EPISODE II
R-PARITY CONSERVING SUSY

MAURIZIO PIERINI
CERN

European Research Council

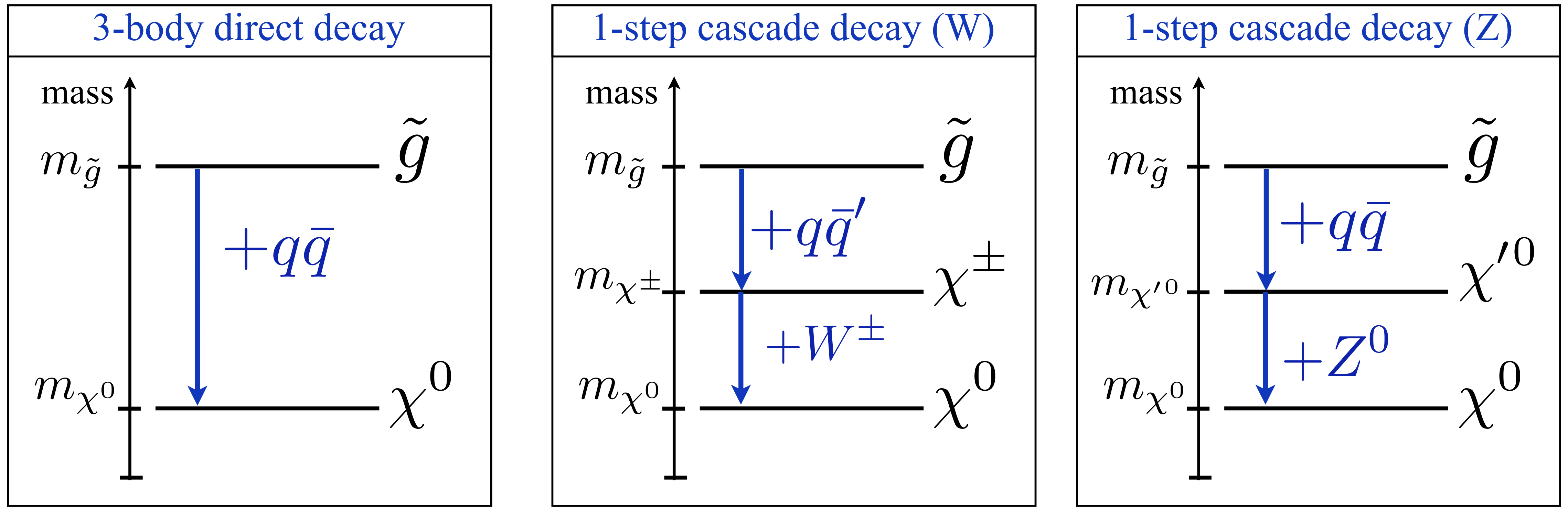
SUMMARY OF EPISODE I

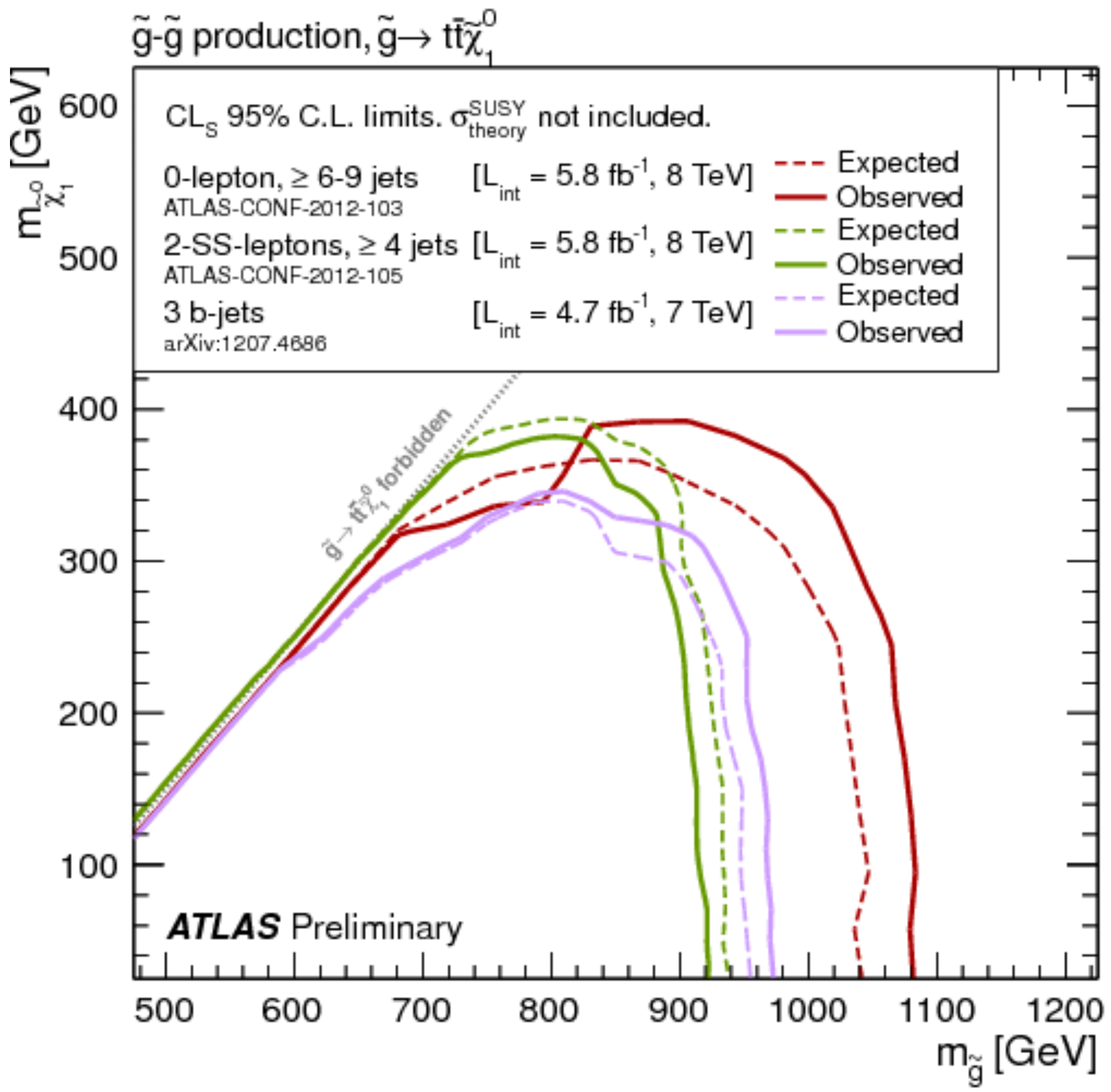
- *LHC has all it needs to be a SUSY discovery machine*
  - *can produce full spectra of particles*
  - *can observe many final states for any particle*
- *Practical limitations (e.g., trigger) should come into consideration when designing the analysis*
- *Data control samples are a key ingredient (a 100% MC-based background prediction would not be considered acceptable at LHC)*
- *Statistical tools in place from Higgs discovery*
- ***Simplified models great guidance to interpret and improve searches, when taken with a grain of salt***


CHECK LIST FOR
MY SUSY ANALYSIS
SIGNATURE
TRIGGER
RECONSTRUCTION
ANALYSIS STRATEGY
STATISTICAL ANALYSIS
INTERPRETATION

SIMPLIFIED MODELS

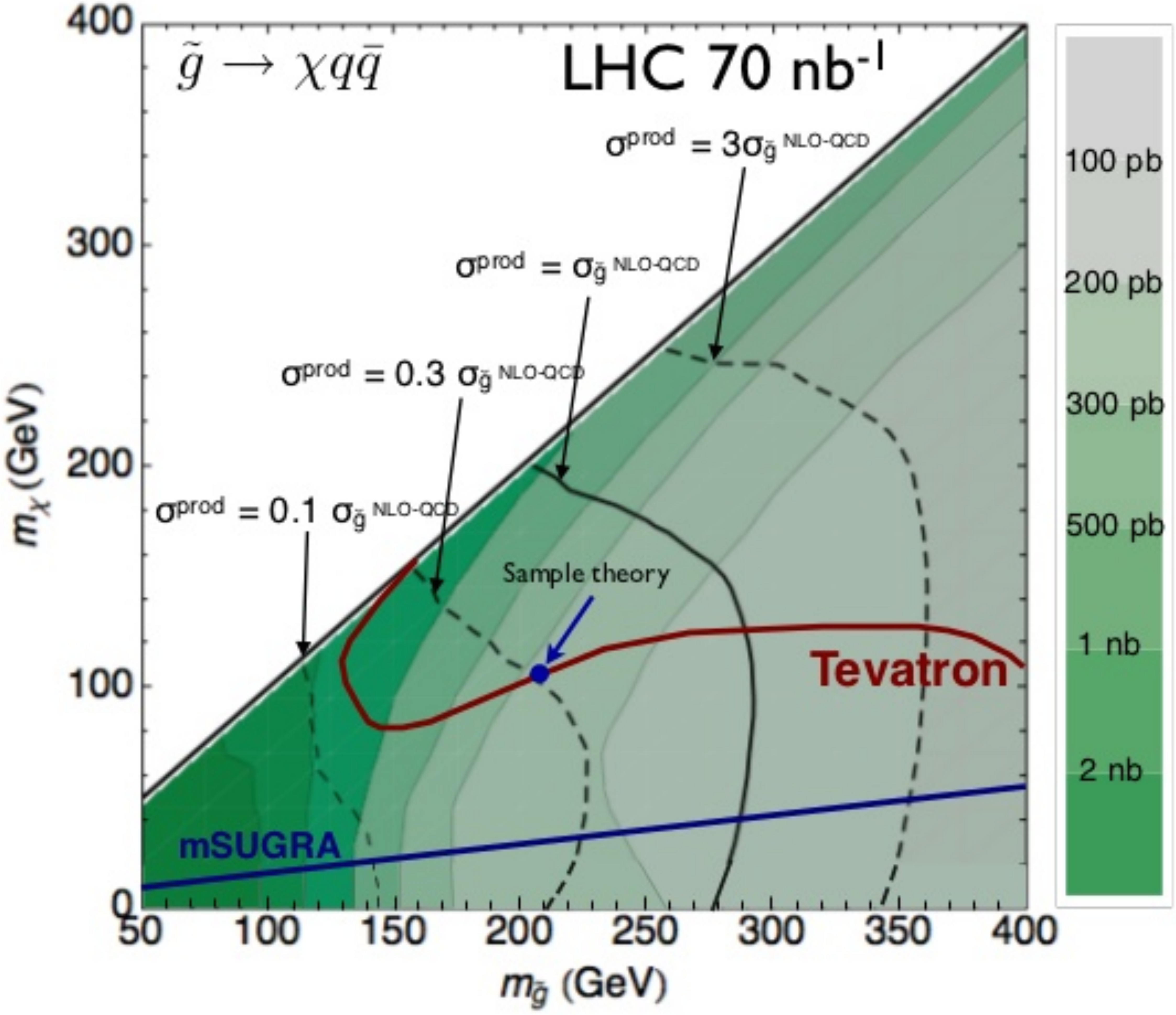
- Before the LHC, searches were entered on full models
- This changed during Run I. Simplified models became the standard
  - Focus on a specific process x decay chain
  - Interpret the analysis in this context





WHY IS THIS NEEDED? YES!

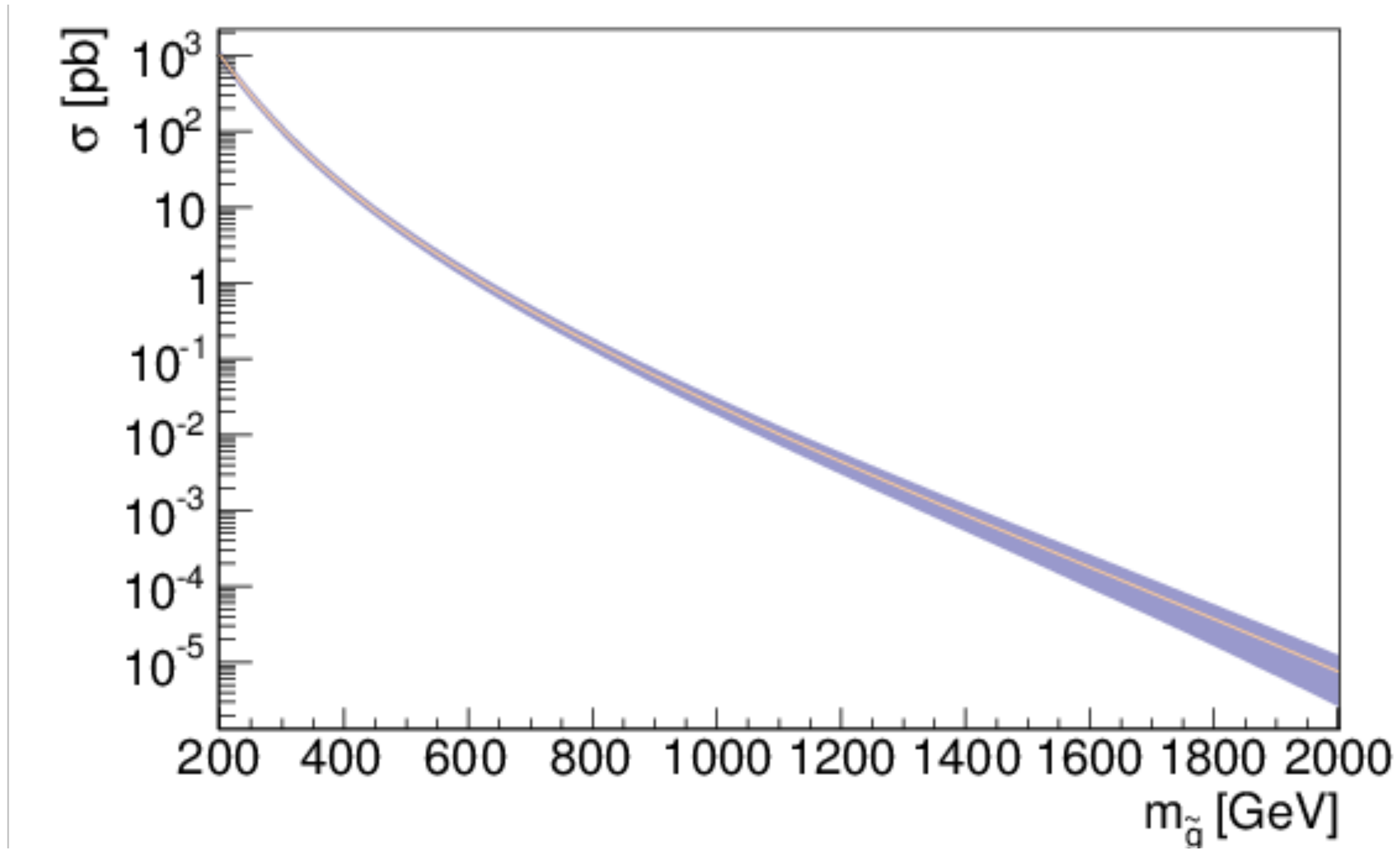
- *At the beginning of the LHC, many pre-LHC-data analyses were actually found to be too much tailored on the benchmark models*
- *Simplified models allowed to go beyond certain implicit assumptions*
- *This new paradigm allowed to discover weakness in the search program and design a next-generation set of analyses*
- *In general, the use of simplified models made our search strategy more robust*

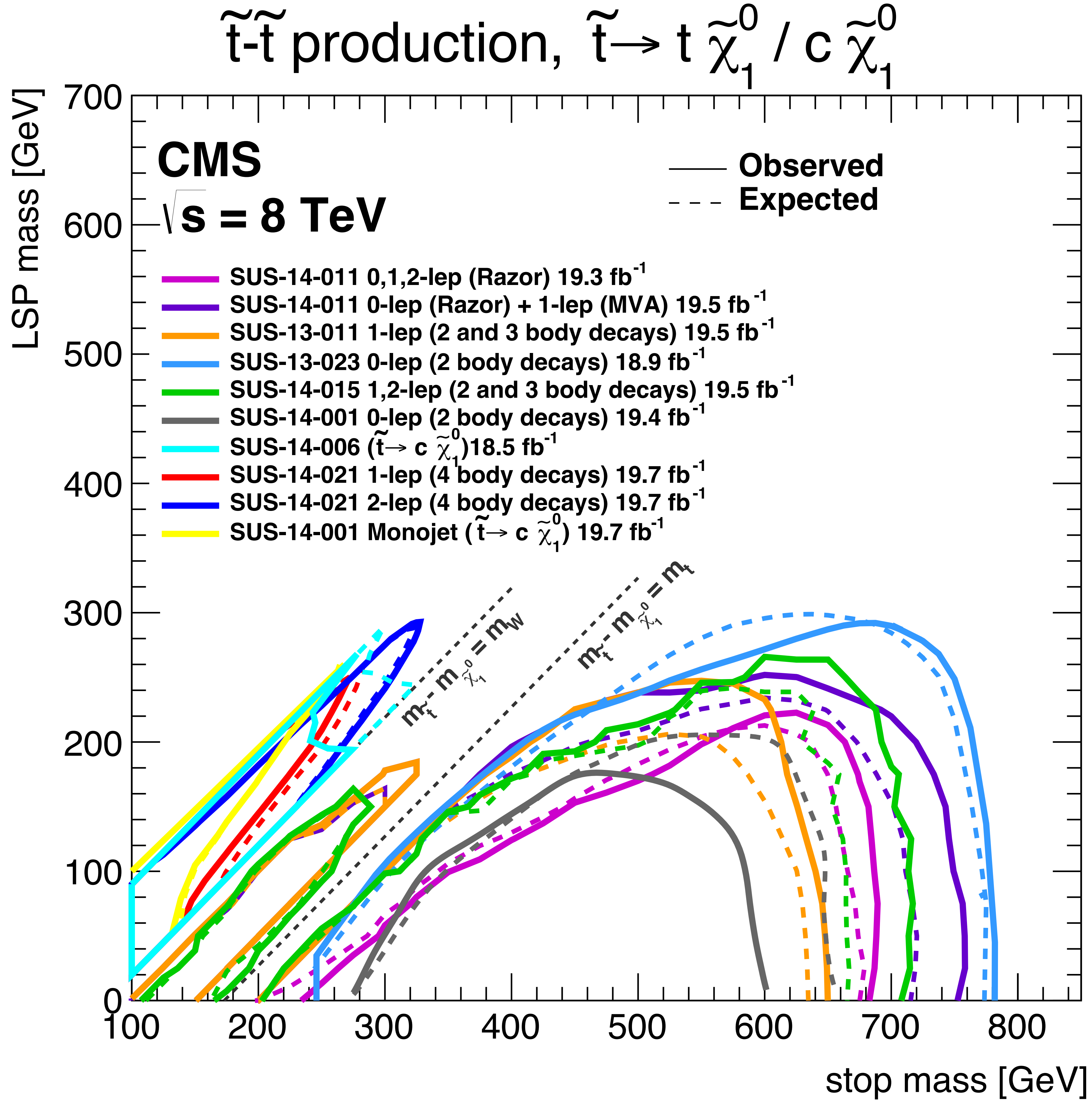


MIND THE HIDDEN ASSUMPTIONS

- *BRs are usually assumed to be 100%. This means that every line in a summary plot is implicitly excluding the others*

- *Cross sections are sometimes computed under special assumptions (e.g., decoupling limit) and don't hold in general*

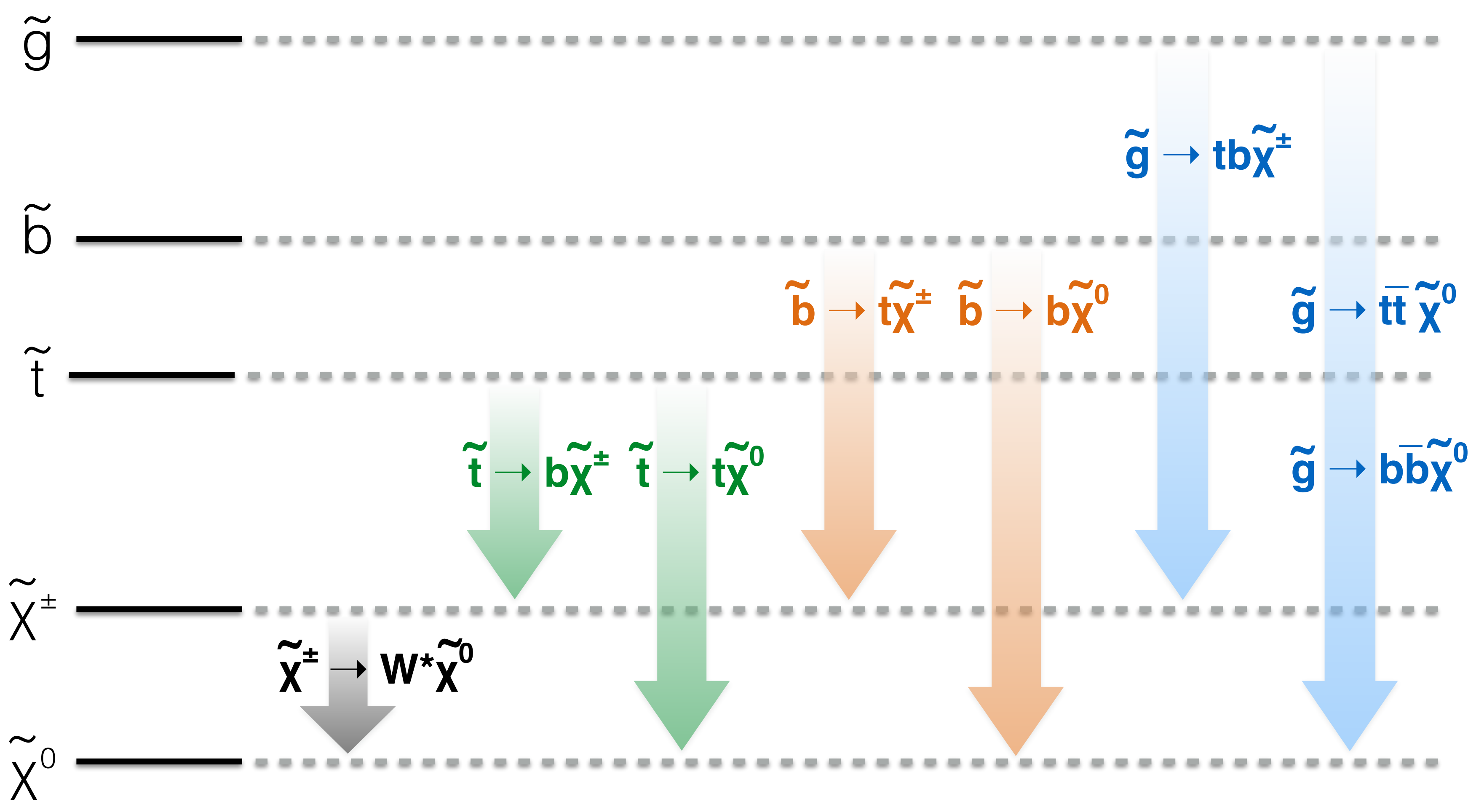


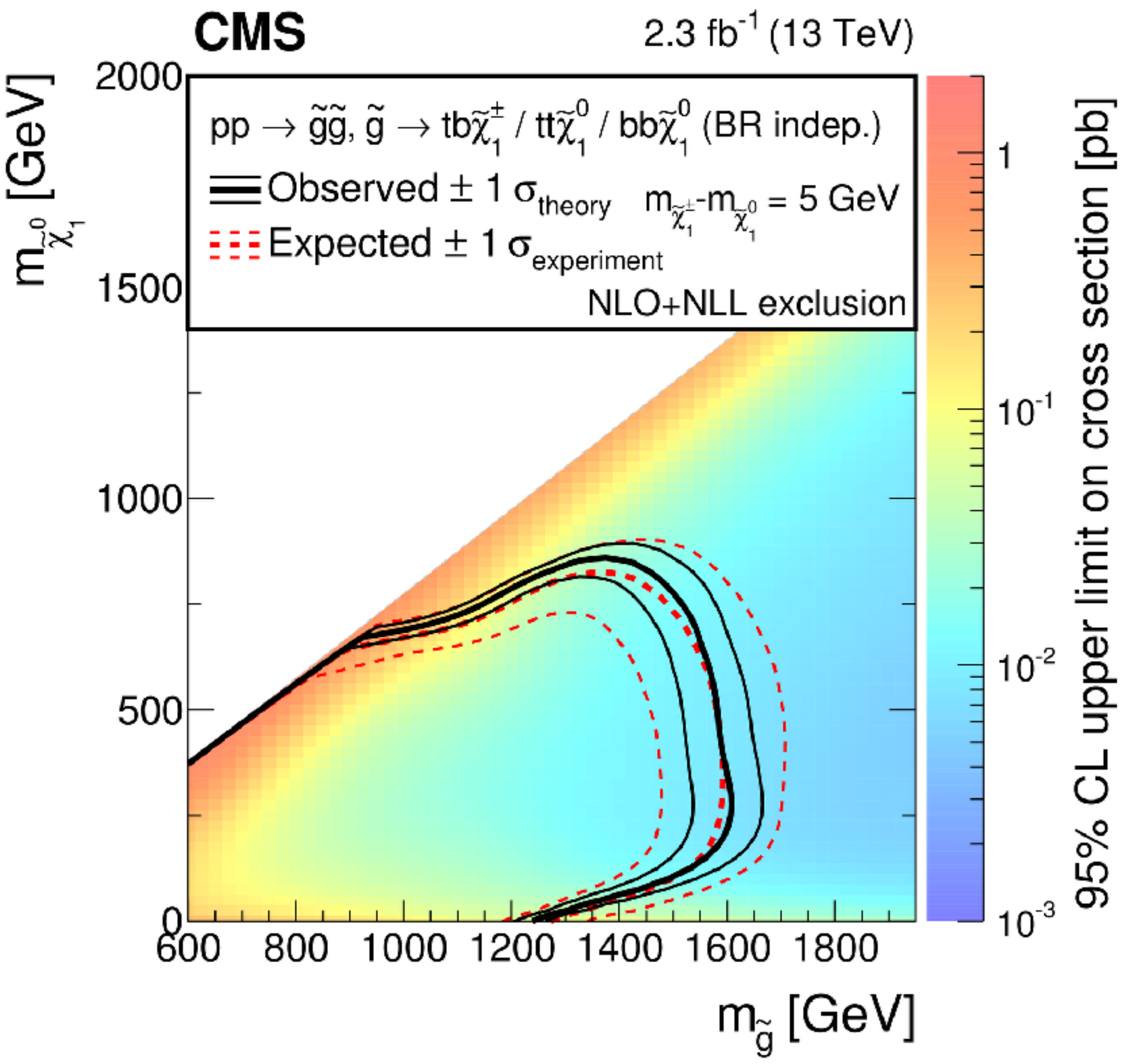


European Research Council

GOING BEYOND ASSUMPTION

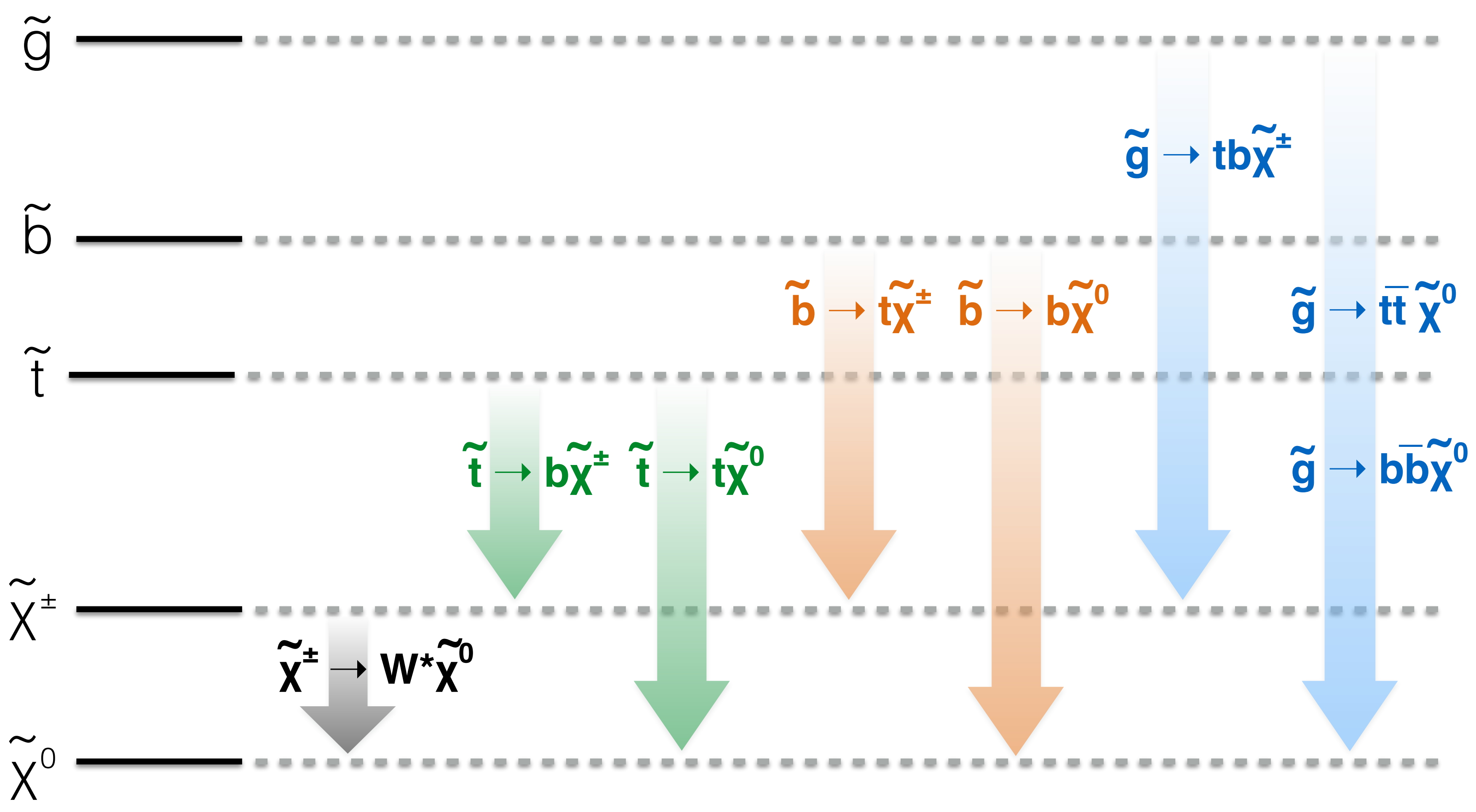
- *Recent tendency is to present BR-independent results*
  - *easy in specific scenarios like Natural SUSY*
  - *e.g., done by scanning the BR space and quoting the worst result*

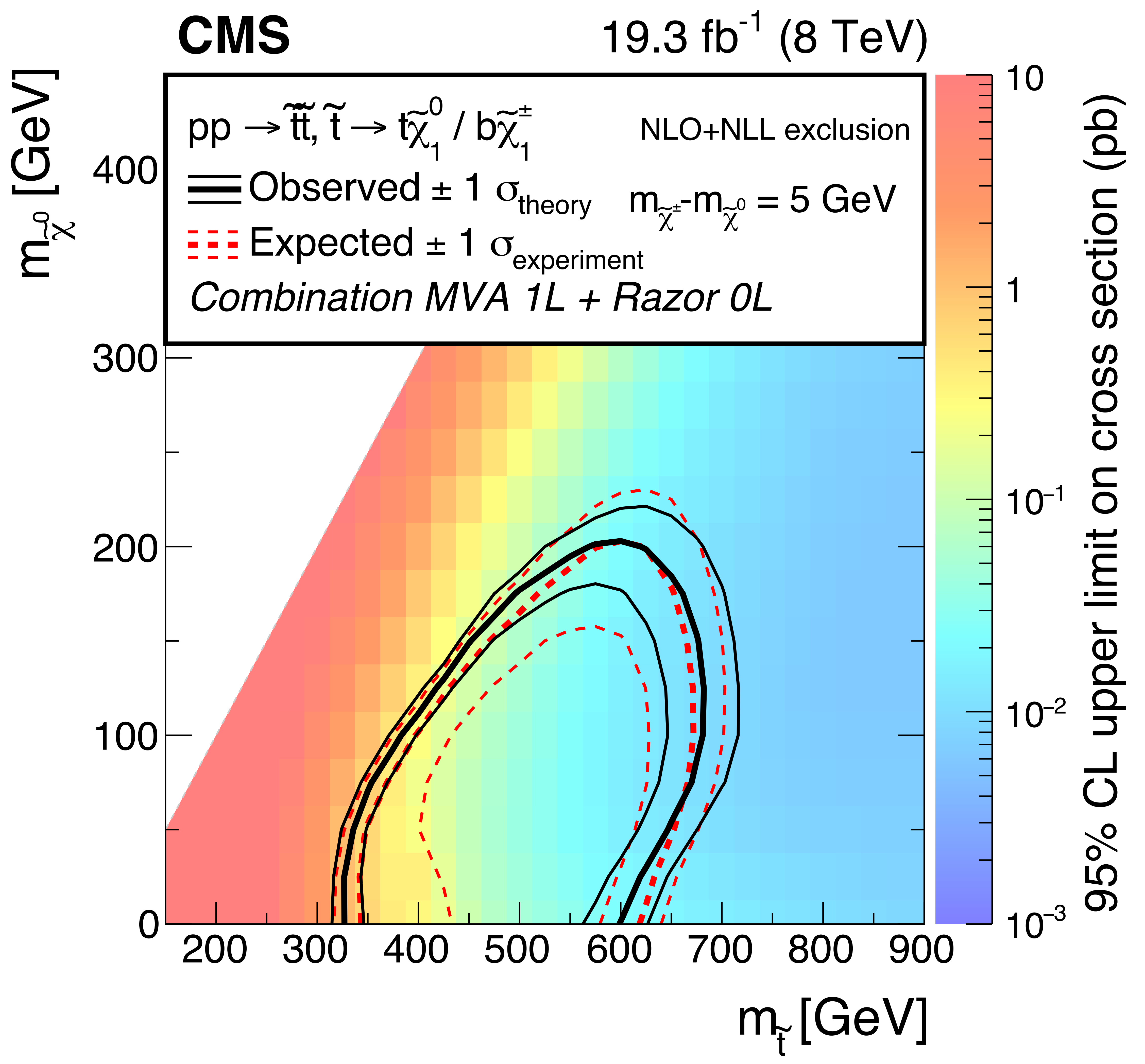




GOING BEYOND ASSUMPTION

- Recent tendency is to present BR-independent results
    - easy in specific scenarios like Natural SUSY
    - e.g., done by scanning the BR space and quoting the worst result





THE PLAN

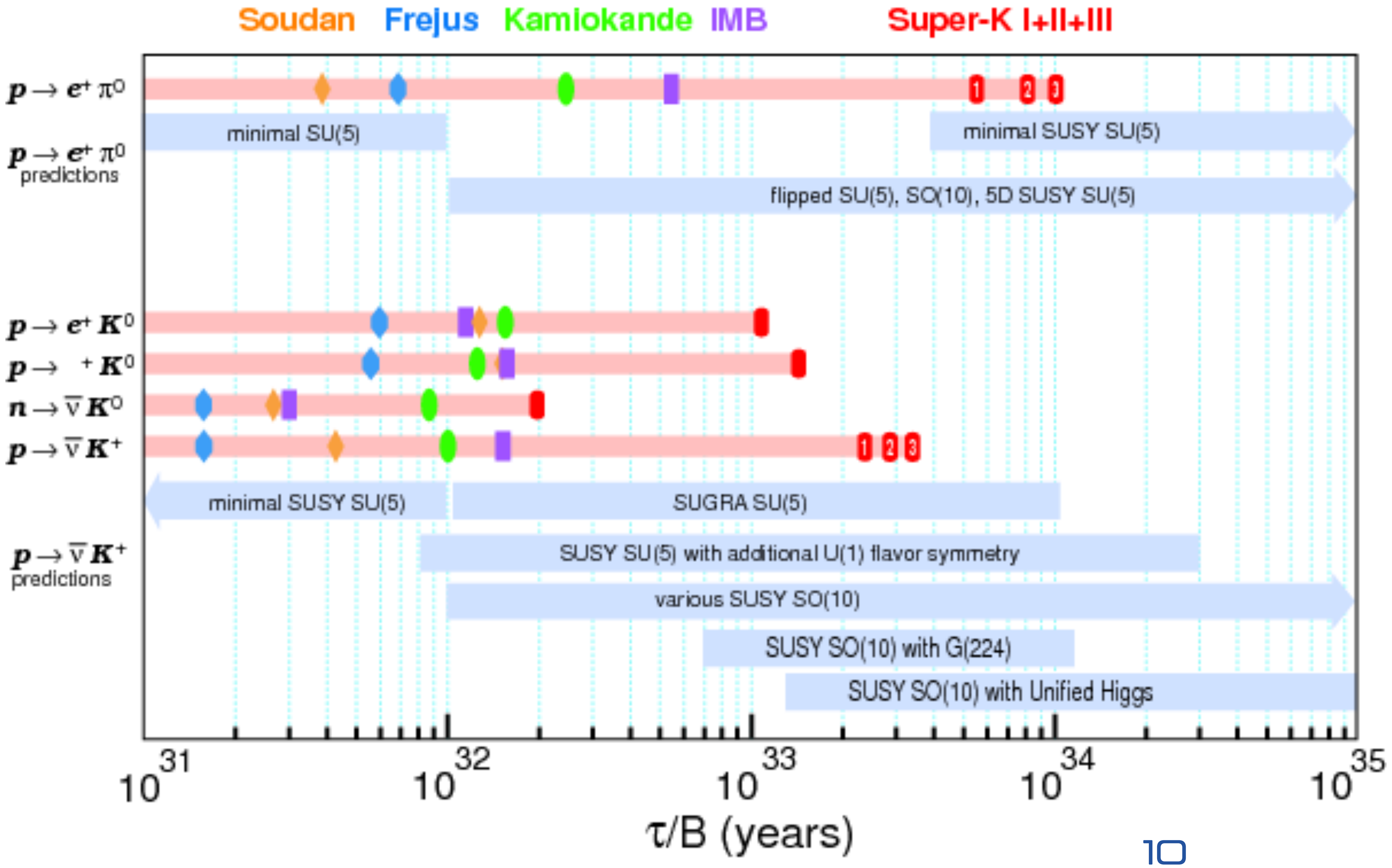
- Lecture 1: **SETTING UP A SEARCH AT THE LHC**
  - Searching for SUSY in practice: strategy, trigger, reconstruction
  - Designing a search: Simplified Models
  - Building a search: signal region, control regions, statistics tools

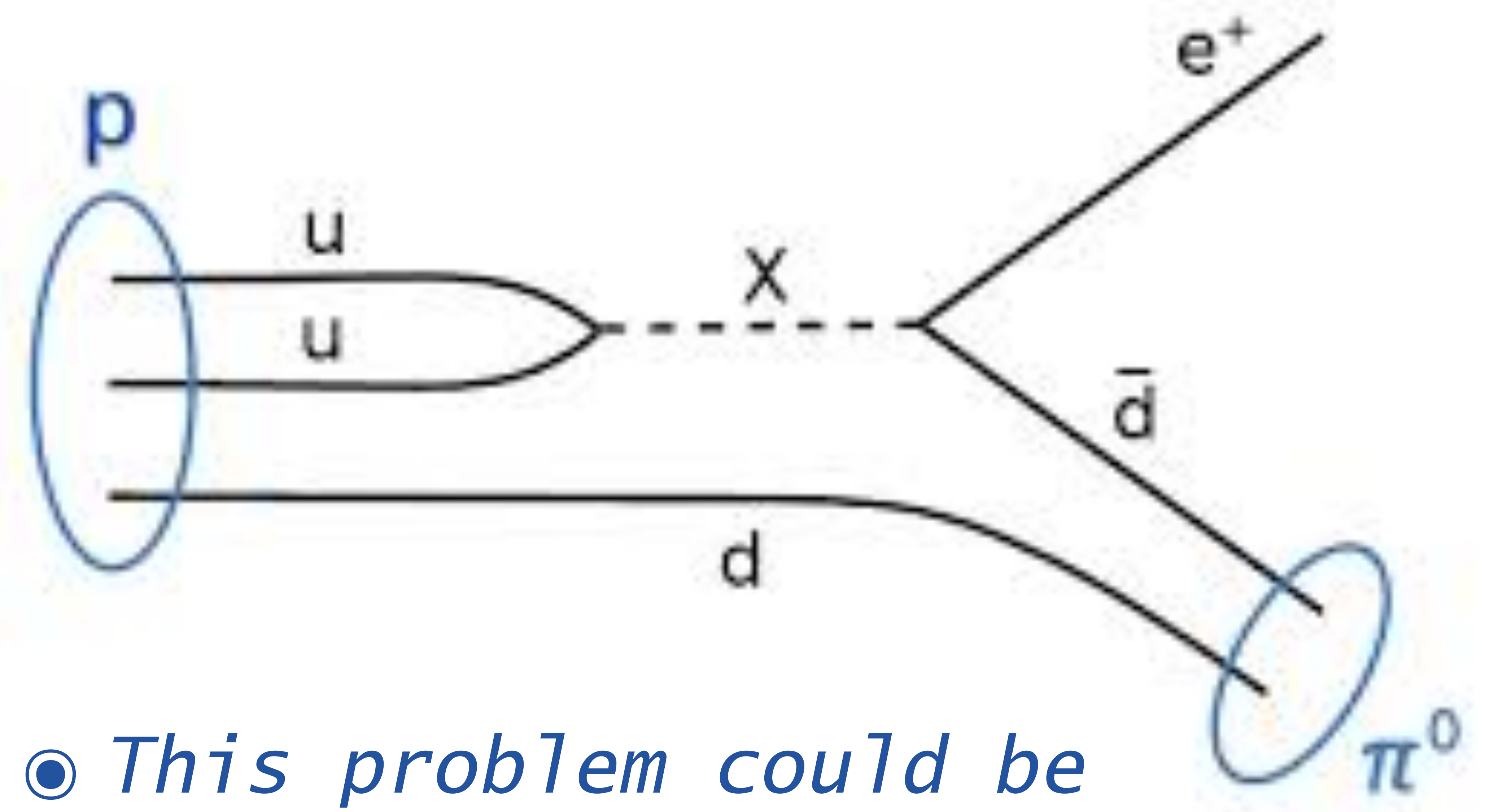
- Lecture 2: **R-PARITY CONSERVING SUSY**
  - DM direct production
  - DM cascade production

