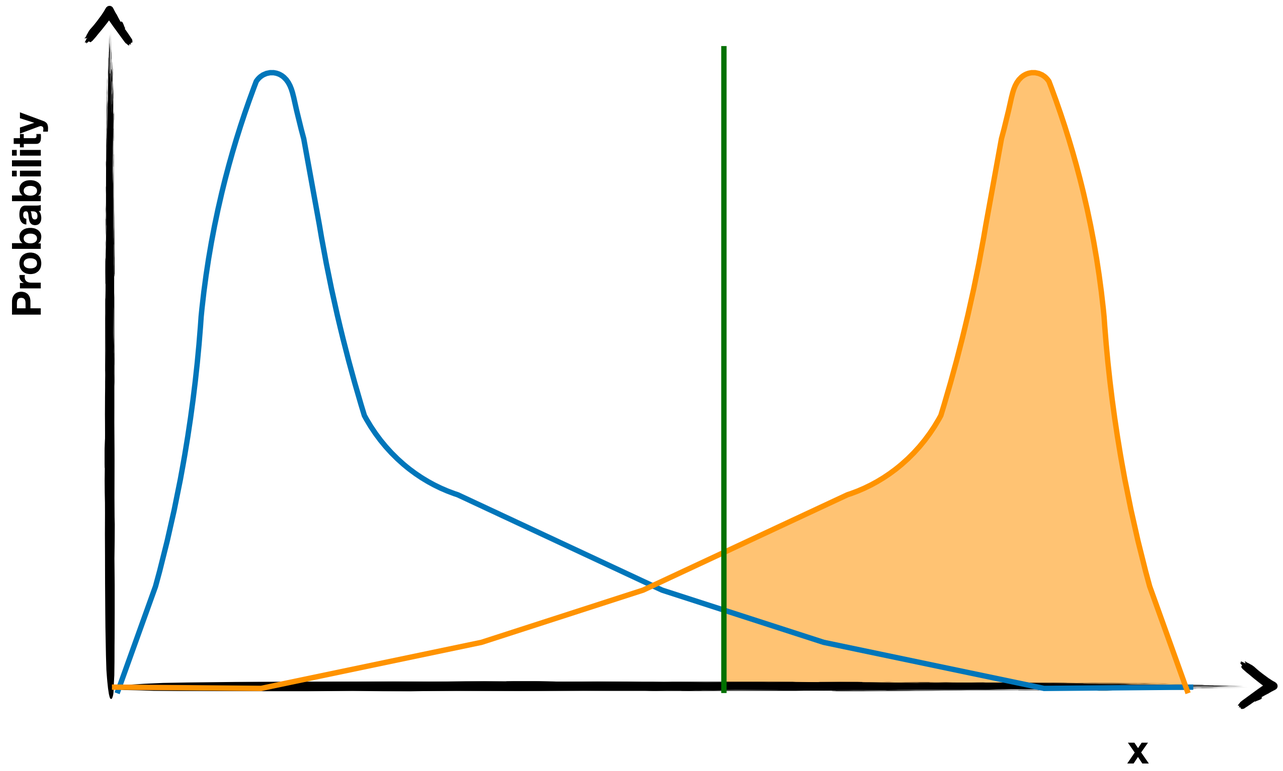
Two main issues:

- Bad behavior for small exerted background
- Computing the expected significance requires to specify a signal cross section (but the optimal can be found w/o)

G. Punzi, arXiv:0308063

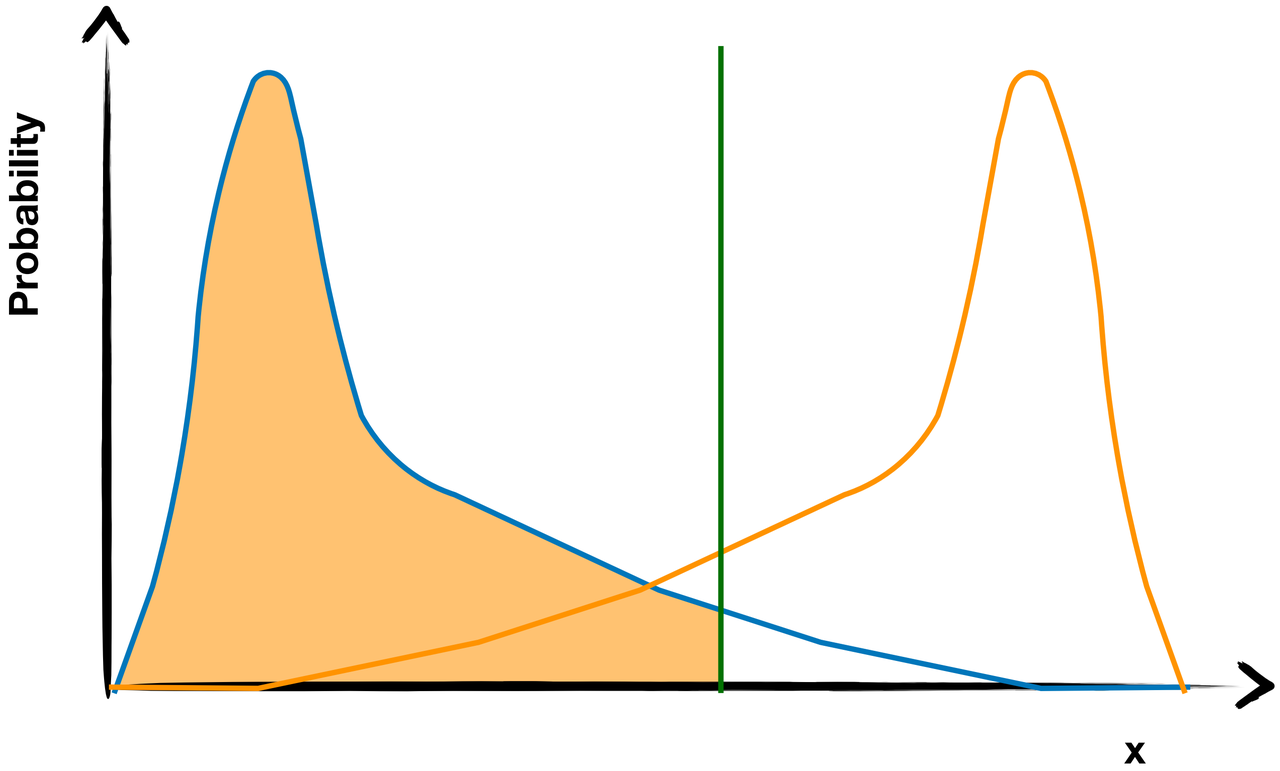
Data Science Selection metrics

- *A given threefold defines the following qualities*
  - *True-positives: Class-1 events above the threshold*
  - *True-negatives: Class-0 events below the threshold*
  - *False-positives: Class-0 events above the threshold*
  - *False-negatives: Class-1 events below the threshold*