- Lecture 3: **BEYOND MET-BASED SEARCHES**
  - RPV SUSY
  - Displaced particles

R-PARITY AND STABLE PARTICLES

- The problem with SUSY is that it predict proton to decay, on which we have strong limits



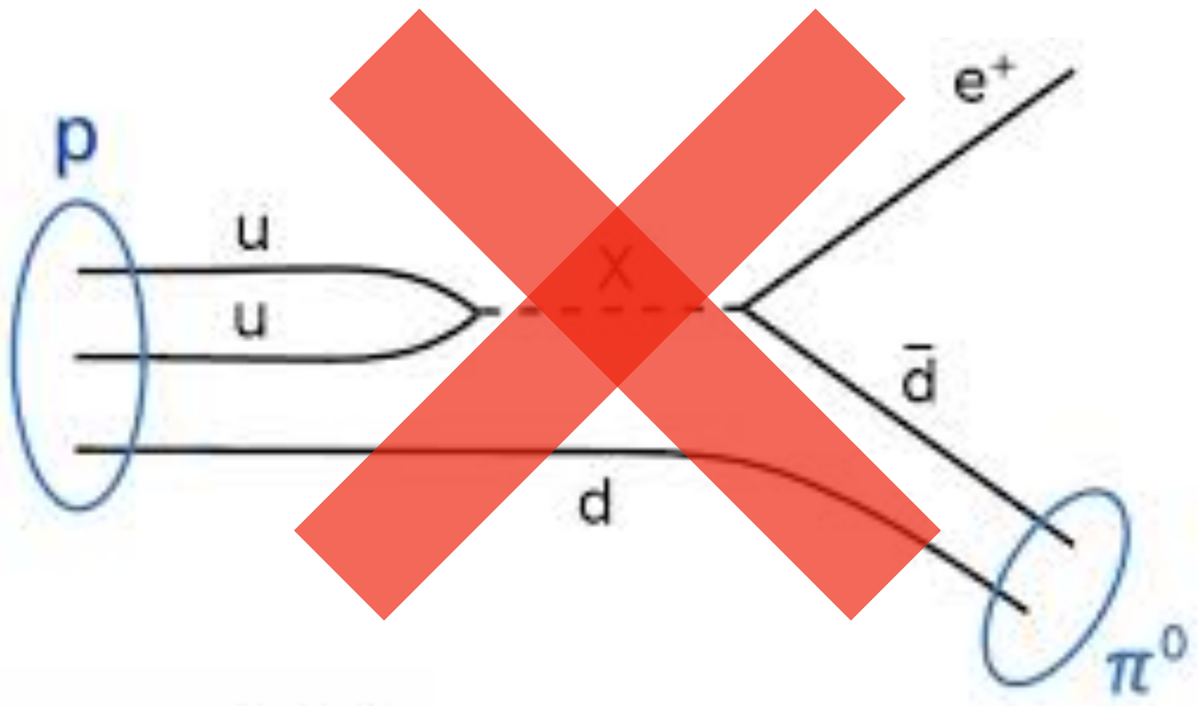


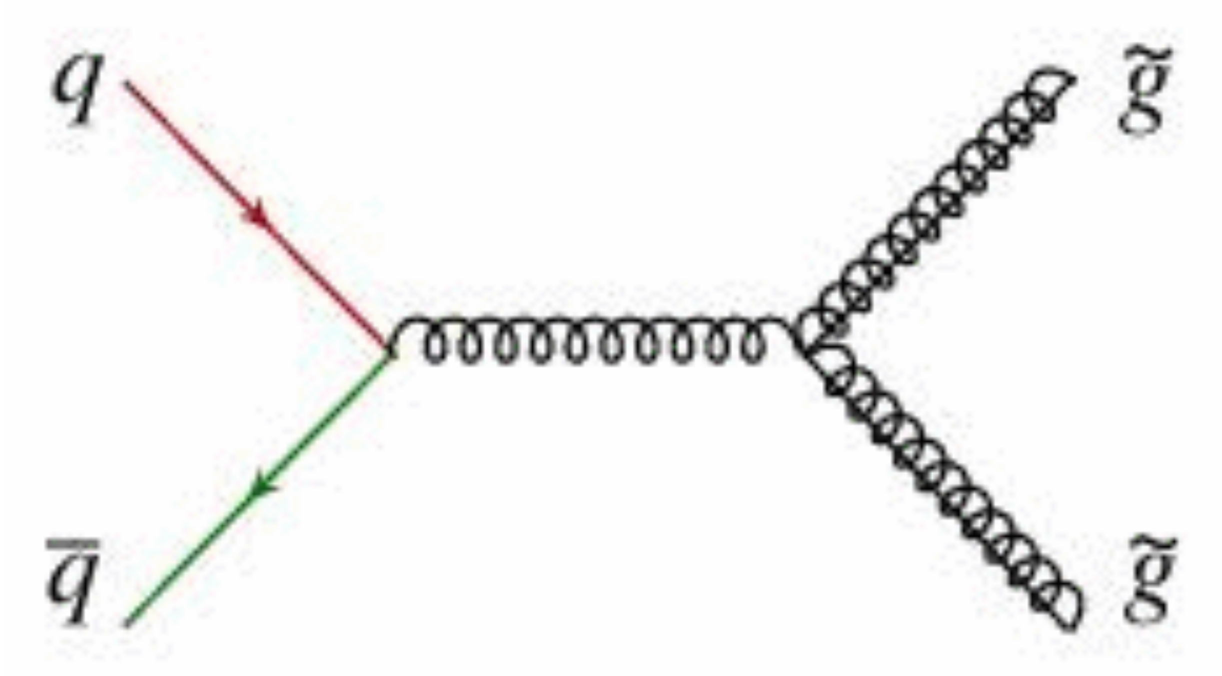
- This problem could be circumvented postulating a new quantum number (R-parity)
- +1 for SM particles
- -1 for SUSY partners

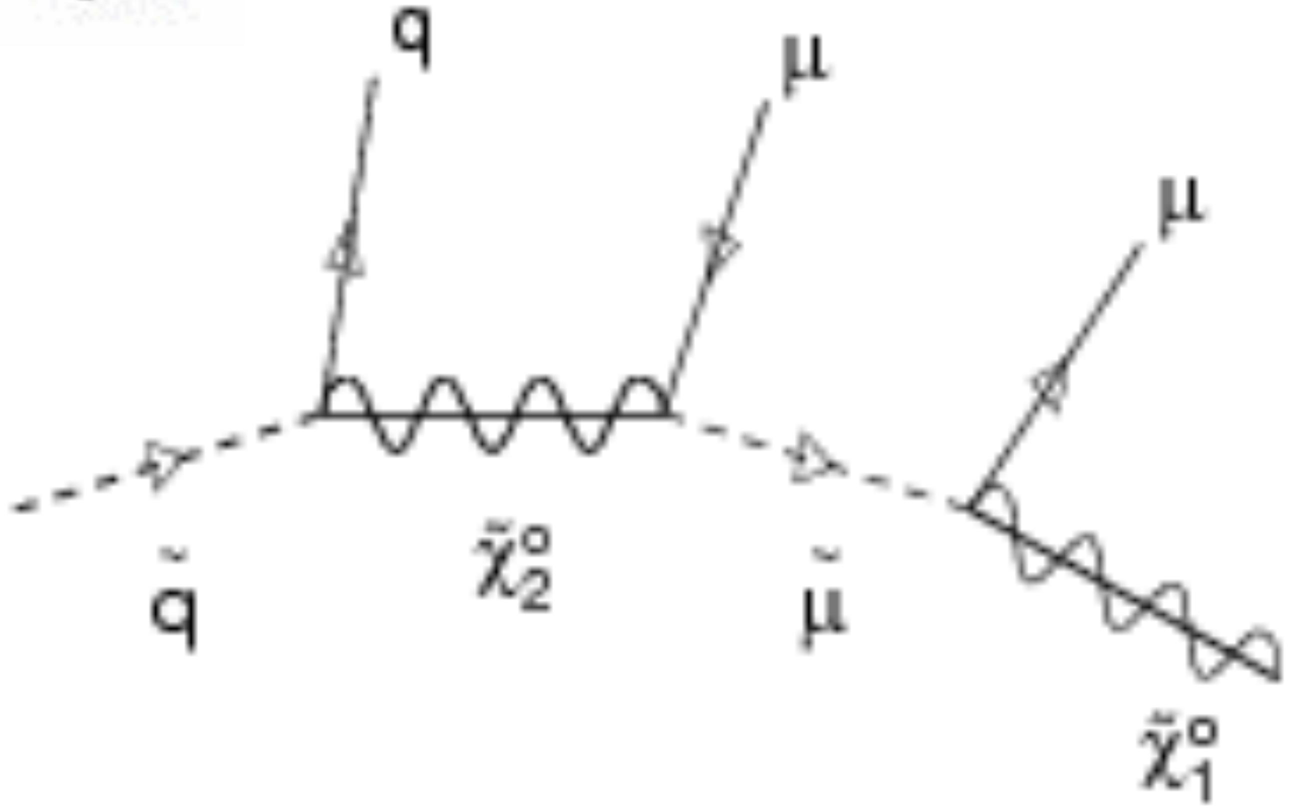
$$R = (-1)^{3(B-L)+2S}$$

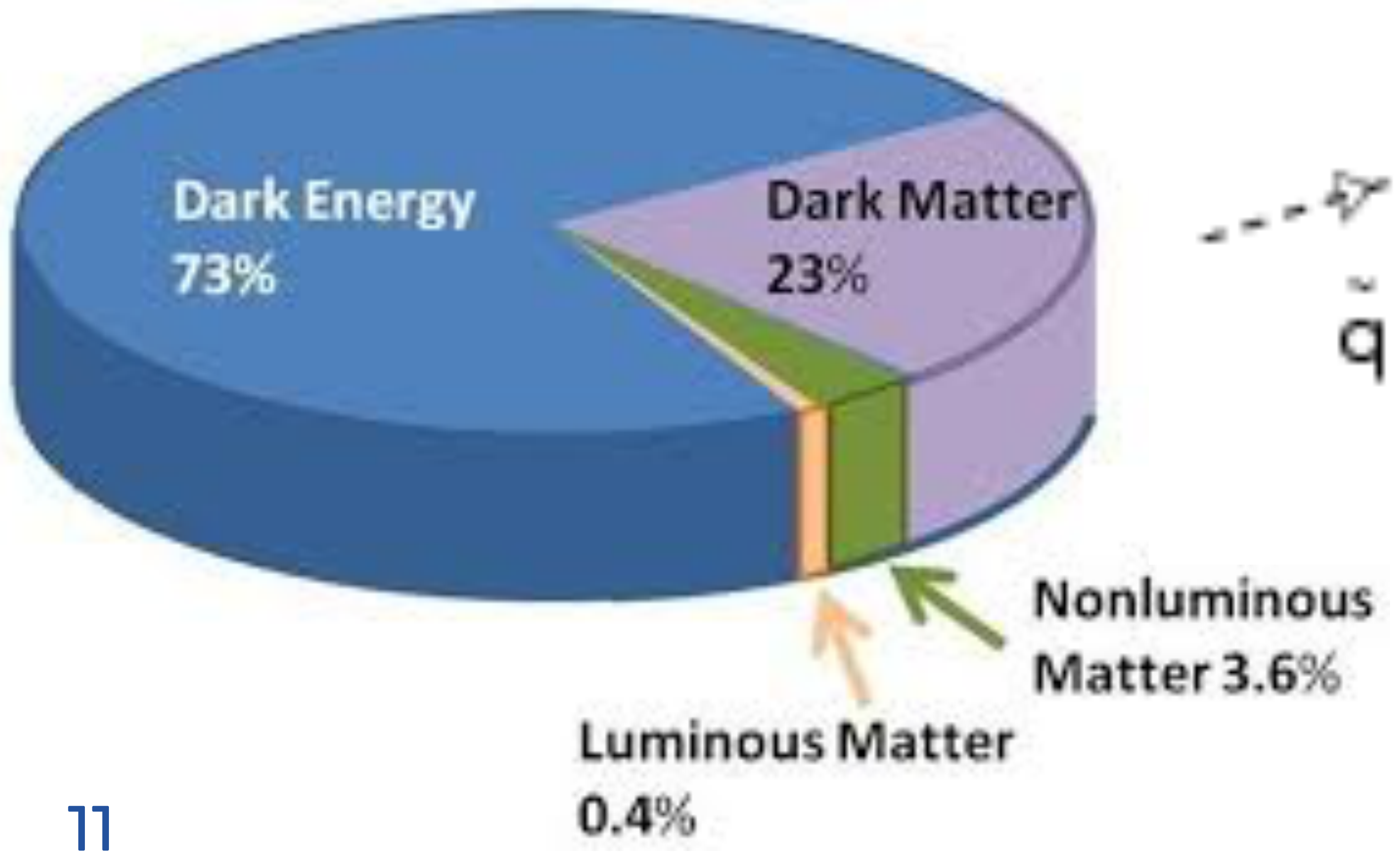
HOW R-PARITY SOLVES THE PROBLEM

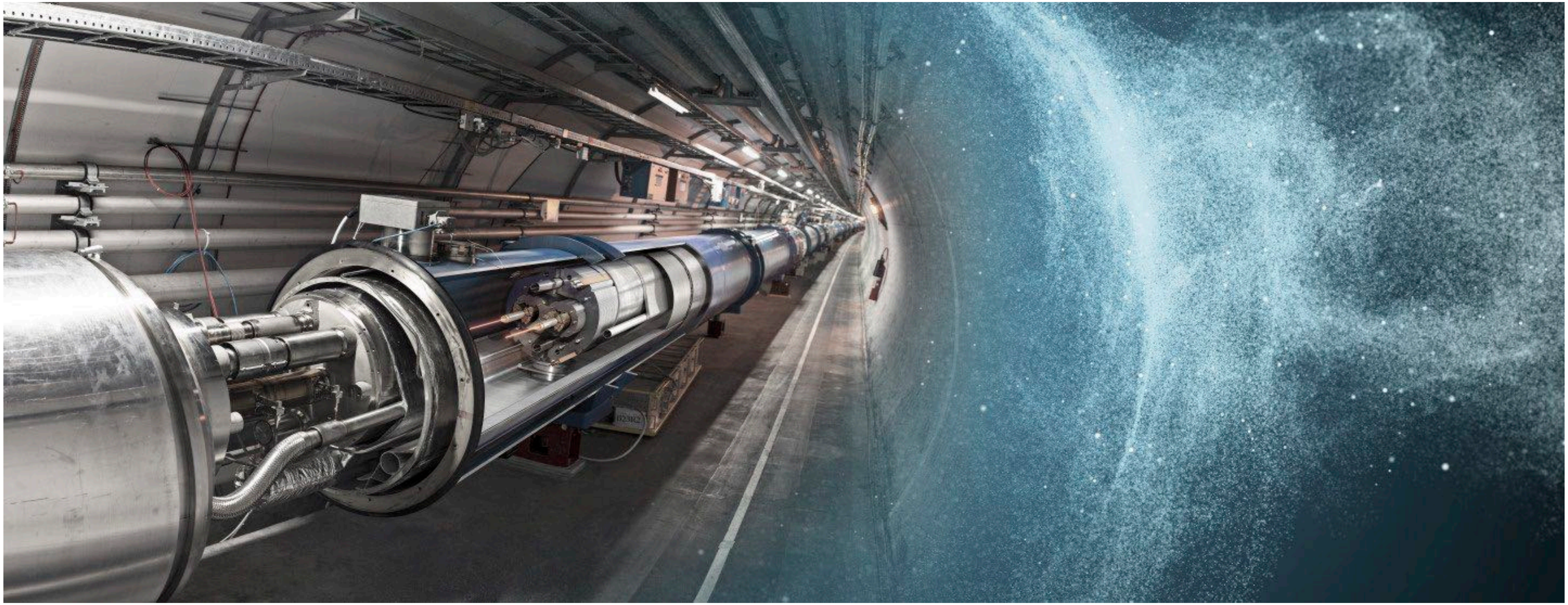
- *R-parity is always positive (negative) for SM particles (SUSY partners)*
- *As a consequence, only an even number of SUSY particles can be created in proton collisions (R=1)*
- *When decaying, a SUSY particle (R=-1) has to produce another SUSY particle*
- *So, the lightest SUSY particle cannot decay:*
    - *you have a Dark matter candidate for free*







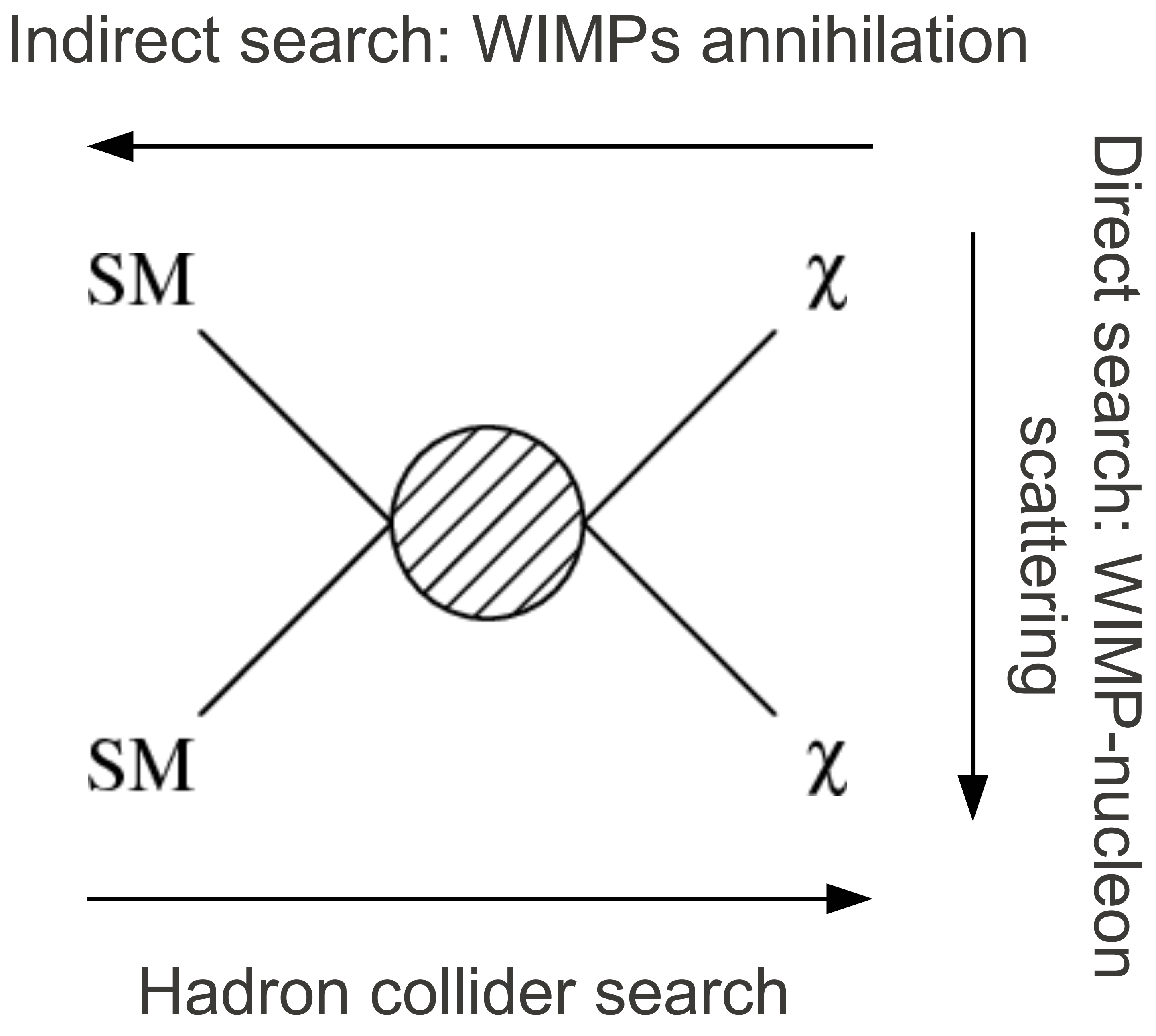


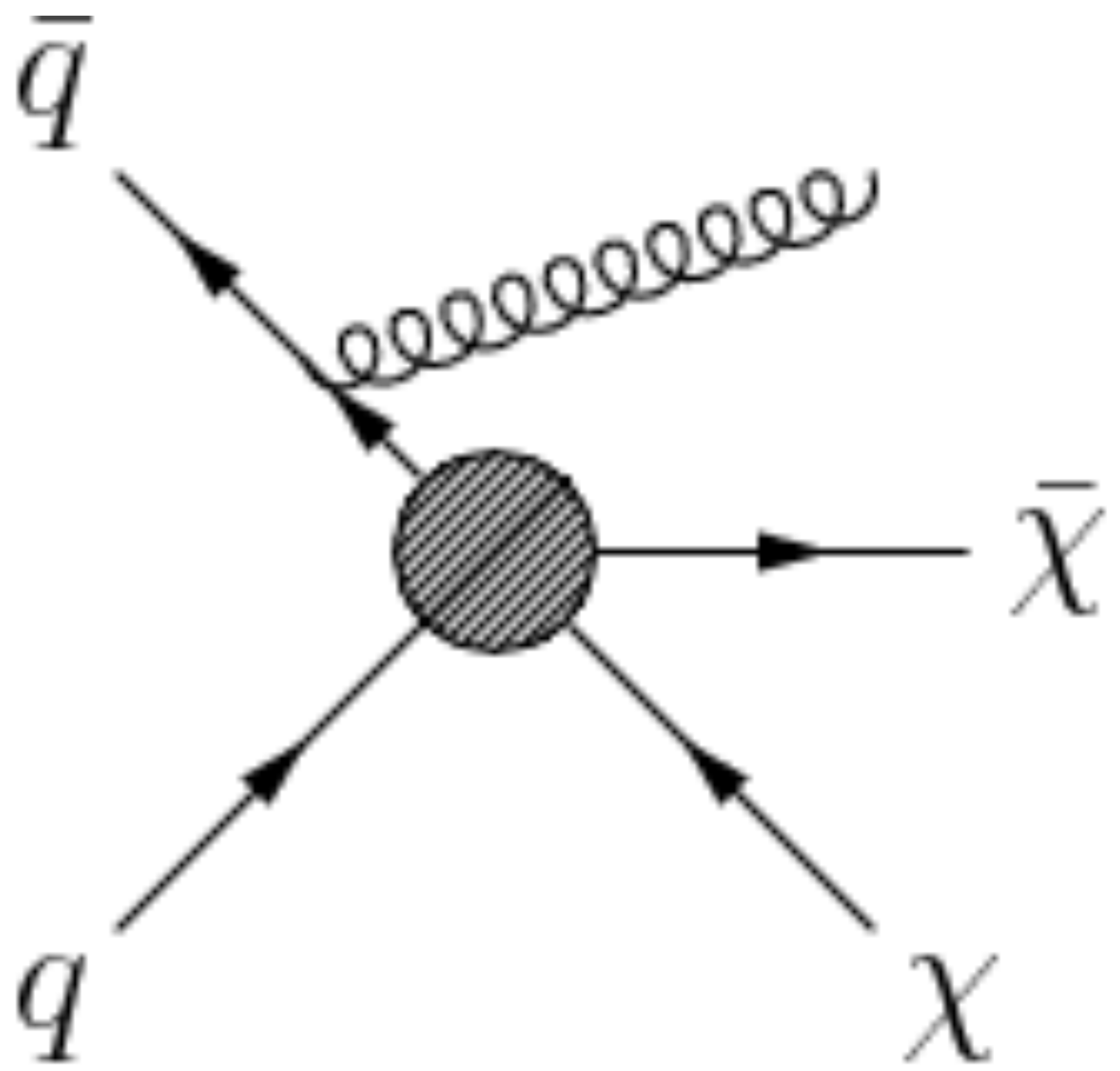


DARK MATTER DIRECT PRODUCTION

INVISIBLE AND MONOJET

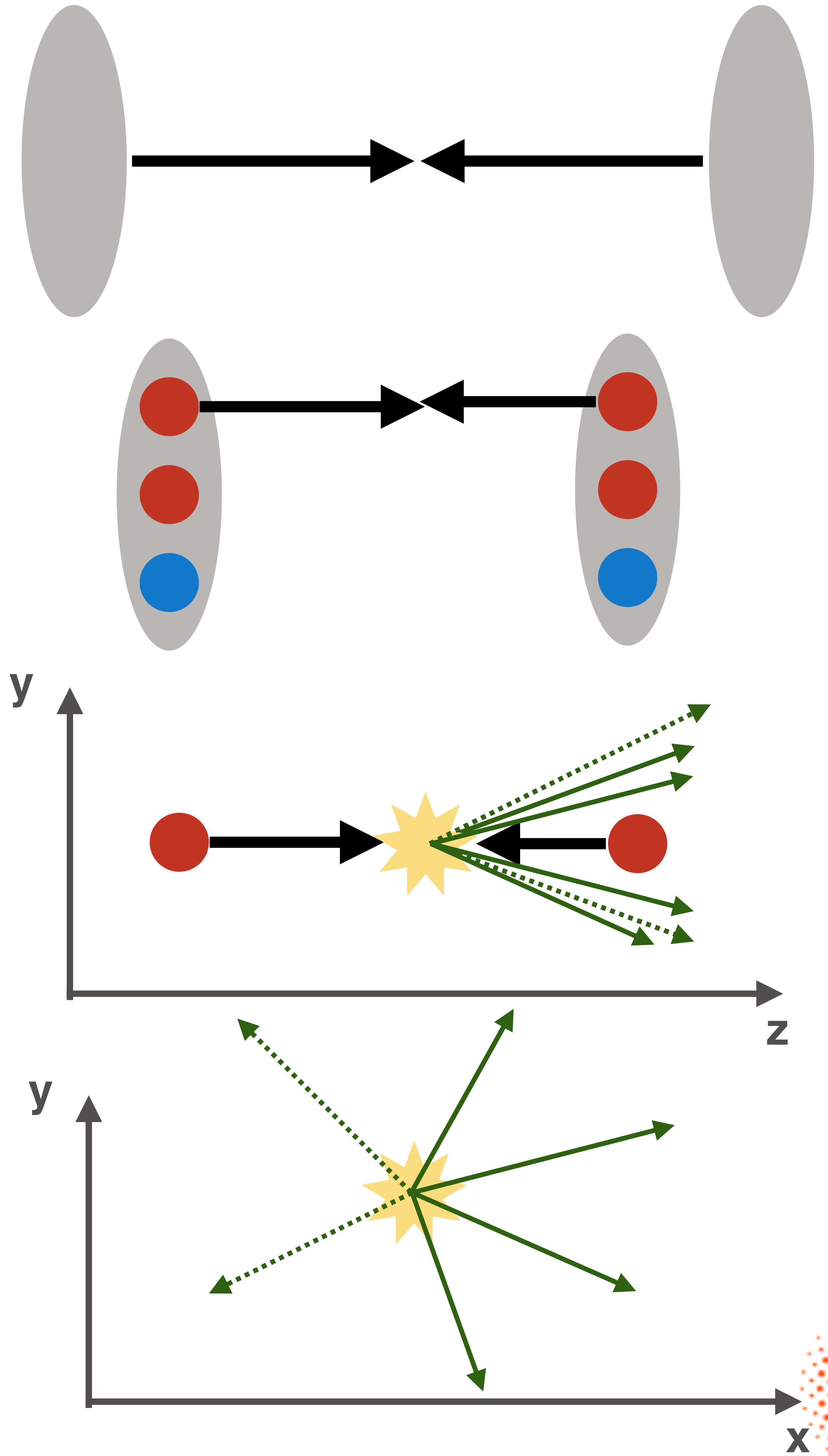
- DM can be produced at the LHC with a process similar to scattering in underground experiments
- But DM is invisible to our detectors, unless something else is produced with DM
- For example, a quark/gluon can be radiated before the collision
- These events look like a single high-pT jet of particles
- Events like this can happen also in the Standard Model. One needs to measure the background



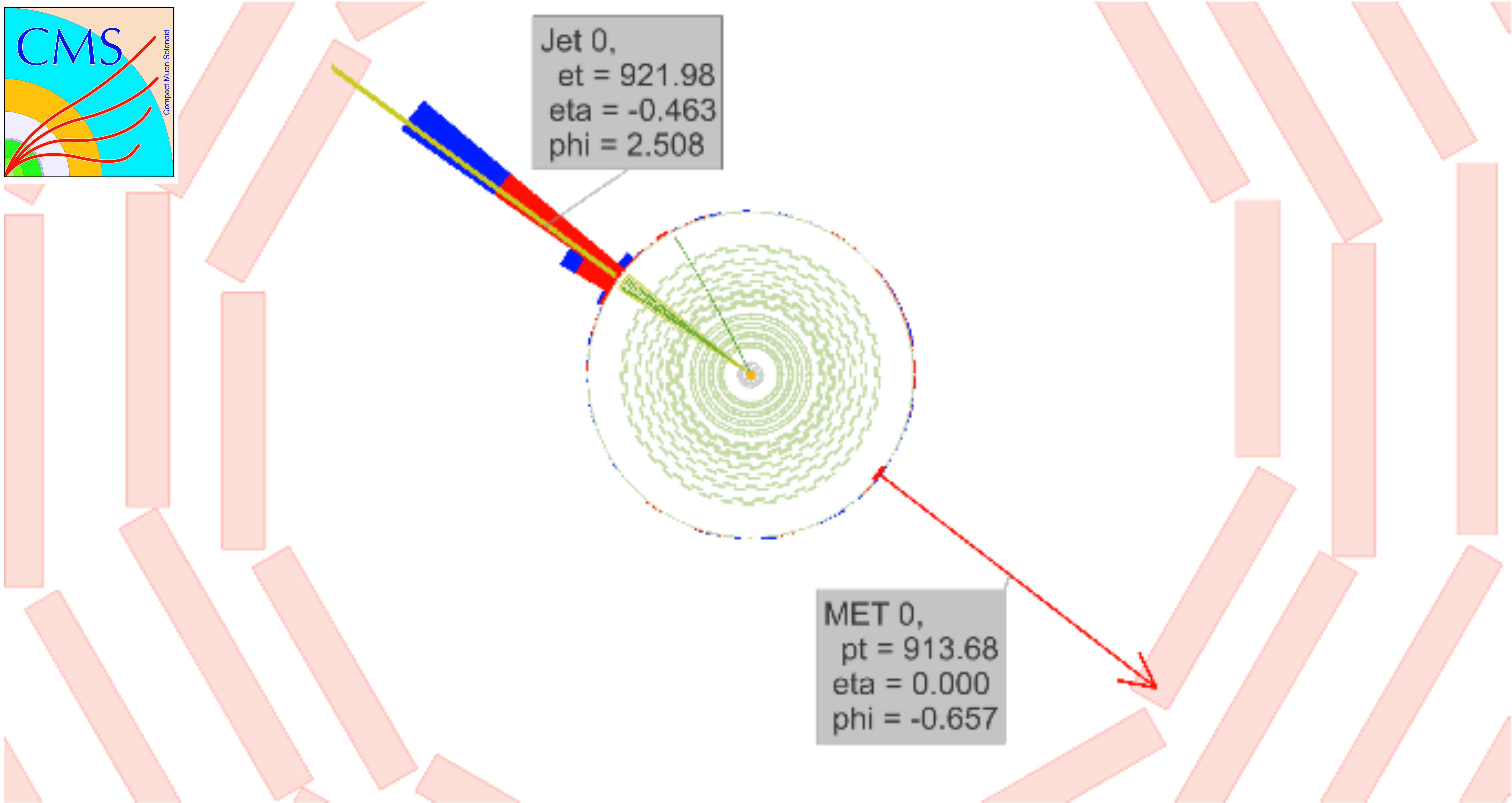


MISSING TRANSVERSE ENERGY

- *Two protons with same energy collide*
- *Actually, the collision is between quarks/gluons in the proton. They carry different fractions of the proton momentum*
- *As a result, there is a momentum imbalance ~ along the beam axes, but not in the transverse plane*
- *Transverse momenta should then balance. If some particle escaped undetected, the balance will be broken*



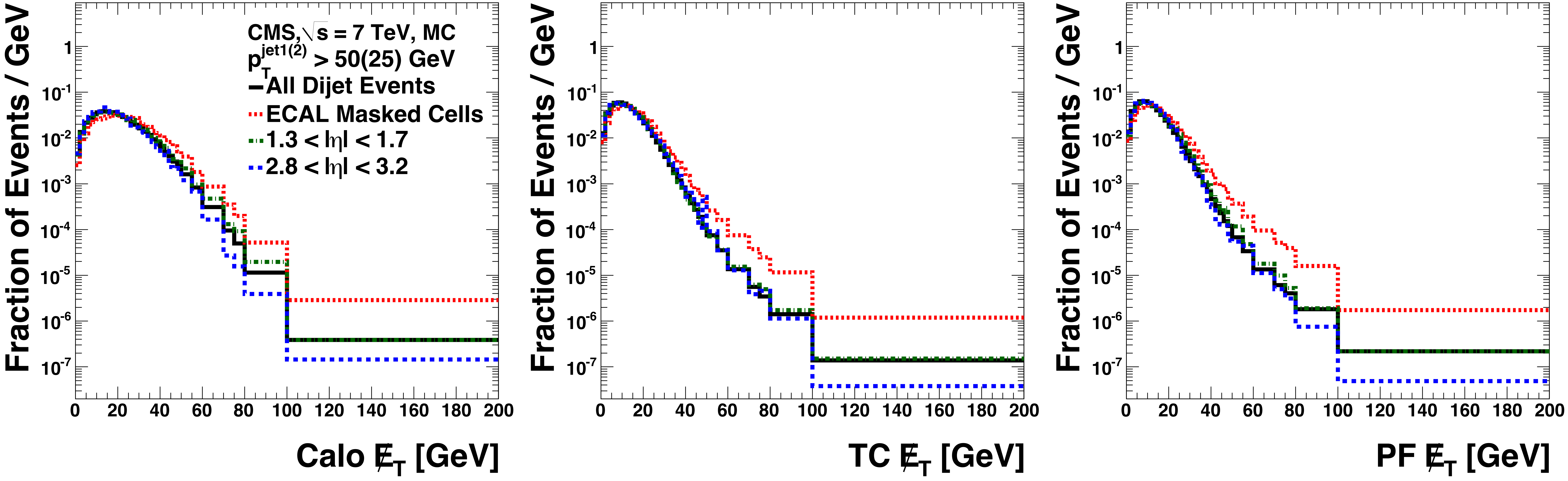
MISSING TRANSVERSE ENERGY



European Research Council

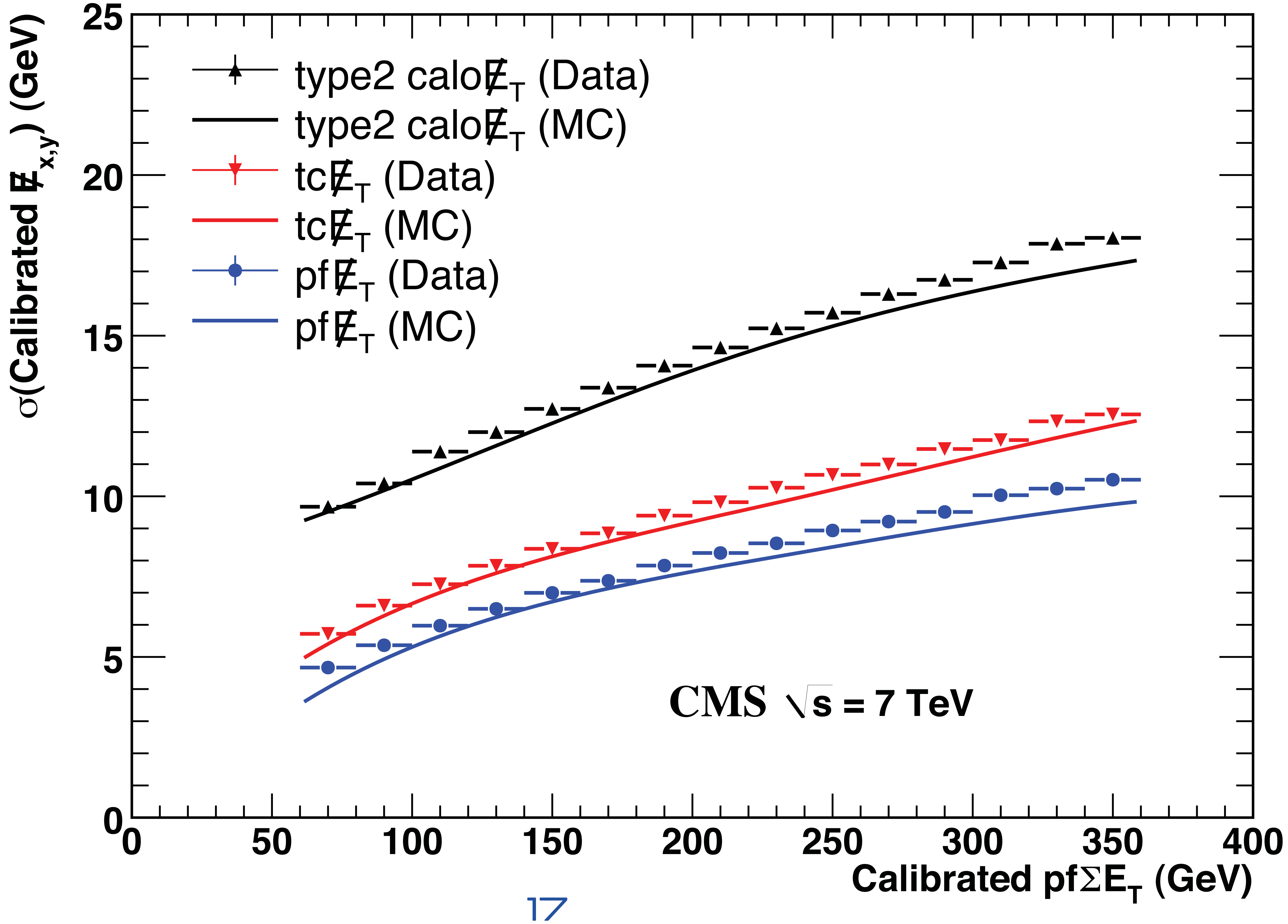
MISSING TRANSVERSE ENERGY

- *MET is computed summing the transverse momenta of reconstructed object* $E_T^{miss} = |-\Sigma_i \vec{p}_t^{\,i}|$
- *As for the jet, which objects? This choice determines the resolution*



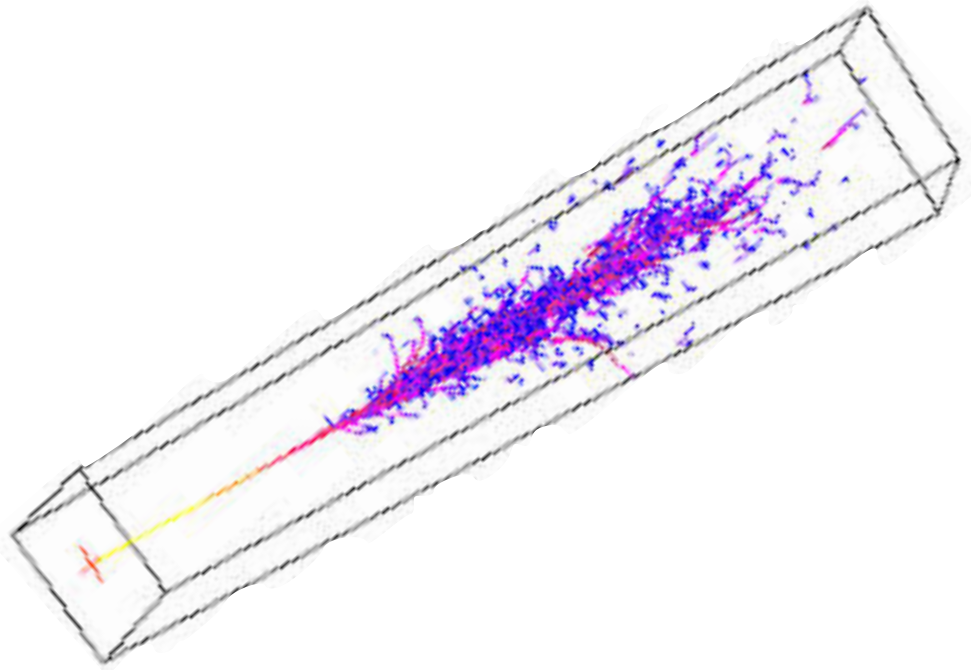
MISSING TRANSVERSE ENERGY

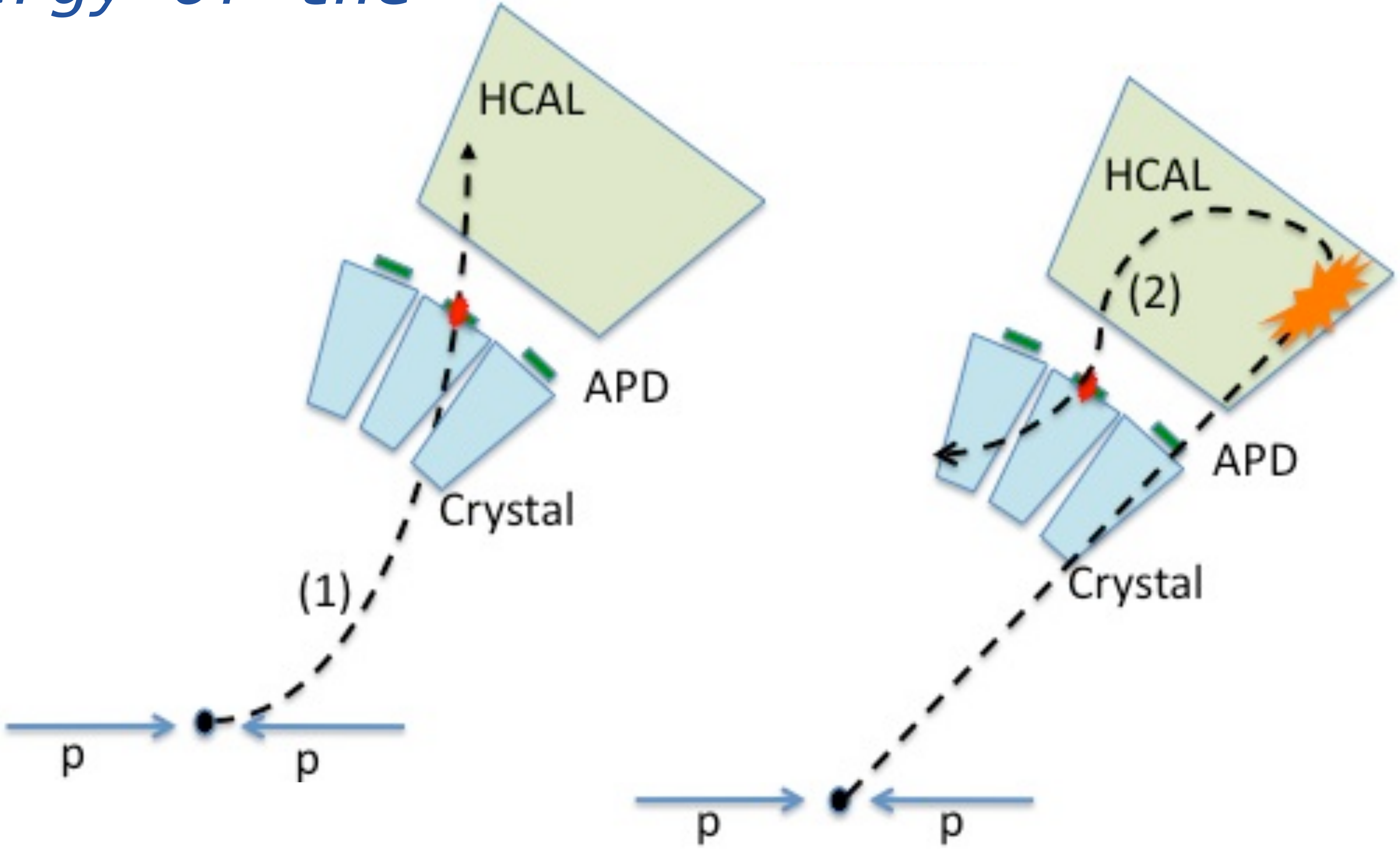
- MET is computed summing the transverse momenta of reconstructed object $E_T^{miss} = |-\Sigma_i \overrightarrow{p}_t^i|$
- As for the jet, which objects? This choice determines the resolution



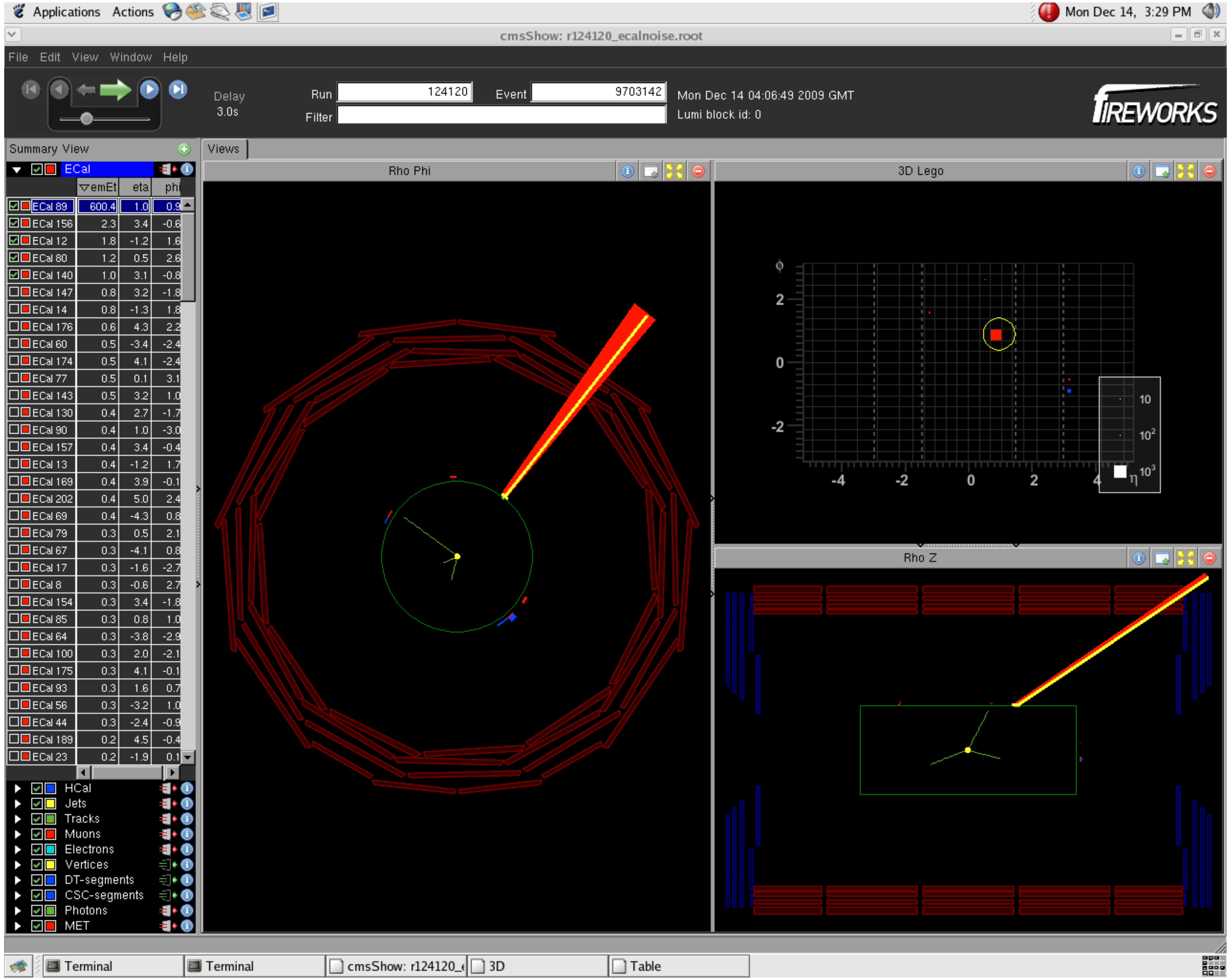
DETECTOR NOISE

- Energy in crystal calorimeters is collected by photomultipliers in the back
- Due to quantum efficiency (<<1) the collected energy is smaller than the energy of the incoming particle
- An incoming particle can deposit energy directly to the photomultiplier directly.
- The calibration constant (E collected Voltage of the signal E original particle) mistakenly translates this into a large deposit

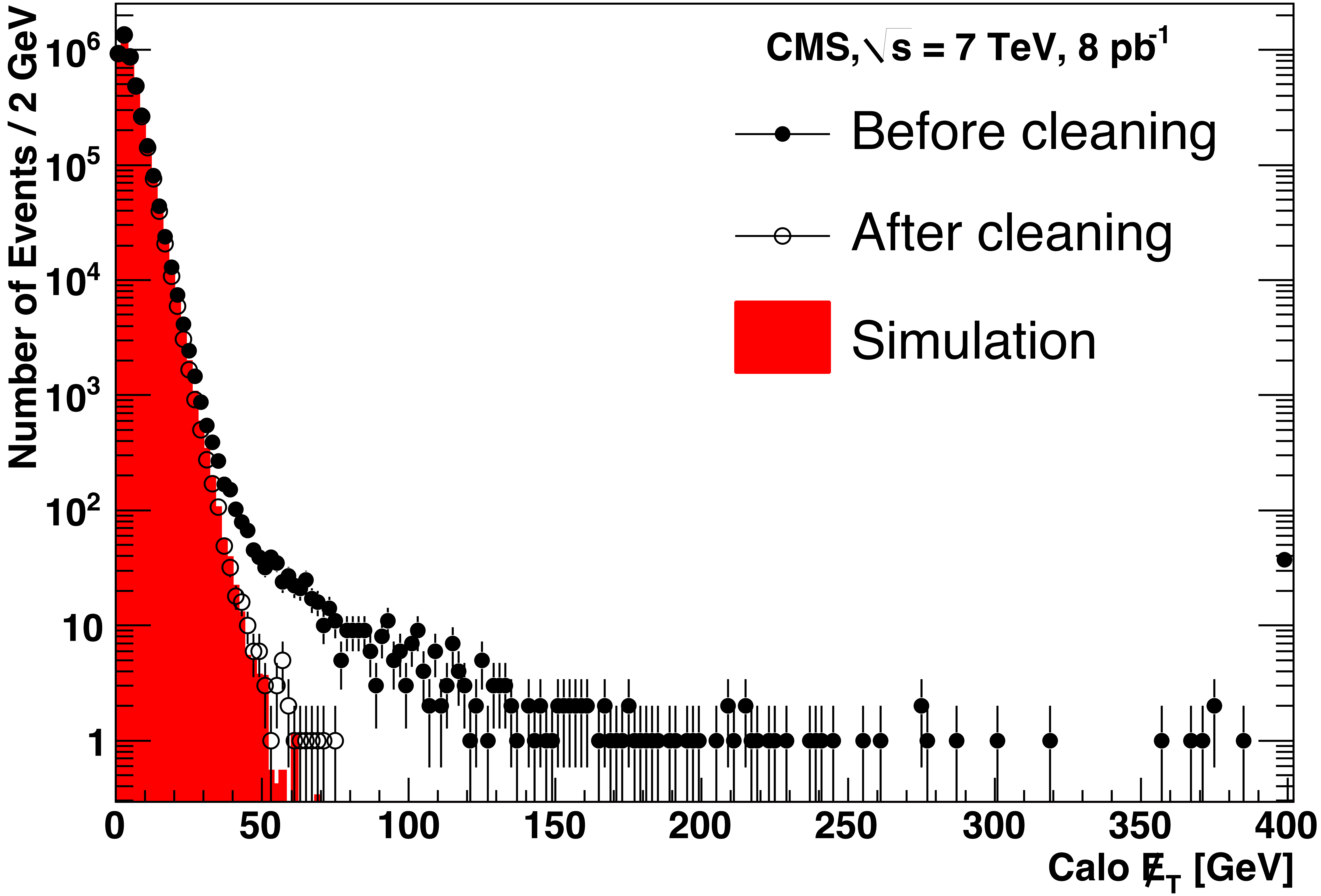




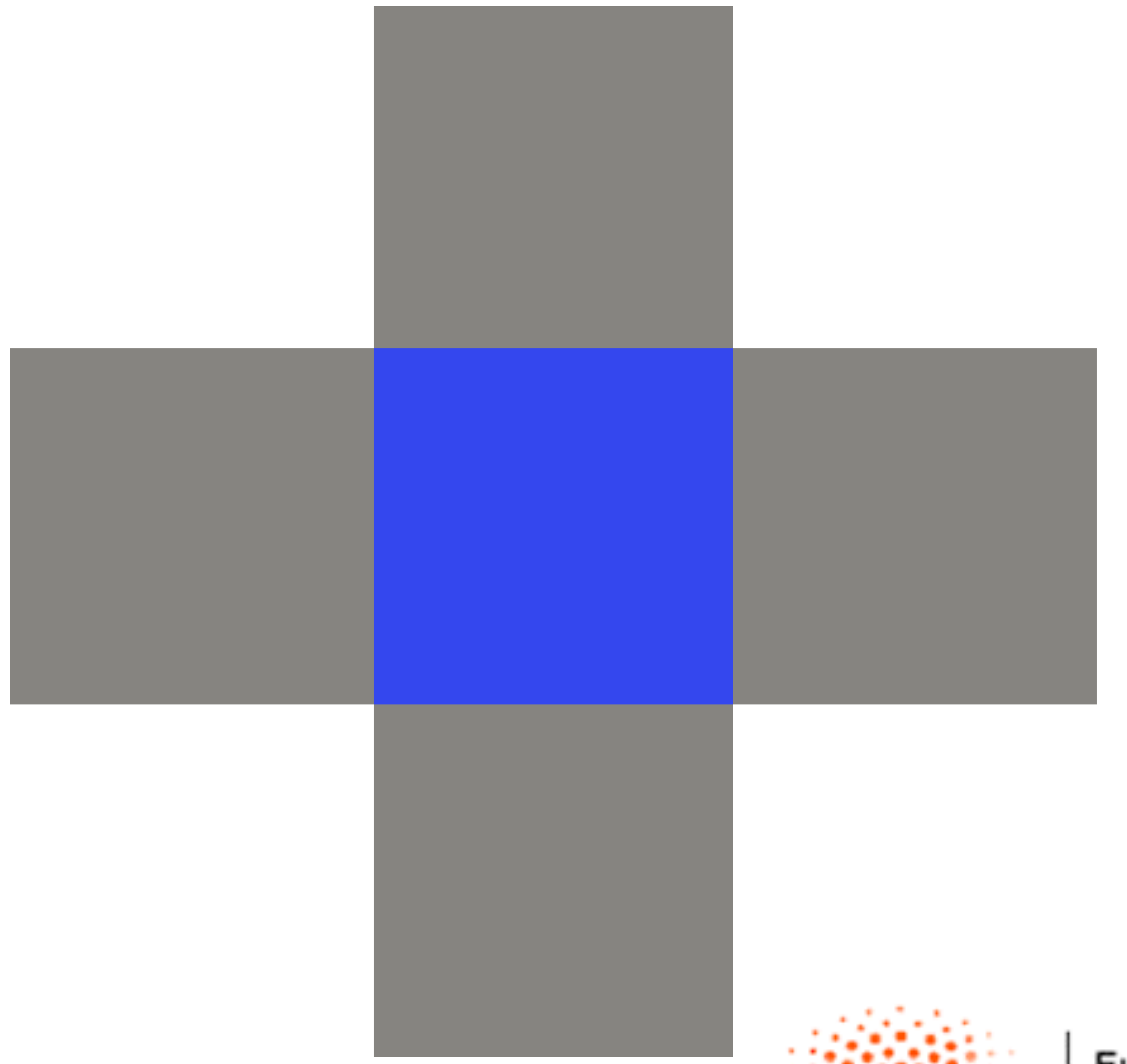
DETECTOR NOISE



CLEANING FILTERS

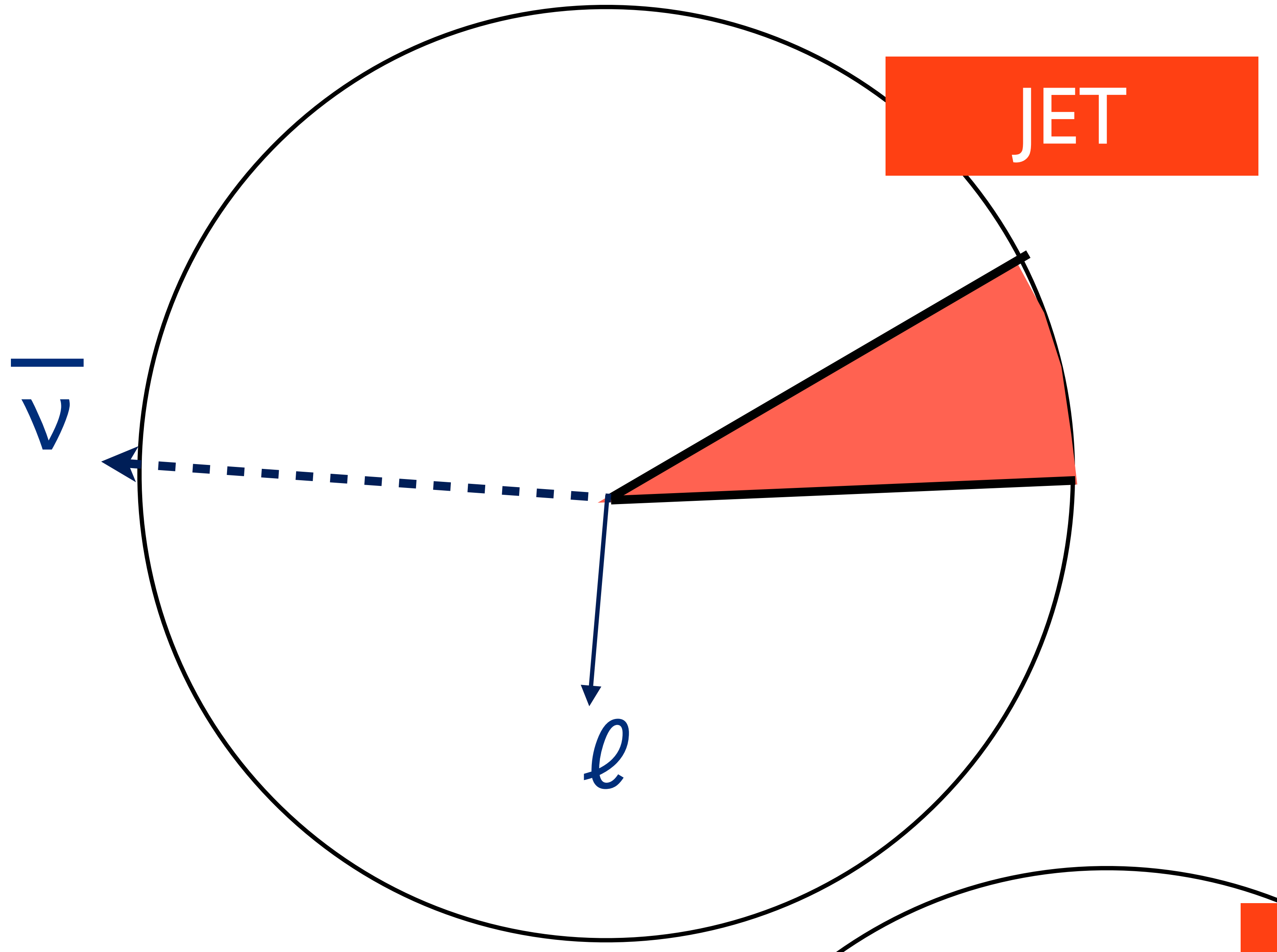


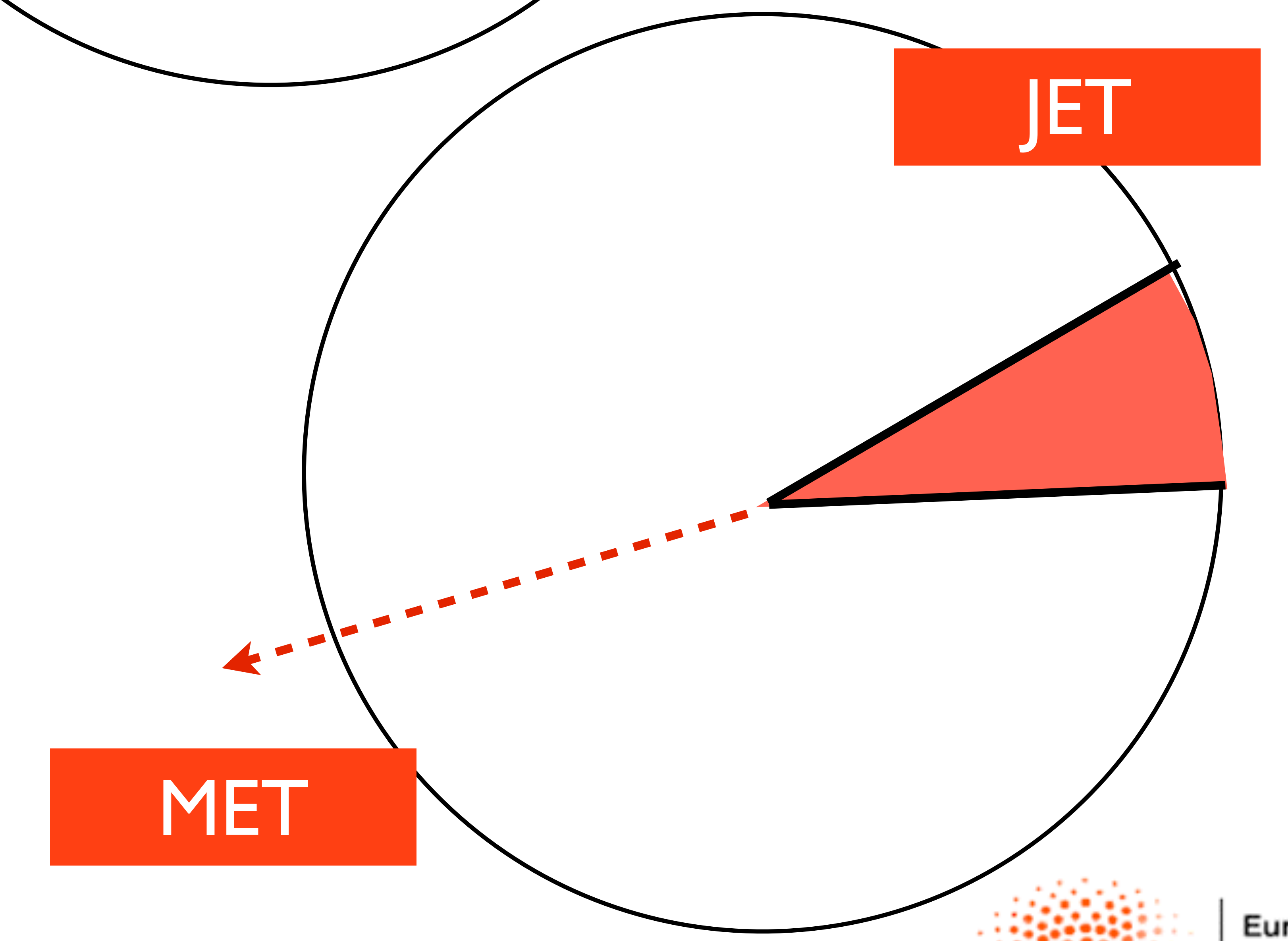
Filter example:
if the energy in the crystals around have small energy compared to central crystal, then this is (most likely) a spike (Swiss cross filter)



SM BACKGROUND: W+JETS

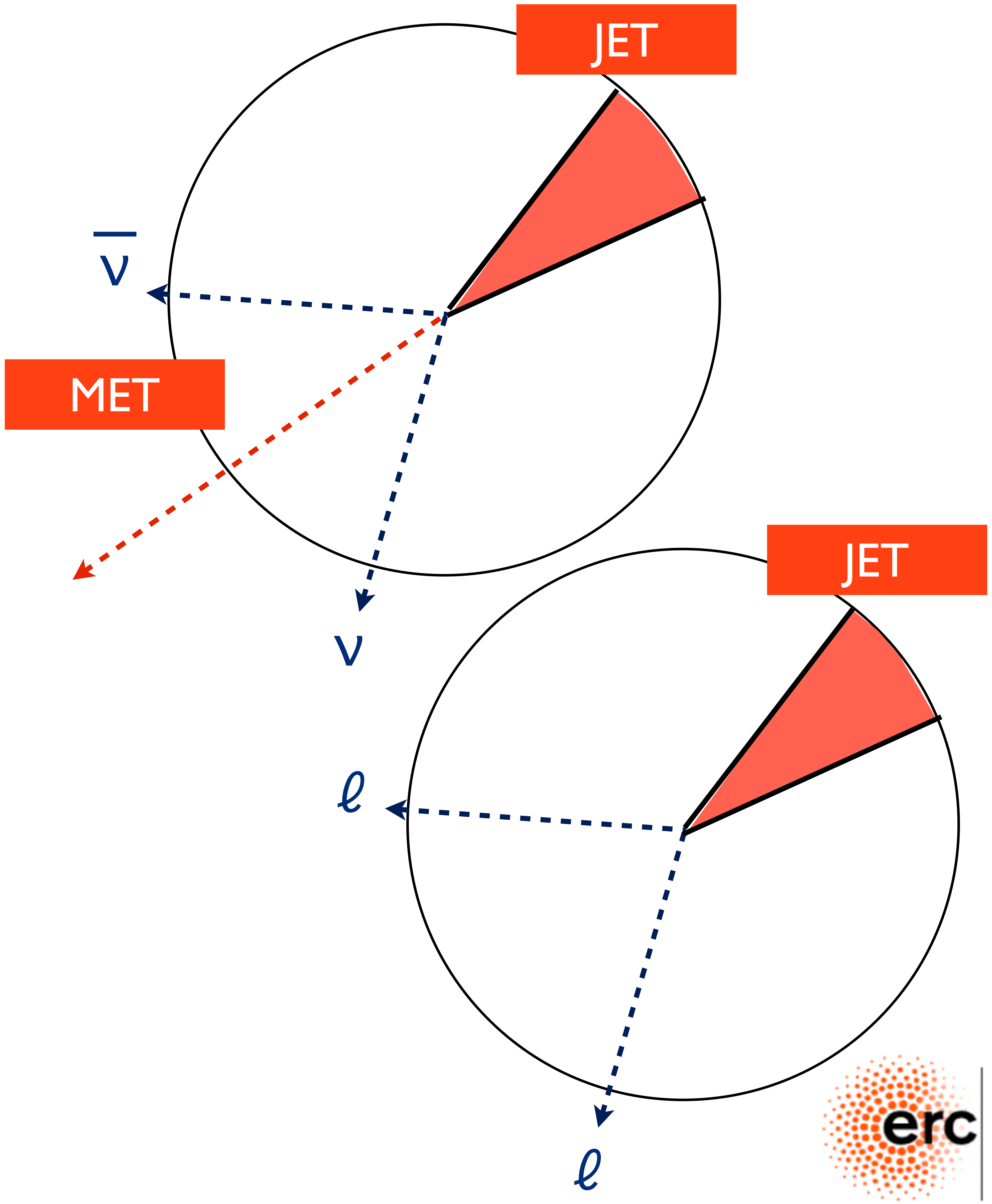
- W bosons are produced in LHC collisions, recoiling against a jet
- 30% of the time the W decays to a lepton and a neutrino
- Sometimes the lepton is not reconstructed. The event looks like monojet
- We can predict how many events like this we expect, counting the events for which we see the leptons (we know the probability of missing the lepton)





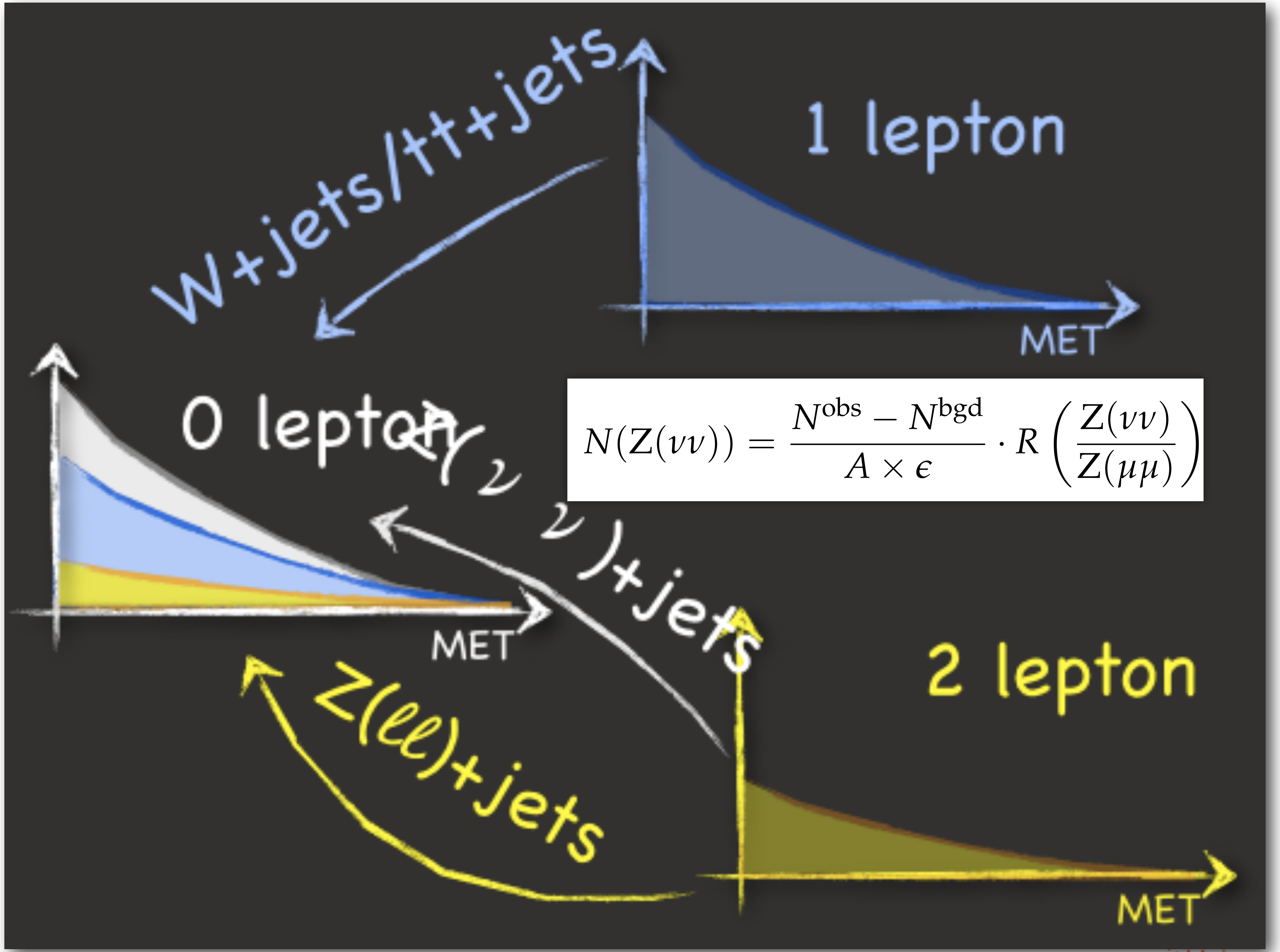
SM BACKGROUND: Z+JETS

- *Z bosons are produced in LHC collisions, recoiling against a jet*
- *~20% of the time the Z decays to a neutrino pair*
- *~10% of the time the Z decays to a lepton pair*
- *We can predict how many events with neutrinos we expect, counting the events for which we see the leptons (we know the probability of missing the lepton)*



SEARCH STRATEGY

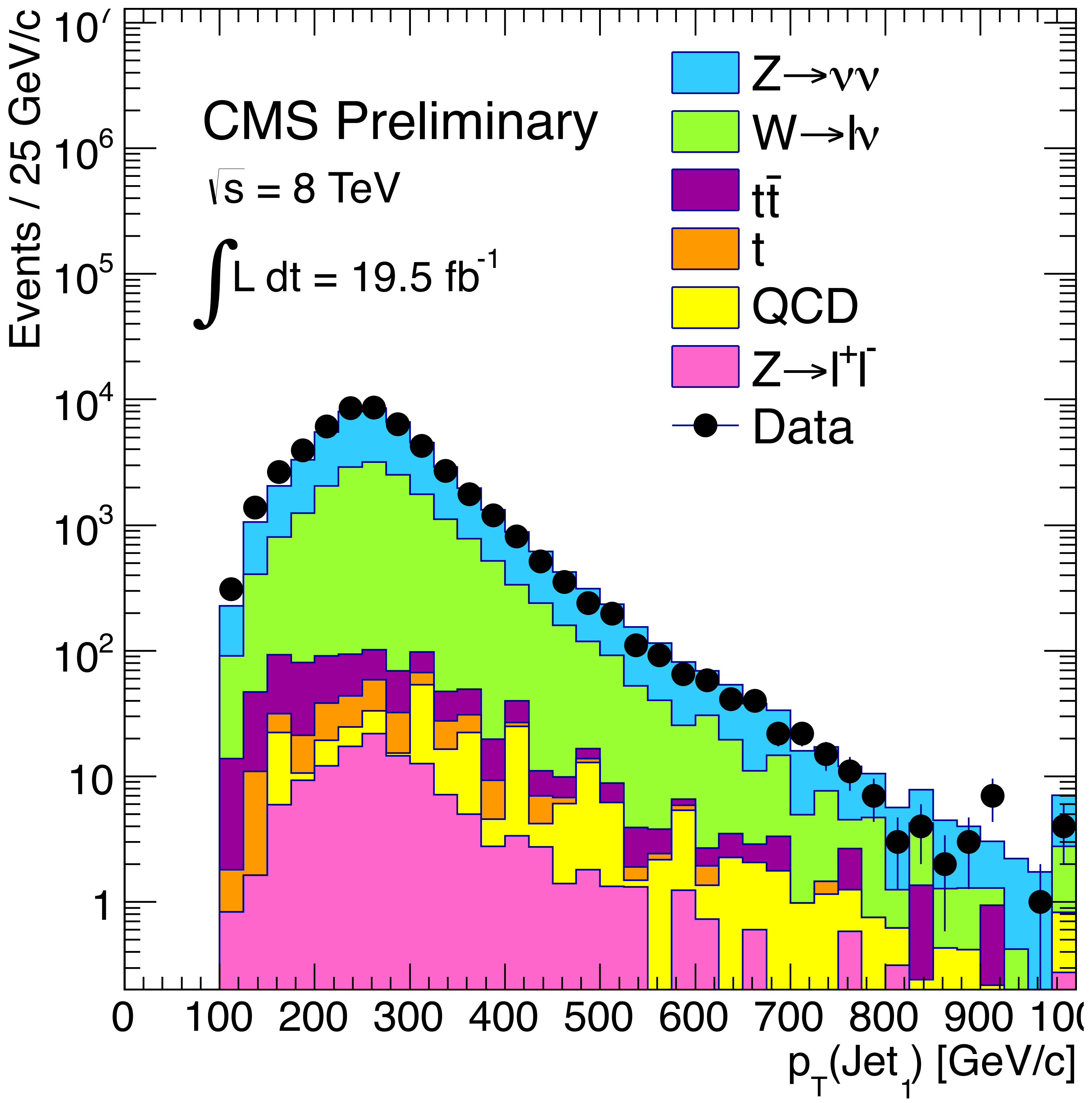
- *Events are selected requiring a high-energy jet and large missing transverse energy*
- *The same selection is applied to events with 1 or 2 leptons*
- *Simulation is use to connect the samples*

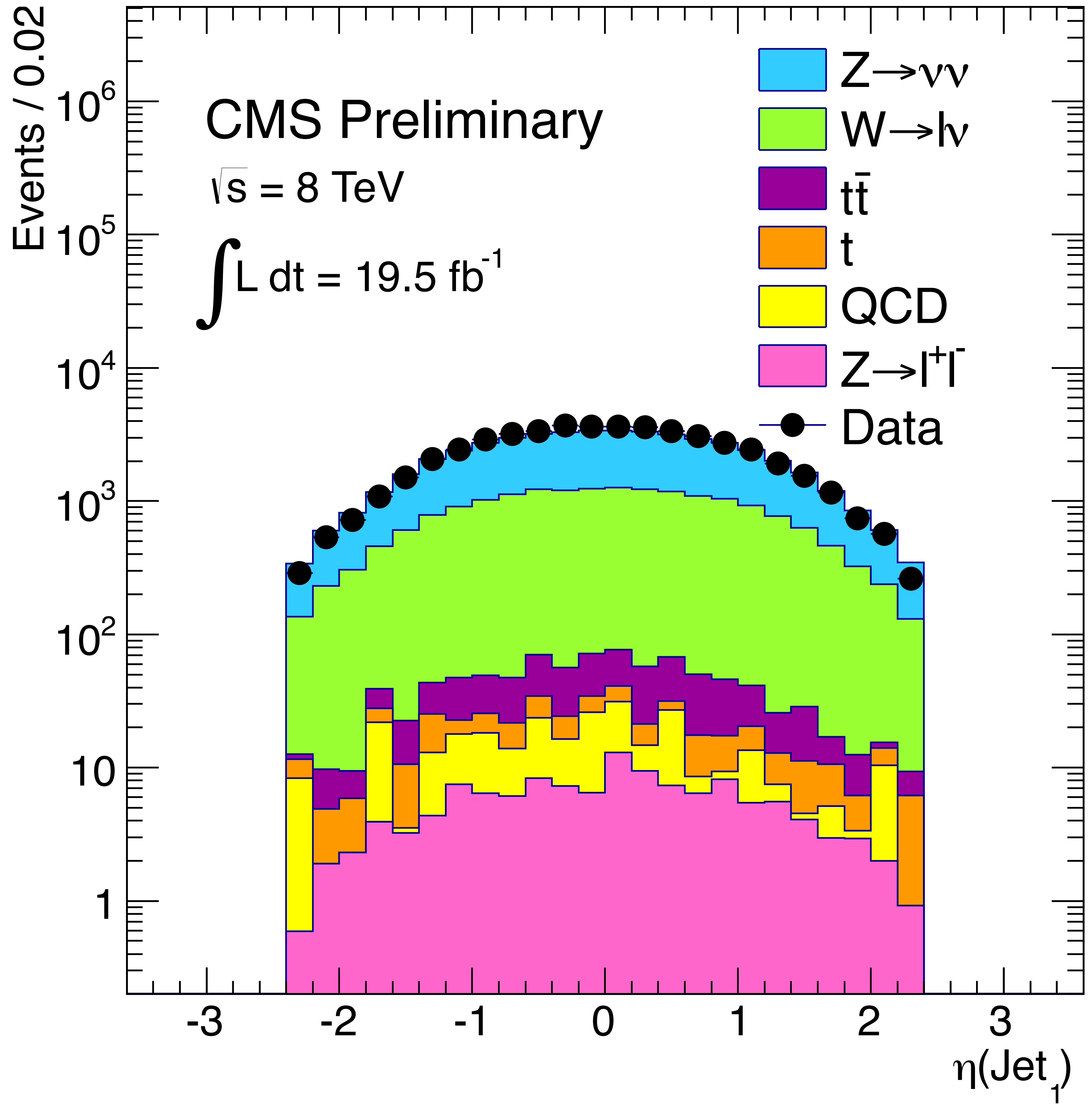


$$N(Z(\nu\nu)) = \frac{N^{obs} - N^{bgd}}{A \times \epsilon} \cdot R\left(\frac{Z(\nu\nu)}{Z(\mu\mu)}\right)$$

SEARCH STRATEGY

- Events are selected requiring a high-energy jet and large missing transverse energy
- The same selection is applied to events with 1 or 2 leptons
- Simulation is use to connect the samples

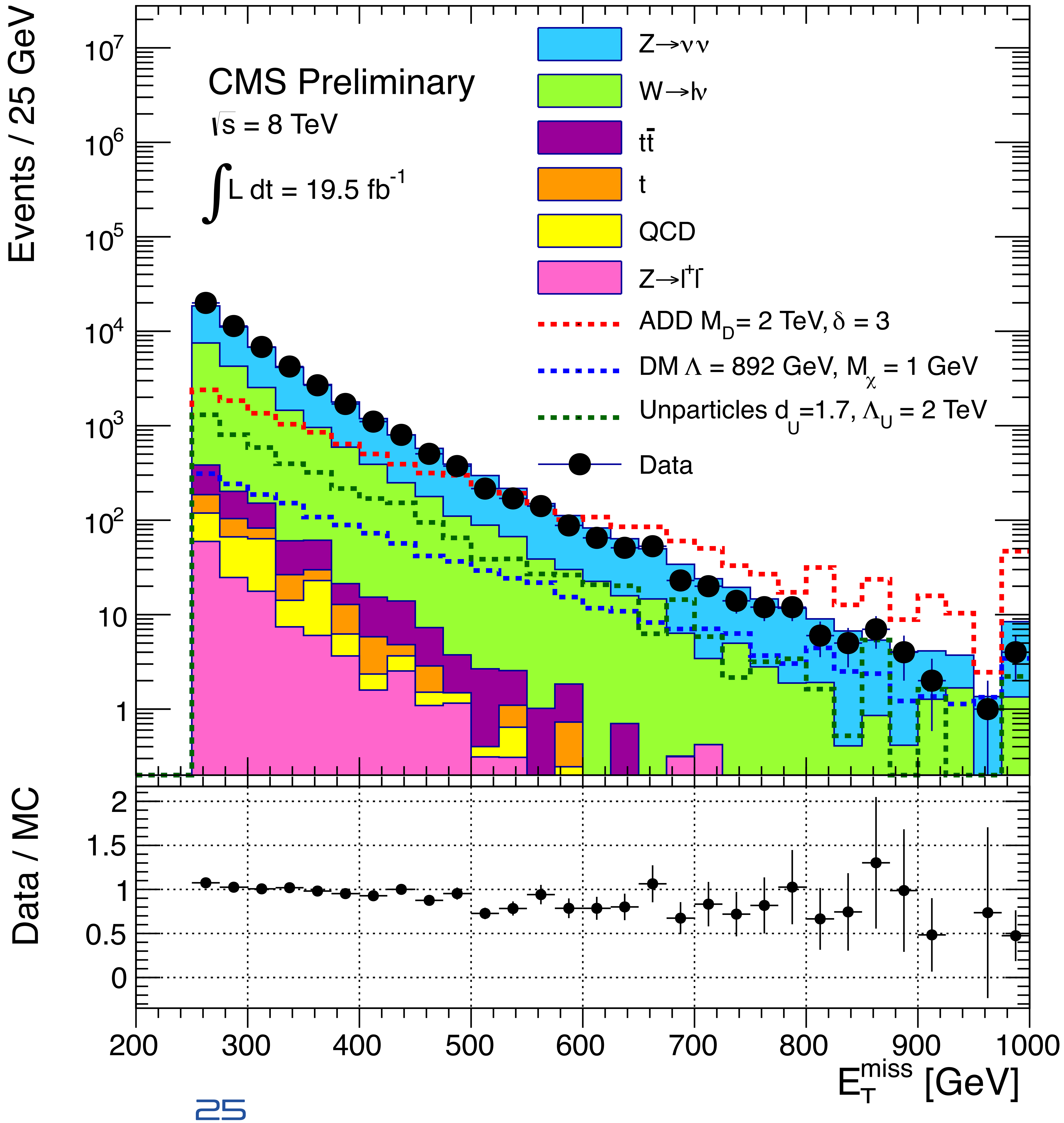




	Z+jets	W+jets	$Z(\nu\nu)$	$t\bar{t}$	Single t	QCD	All MC	Data
$E_T^{miss} > 250$	3405.2	0.5	0.0	27.4	10.6	0.0	3444	3626
$E_T^{miss} > 300$	1493.7	0.0	0.0	8.8	4.1	0.0	1507	1485
$E_T^{miss} > 350$	696.2	0.0	0.0	4.4	3.1	0.0	704	663
$E_T^{miss} > 400$	344.9	0.0	0.0	0.6	0.3	0.0	346	323
$E_T^{miss} > 450$	177.1	0.0	0.0	0.0	0.0	0.0	177	173
$E_T^{miss} > 500$	97.1	0.0	0.0	0.0	0.0	0.0	97	84
$E_T^{miss} > 550$	54.4	0.0	0.0	0.0	0.0	0.0	54	47