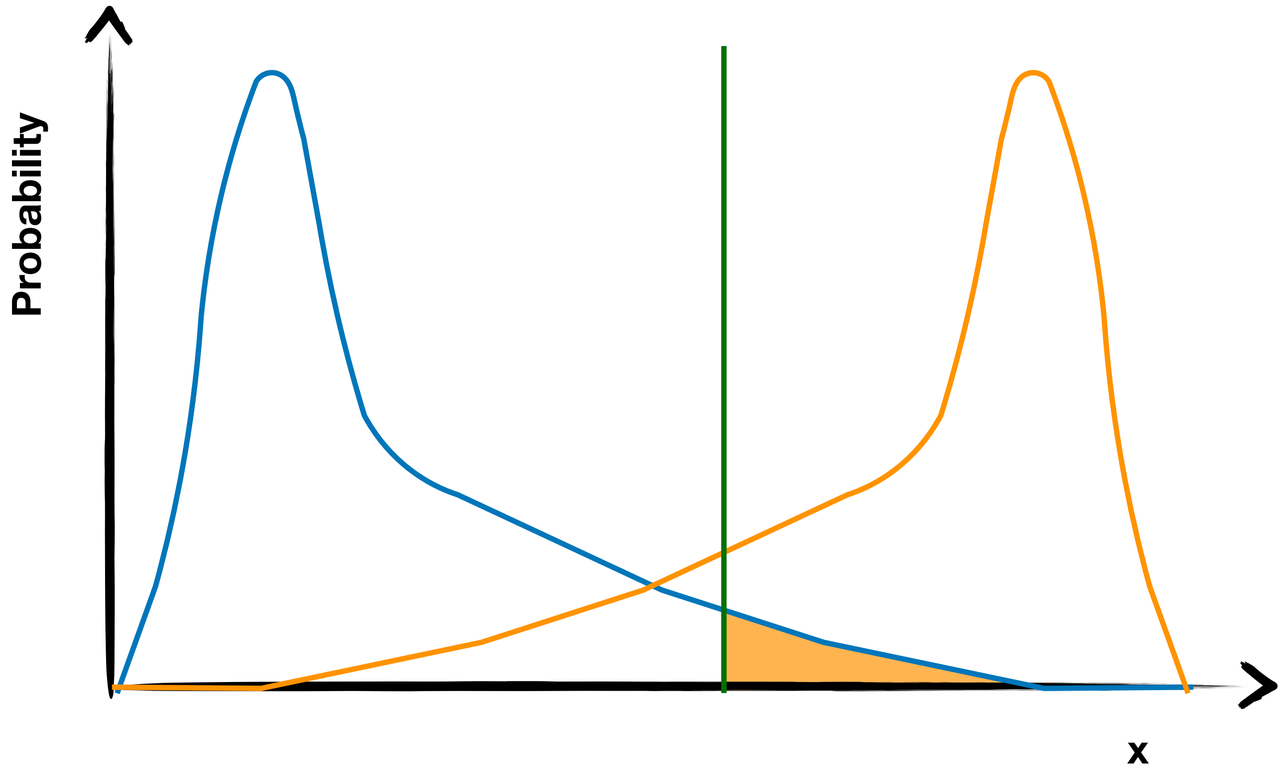
Data Science Selection metrics

- *A given threefold defines the following qualities*
  - *True-positives: Class-1 events above the threshold*
  - *True-negatives: Class-0 events below the threshold*
  - *False-positives: Class-0 events above the threshold*
  - *False-negatives: Class-1 events below the threshold*



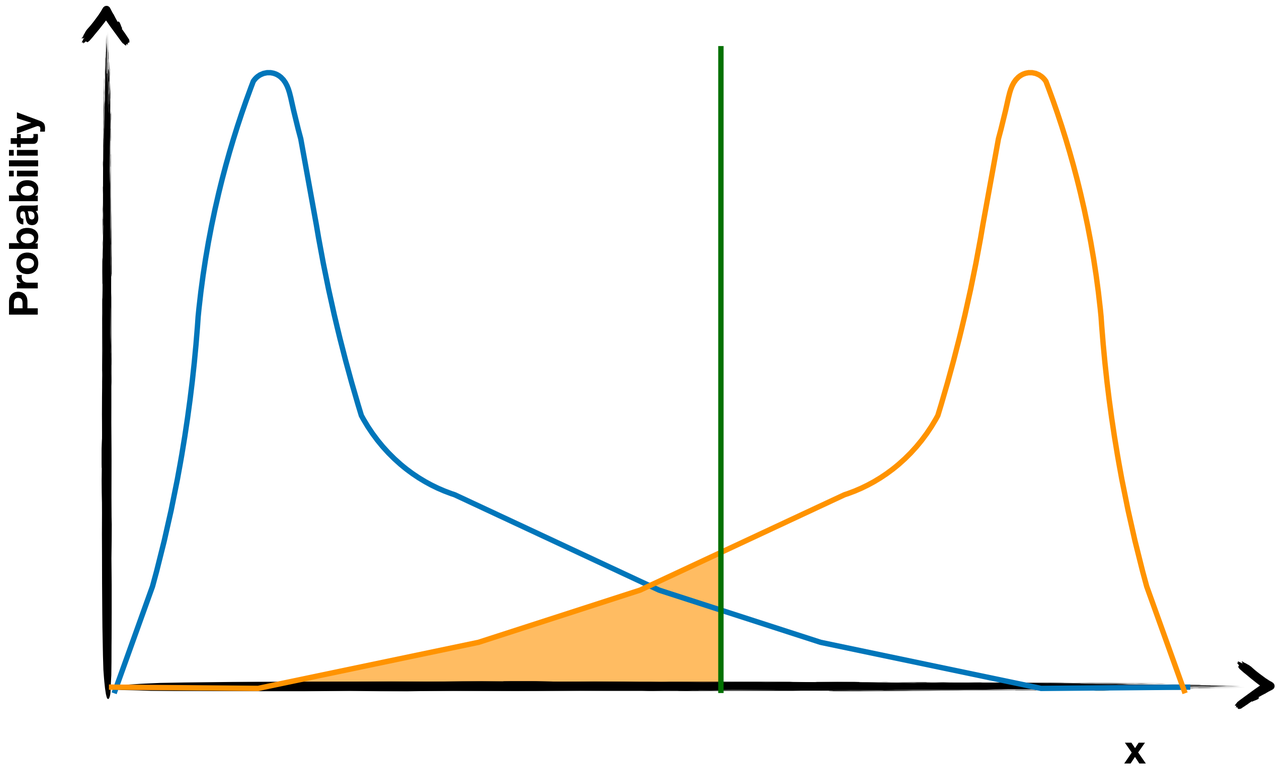
Data Science Selection metrics

- *A given threefold defines the following qualities*
  - *True-positives: Class-1 events above the threshold*
  - *True-negatives: Class-0 events below the threshold*
  - *False-positives: Class-0 events above the threshold*
  - *False-negatives: Class-1 events below the threshold*



Data Science Selection metrics

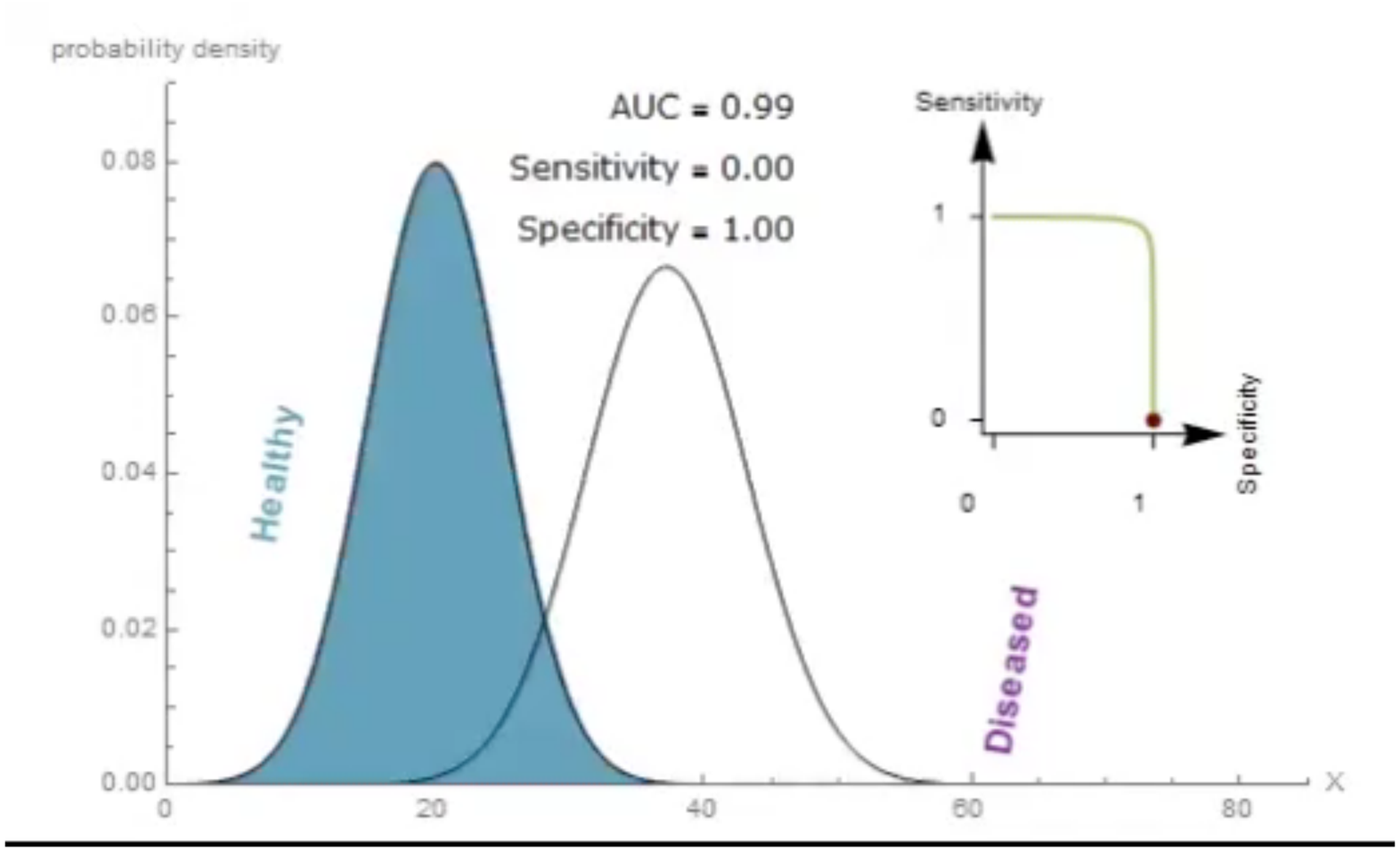
- *A given threefold defines the following qualities*
  - *True-positives: Class-1 events above the threshold*
  - *True-negatives: Class-0 events below the threshold*
  - *False-positives: Class-0 events above the threshold*
  - *False-negatives: Class-1 events below the threshold*



Classifier metrics

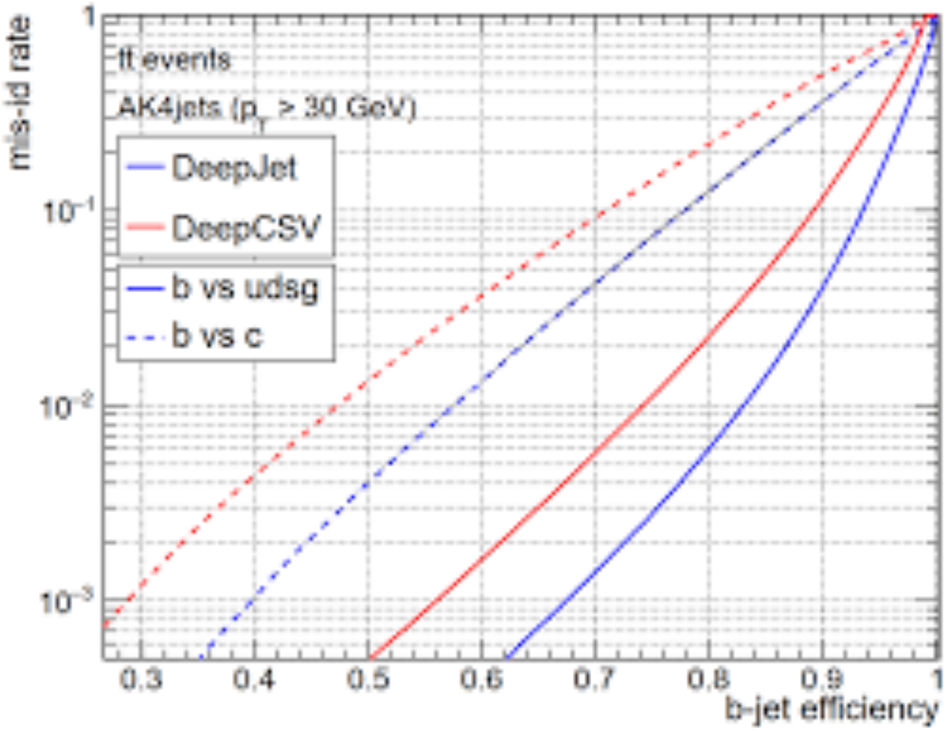
- *Starting ingredients are true positive (TP) and true negative (TN) rates*
- *Accuracy: (TP+TN)/Total*
  - *The fraction of events correctly classified*
- *Sensitivity: TP/(Total positive)*
  - *AKA signal efficiency in HEP*
- *Specificity: TN/(Total negative)*
  - *AKA mistag rate in HEP*
- *Depending on which quantity you prioritise, you would cut at a different place*