SEARCH STRATEGY

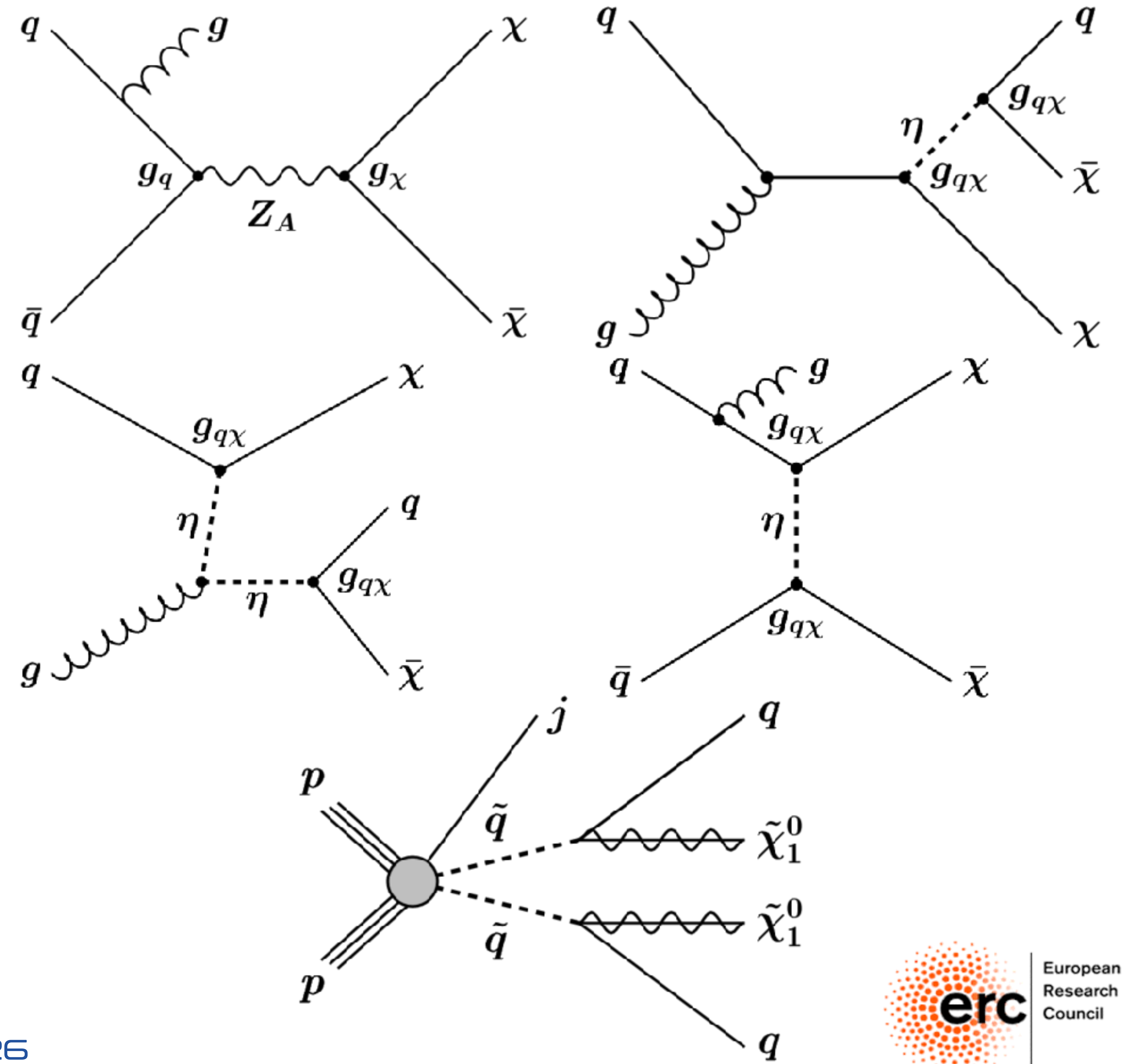
- *Events are selected requiring a high-energy jet and large missing transverse energy*
- *The same selection is applied to events with 1 or 2 leptons*
- *Simulation is use to connect the samples*
- *The prediction agrees with the expectation, ie. no signal is found*



INTERPRETATION

- *In absence of a signal, the result is interpreted as an exclusion limit (at 95% confidence level) on the existence of Dark Matter*

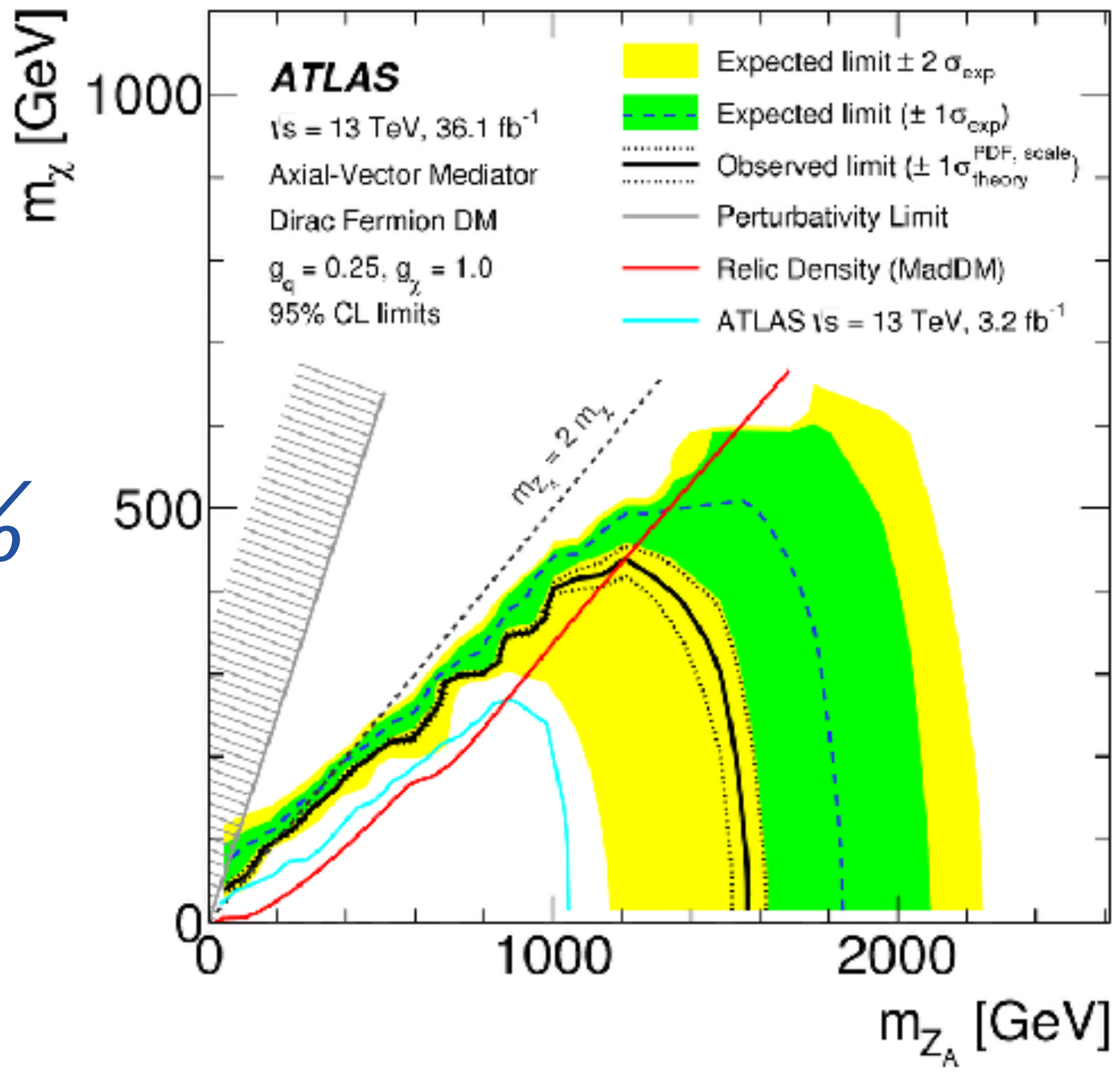
- *Several scenarios considered, depending on the nature of the mediator (simplified models not directly connected to SUSY)*

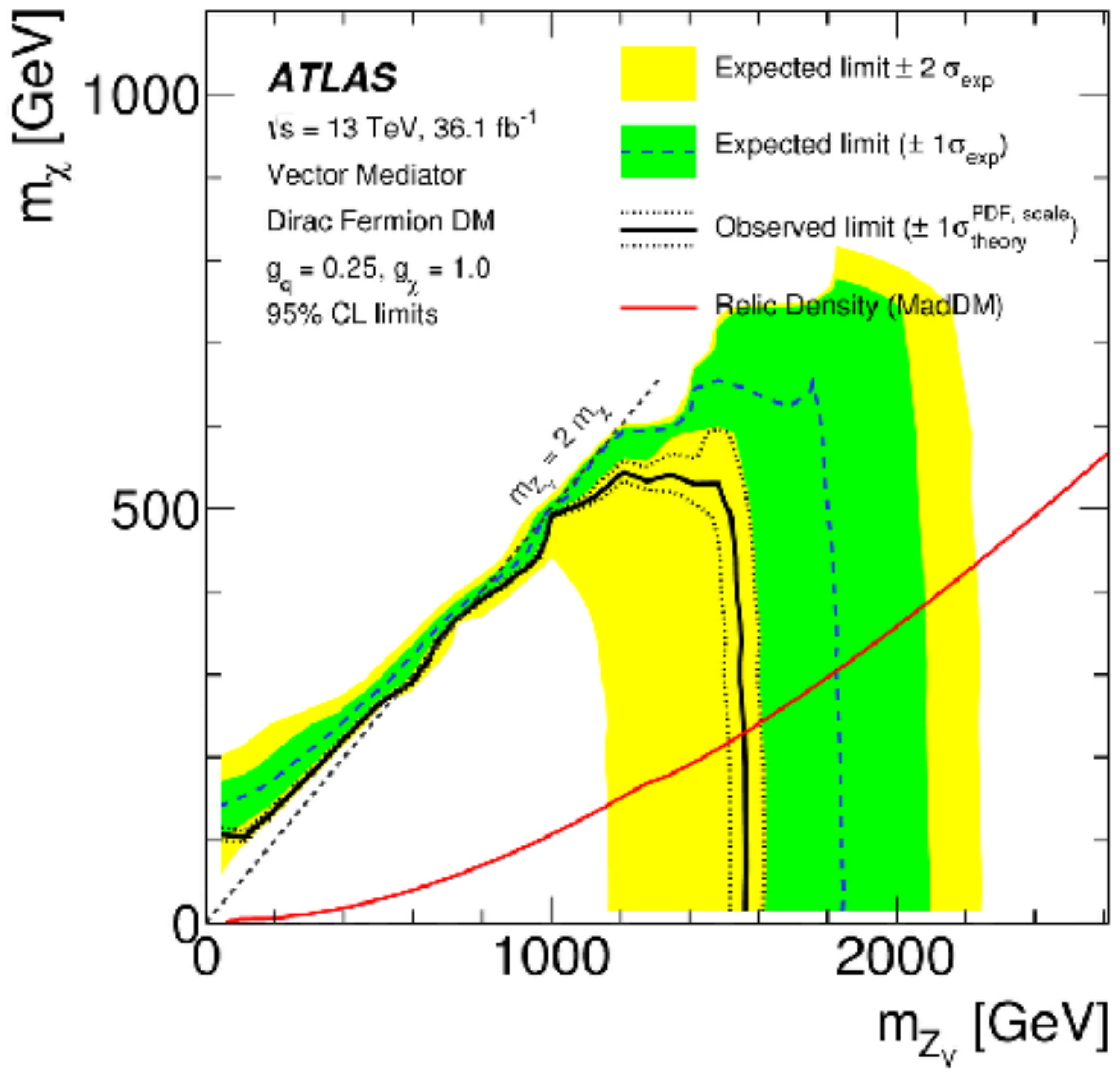


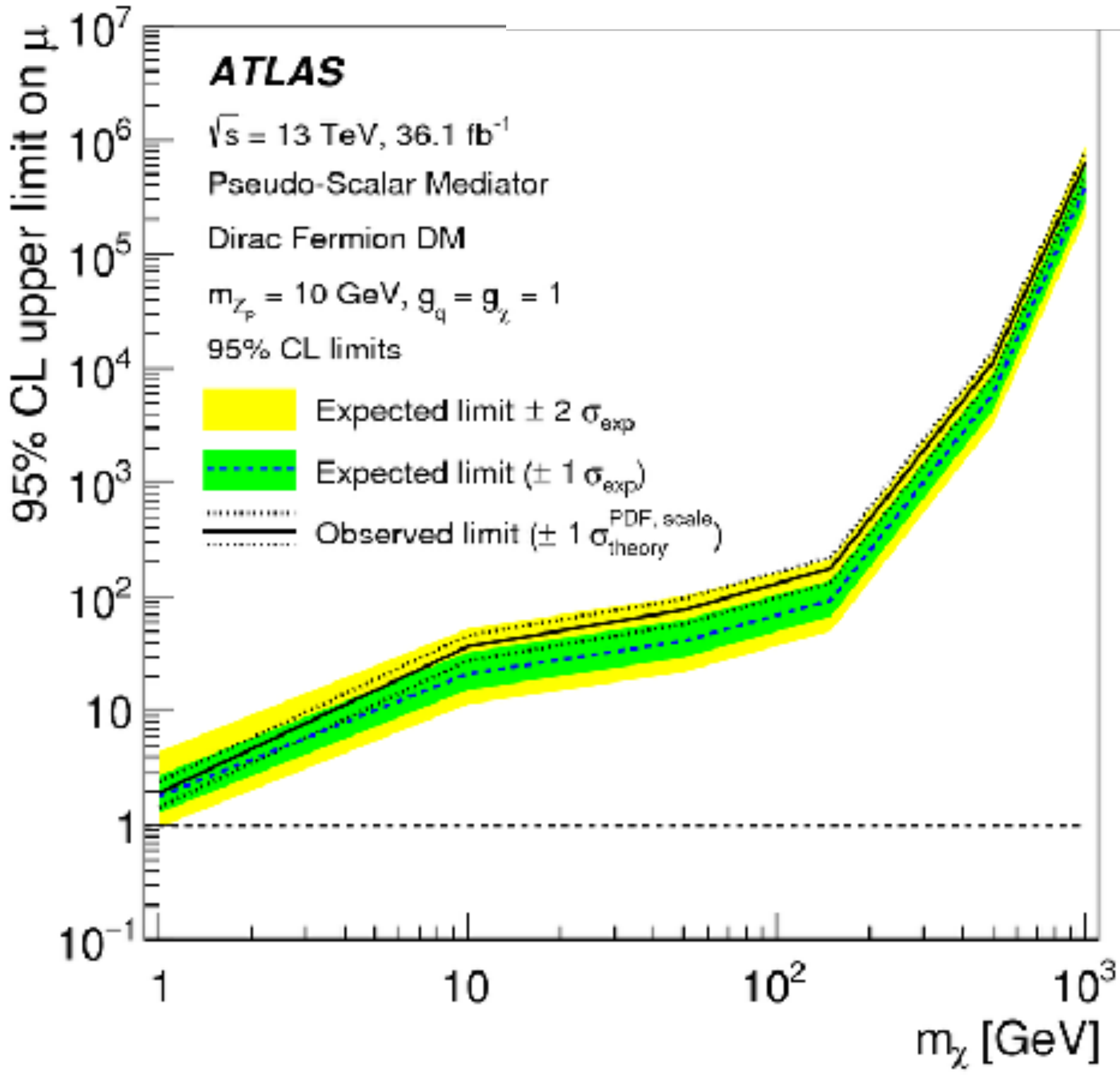
INTERPRETATION

- *In absence of a signal, the result is interpreted as an exclusion limit (at 95% confidence level) on the existence of Dark Matter*

- *Several scenarios considered, depending on the nature of the mediator (simplified models not directly connected to SUSY)*

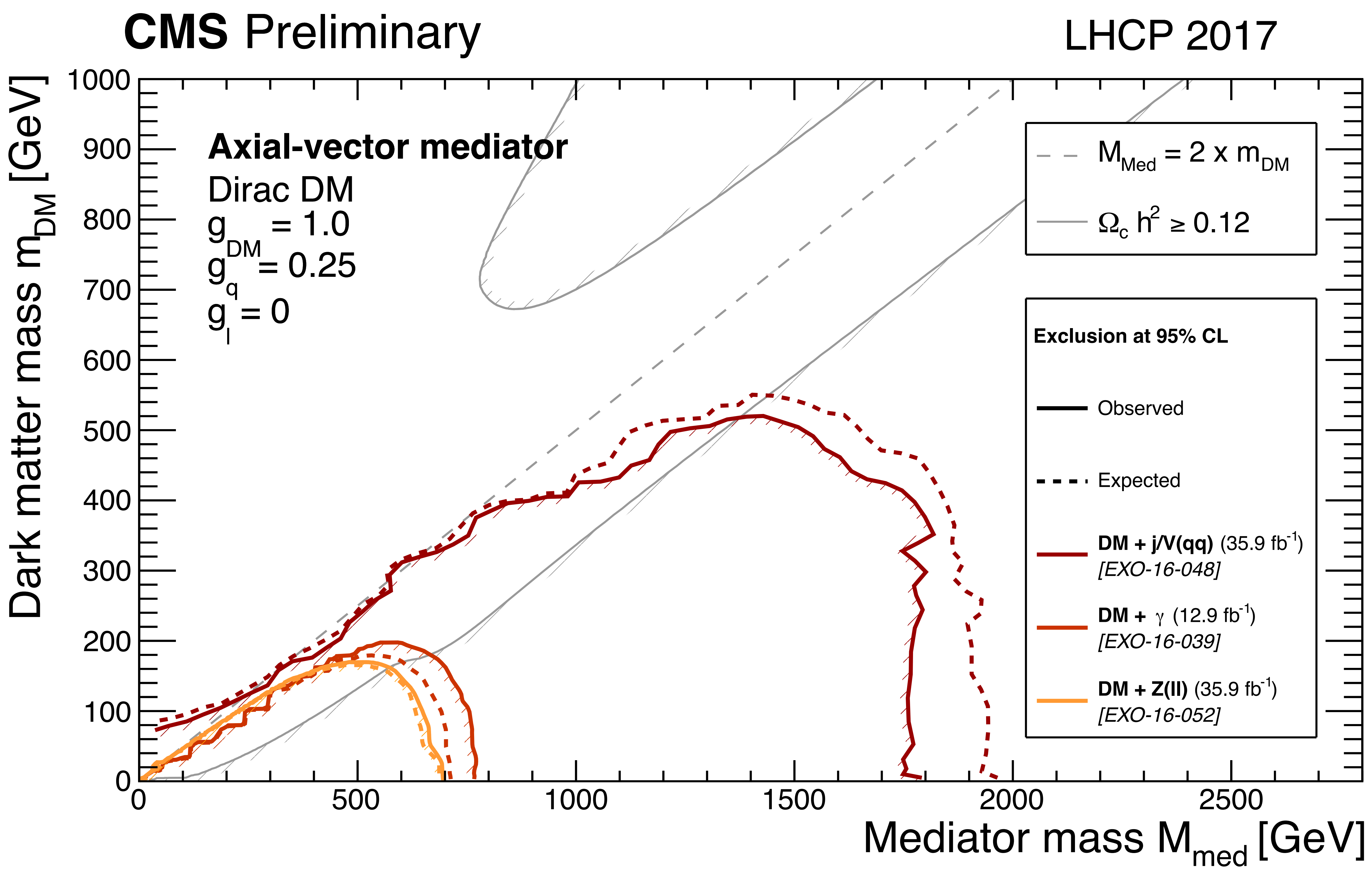


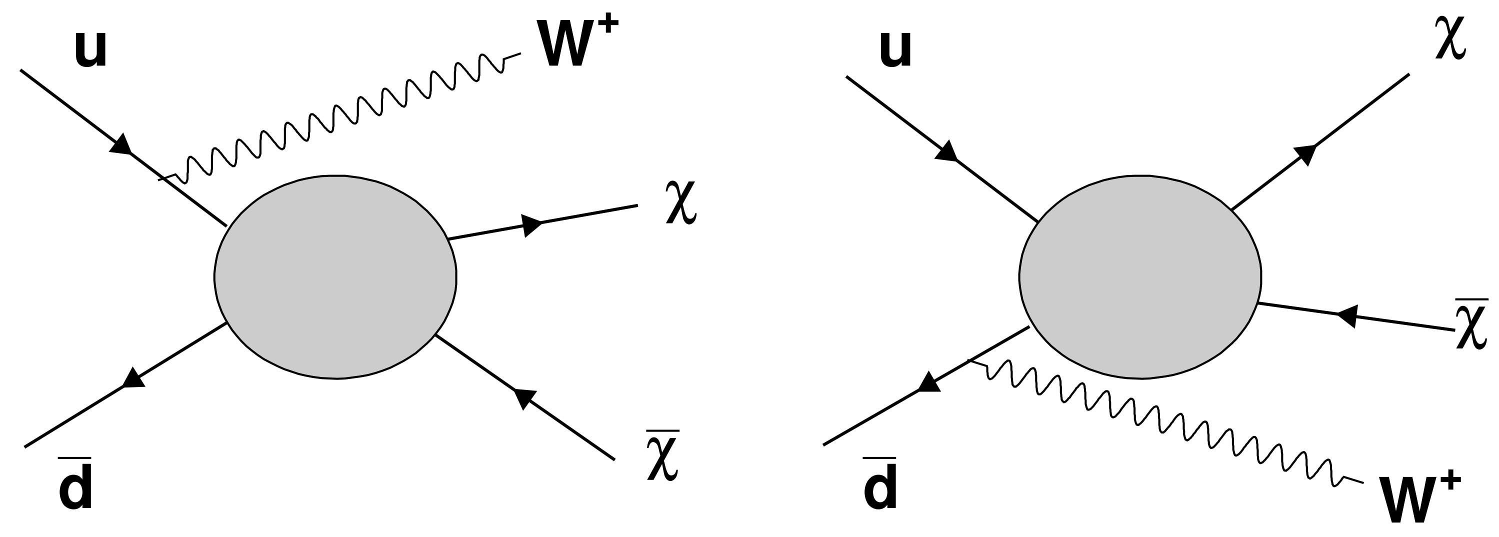


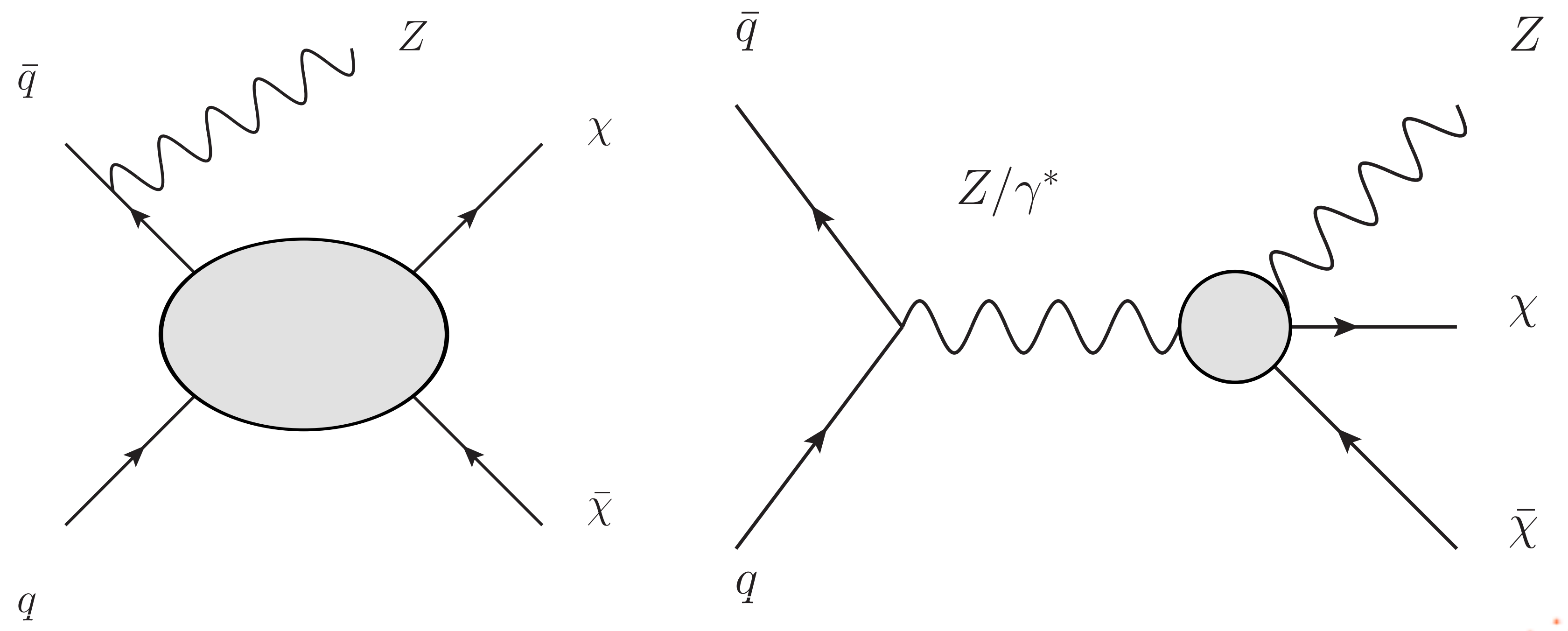


MORE THAN MONOJET

- *Depending on the nature of the mediator, DM could couple to more than quarks*
- *Different mediators give different LHC phenomenology*
- *Useful to test the assumptions behind the xsec scattering plot (and possibly clarify the situation). Eg W emission tests differences between u and d quarks (hence, neutrons and protons)*

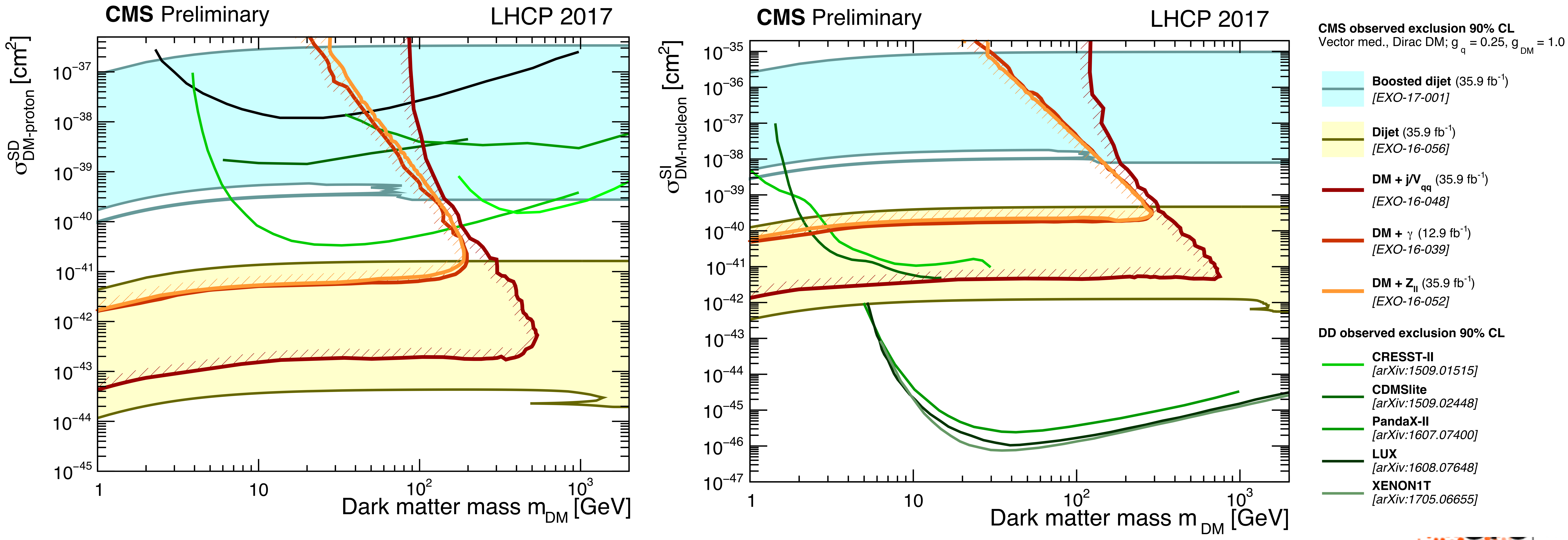


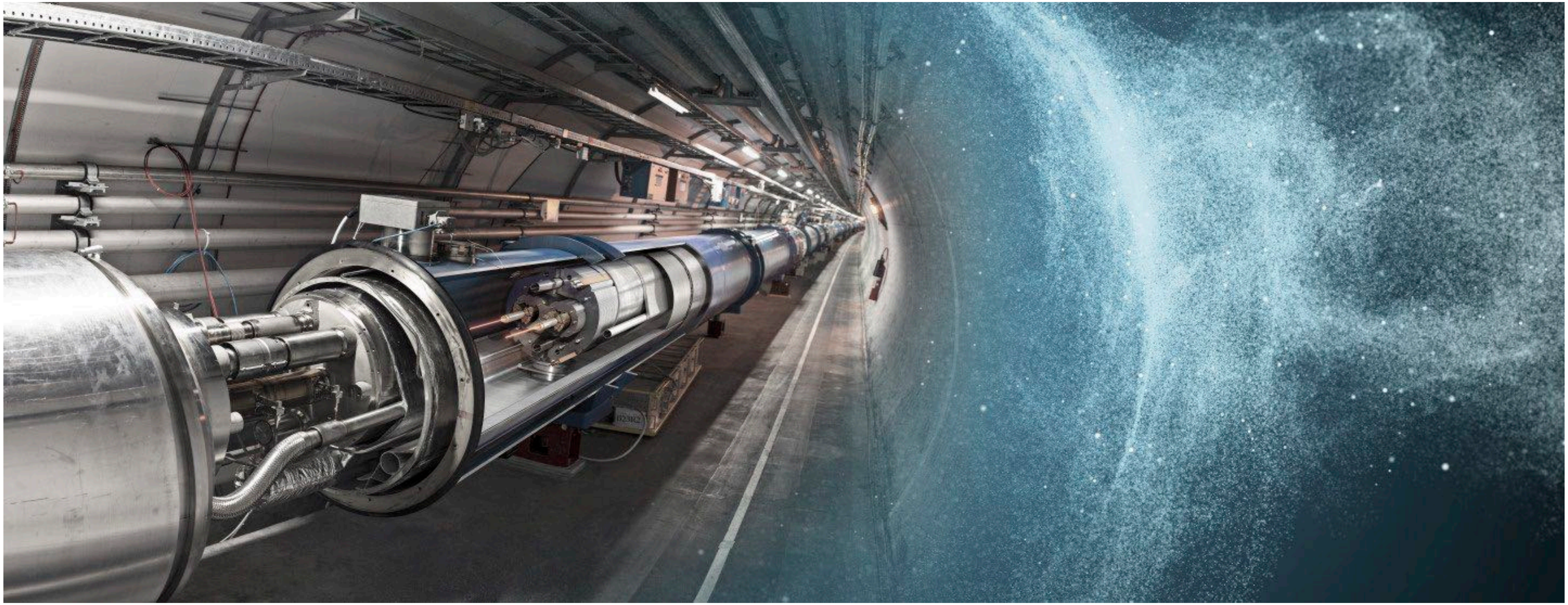




LHC & DIRECT DETECTION

- *Within **the model-dependent assumptions** of these simplified models, one can draw the LHC bounds on the same plane as the underground experiments*
- *Sensitivity at low masses*
- *Complementarity with different experiments in different scenarios*





DARK MATTER CASCADE PRODUCTION

A "CLASSIC" SUSY SEARCH

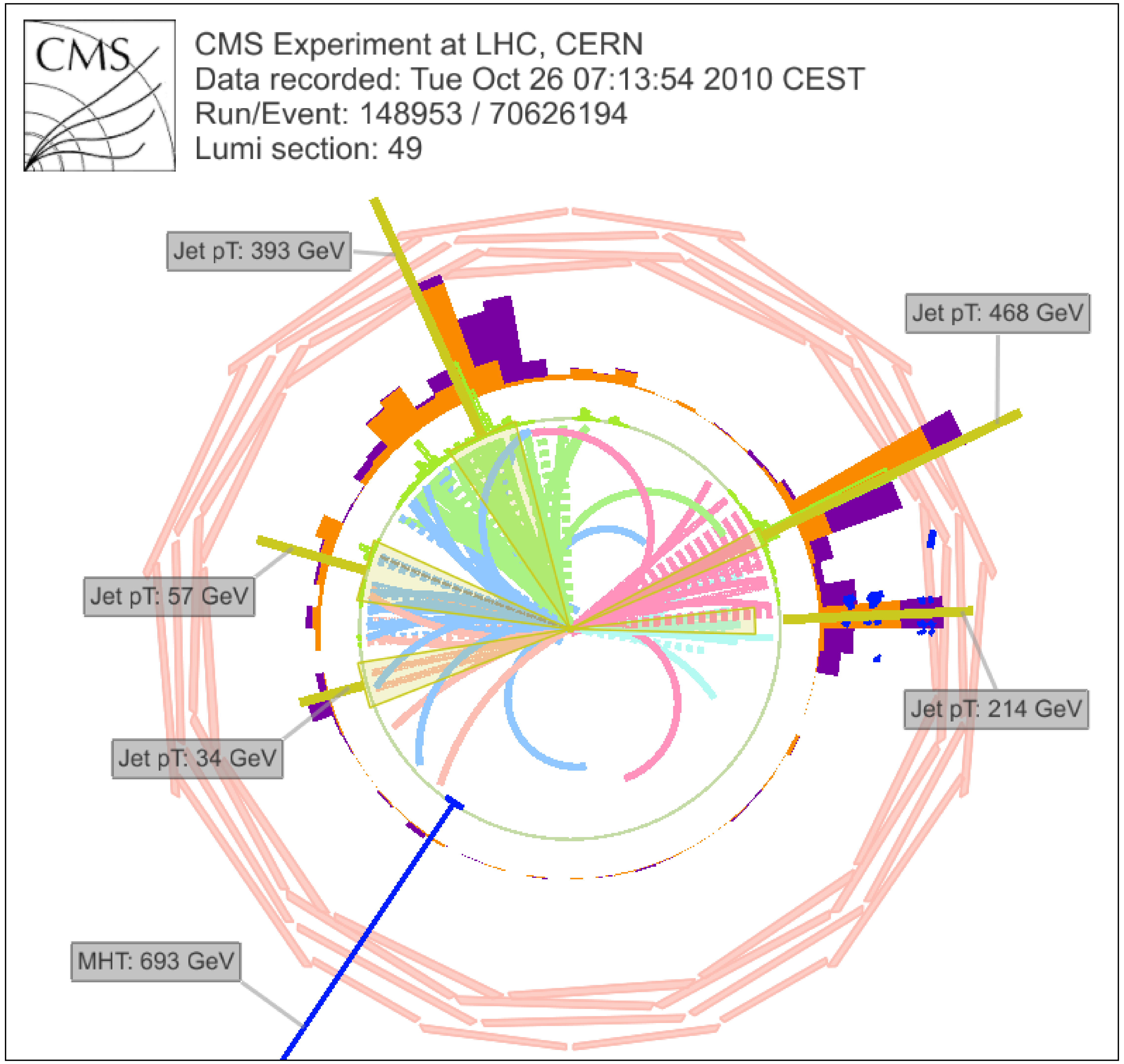
- *The typical signature: a lot of energy seen in the detector, recoiling against a lot of MET*
- *Several variables to quantify this behavior:*