Receiver operating characteristic



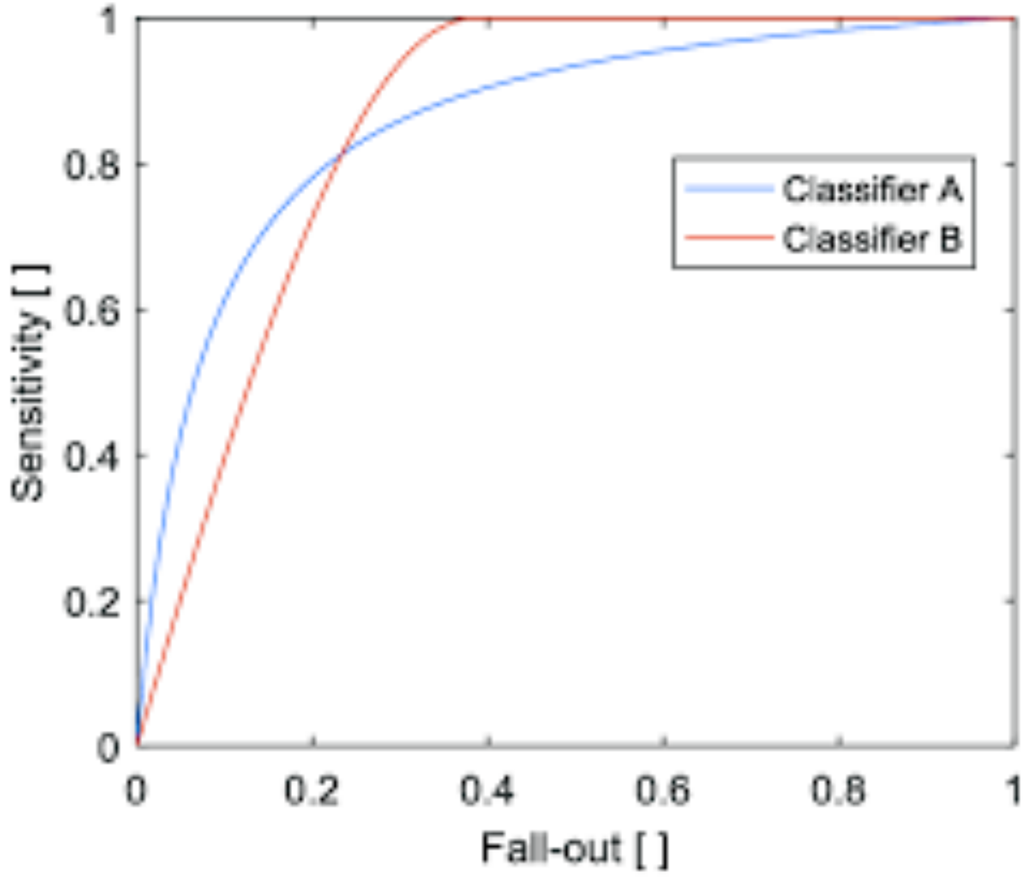
Selection power

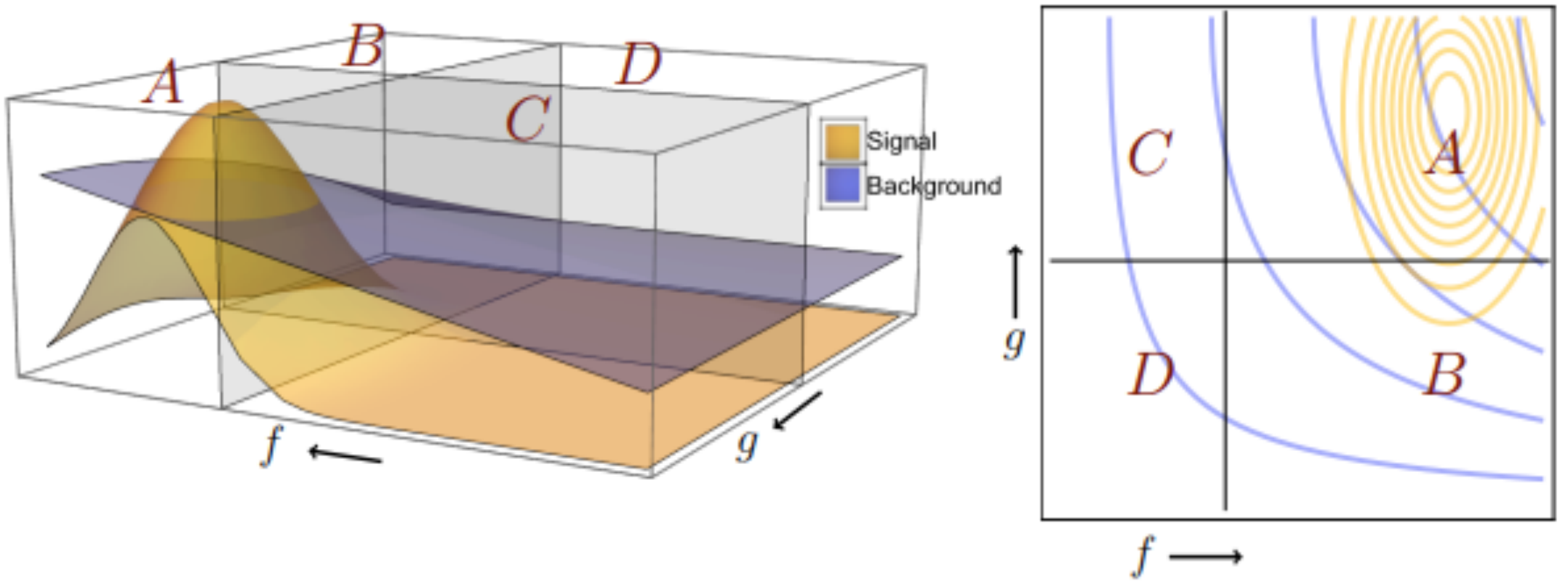
- One can use the ROC curve to quantify a selection power without a specific cut choice
  - Can be used to compare with N-dim cut based algorithms
  - Can be used to compare different algorithms (architectures, input features, etc)
- In practice, one selects a working point to use for a specific search
  - Where to cut depends on the specific case
  - Custom figures of merits are used to choose



Think Beyond the ROC

- *A ROC curve and the area under the curve (arc) are often used to comprare classifiers*
- *This is an unquestionable criterion when there is separation*
- *This is extremely misleading when the separation is less obvious (e.g., crossing lines)*
- *What matters is which classifier is better where you intend to cut not in average*

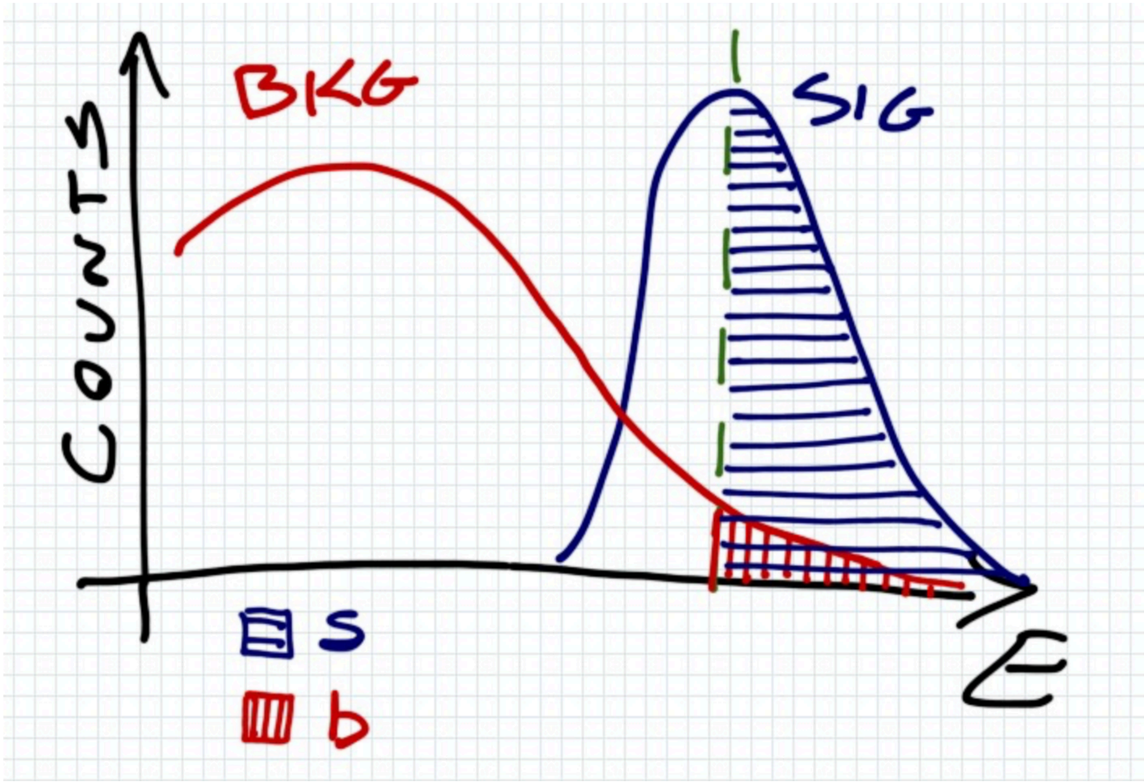




Background Estimate

Back to our counting experiment

- *The full likelihood including systematics has three terms*
  - *The "real" likelihood*
  - *The constraint on the signal expected yield (typically from MC or similar data, e.g., Z→ee for H→γγ) (*)*
  - *The constraint on the background expected yield. This is where troubles start*

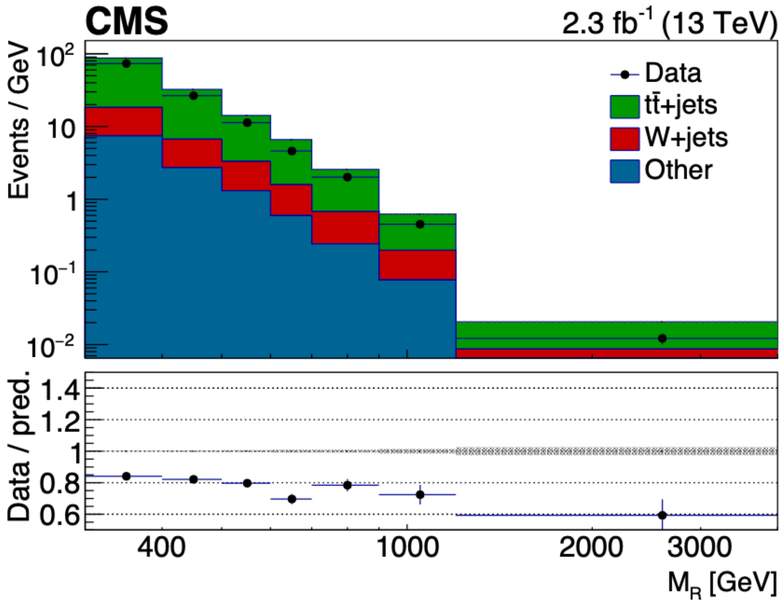


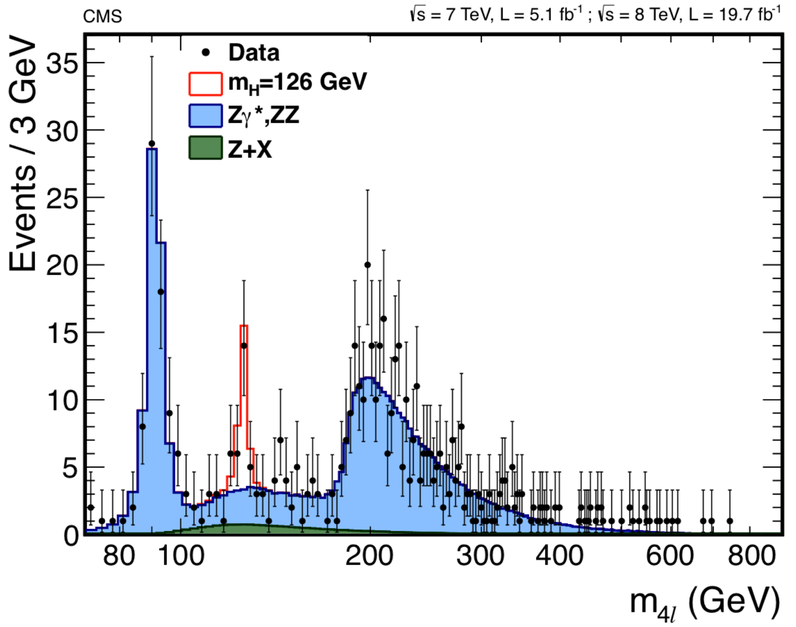
$$\mathscr{L} = P(n \mid \lambda_B + \lambda_S) G(\bar{\lambda}_S \mid \lambda_S, \sigma_{\lambda_S}) G(\bar{\lambda}_B \mid \lambda_B, \sigma_{\lambda_B})$$

(*) In the following slides I will drop the signal, since the discussion is about controlling the bkg uncertainty

MC-based background estimate

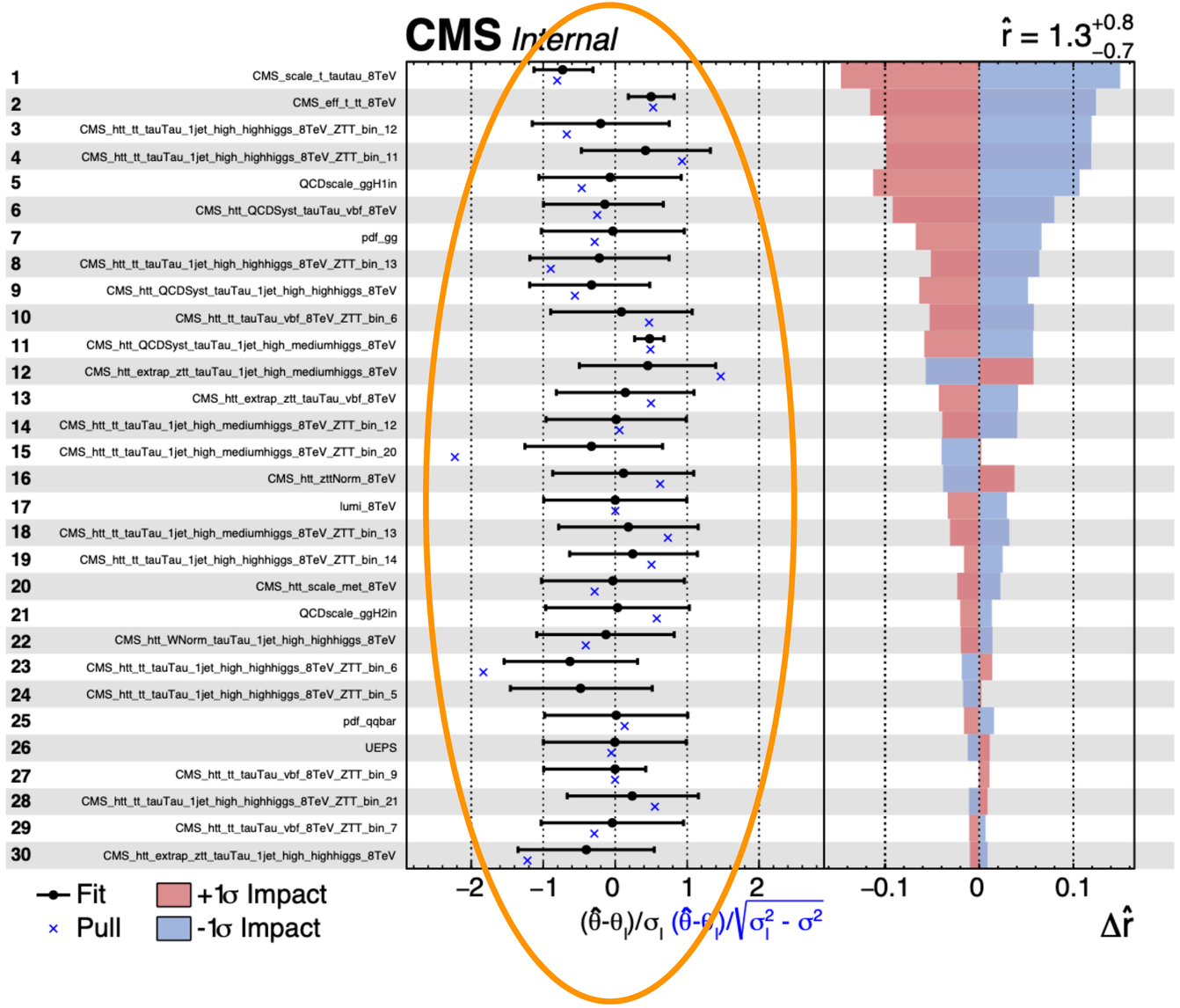
- ***Monte Carlo simulation***
  - *One could predict the background with simulation. Not reliable per se (uncertainty on simulation accuracy difficult to estimate).*
  - *A good baseline for a more accurate prediction*
- ***Fitting the uncertainties***
  - *Sometime one can estimate uncertainties of the simulation and model them through nuisance parameters*
  - *Data/MC agreement is then improved with profiling while fitting for the signal*
  - *Uncertainties on nuisances propagates to uncertainties on signal through correlation*