$$HT = \Sigma_{jet} |p_T^{jet}|$$

$$|MHT| = |\Sigma_{jet} p_T^{jet}|$$

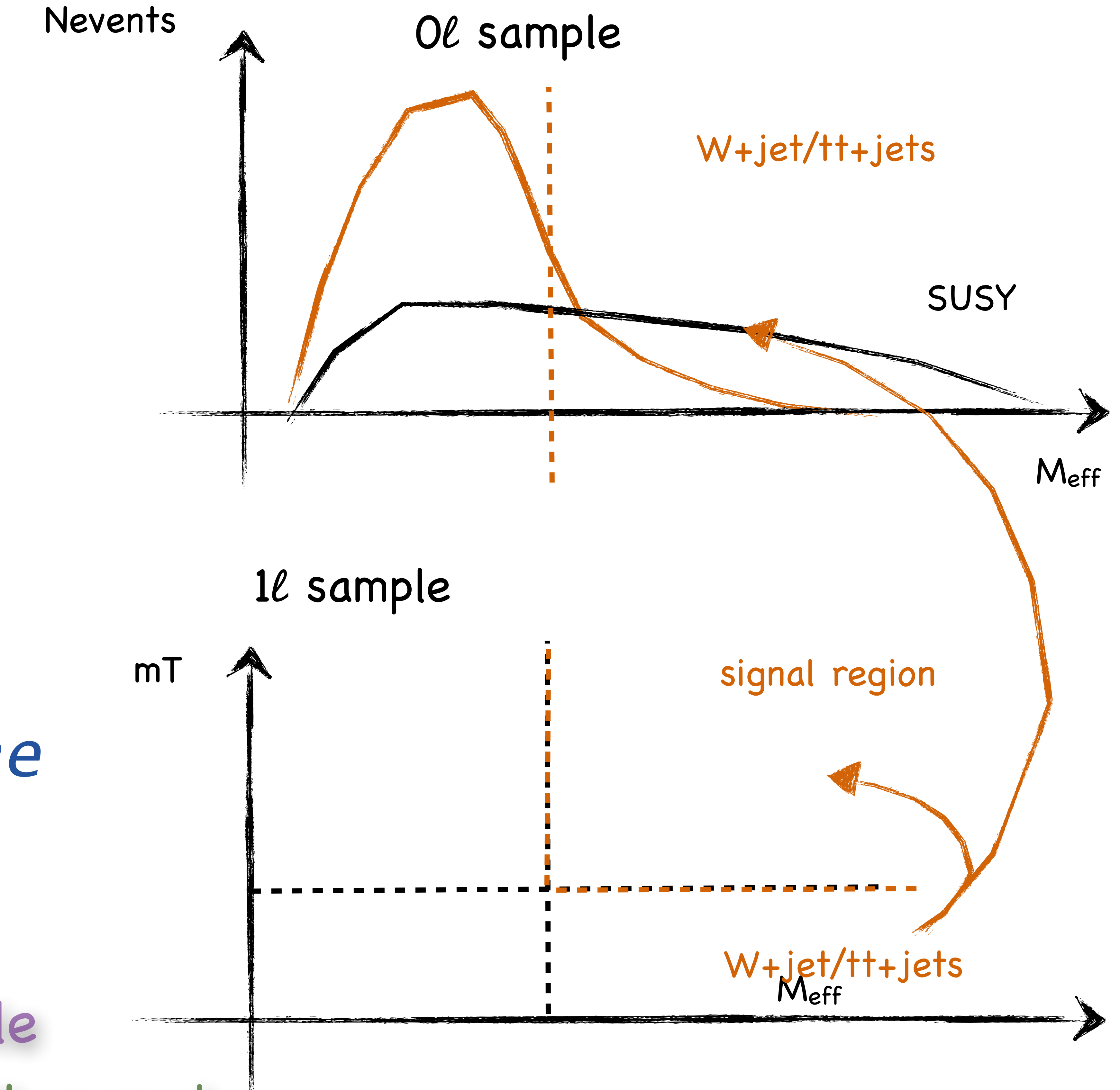
$$|MET| = |\Sigma_{cell} E_T^{cell}|$$

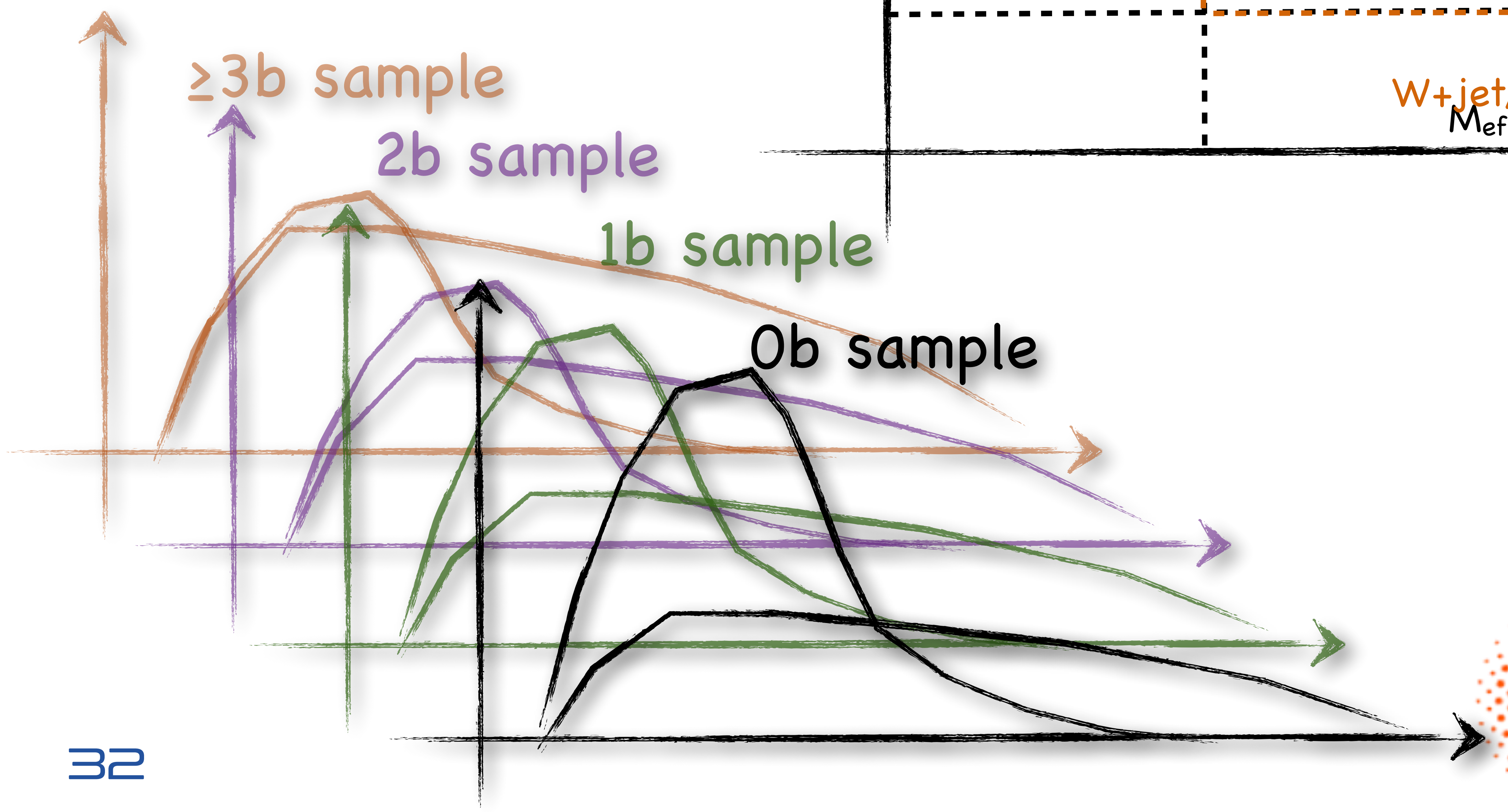
$$m_{eff} = HT + |MET|$$



SEARCH STRATEGY

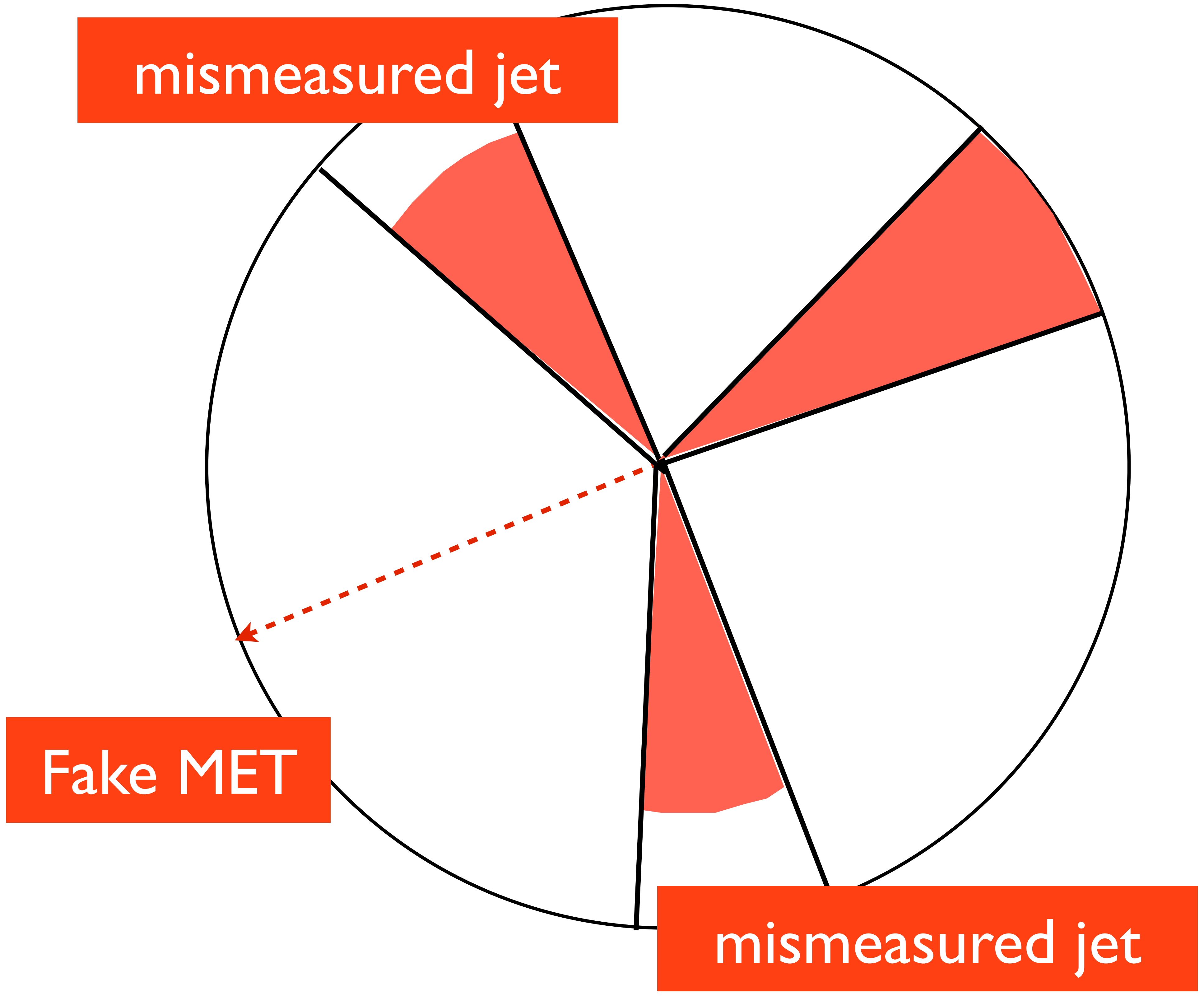
- The main ingredient to the search is a kinematic plane.
  - A measurement of the event energy (HT, Meff, MR)
  - A measurement of the unbalancing (MET,αT,MT2,R)
- The background on the tail is measured from the core, using transfer factors from MC
- The searches are repeated for different final states
- Searching for specific signals (e.g. stop production) advanced techniques (e.g. BDT) could be used





MAIN BACKGROUNDS

- For final states with leptons, situation is similar to monojet (W+jets, tt, ..)
- QCD is an extra problem for all-jets final states
  - when jet pT mismeasured, fake MET arises and event ends in tails of kinematic distributions
  - Need to "measure" the probability of this from data
  - MC statistics usually insufficient → large errors on MC scale factors
  - Often, analytic extrapolation from sidebands used instead

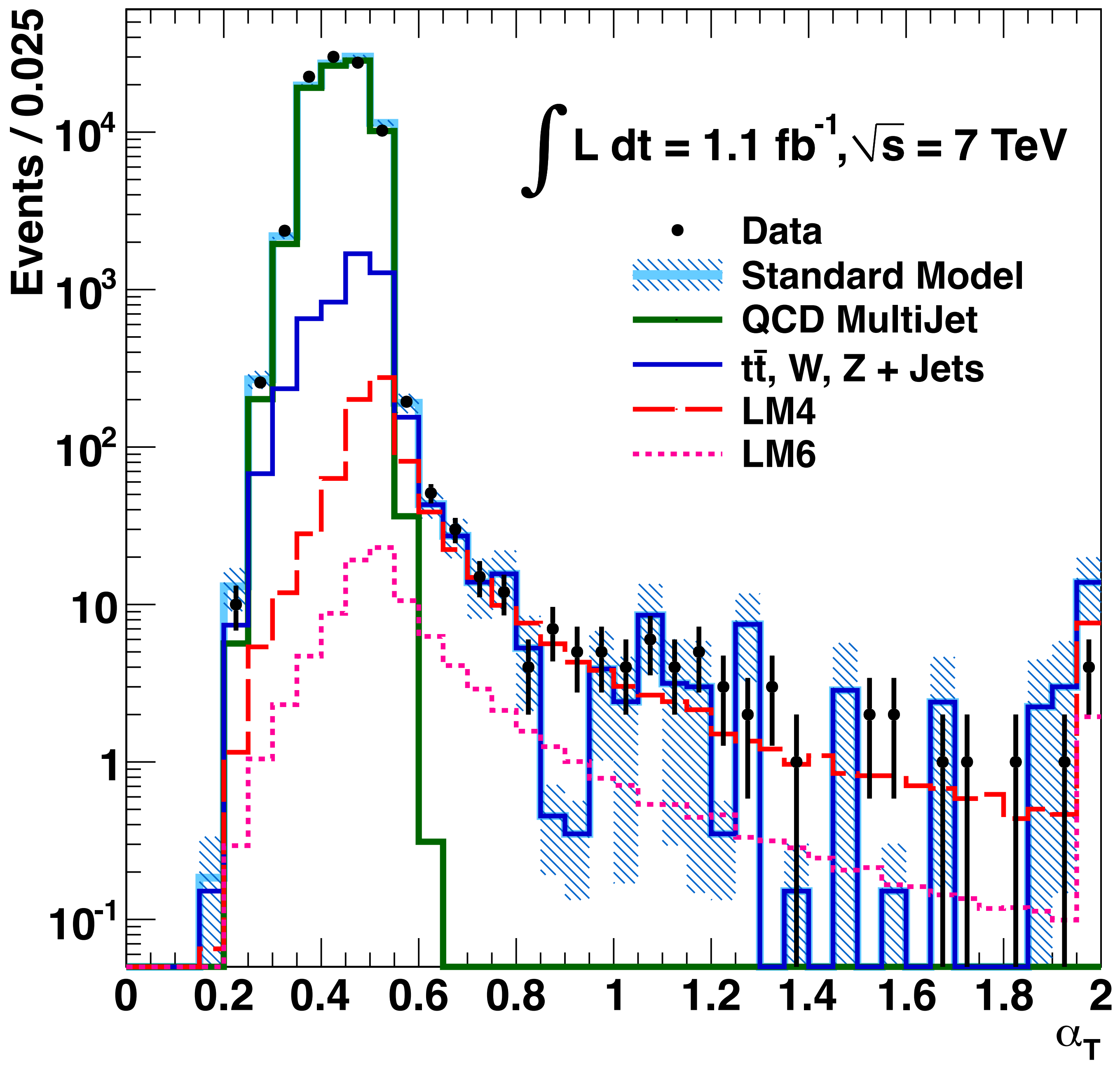


α_T: REJECTING QCD

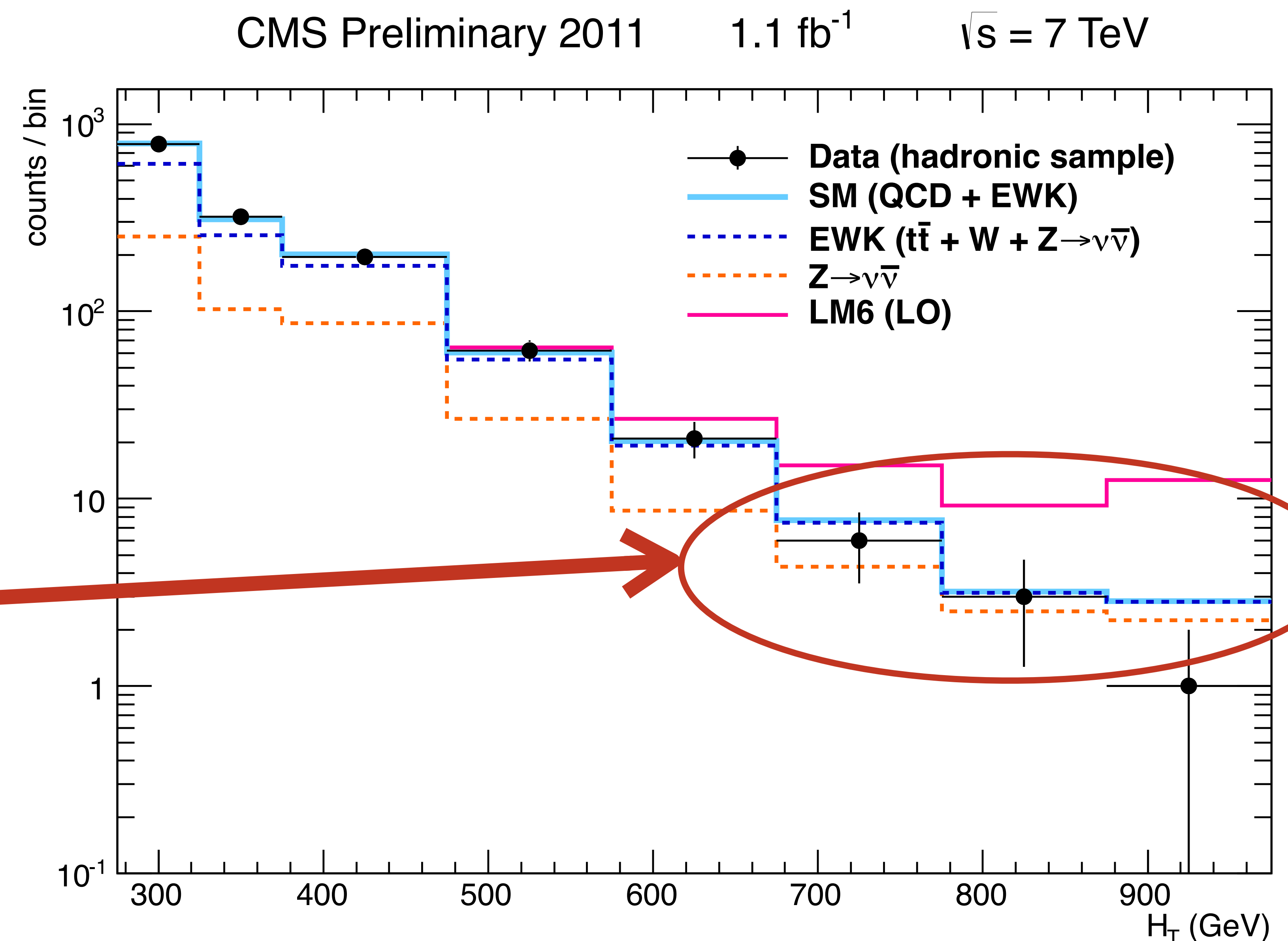
Randall & Tucker-Smith

$$\alpha \equiv \frac{p_{T2}}{m_{jj}}.$$

$$\alpha_{\mathrm{T}} = \frac{E_{\mathrm{T}}^{\mathrm{jet}_2}}{M_{\mathrm{T}}} = \frac{E_{\mathrm{T}}^{\mathrm{jet}_2}}{\sqrt{\left(\sum_{i=1}^{2} E_{\mathrm{T}}^{\mathrm{jet}_i}\right)^2 - \left(\sum_{i=1}^{2} p_x^{\mathrm{jet}_i}\right)^2 - \left(\sum_{i=1}^{2} p_y^{\mathrm{jet}_i}\right)^2}},$$



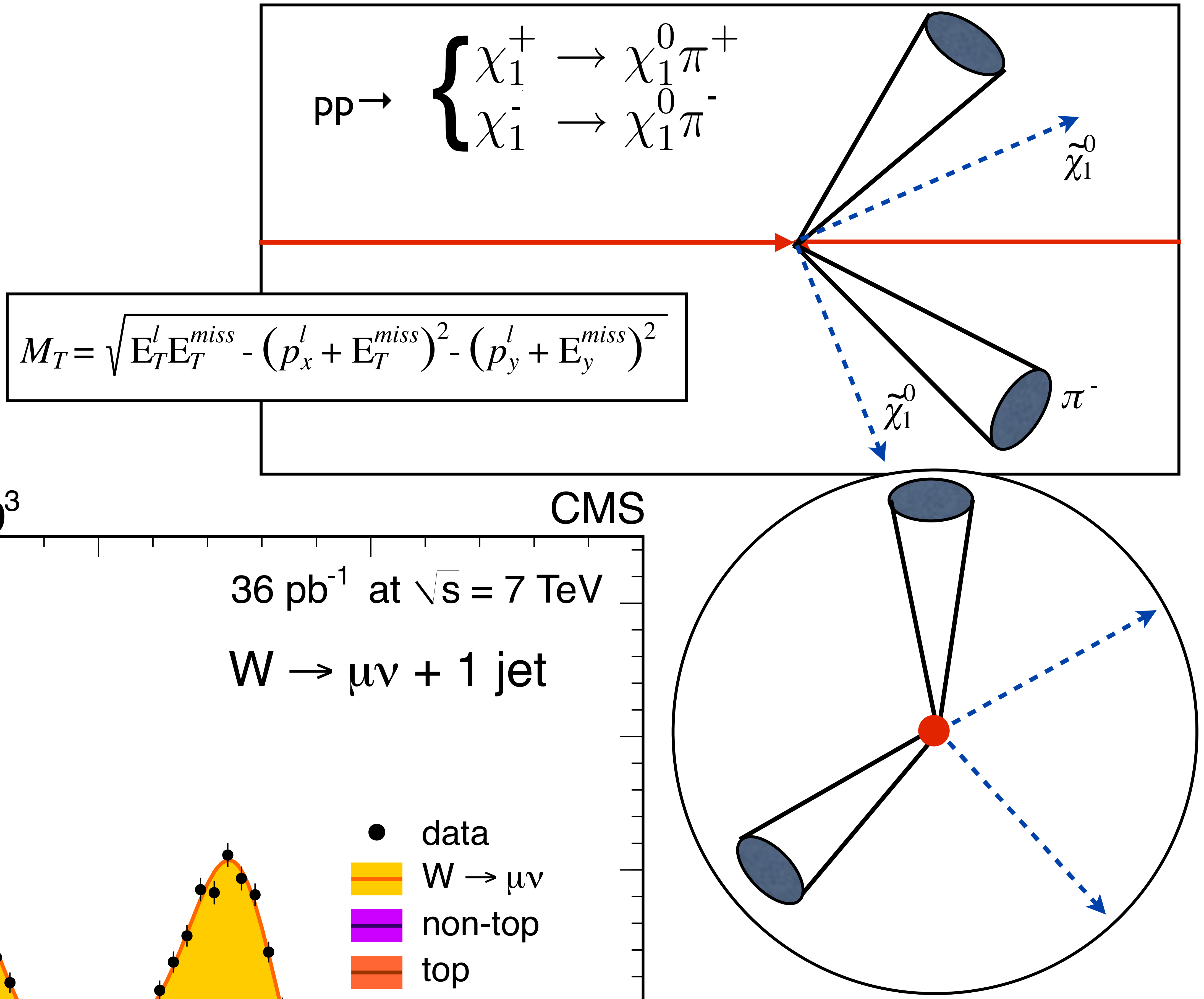
- αT = 0.5 for perfectly balanced dijet events / <0.5 for dijet + mismeasurements
- EW main bkg after αT cut
- QCD events could leak to αT>0.5 because of detector effects (rare)
- large fraction of signal events removed (efficiency vs purity)

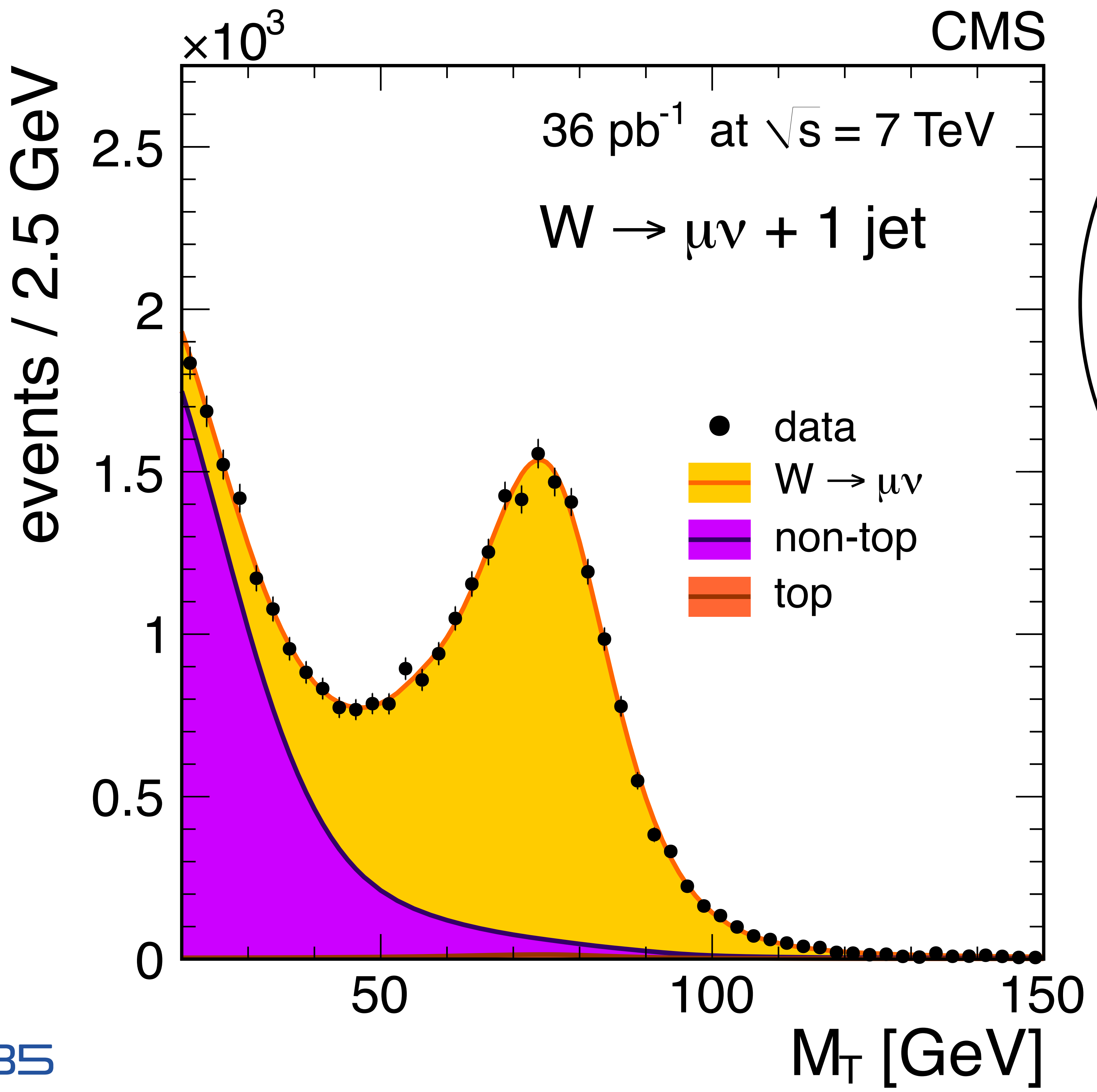


- After αT cut the signal looks similar to bkg in αT
- another variable needs to be used to characterize the signal
- HT used by CMS

M_{T2}: TWO MISSING PARTICLES

- We are looking for events with two undetected neutral particles leaving the detector
- We measure the sum of their pT as MET
- This is similar to the detection of the W, for which the edge of the mT distribution is used
- The presence of two missing particles make the picture more complicated. But the physics intuition holds





M_{T2}: TWO MISSING PARTICLES

- If we could see all the particles, we could compute

$$m^2_{\chi_1^+} = m^2_\pi + m^2_{\chi_1^0} + 2\left[E_T^\pi E_T^{\chi_1^0}\cosh(\Delta\eta) - \mathbf{p}_T^\pi \cdot \mathbf{p}_T^{\chi_1^0}\right]$$

- If we could measure pT(X0), but not pz(X0), the best we could do would be

$$m_T^2(\mathbf{p}_T^\pi, \mathbf{p}_T^{\chi_1^0}; m_{\chi_1^0}) \equiv m^2_{\pi^+} + m^2_{\chi_1^0} + 2(E_T^\pi E_T^{\chi_1^0} - \mathbf{p}_T^\pi \cdot \mathbf{p}_T^{\chi_1^0})$$

- Since cosh>1, mT≤m, the equality holding for both pz(X0)=0. This means that max(mT) has an “edge” at m

- For each event we have two values of mT (two copies of the same decay). Both are such that mT<m. This means that max(mT(1), mT(2))<m

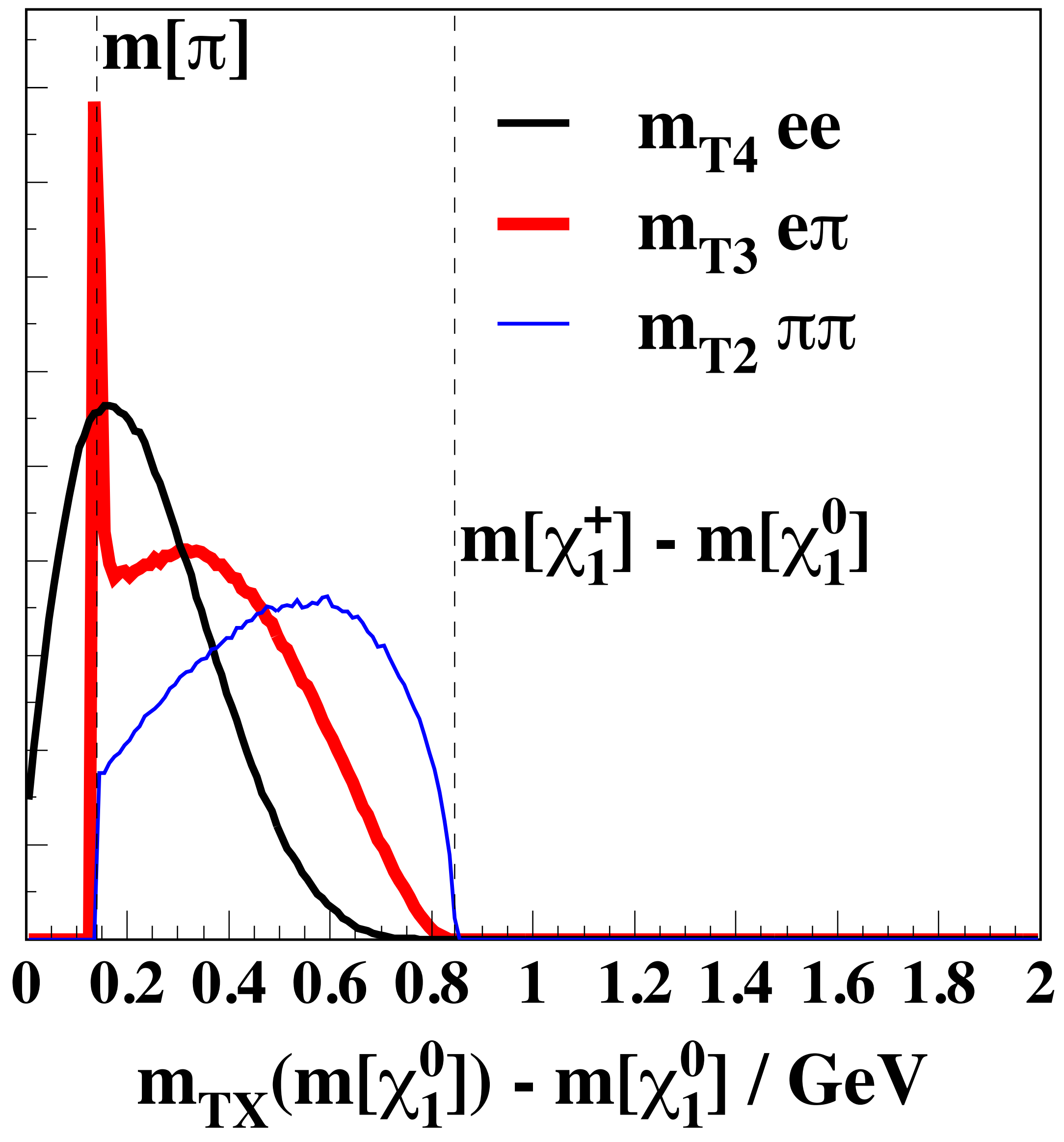
- We only know pT(X01)+ pT(X02)=ETmiss. A wrong assignment of the missing momenta brakes the mT<m condition. But the condition would hold for the correct assignment. This means that min(mT)<mT(true)<m.

- This defined mT2 as

$$m^2_{T2}(\chi) \equiv \min_{\not{q}_T^{(1)}+\not{q}_T^{(2)}=\not{p}_T}\left[\max\left\{m_T^2(\mathbf{p}_T^{\pi^{(1)}}, \not{q}_T^{(1)}; \chi),\ m_T^2(\mathbf{p}_T^{\pi^{(2)}}, \not{q}_T^{(2)}; \chi)\right\}\right]$$

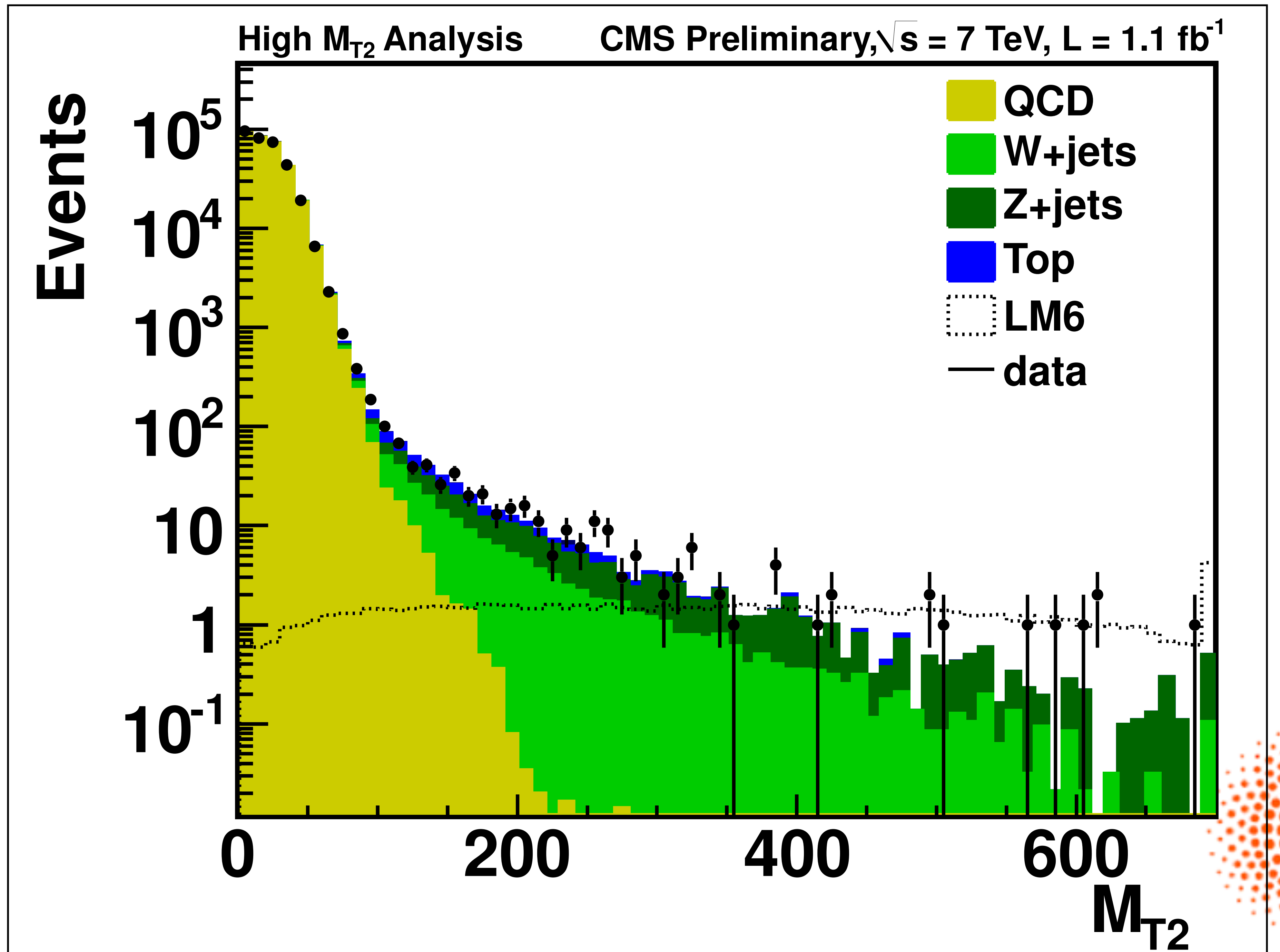
M_{T2}: TWO MISSING PARTICLES

- *The variable we have is a function of the mass of the LSP*



- ***SUSY characterisation:***
  - *Scan the LSP mass and look for the edge developing in your sample of SUSY events (if you have one...)*
- ***SUSY search:***
  - *Assume a mass value (eg mLSP=0)*
  - *Assume that the visible system in has 0 mass*
  - *An analytical expression for MT2 is found*

$$(M_{T2})^2 = 2A_T = 2p_T^{vis(1)} p_T^{vis(2)} (1 + cos\phi_{12})$$

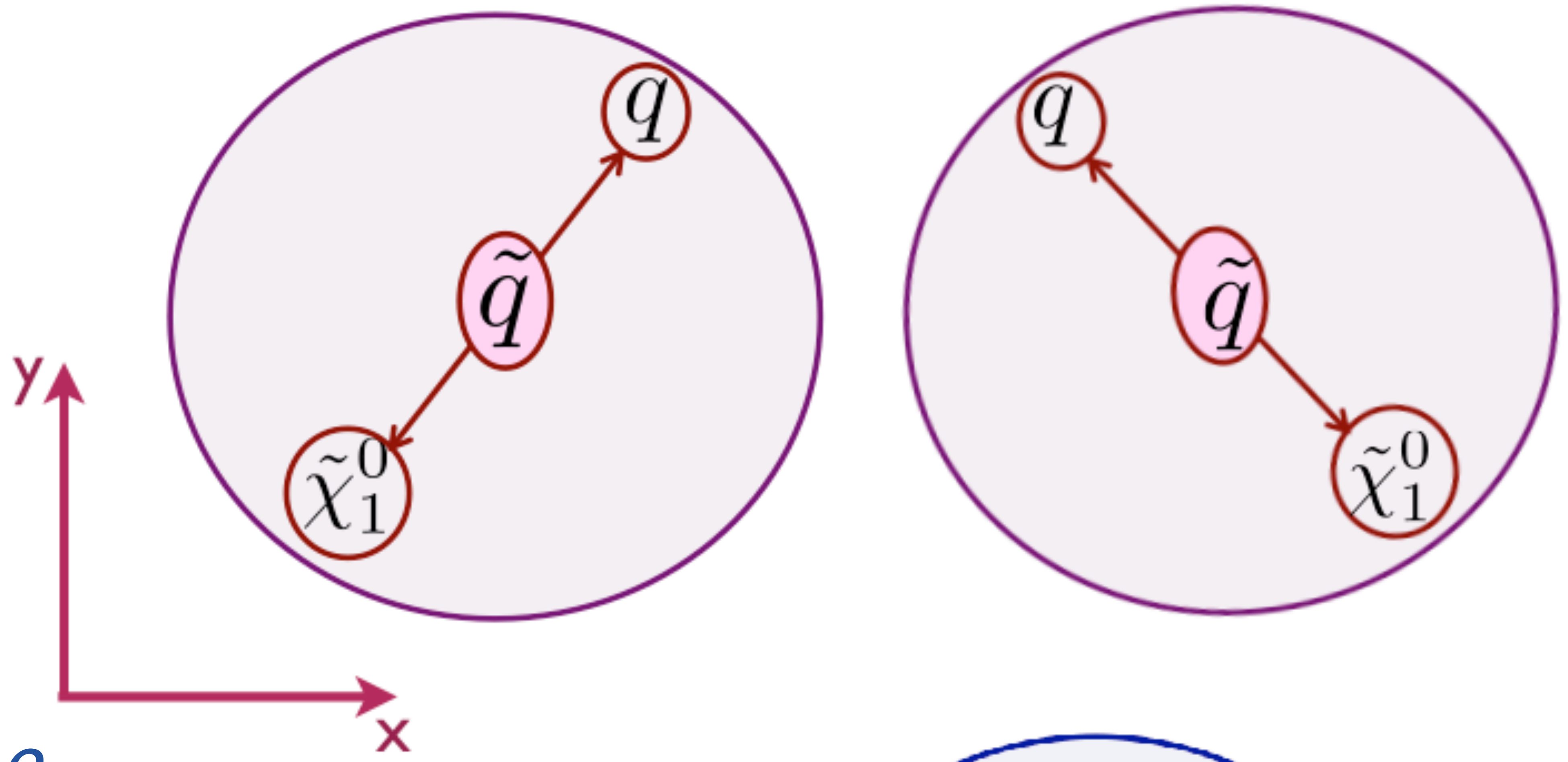


  - *The edge is lost but we have an αT-like variable to kill the QCD*

THE RAZOR FRAME

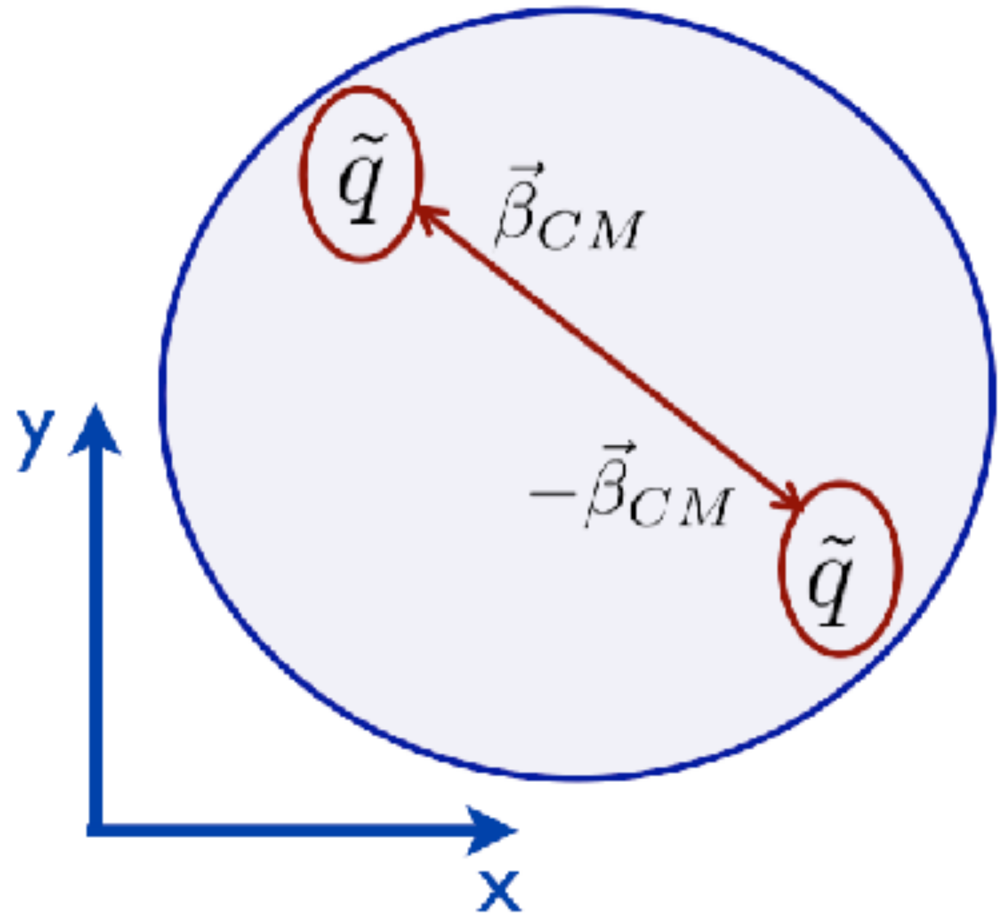
- *Two squarks decaying to quark and LSP. In their rest frames, they are two copies of the same monochromatic decay. In this frame p(q) measures MΔ*