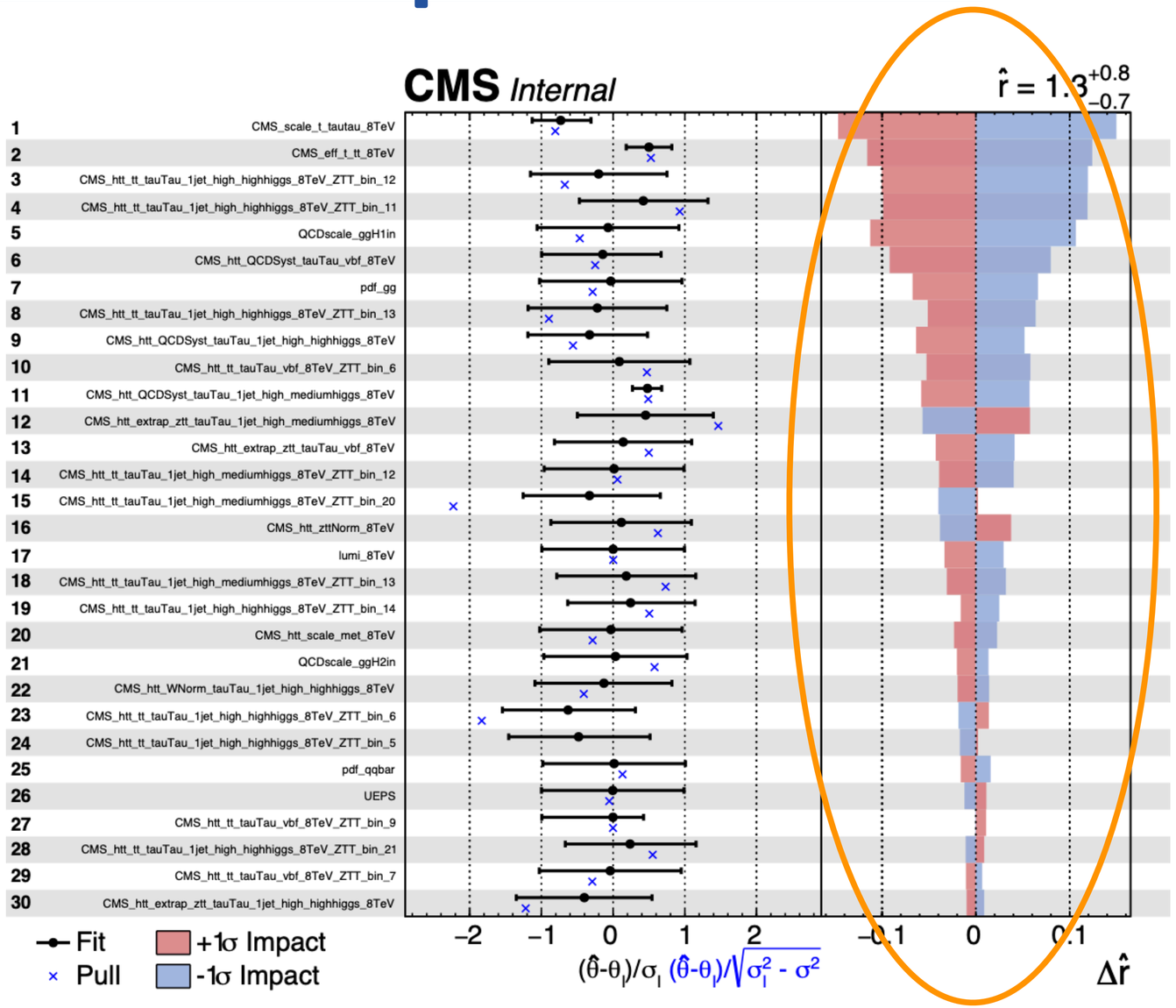
Nuisance pulls

- *A typical plots shown by experiment is the "impact plot"*
  - *It shows the pull of the nuisance parameters*
  - *And the impact that the nuisance variation has on the parameter of interest $\hat{r}$ (e.g., the signal yield)*



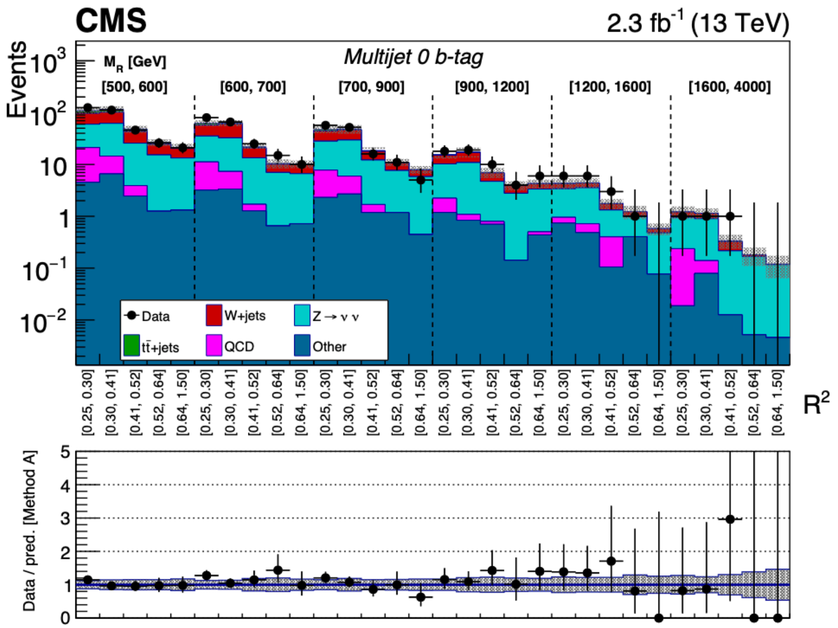
Nuisance pulls

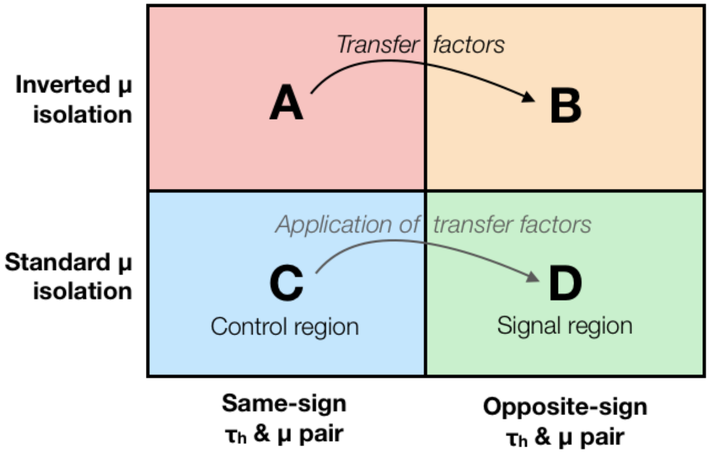
- *A typical plots shown by experiment is the "impact plot"*
  - *It shows the pull of the nuisance parameters*
  - *And the impact that the nuisance variation has on the parameter of interest $\hat{r}$ (e.g., the signal yield)*

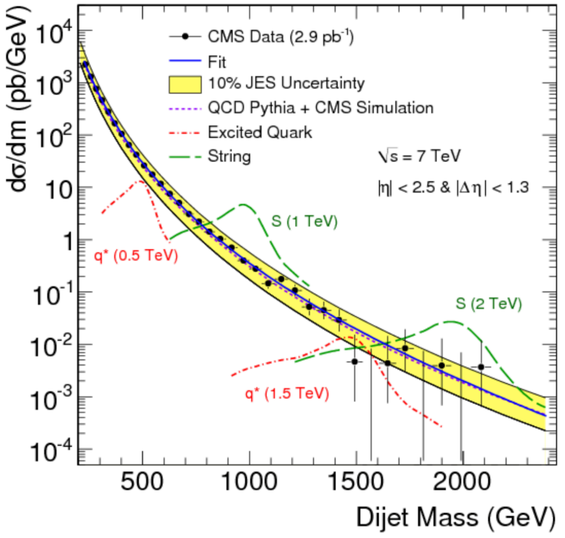


Data-driven methods

- *A control region:*
  - *One can use a bkg control region bridging the observed yield to the signal region using simulation*
- *A 2D sideband*
  - *One can use two independent quantities to define the signal region and scale background from nearby sideband (ABCD)*
- *Connecting the bins*
  - *One can fit the background across adjacent bins with a smooth function*

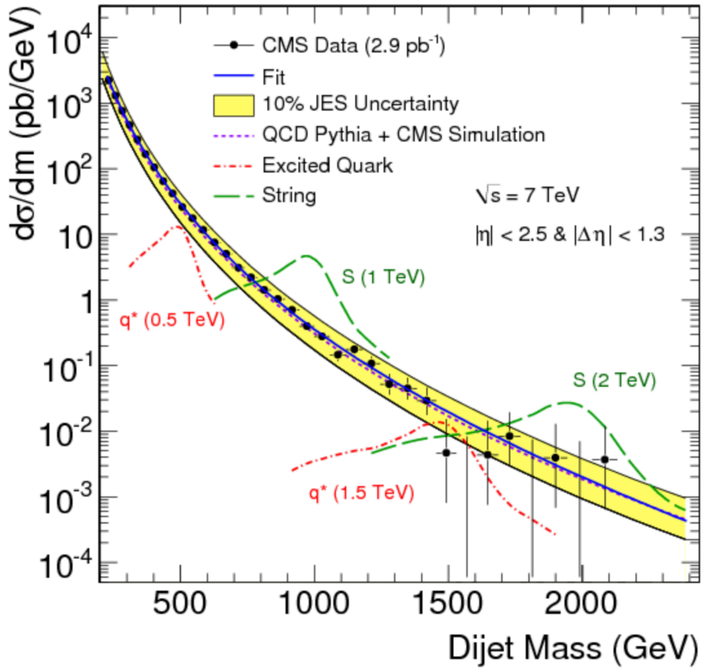






Smooth fit

- *Given a background function f(x) describing the background shape distribution in x, one can predict the expected background*
- *The function has parameters α that one has to determine*
  - *Cannot trust Monte Carlo in a data driven method*
  - *Can use profile the α in the fit*
- *One has to choose a robust function and attach some systematic uncertainty to the choice*



$$E[n_i] = f(x_i \mid \alpha)$$

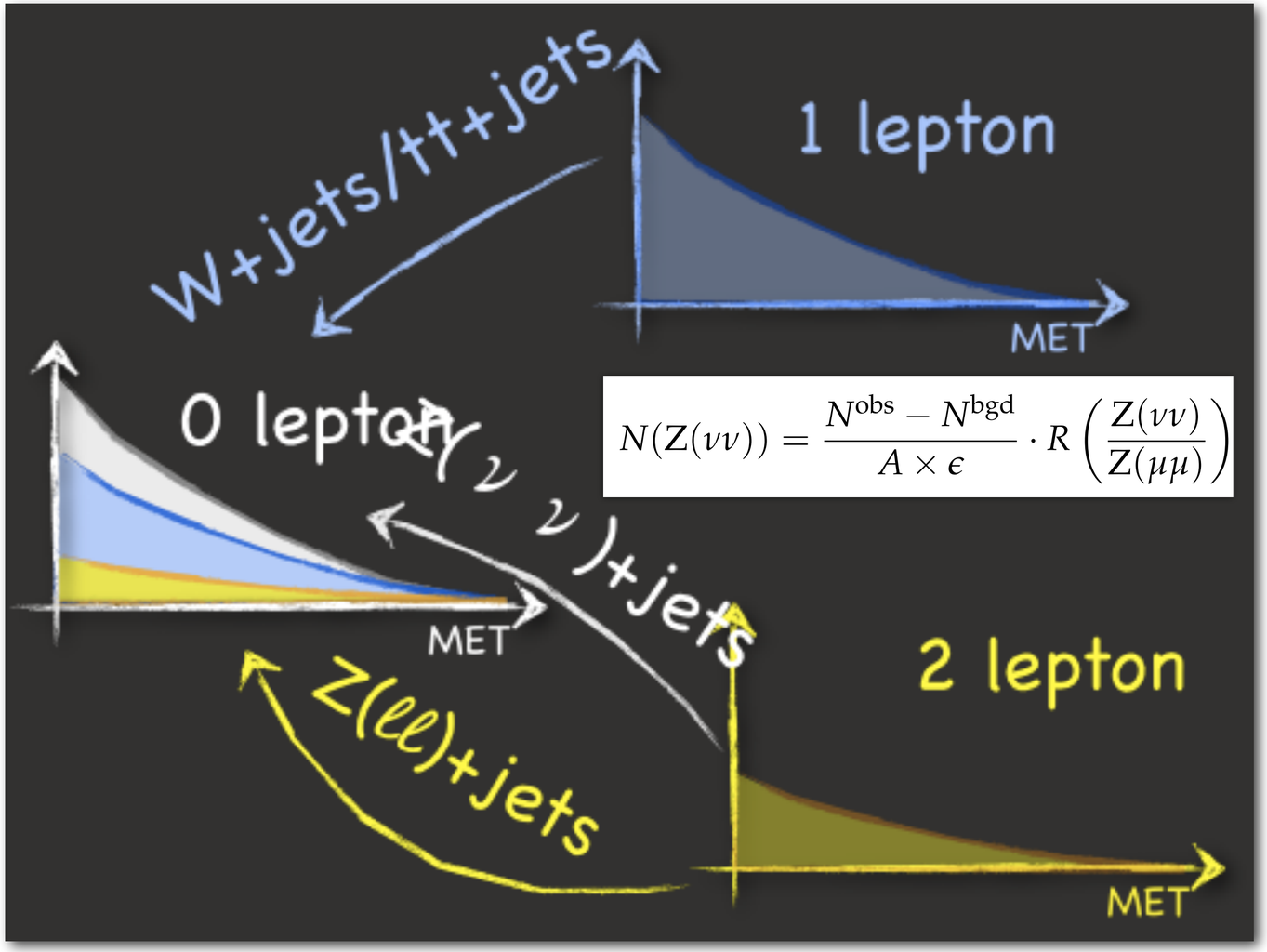
$$\hat{\mathscr{L}} = \max_{\alpha} \prod_{i} P(n_i \mid f(x_i \mid \alpha))$$

Smooth fit

- *PROS:*
  - *Simple to use (e.g., in bump hunts)*
  - *Very little use of MC simulation (typically top test function choice, but data control regions can be used for that*
- *CONS:*
  - *Modeling the un certainty on the functional choice not trivial*
  - *Choice of function tend to be problematic on tail*
  - *With large statistics, poor goodness-of-fit for bkg-only hypothesis can compromise analysis robustness*

MC-assisted prediction

- *Typically used in searches (e.g., SUSY)*
- *Several control regions enriched of specific backgrounds*
  - *Z->ll for Z->nn bkg*
  - *l+jets without b-tag jets for W+jets*
  - *l+jets with b-tag jets for tt*



MC-assisted prediction

- *Common likelihood is defined for signal region (SR) and control regions (CRs)*
- *Monte Carlo samples are added as additional control regions*
- *expected yields in various data regions are connected, using functions of corresponding MC expected yields*
- *The profiles likelihood is obtained maximizing over the λs*
- *One has to add signal and its uncertainties, as discussed already*

$$\mathscr{L} = \prod P_{SR}(n_{SR,i} \mid \lambda_{SR,i}) P_{CR}(n_{CR,i} \mid \lambda_{CR,i}) \qquad \lambda_{SR,i} = \lambda_{CR,i} \frac{\lambda^{MC}_{SR,i}}{\lambda^{MC}_{CR,i}}$$

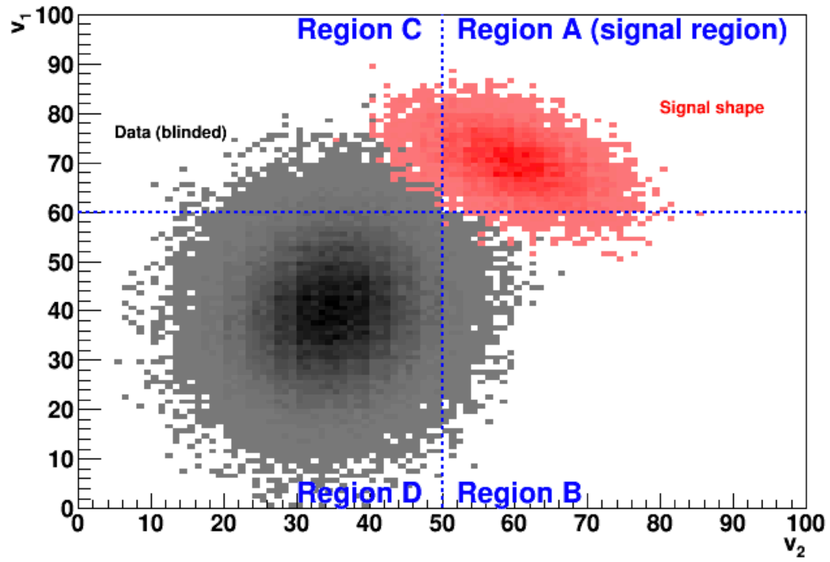
$$\mathscr{L} \rightarrow \mathscr{L} = \max_{\lambda} \prod P_{SR}(n_{SR,i} \mid \lambda_{CR,i} \times \lambda^{MC}_{SR,i} / \lambda^{MC}_{CR,i})$$
$$P_{CR}(n_{CR,i} \mid \lambda_{CR,i}) \times P^{MC}_{SR}(n^{MC}_{SR,i} \mid \lambda^{MC}_{SR,i}) P^{MC}_{CR}(n^{MC}_{CR,i} \mid \lambda^{MC}_{CR,i})$$