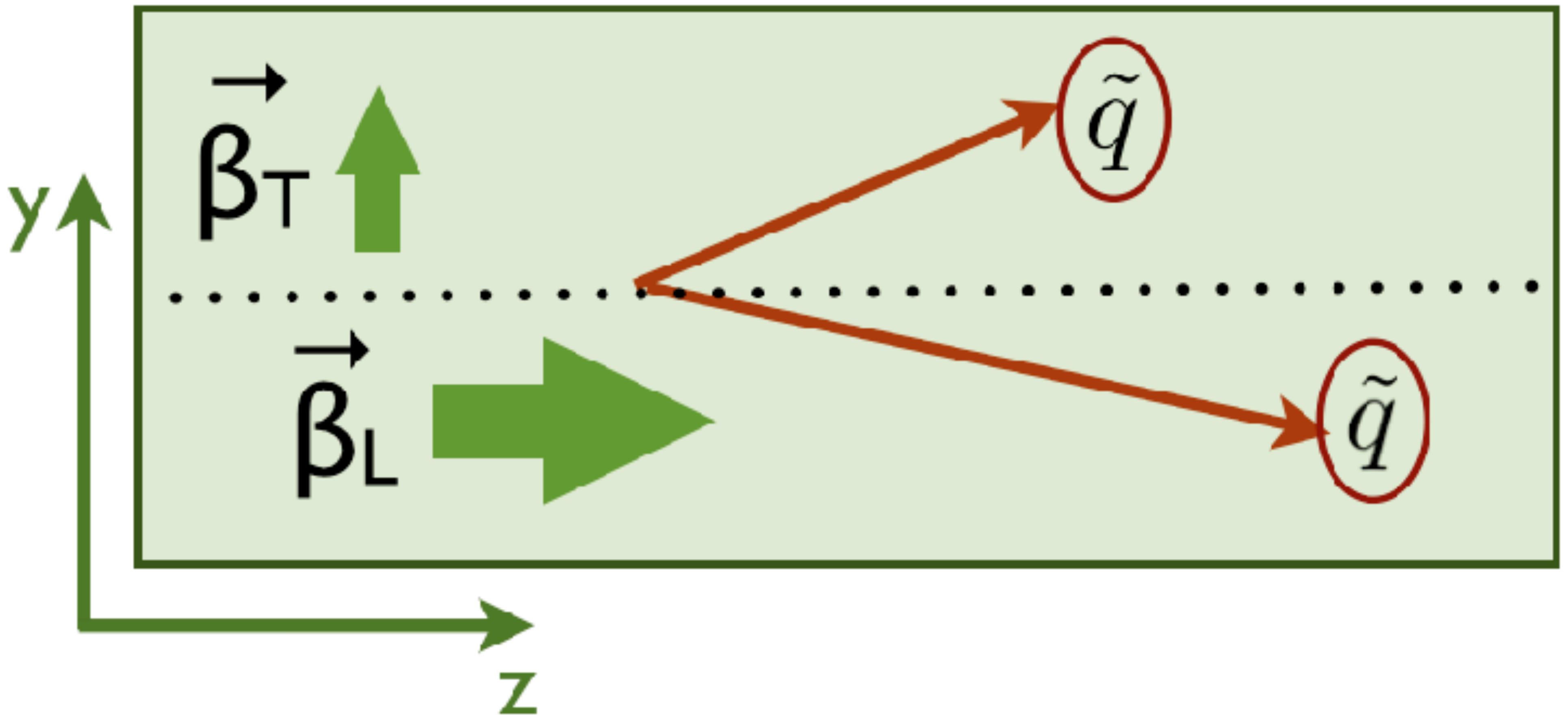
$$M_\Delta \equiv \frac{M_{\tilde{q}}^2 - M_{\tilde{\chi}}^2}{M_{\tilde{q}}} = 2M_{\tilde{\chi}}\gamma_\Delta\beta_\Delta$$



- *In the lab frame, the two squarks are boosted longitudinally. The LSPs escape detection and the quarks are detected as two jets*

- *In the rest frame of the two incoming partons, the two squarks recoil one against each other.*





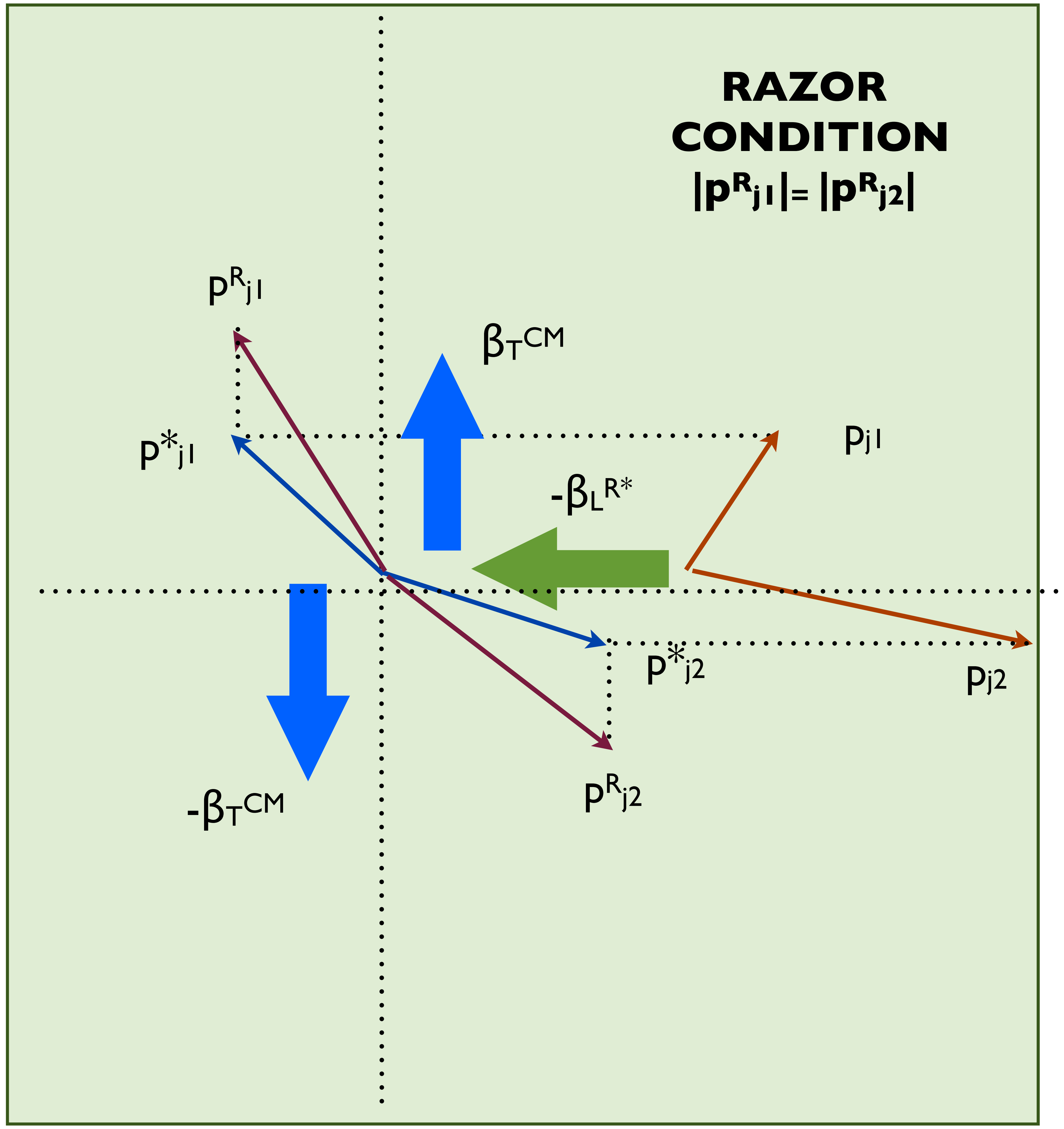
If we could see the LSPs, we could boost back by β_L, β_T, and β_{CM}
In this frame, we would then get
$|p_{j1}| = |p_{j2}|$
Too many missing degrees of freedom to do just this

European Research Council

THE RAZOR FRAME

- *In reality, the best we can do is to compensate the missing degrees of freedom with assumptions on the boost direction*
  - *The parton boost is forced to be longitudinal*
  - *The squark boost in the CM frame is assumed to be transverse*
- *We can then determine the two by requiring that the two jets have the same momentum after the transformation*
- *The transformed momentum defines the MR variable*

$$M_R \equiv \sqrt{(p_{j_1}+p_{j_2})^2 - (p_z^{j_1}+p_z^{j_2})^2}$$



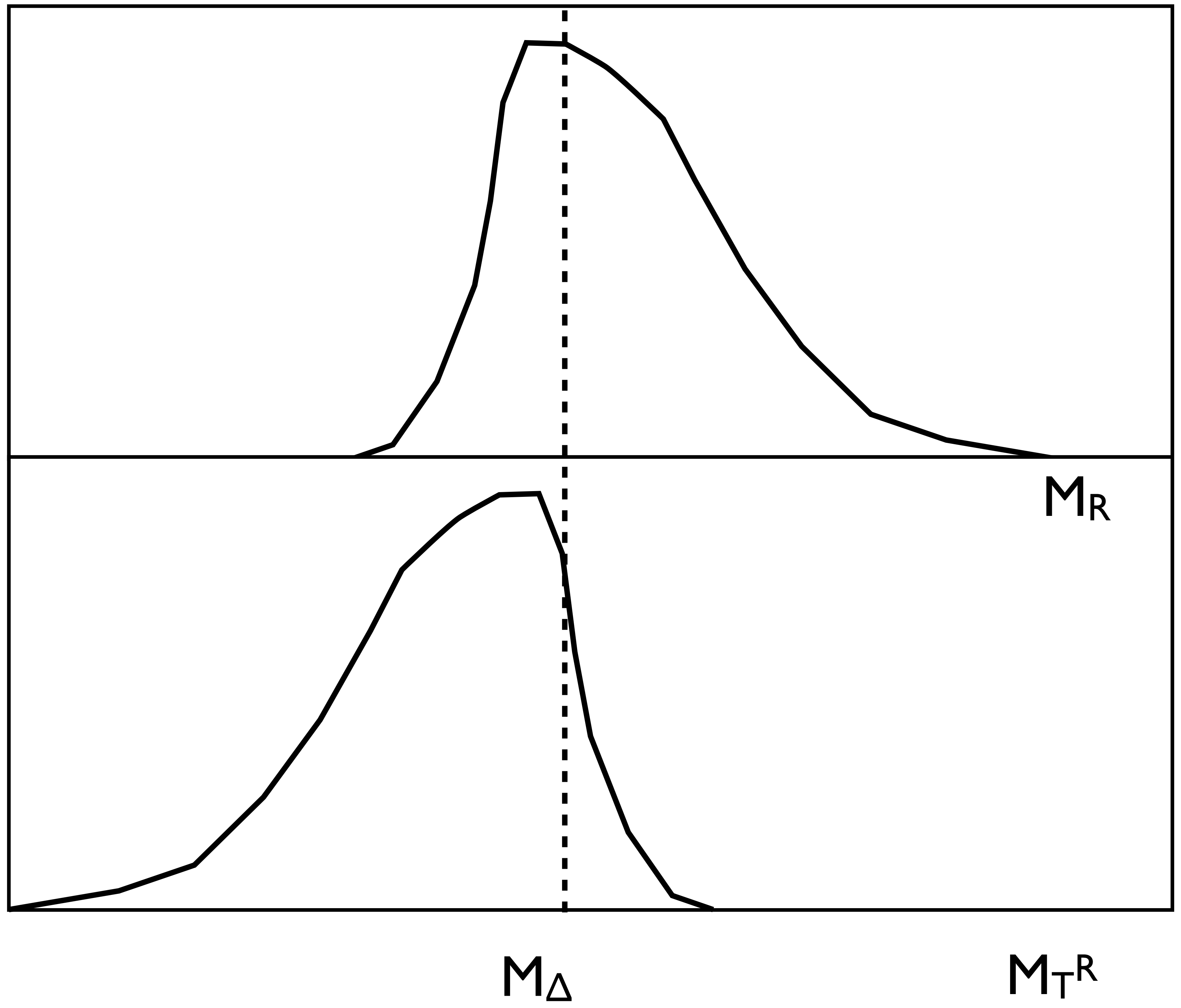
European Research Council

THE RAZOR VARIABLE

- MR is boost invariant, even if defined from 3D momenta
- No information on the MET is used
- The peak of the MR distribution provides an estimate of MΔ
- MΔ could be also estimated as the "edge" of MTR

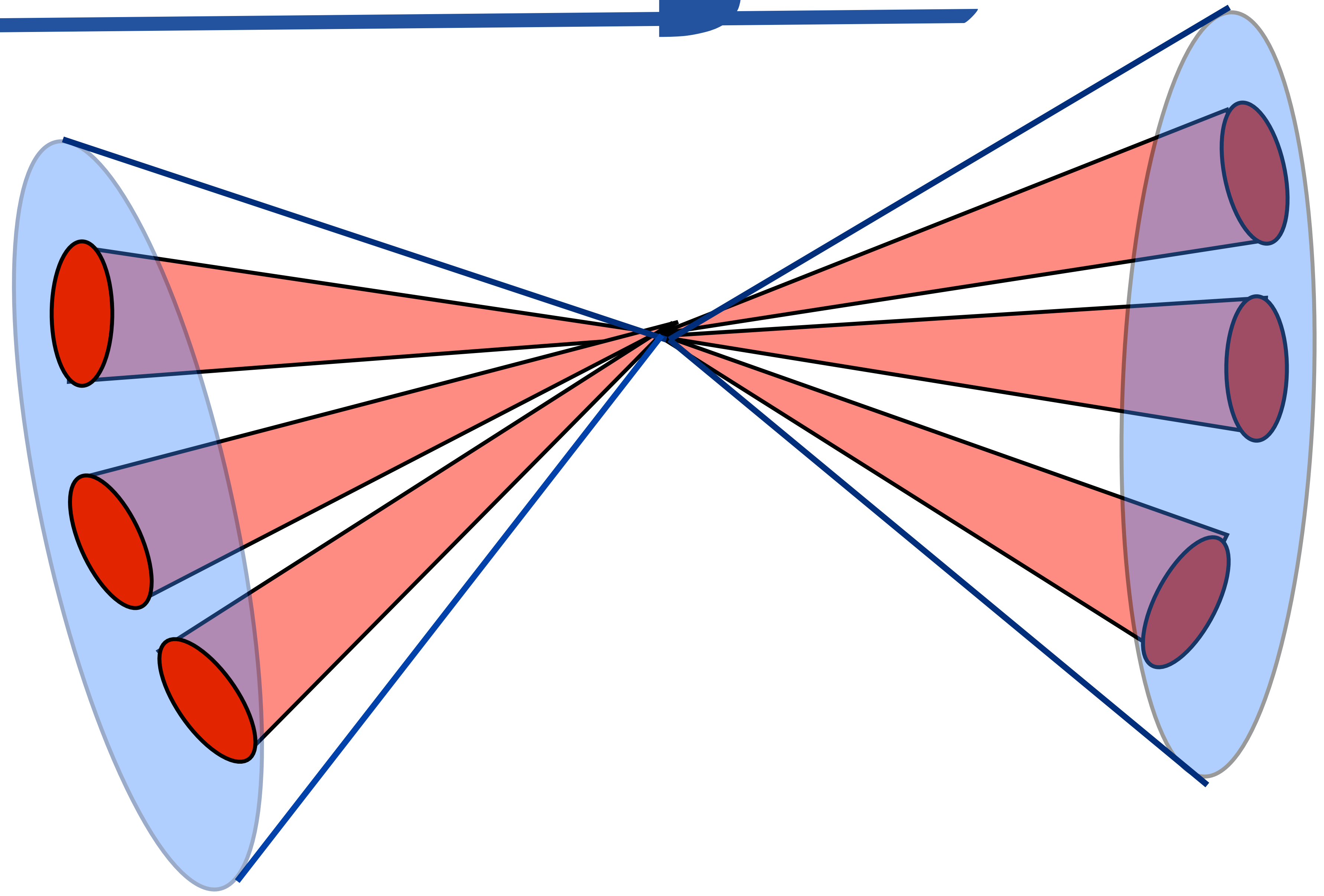
$$M_T^R \equiv \sqrt{\frac{E_T^{miss}(p_T^{j1}+p_T^{j2}) - \vec{E}_T^{miss}\cdot(\vec{p}_T^{j1}+\vec{p}_T^{j2})}{2}}$$

- MTR is defined using transverse quantities and it is MET-related
- The Razor (aka R) is defined as the ratio of the two variables



$$R \equiv \frac{M_T^R}{M_R}$$

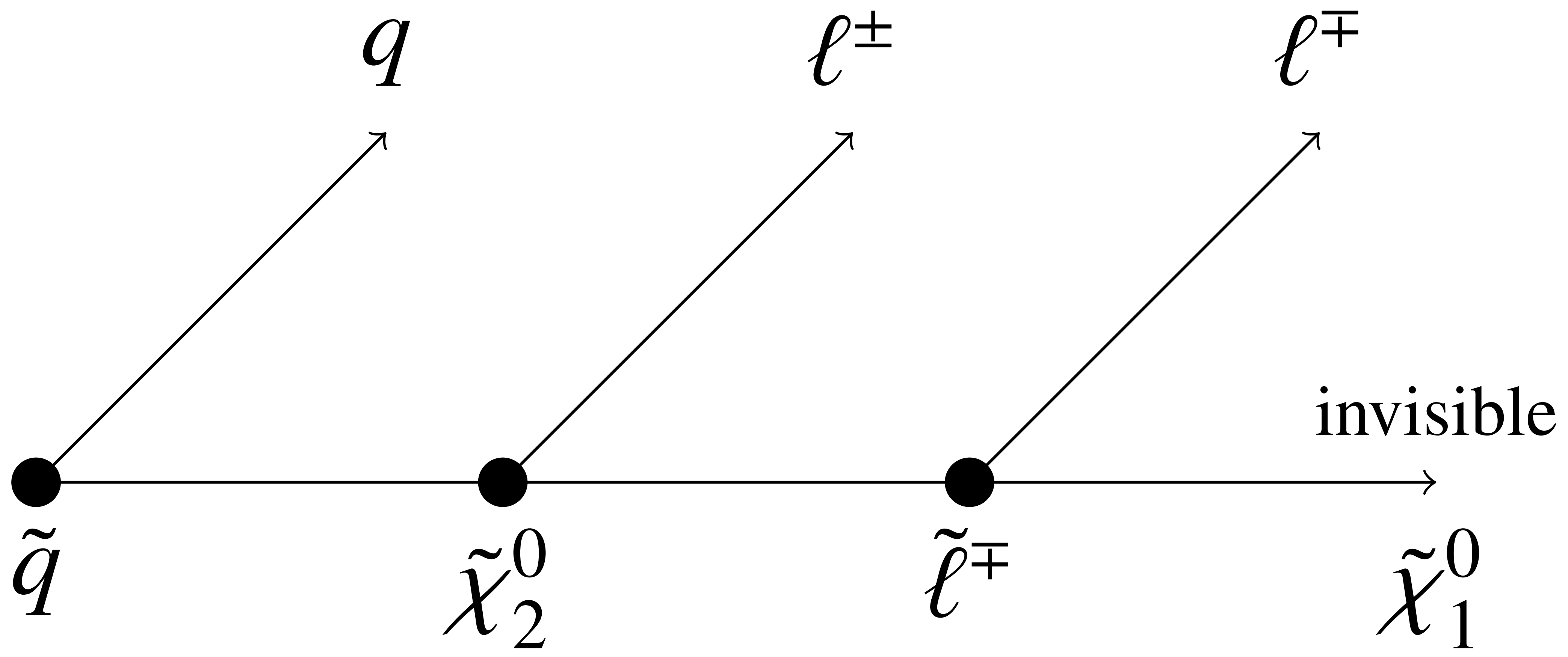
FROM DIJET TO MULTIJETS

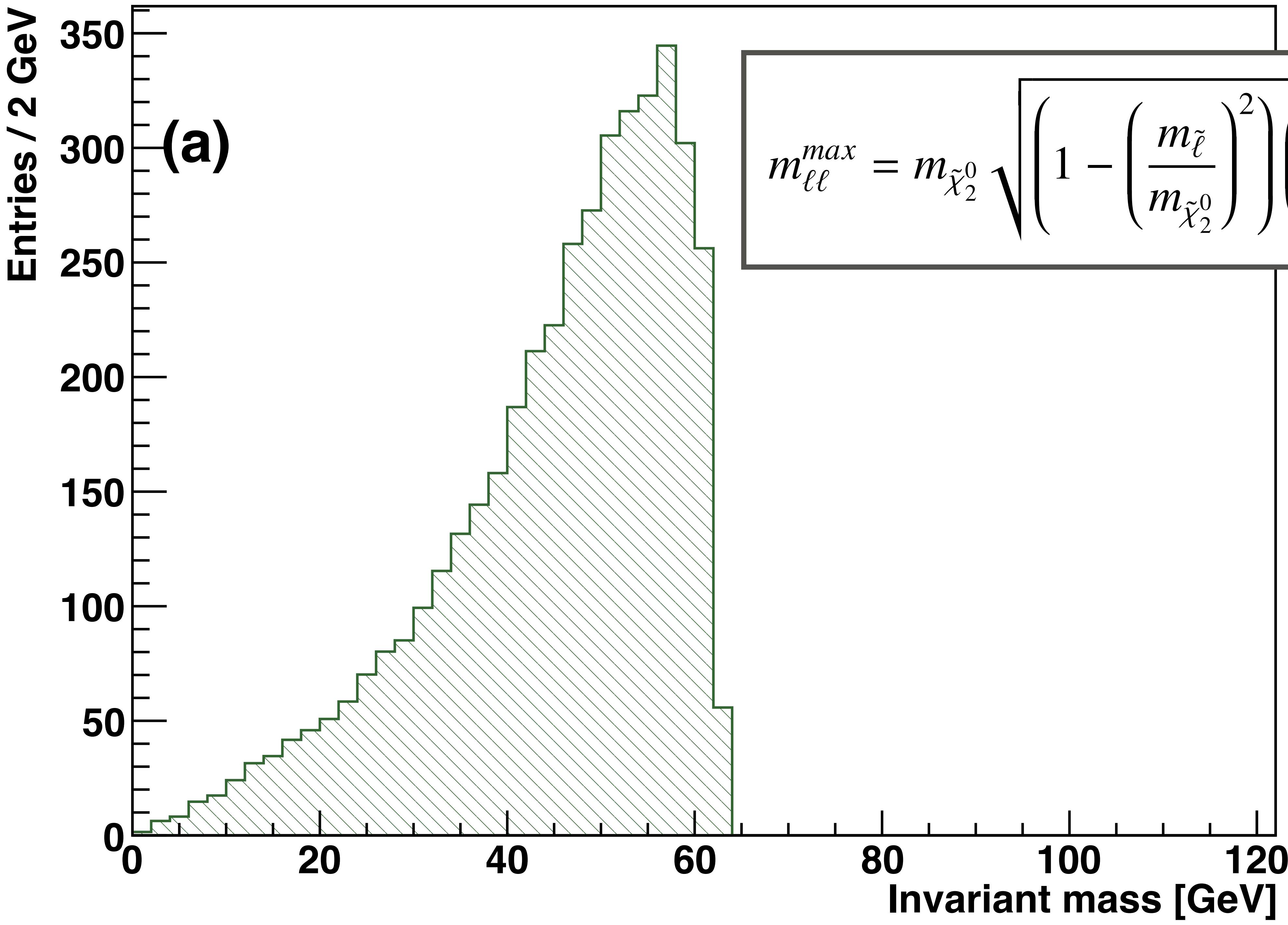


- The "new" variables rely on the dijet+MET final state as a paradigm
- All the analyses have been extended to the case of multijet final states clustering jets in two hemispheres (aka mega-jets)
- Several approaches used
  - minimizing the HT difference between the mega-jets (aT CMS)
  - minimizing the invariant masses of the two jets (Razor CMS)
  - minimizing the Lund distance (MT2 CMS) $$(E_i - p_i cos\theta_{ik})\frac{E_i}{(E_i+E_k)^2} \le (E_j - p_j cos\theta_{jk})\frac{E_j}{(E_j+E_k)^2}$$
  - …
- Eventually, all this has to be replaced by exclusive jet algorithms (e.g., XCone)

DILEPTON EDGE

- For long decay chains with two leptons in cascade, an edge develops in the dilepton invariant mass
  - Can be used to highlight the presence of a signal
  - At same time, allows to characterise the underlying SUSY spectrum
- This is the cleanest possible SUSY signature at the LHC
  - edge position measured with a few GeV uncertainties
  - dilepton final states quite clean from background
  - shape information adds strong kinematic discrimination

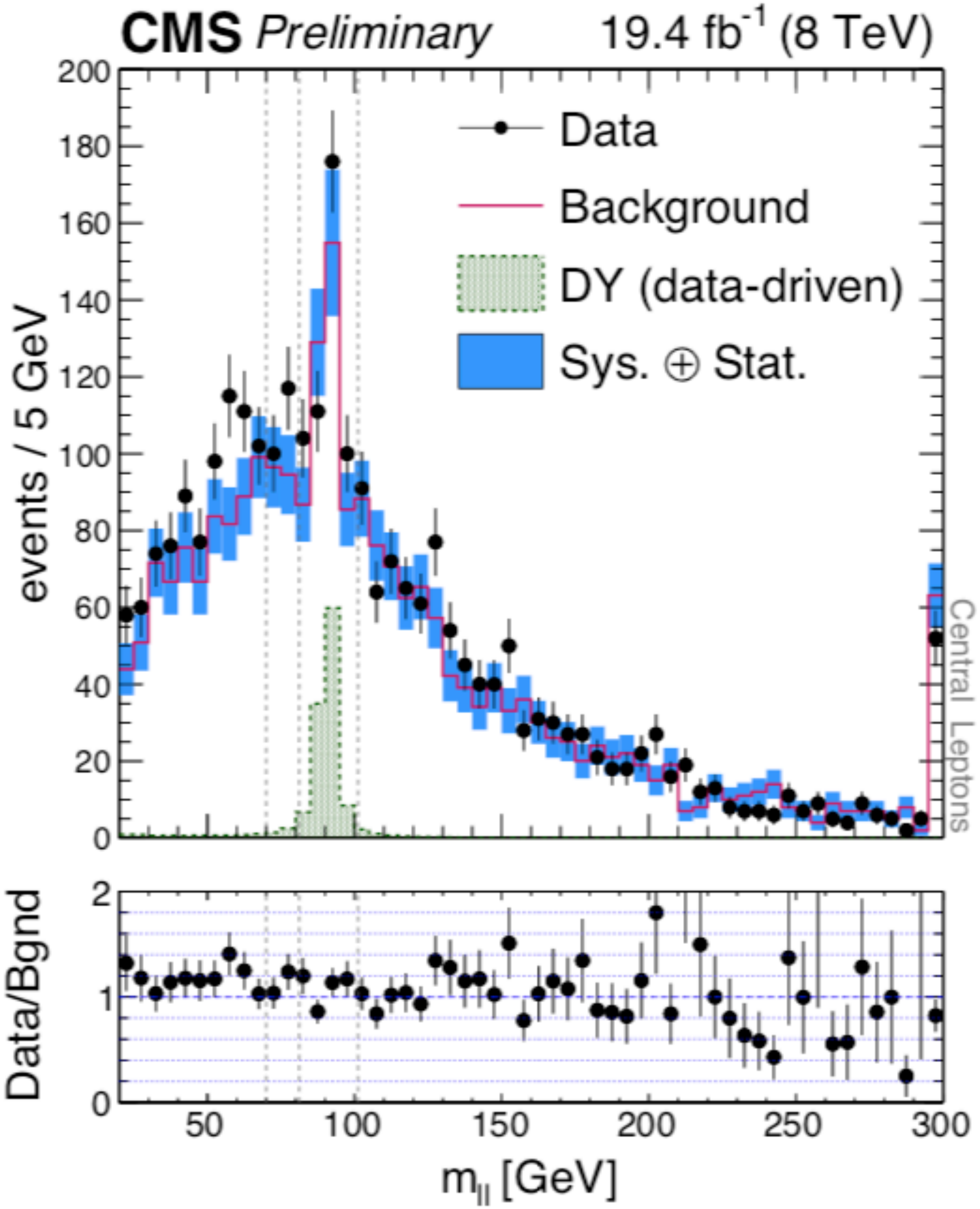




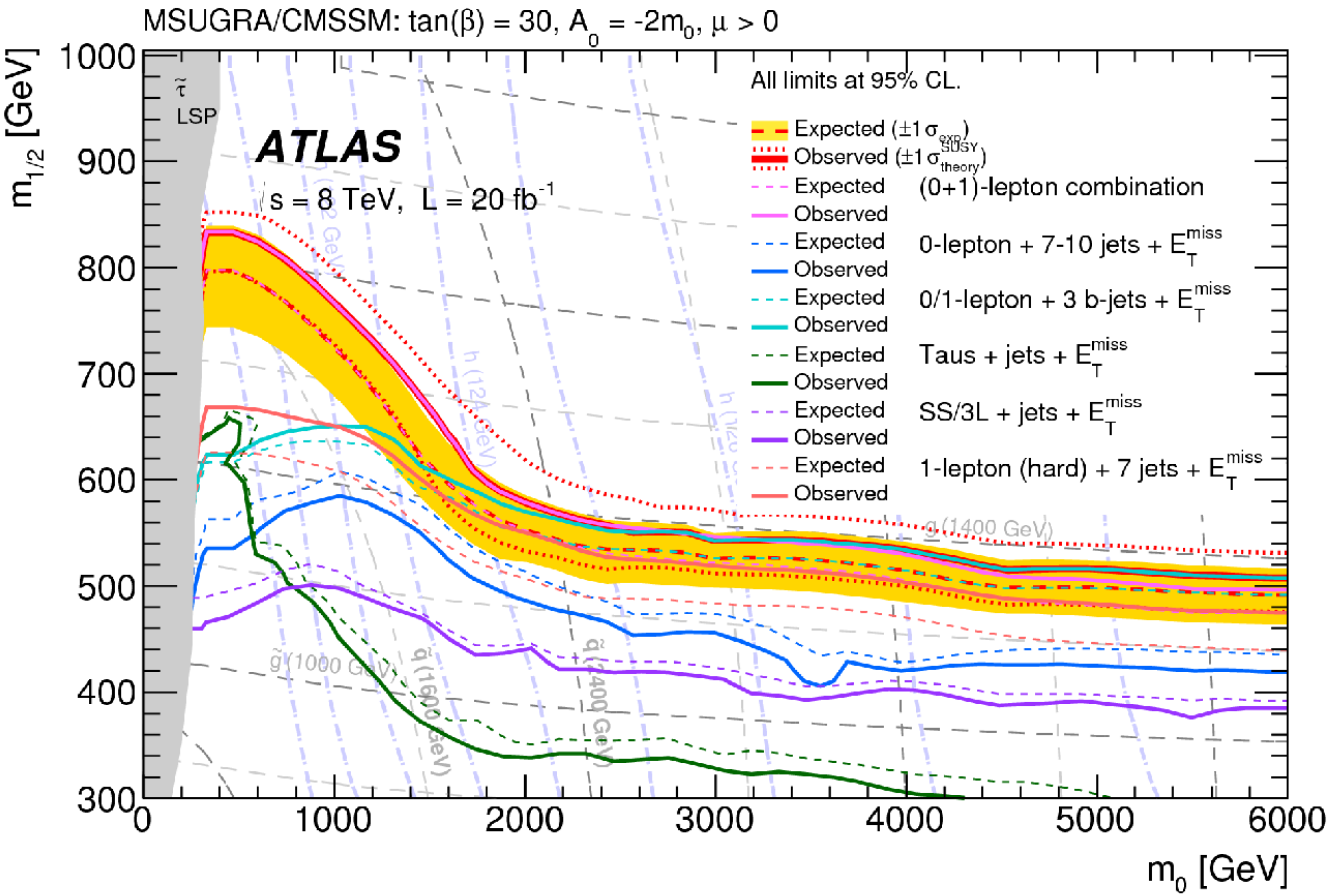
$$m_{\ell\ell}^{max} = m_{\tilde{\chi}_2^0}\sqrt{\left(1-\left(\frac{m_{\tilde{\ell}}}{m_{\tilde{\chi}_2^0}}\right)^2\right)\left(1-\left(\frac{m_{\tilde{\chi}_1^0}}{m_{\tilde{\ell}}}\right)^2\right)}$$

A LESSON TO BE LEARNED

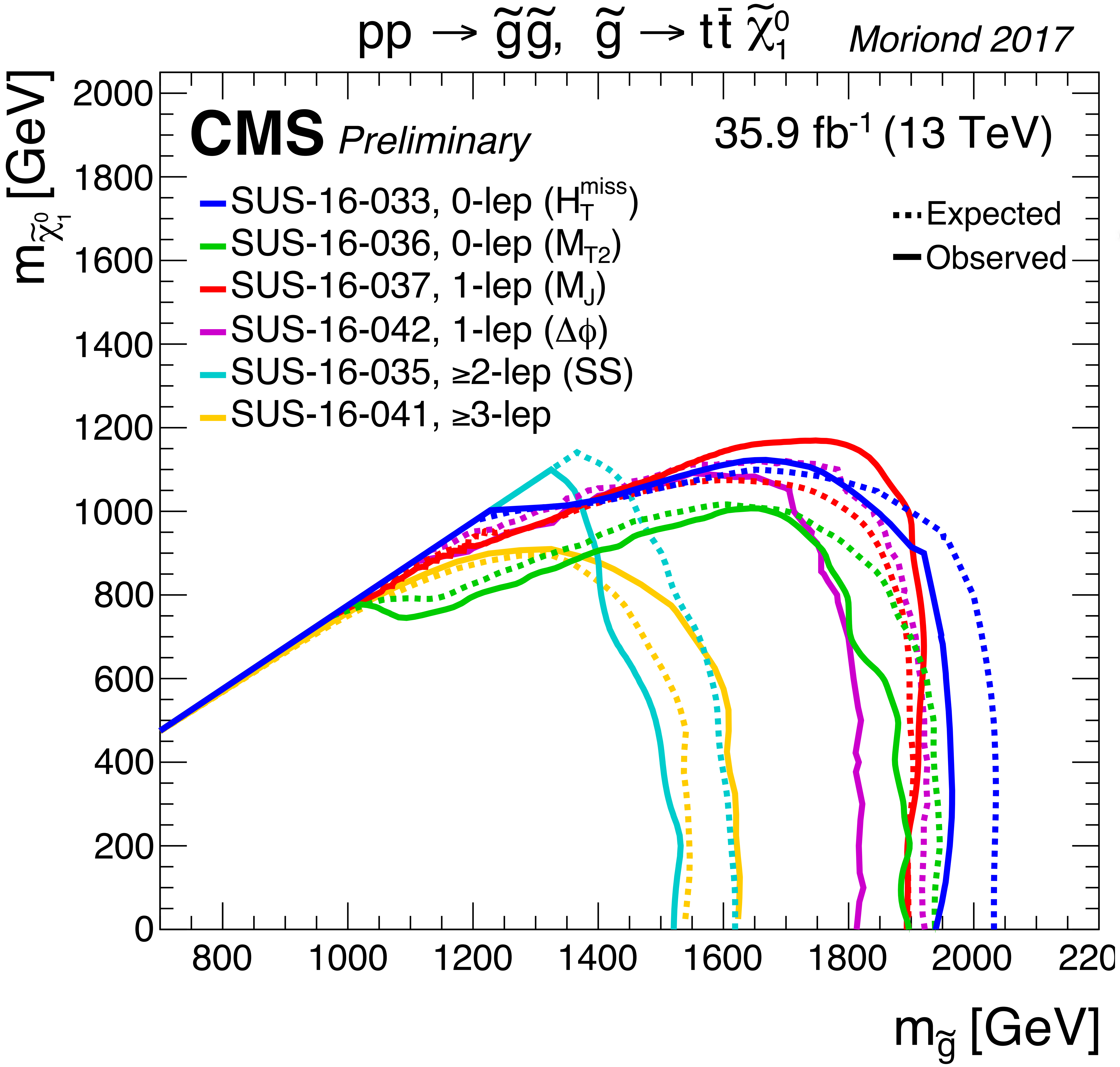
- *At some point, CMS saw an excess with this search*
  - *Regardless of the clean signature, it took 2 years to take out the result (which was then disproved by new data + ATLAS)*
  - *Just because there is also background, signal can be small and life is complicated*
- ***Establishing a discovery at the LHC which is NOT a more-than-expected resonance has sociological implications that we underestimate***

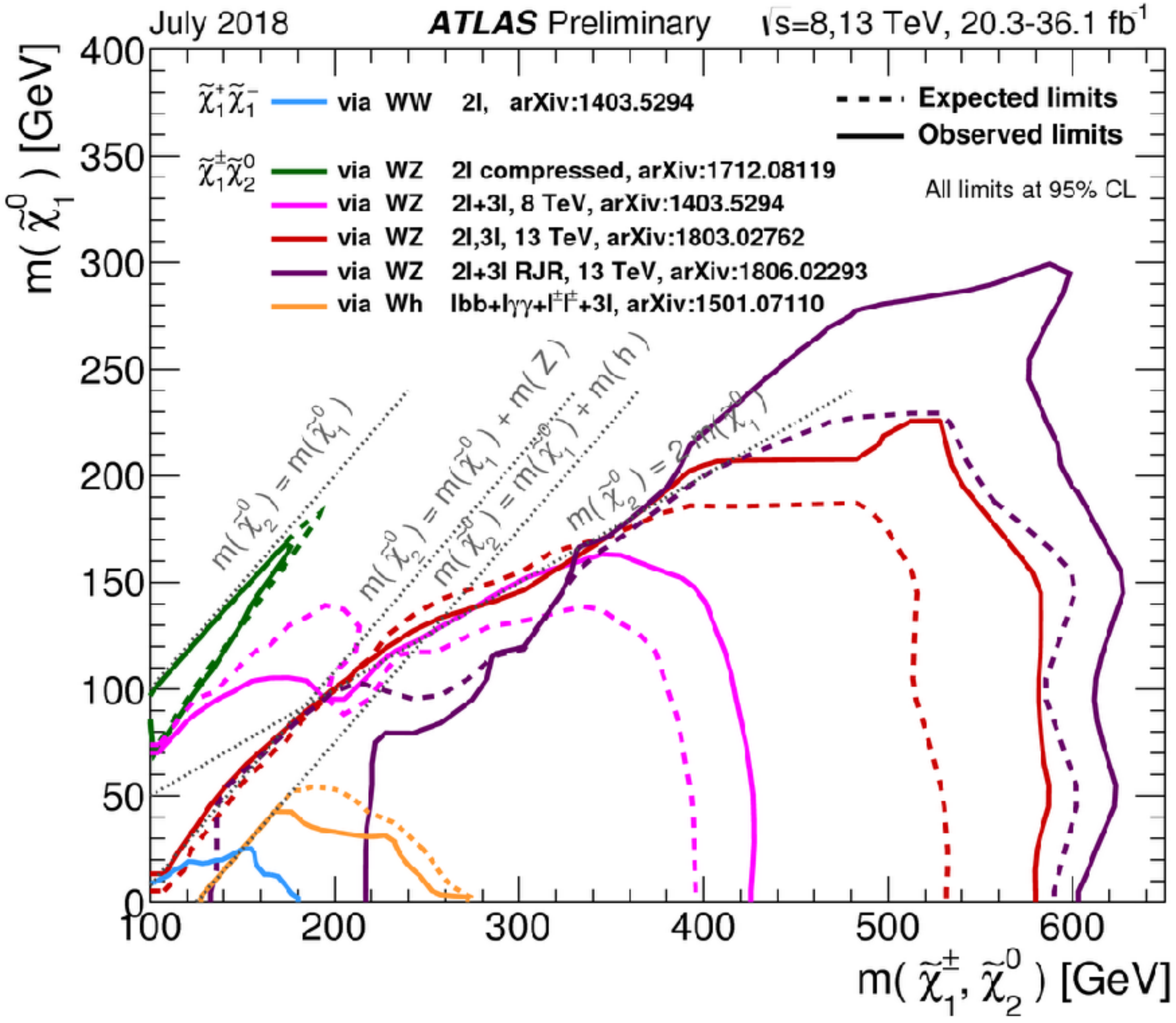


WHERE DO WE STAND?

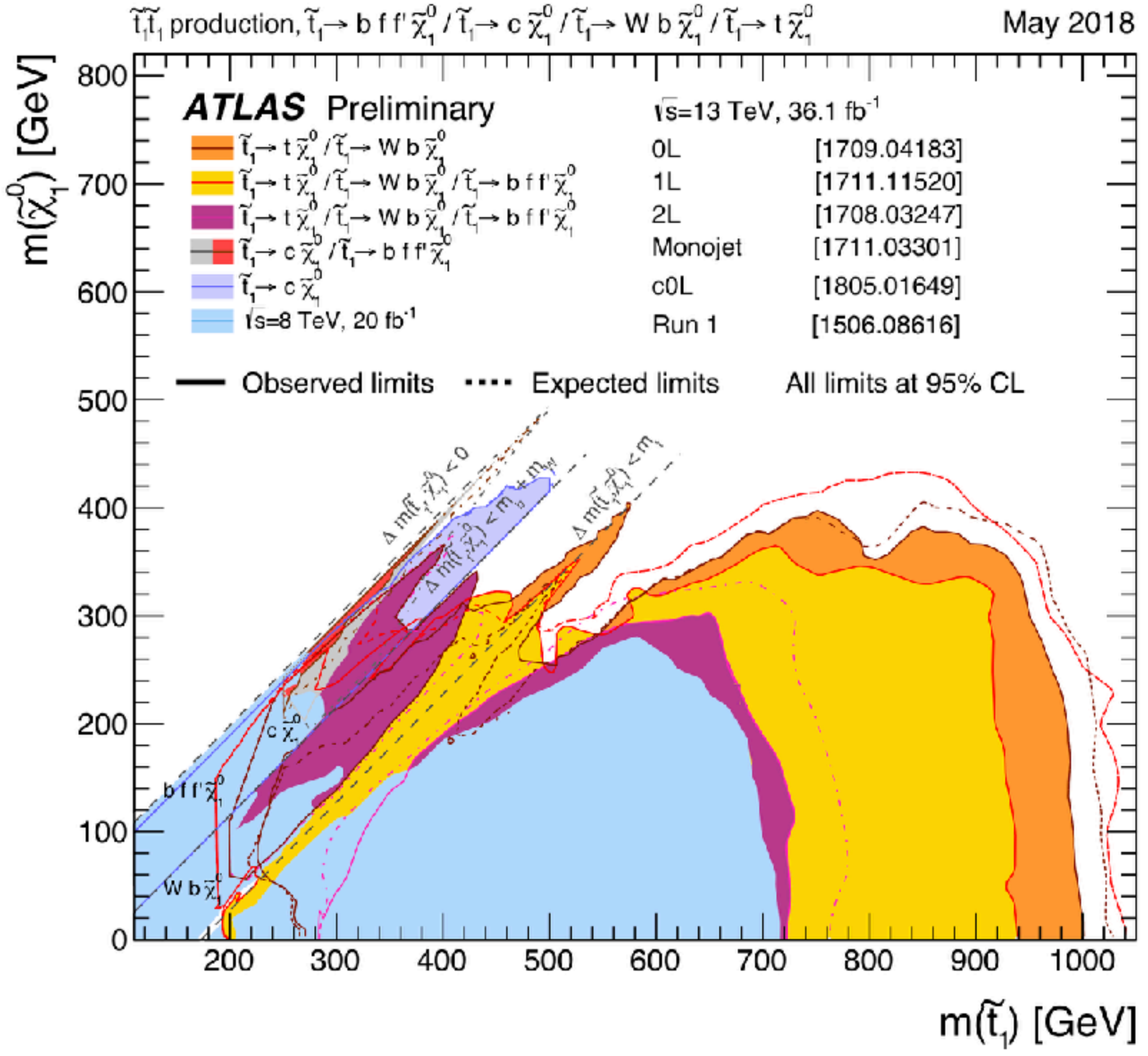


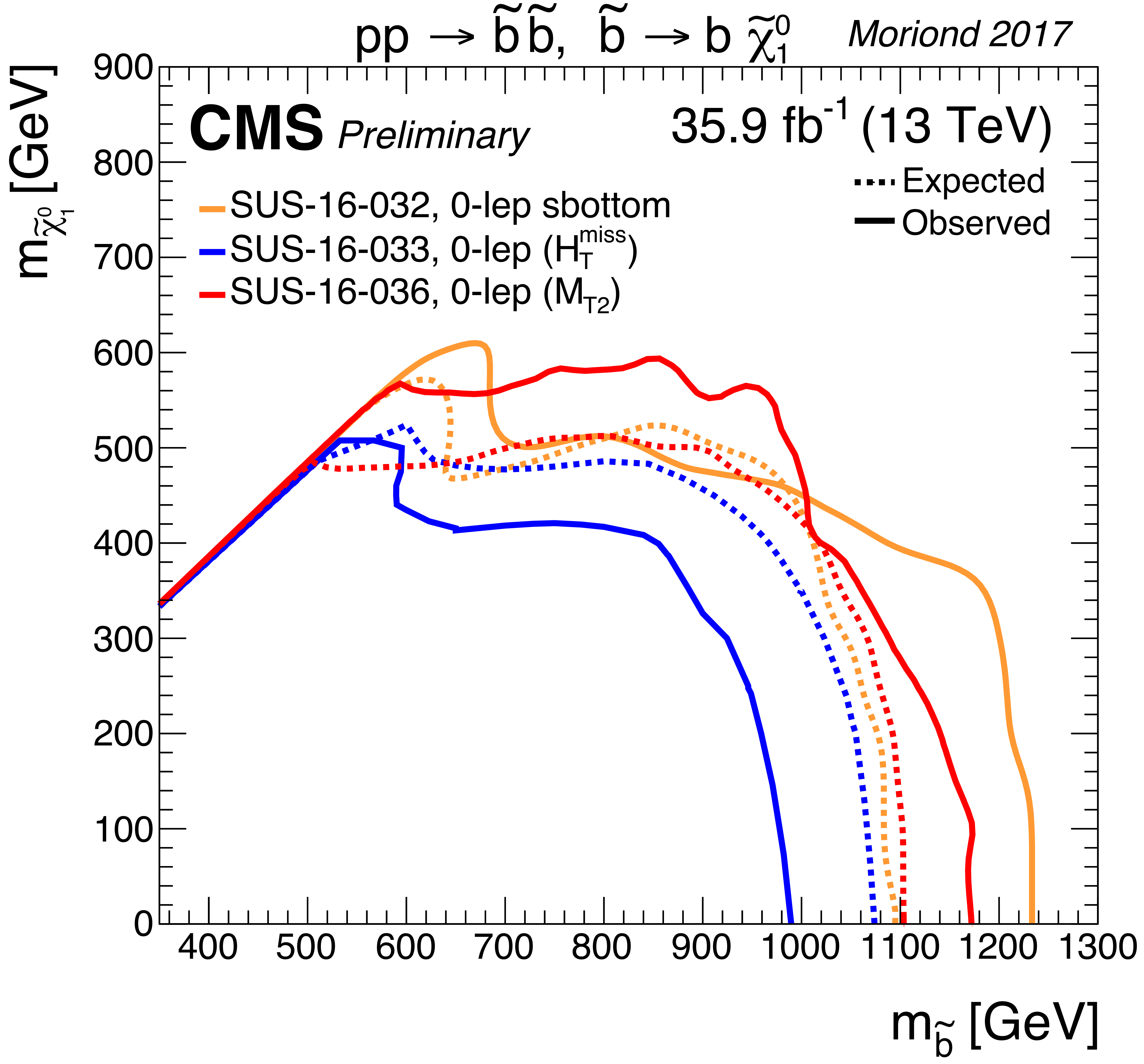
WHERE DO WE STAND? GAUGINOS





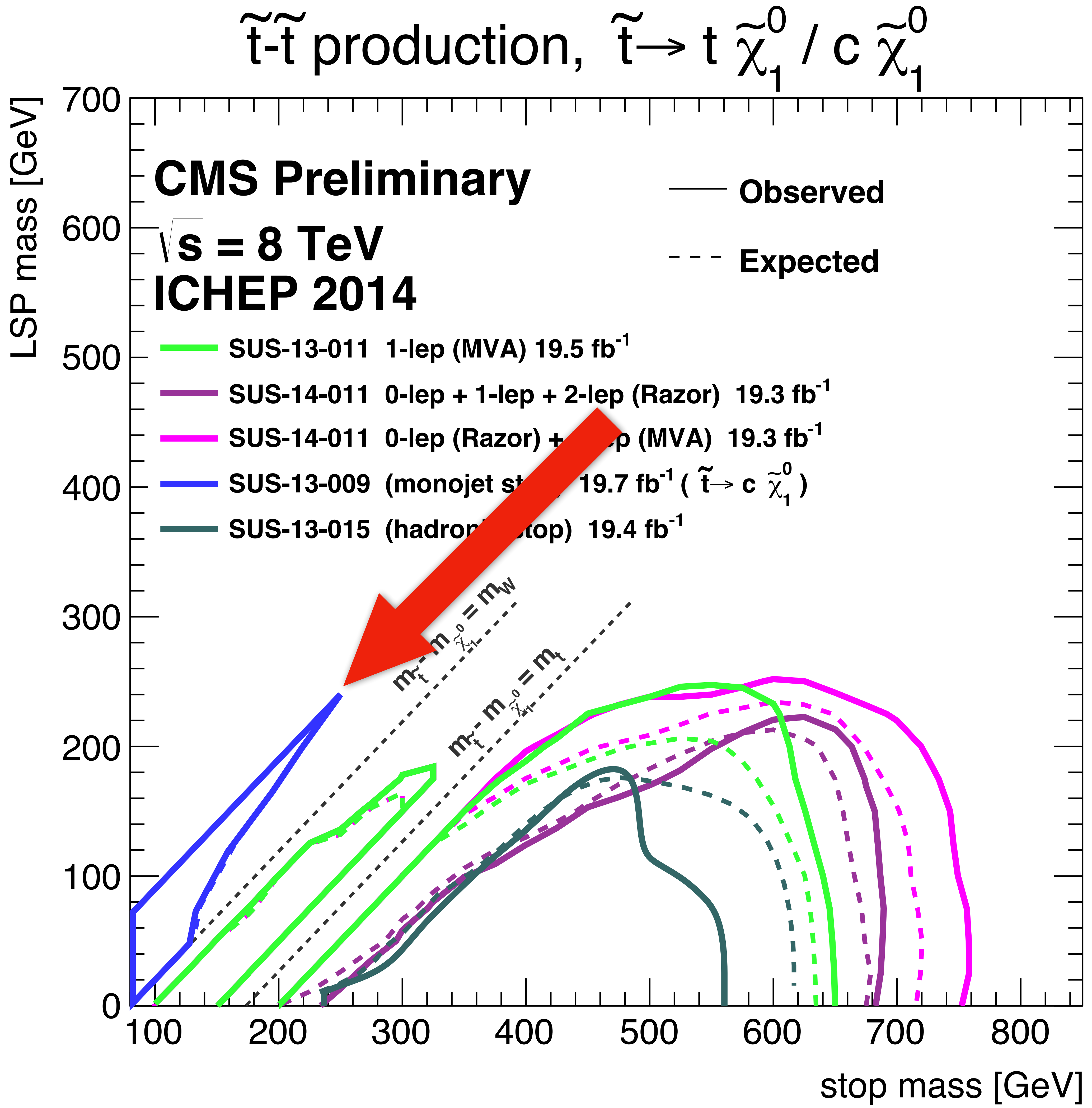
WHERE DO WE STAND? SQUARKS





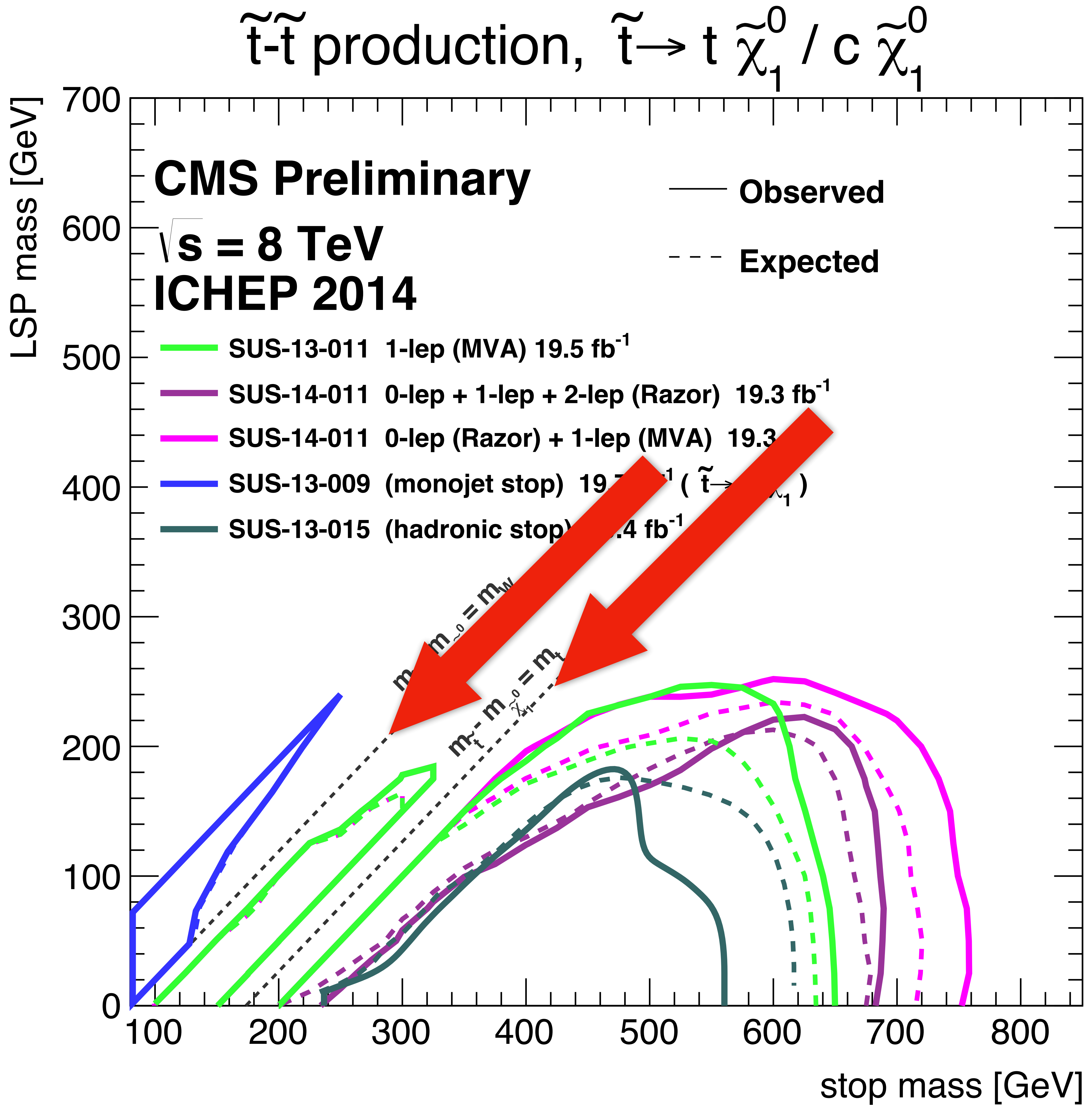
LOOKING IN THE CORNERS

- *There regions which are complicated to explore*
  - *compressed spectra → soft objects in the final state*
  - *mass differences ~ SM particle masses → valleys of reduced sensitivity*
- *One can approach these regions*
  - *directly, dealing with soft objects*
  - *indirectly, exploring cascade production (as with DM)*



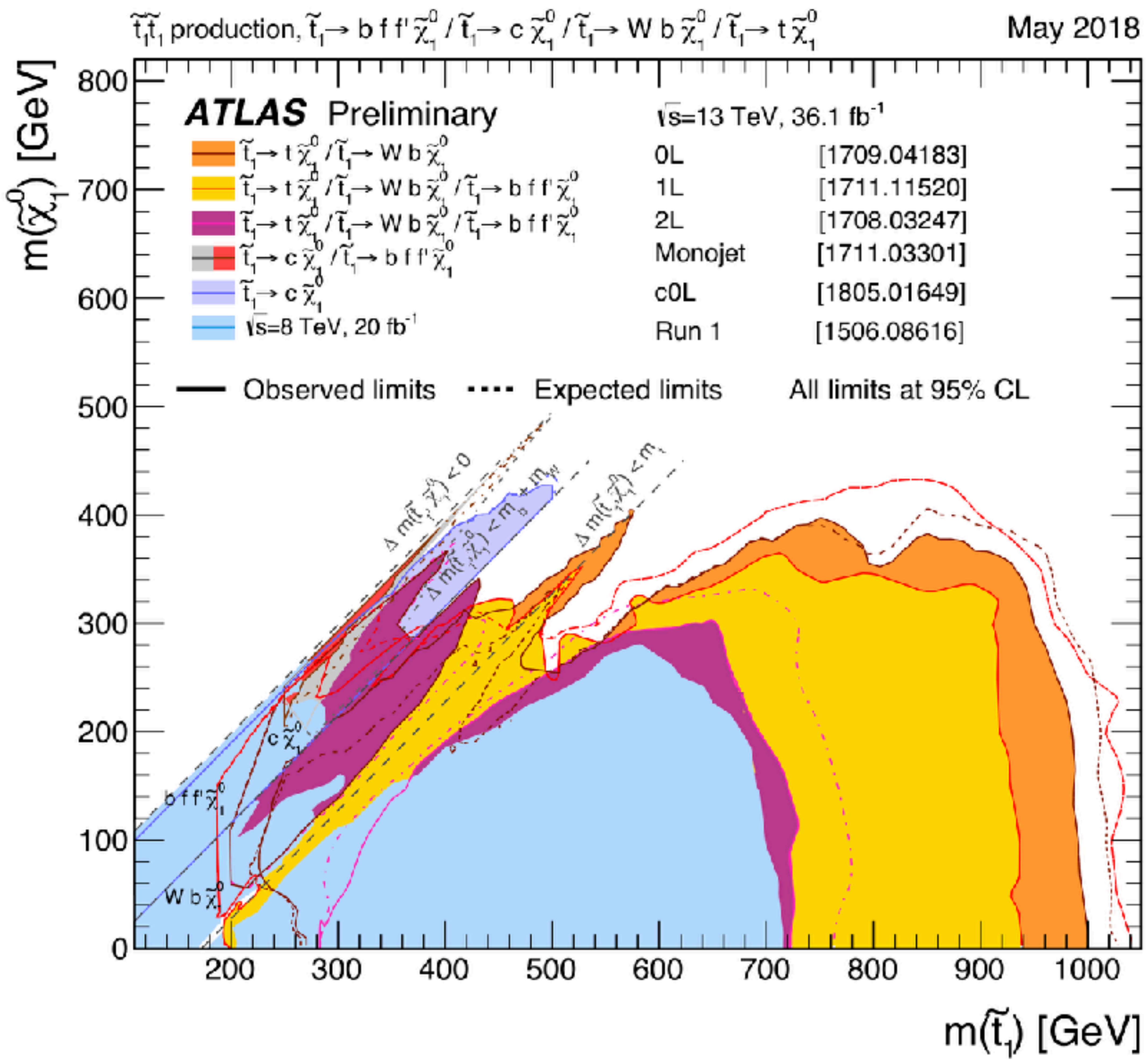
LOOKING IN THE CORNERS

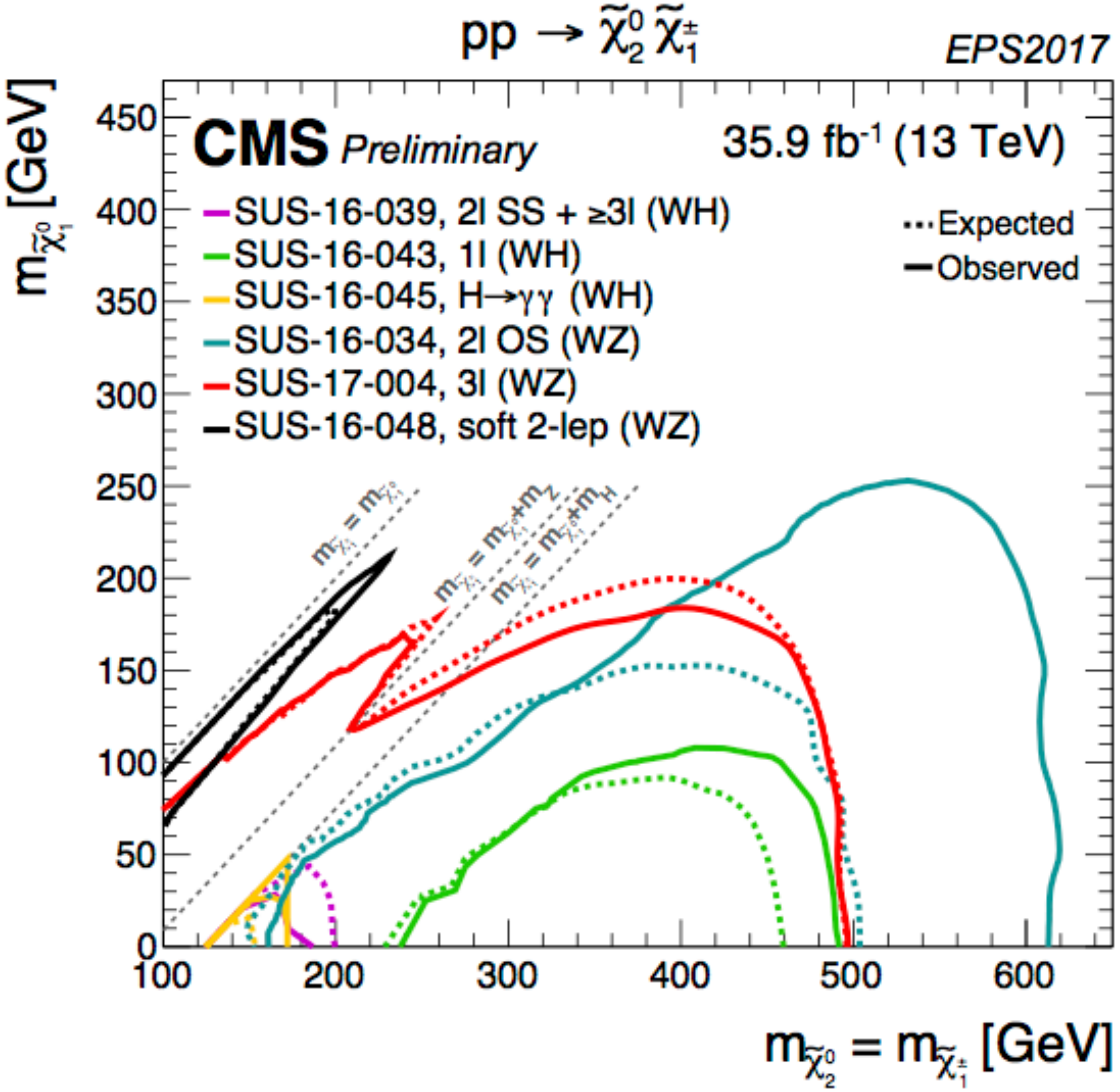
- *There regions which are complicated to explore*
  - *compressed spectra → soft objects in the final state*
  - ***mass differences ~ SM particle masses*** *→ valleys of reduced sensitivity*
- *One can approach these regions*
  - *directly, dealing with soft objects*
  - *indirectly, exploring cascade production (as with DM)*



LOOKING IN THE CORNERS

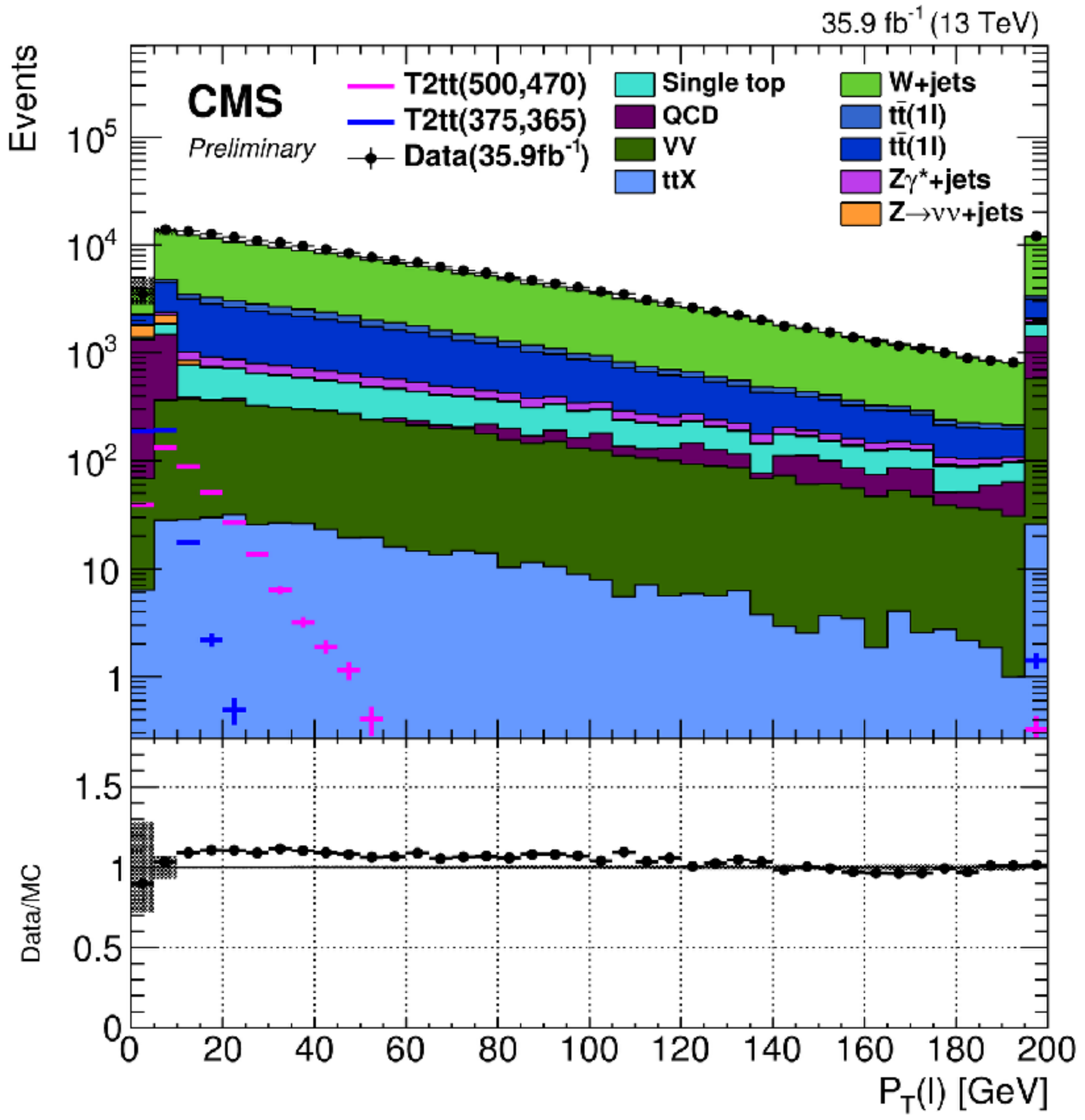
- *This is where a lot of the RunII work went*





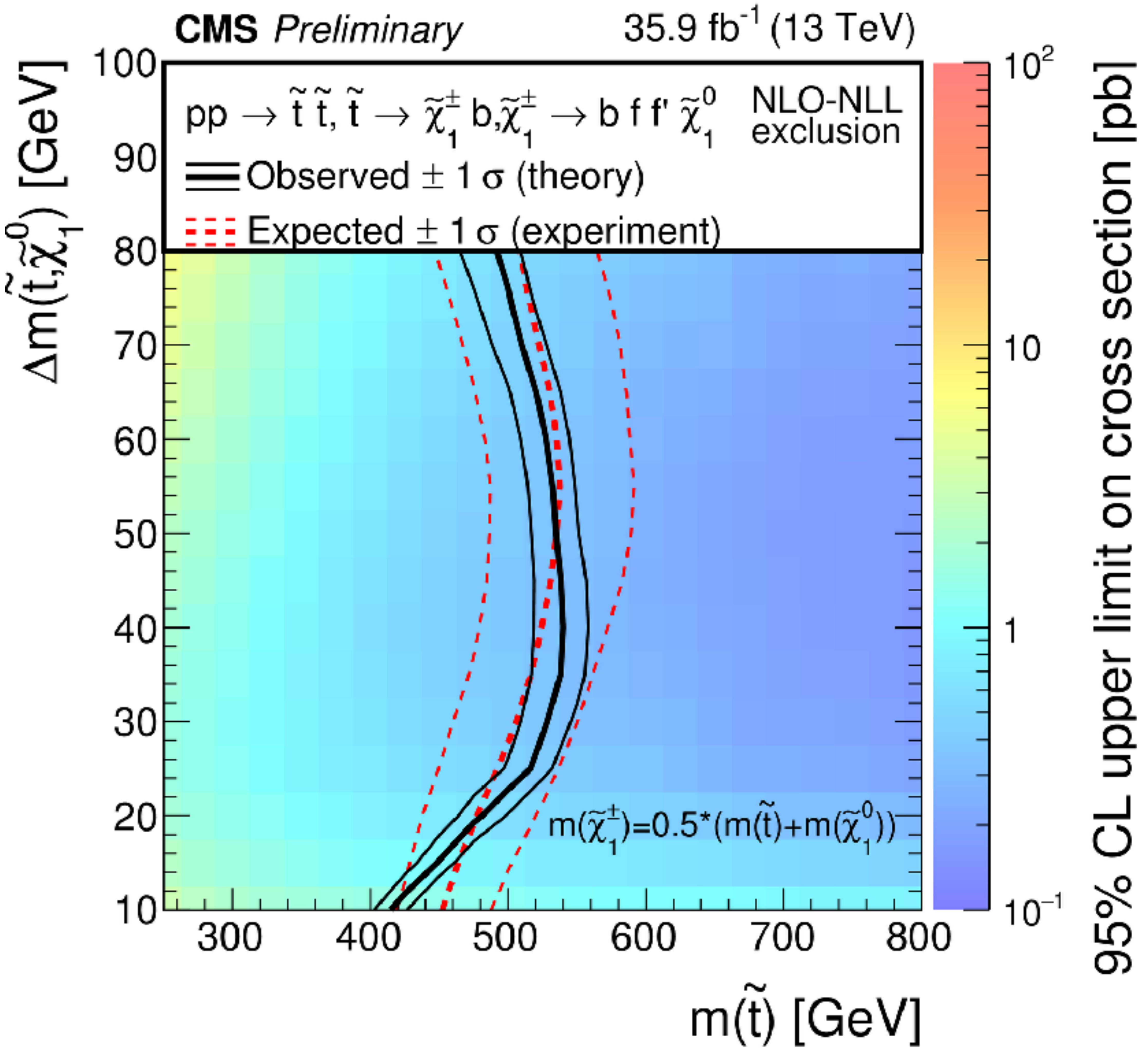
SOFT LEPTONS

- To probe SUSY models with ~GeV leptons, one has to start from the trigger
  - dedicated L1 and HLT seeds looking at lepton+jets
- If this is place, the analysis is quite similar to the others
  - but new backgrounds appear, e.g. QCD (usually negligible in leptonic final states
- The challenge will be to keep sensitivity there with HL-LHC (maybe using again the ISR trick?)



SOFT LEPTONS

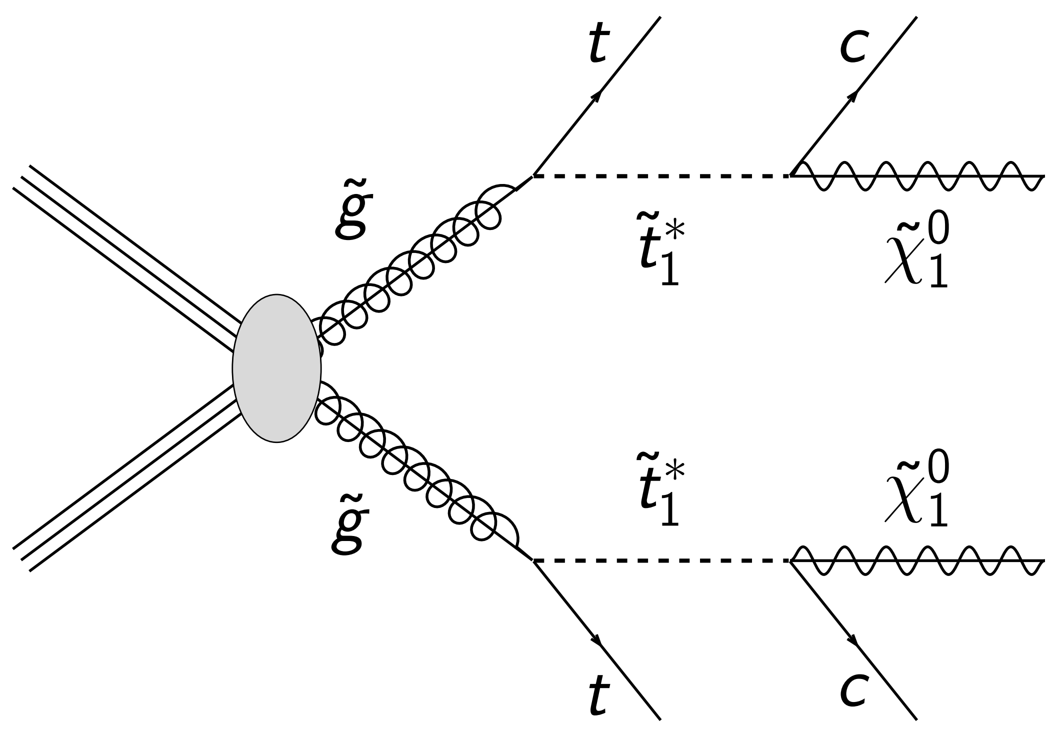
- To probe SUSY models with ~GeV leptons, one has to start from the trigger
  - dedicated L1 and HLT seeds looking at lepton+jets
- If this is place, the analysis is quite similar to the others
  - but new backgrounds appear, e.g. QCD (usually negligible in leptonic final states
- The challenge will be to keep sensitivity there with HL-LHC (maybe using again the ISR trick?)

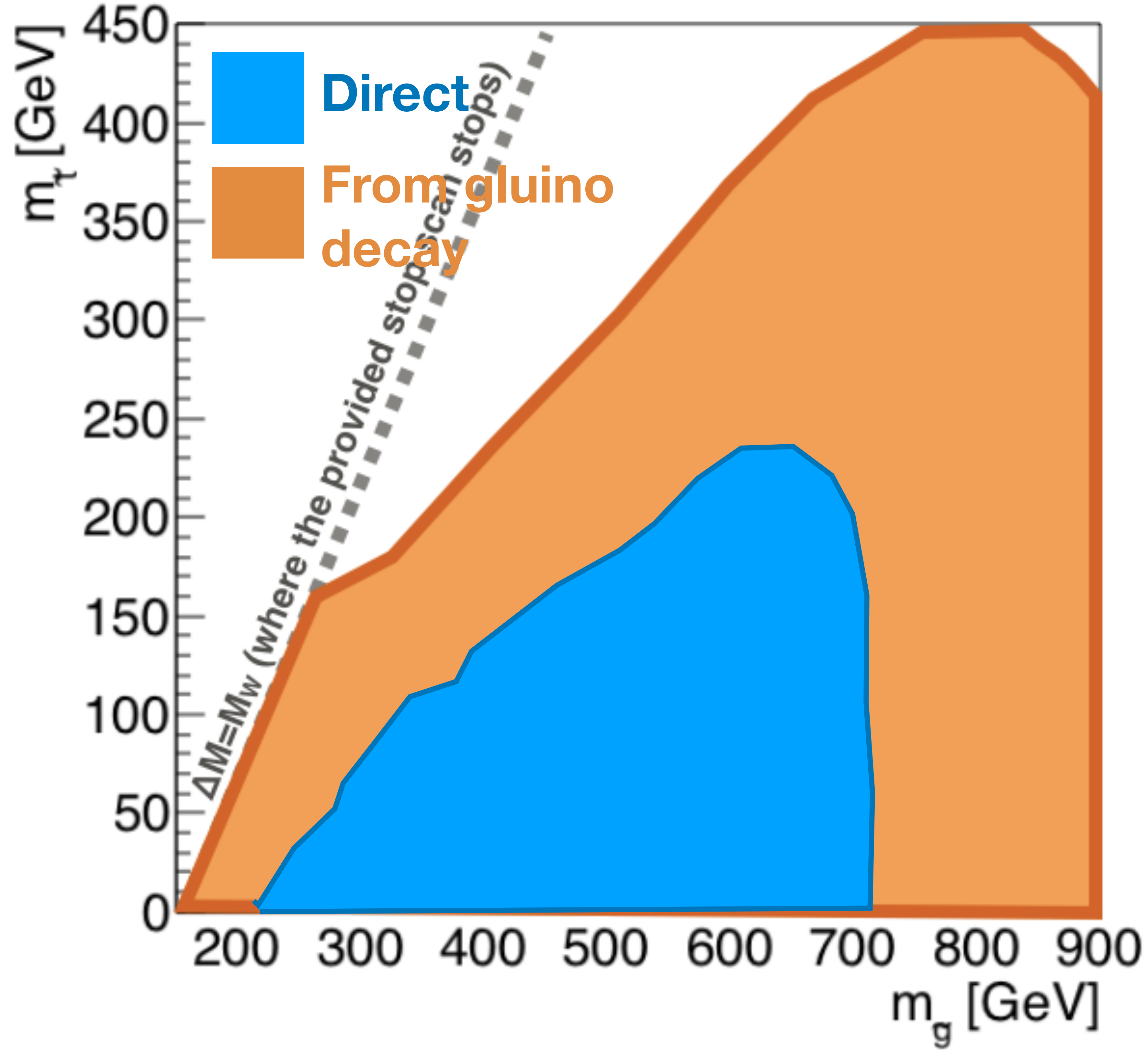


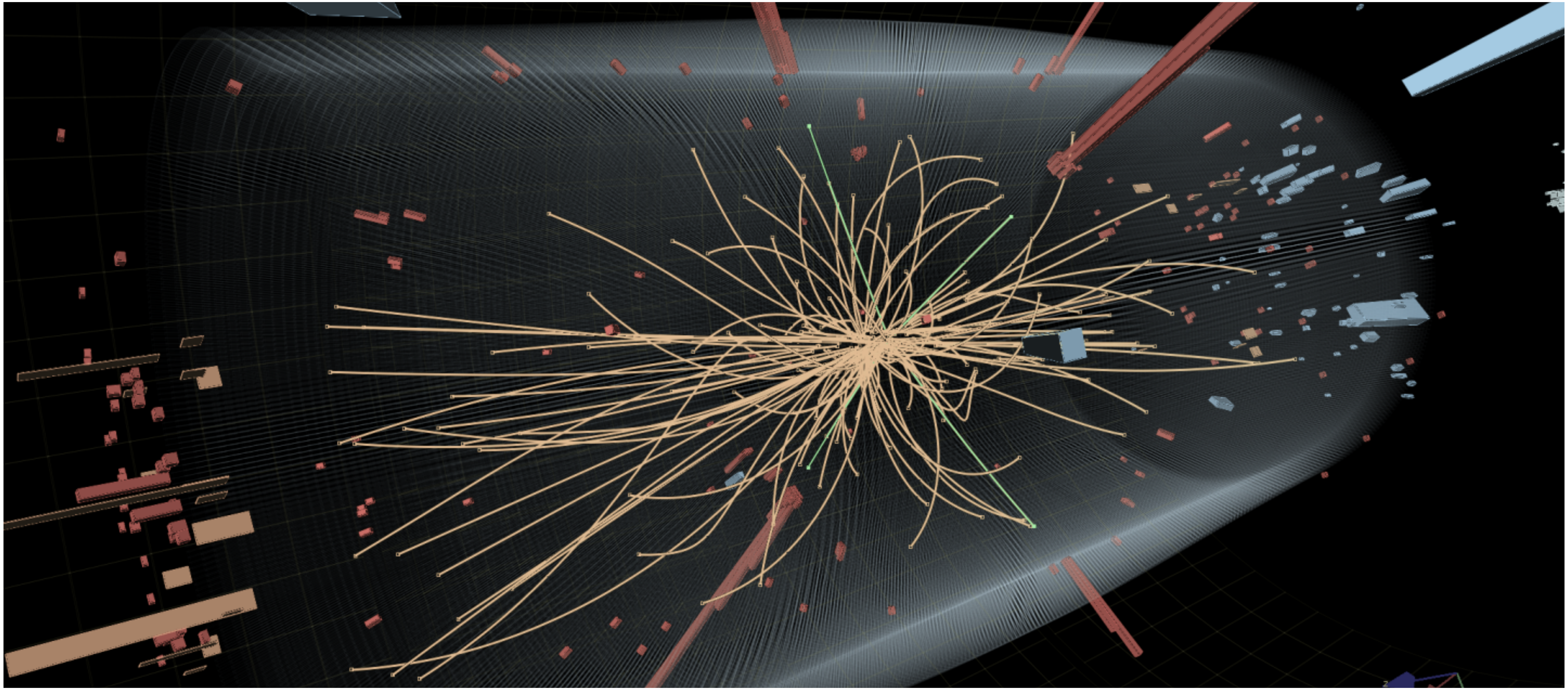
DIRECT VS CASCADE PRODUCTION

- *One can use cascade production to probe compress spectra:*
  - *for example, compressed stop*
  - *when produced directly, difficult to observe soft charm jets*
  - *when produced from gluinos, can see the top quark*
  - ***Interesting new regime: for very heavy gluinos, the Ws from t (and eventually the full t) could merge into a single heavy jet → substructure***

$$\tilde{t} \rightarrow c\chi_1^0$$







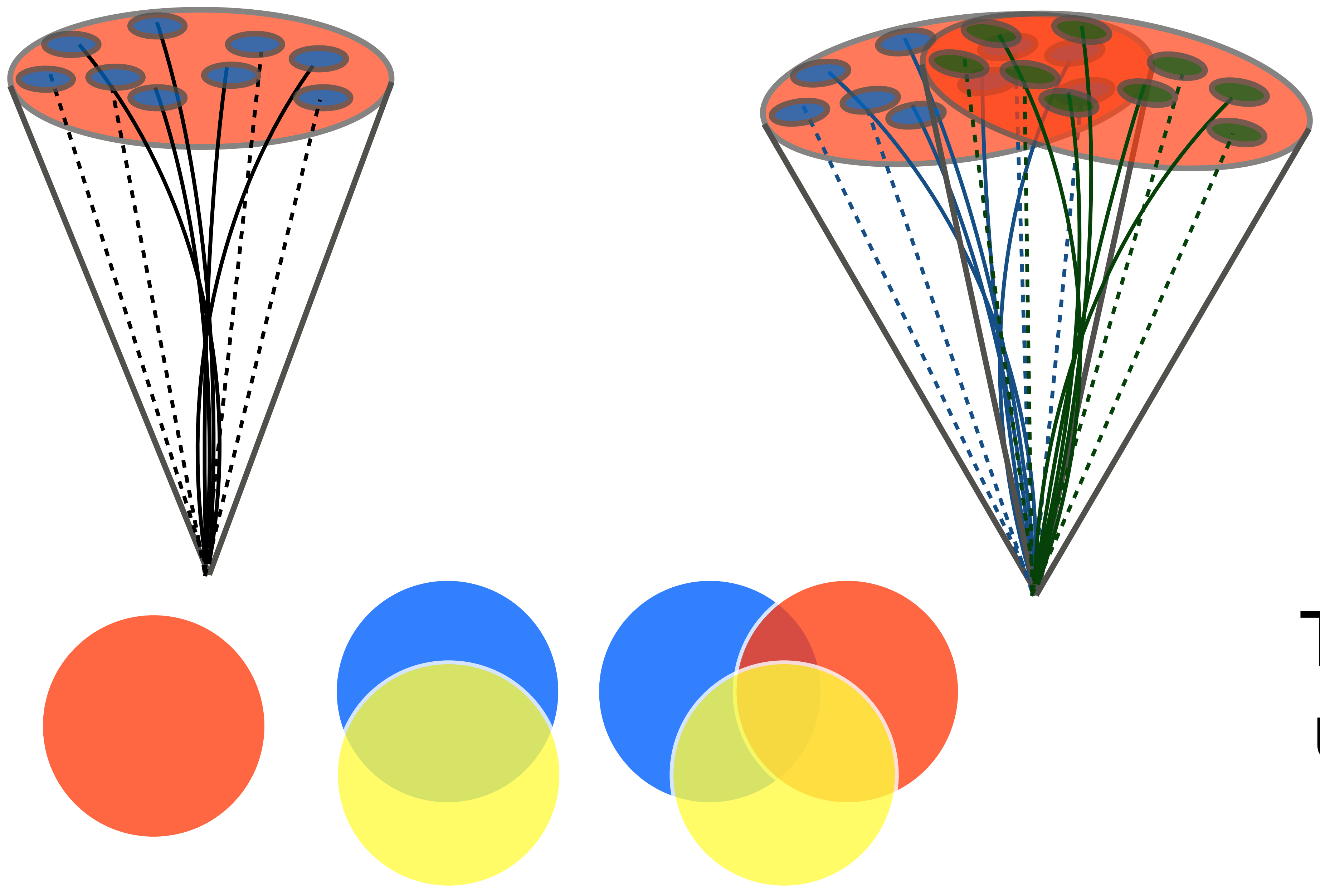
I open a Parenthesis here to tell you about jet substructure