MC-assisted prediction

- *PROS:*
  - *more robust vs MC simulation since only ratios of MC yields are used*
  - *Generalizes very well to multi-bin fits*
- *CONS*
  - *Available MC statistics becomes a crucial factor that can limit the precision*
  - *MC modeling can be different if extrapolation is across regions with different kinematic properties (e.g., from low-momentum to high-momentum)*

ABCD method

- *Similar to our initial counting experiment, but 2D*
- *Two quantities defining ABCD plane are independent for the background*
- *Selection factorizes, so one can obtain a bkg prediction from data using three sidebands*



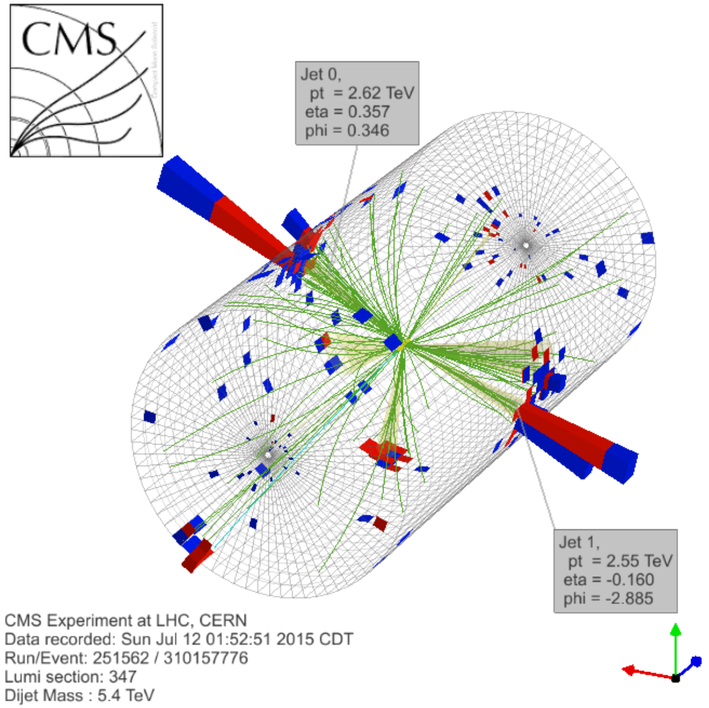
$$\mathscr{L} = P(n_A \mid \lambda_A) P(n_B \mid \lambda_B) P(n_C \mid \lambda_C) P(n_D \mid \lambda_D)$$

$$\lambda_A = \lambda_C \times k_{C \to A} = \lambda_C \times k_{D \to B} = \lambda_C \frac{\lambda_B}{\lambda_D}$$

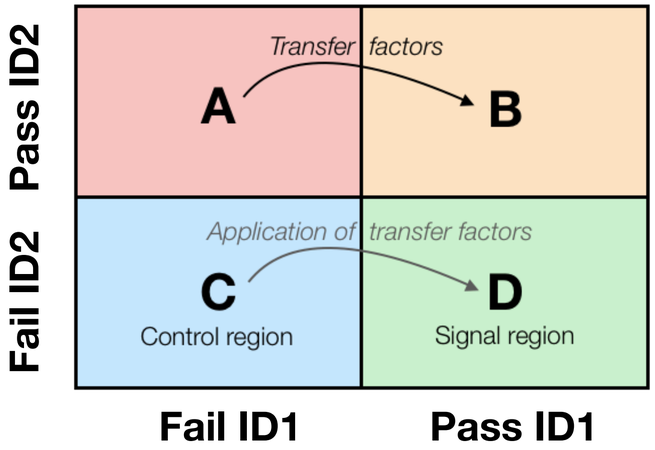
$$\hat{\mathscr{L}}(n_A) = \max_{\lambda_B, \lambda_C, \lambda_D} P(n_A \mid \lambda_C \frac{\lambda_B}{\lambda_D}) P(n_B \mid \lambda_B) P(n_C \mid \lambda_C) P(n_D \mid \lambda_D)$$

ABCD method

- *PROS*
  - *In principle very clean methods*
  - *Easy to n-bin generalization (ABCD per bin)*
- *CONS*
  - *In practice, difficult to find two uncorrelated quantities (but notable cases exist)*
  - *Residual correlation hard to model (with MC?) and associated systematic can be limiting factor*
  - *transfer factor can depend on other quantities (e.g., p_T etc.)*

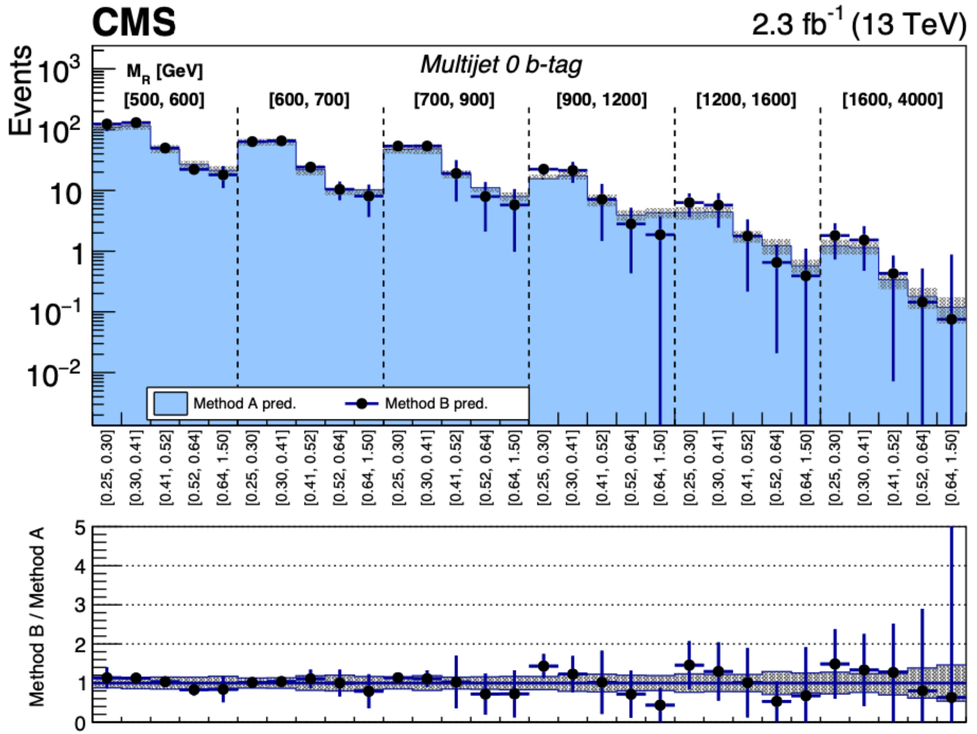


Whenever signature has to events with independent ID



When using what

- *SM process measurements typically use MC-based profile likelihood fits*
  - *With background control regions added to make fit robots*
- *Searches for new resonances typically use bump hunts with functional fit of the background*
  - *Systematic on function choice is becoming factor, so the other methods are becoming popular here too*
- *Searches for exotics (e.g., long-lived particles) use ABCD a lot*
  - *Usually exploiting exotic-signature ID for ABCD plance*
- *Searches with traditional objects in final states (leptons, jets, MET) use MC-assisted data-driven predictions*
  - *SUSY, Dark Matter searches, etc.*
- *All methods have hidden assumptions and associated systematics*
- *Robustness comes from using (and comparing) different methods*



Summary

- *We reviewed how to define an event selection*
  - *what to cut on*
  - *how to cut*
  - *where to cut*
- *We saw the implications of online vs offline selection*
- *We discuss a few of the most popular background prediction methods*

Backup

Efficiency, cross section, luminosity

- At collider, the expected number of events (S or B) is the product between
  - The number of produced events for a given process $N = \sigma\mathscr{L}$, where $\mathscr{L}$ is the luminosity and σ is the cross section
  - The probability of a sample from a given process to pass the cuts (i.e., the efficiency ε we defined before)
- In other experimental setups, the luminosity is traded for the corresponding time (e.g., time of exposure of a target of a given size, etc.)
- σ is computed from theory, $\mathscr{L}$ is measured at the experiment
- But what if σ is not known (e.g., in searches for new physics)?