FROM JETS TO BOOSTED JETS

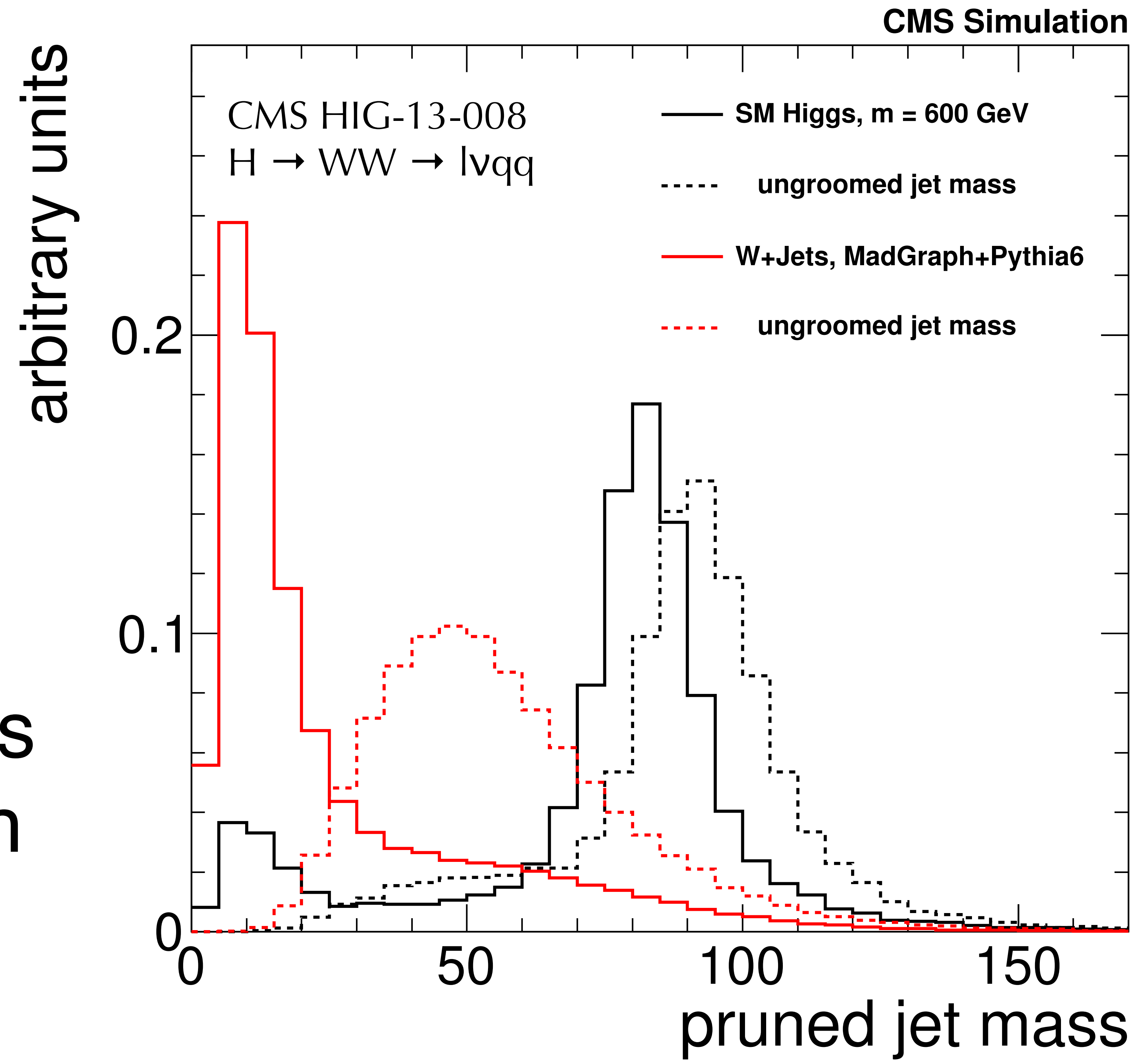
- *Heavy particles (e.g., W, Z, and H bosons, top quark) can decay to 2q/3q/4q final states, giving normally multi-jet signatures*

- *For large enough pT, the decay products might merge into a single jet*

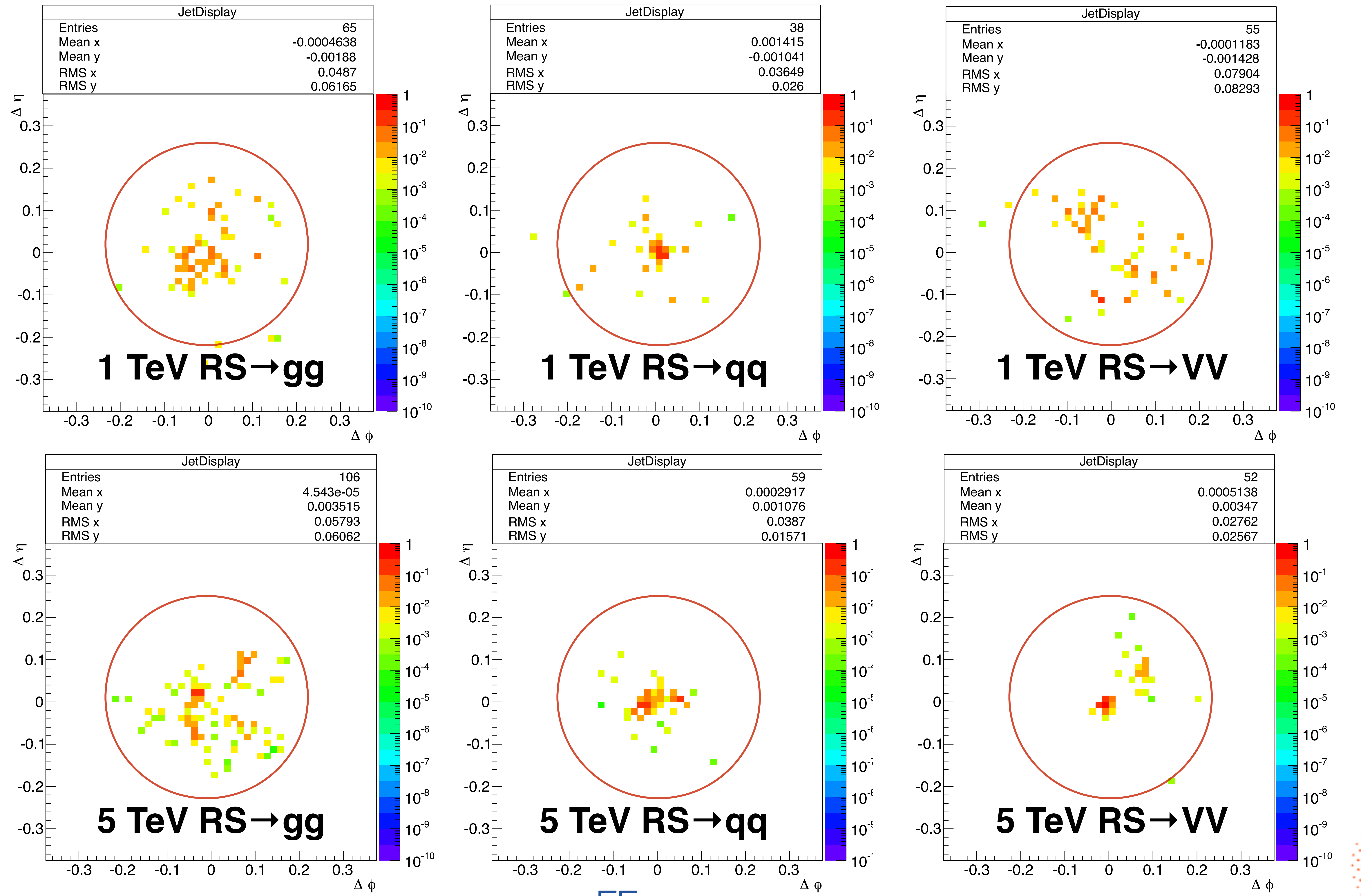
- *These jets are special: the mass of the jet peaks at the "right" value (unlike QCD jets, for which large mass values are generated by QCD)*



$\Delta R \sim 2\, M_W/p_T$
(to be compared with jet size R)
Typically large jets used (Anti-Kt with R=0.8)



JET SUBSTRUCTURE



SUBSTRUCTURE VARIABLES

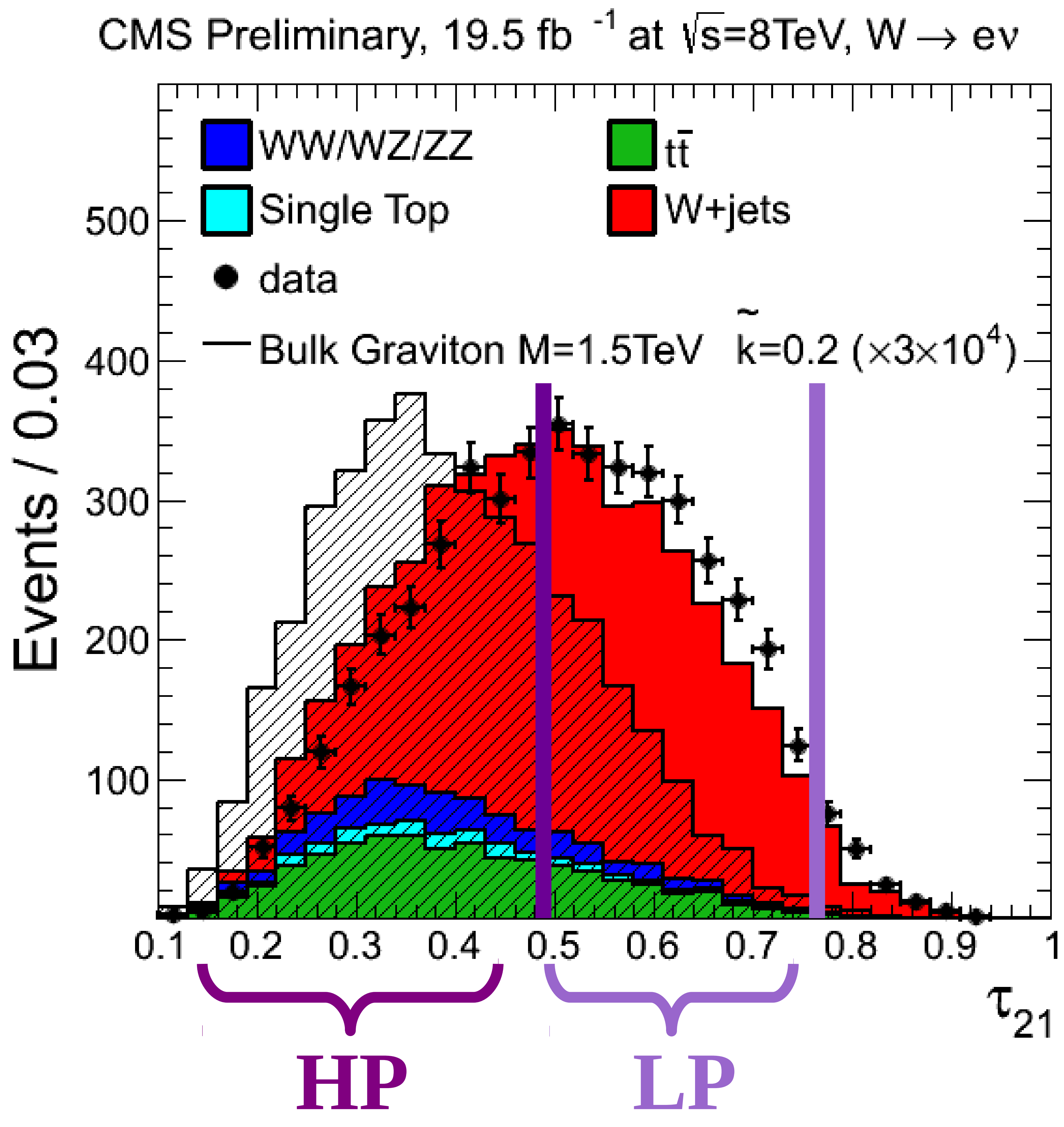
- *Several jet-shape variables proposed to quantify this behaviour (see ongoing [BOOST conference]() for a full overview)*

- *N-subjettiness is among the most popular*

  - *Quantify how well the constituents of a jet can be arranged in N subjects*

$$\tau_N^{(\beta)} = \frac{1}{p_{TJ}} \sum_{i \in \mathrm{Jet}} p_{Ti} \min\left\{R_{1i}^{\beta}, R_{2i}^{\beta}, \ldots, R_{Ni}^{\beta}\right\}$$

J. Thaler and K. Van Tilburg http://arxiv.org/abs/1011.2268



2-body: $\tau_1^{(1)}, \tau_1^{(2)}$

3-body: $\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(1)}, \tau_2^{(2)}$

4-body: $\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(1)}, \tau_3^{(2)}$

5-body: $\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(0.5)}, \tau_3^{(1)}, \tau_3^{(2)}, \tau_4^{(1)}, \tau_4^{(2)}$

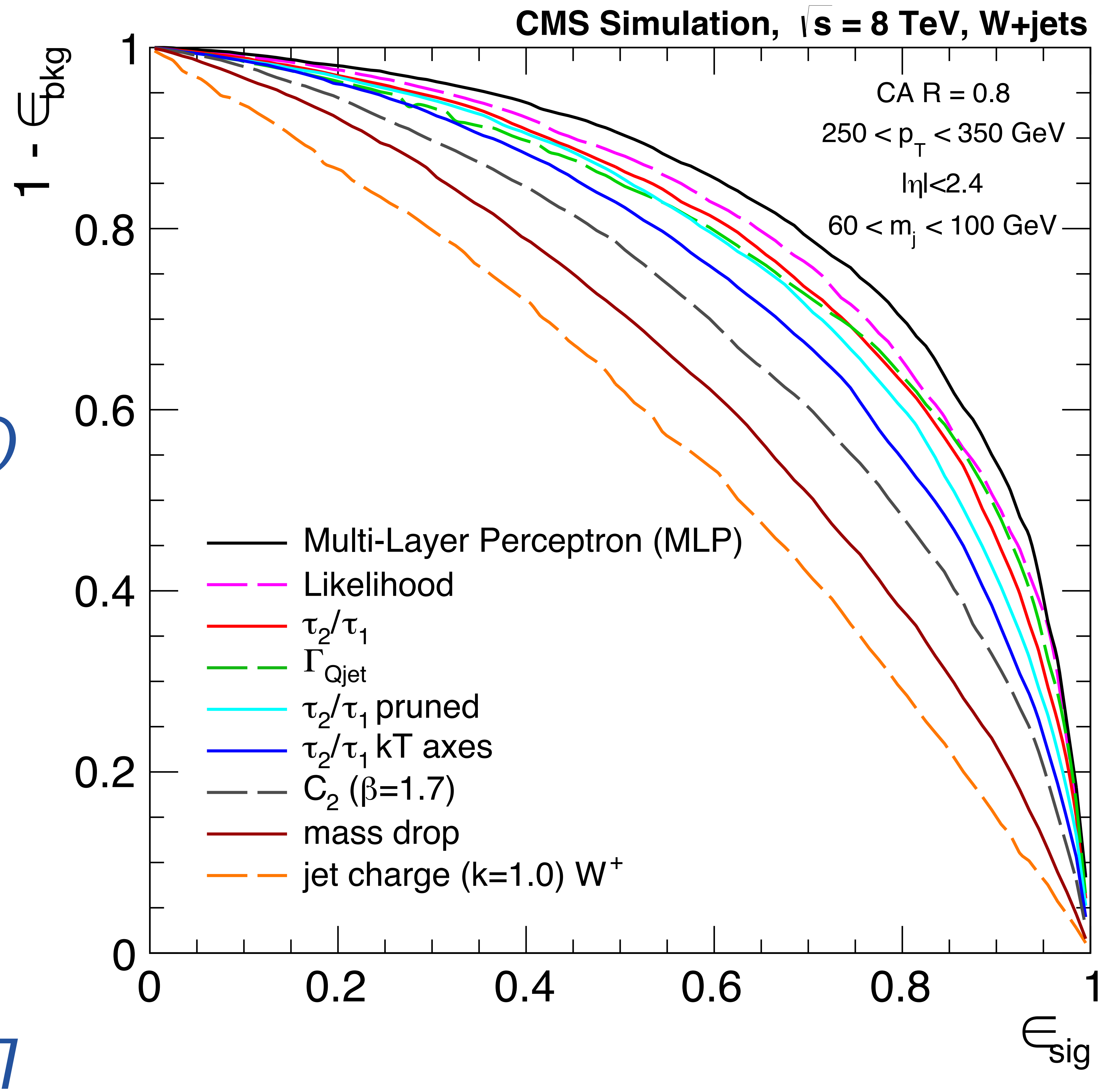
6-body: $\tau_1^{(0.5)}, \tau_1^{(1)}, \tau_1^{(2)}, \tau_2^{(0.5)}, \tau_2^{(1)}, \tau_2^{(2)}, \tau_3^{(0.5)}, \tau_3^{(1)}, \tau_3^{(2)}, \tau_4^{(0.5)}, \tau_4^{(1)}, \tau_4^{(2)}, \tau_5^{(1)}, \tau_5^{(2)}$

  - *Can construct a complete basis by computing τN it for several N*

European Research Council

A DIJET TAGGER

- A typical tagger would consist of
  - A jet grooming procedure (trimming, pruning, soft drop) to remove soft radiation in the jet (and pileup, to some extent)
  - A (post-grooming) jet mass cut
  - A cut on an appropriate set of substructure variables
  - For instance, S vs B discrimination in CMS is optimal for di-subjet (W/Z/H) when τ2/τ1 ratio is considered

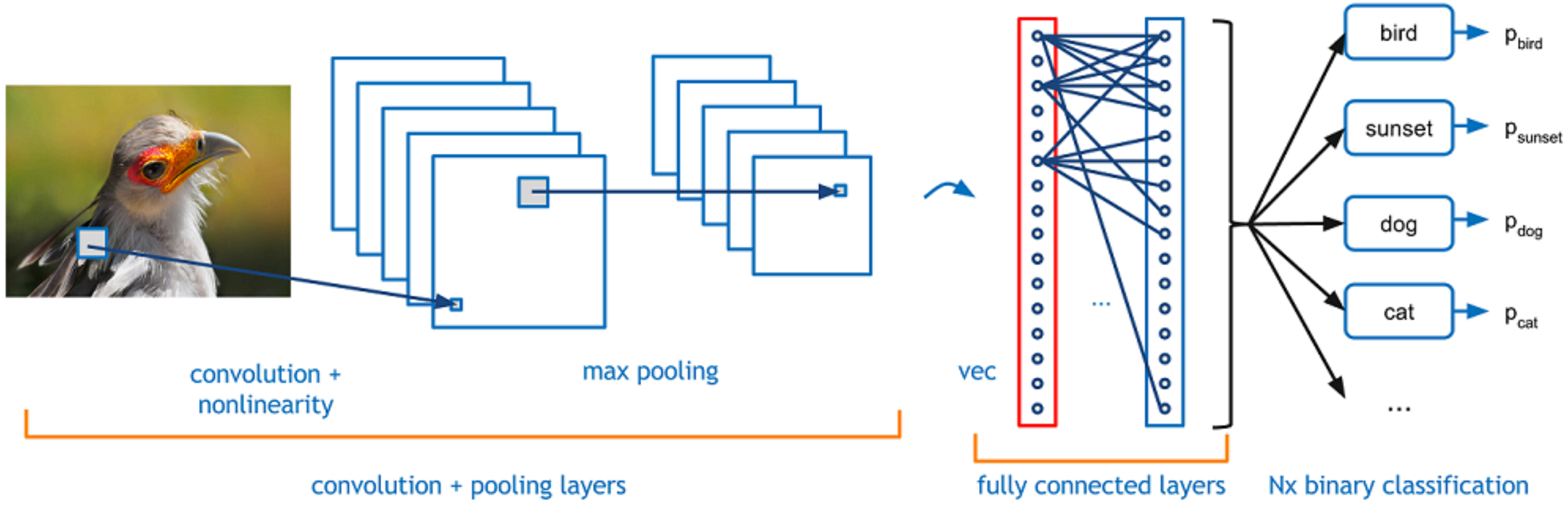


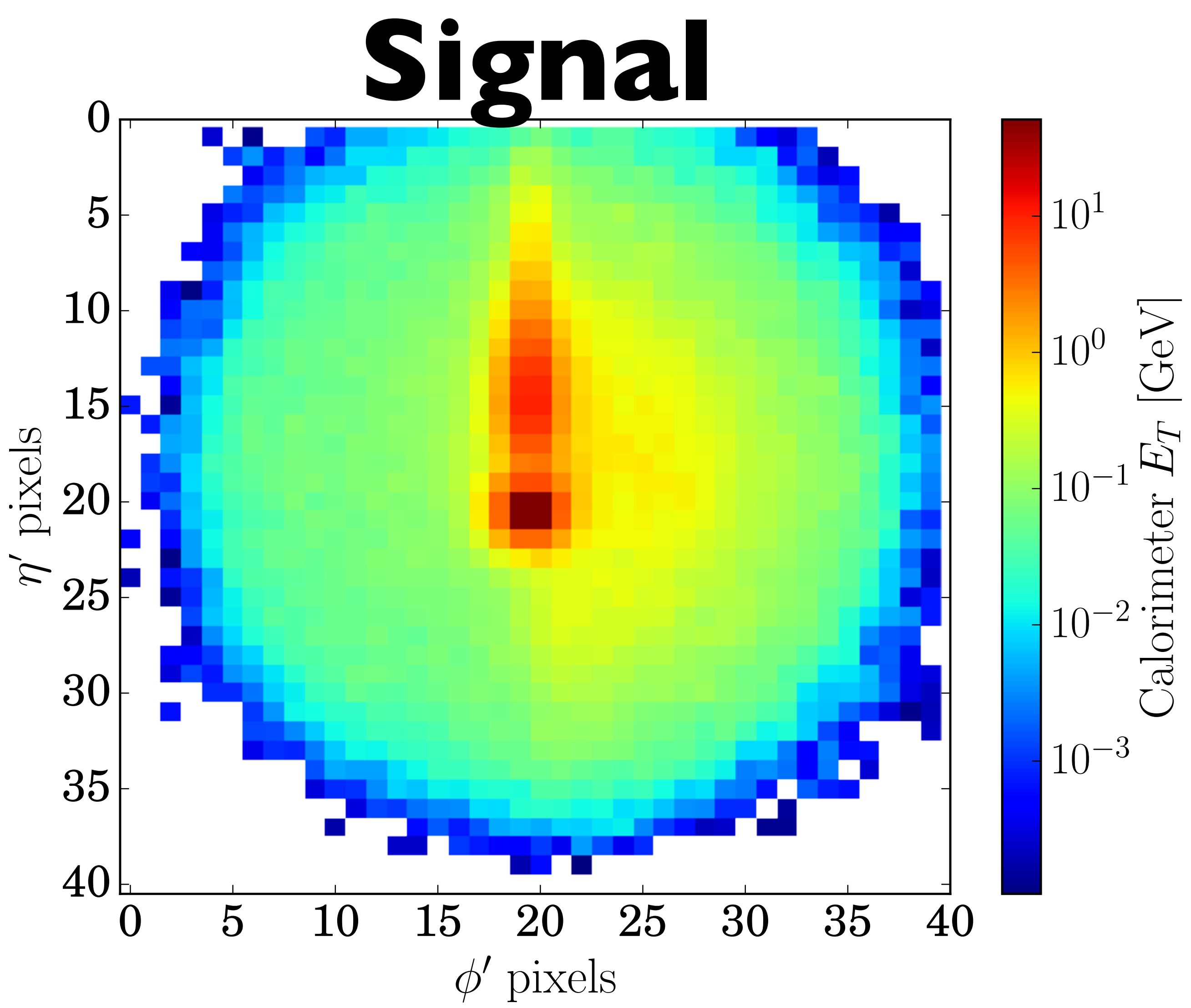
CMS-PAS-JME-13-006

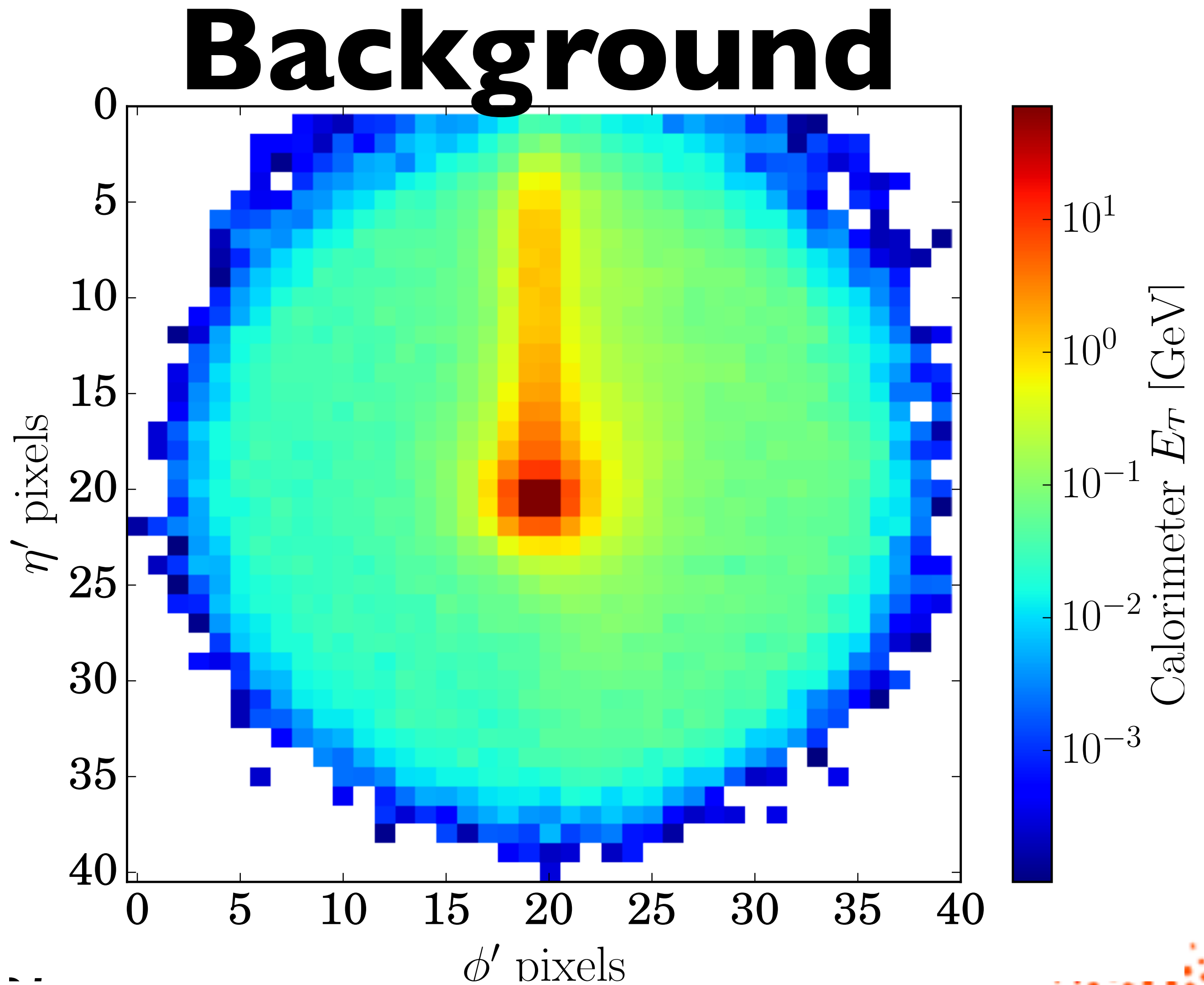
DEEP LEARNING TAGGING

- One can imagine a jet as an image impressed by energy deposits on calorimeters

- On this image, one can apply modern computing-vision techniques, e.g., Convolutional Neural networks





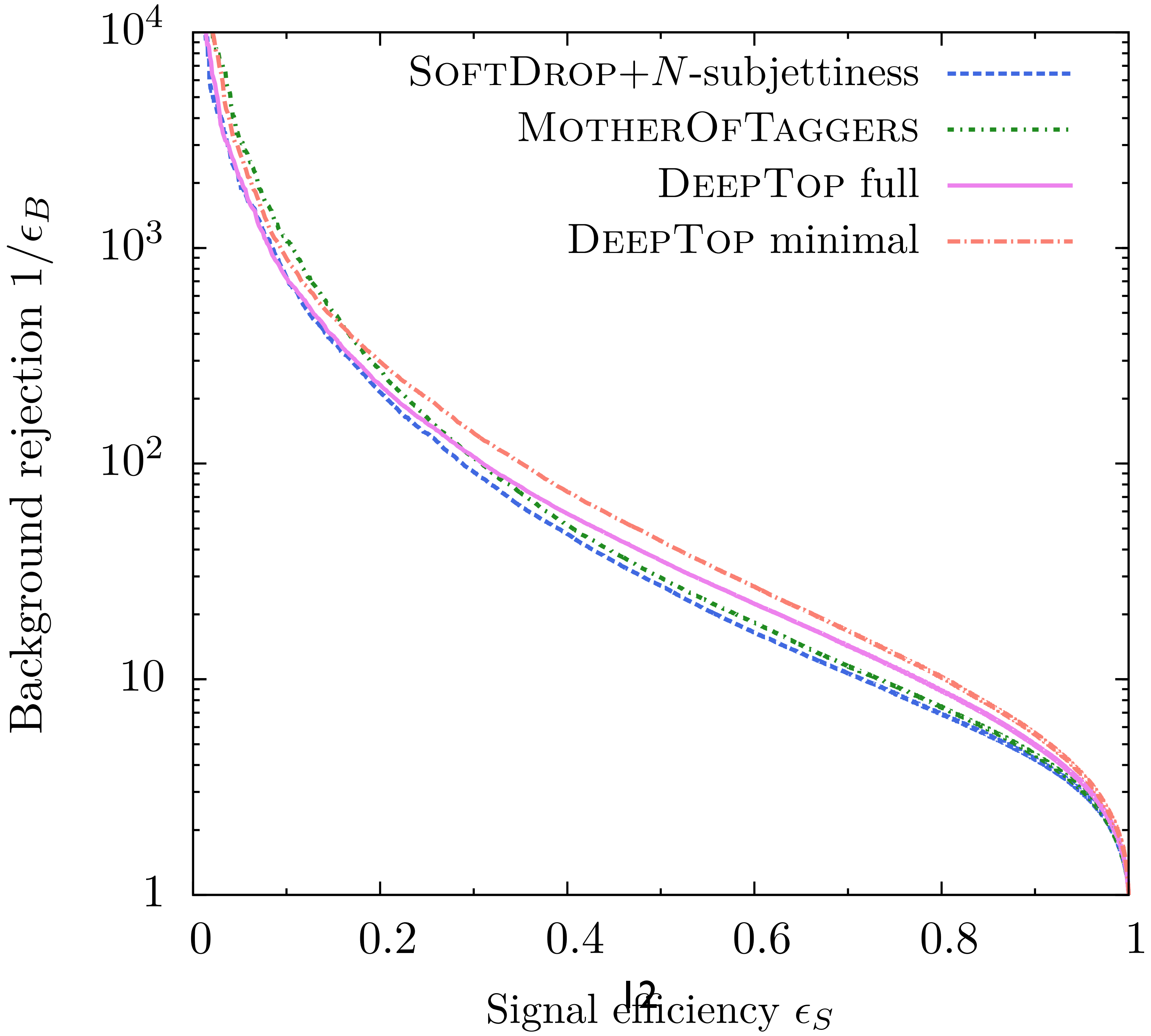


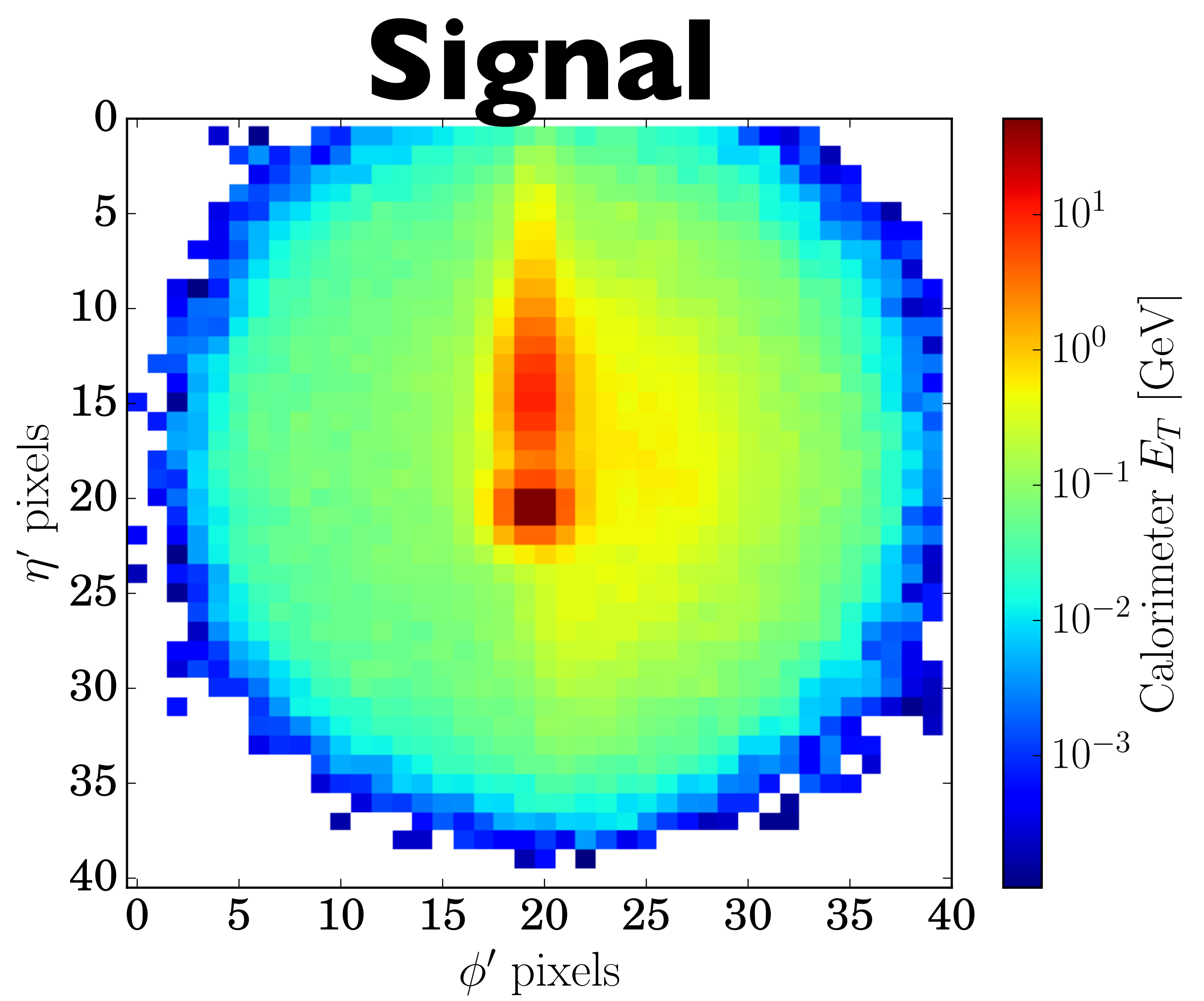
arXiv: 1701.08784

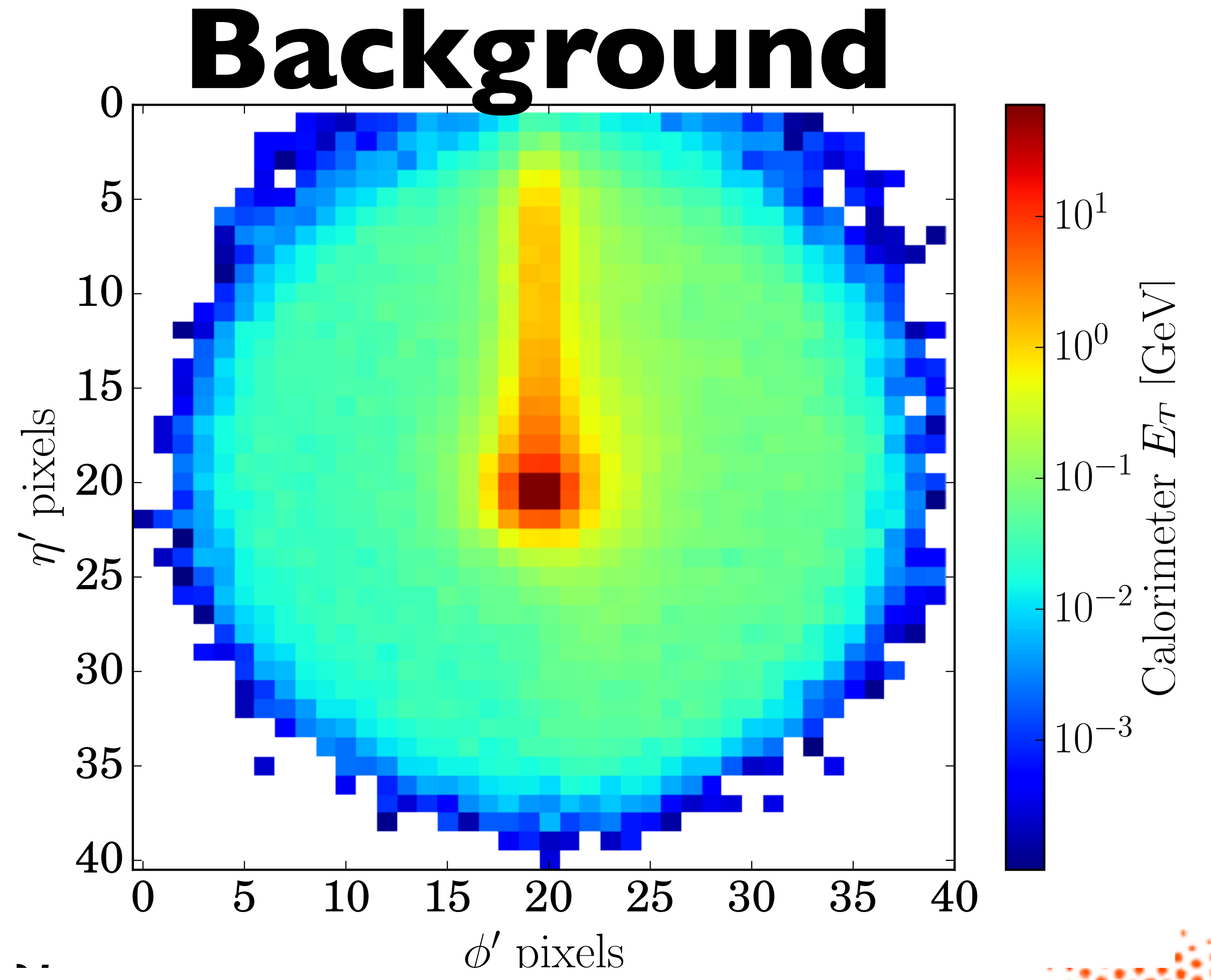
European Research Council

DEEP LEARNING TAGGING

- *One can imagine a jet as an image impressed by energy deposits on calorimeters*
- *On this image, one can apply modern computing-vision techniques, e.g., Convolutional Neural networks*



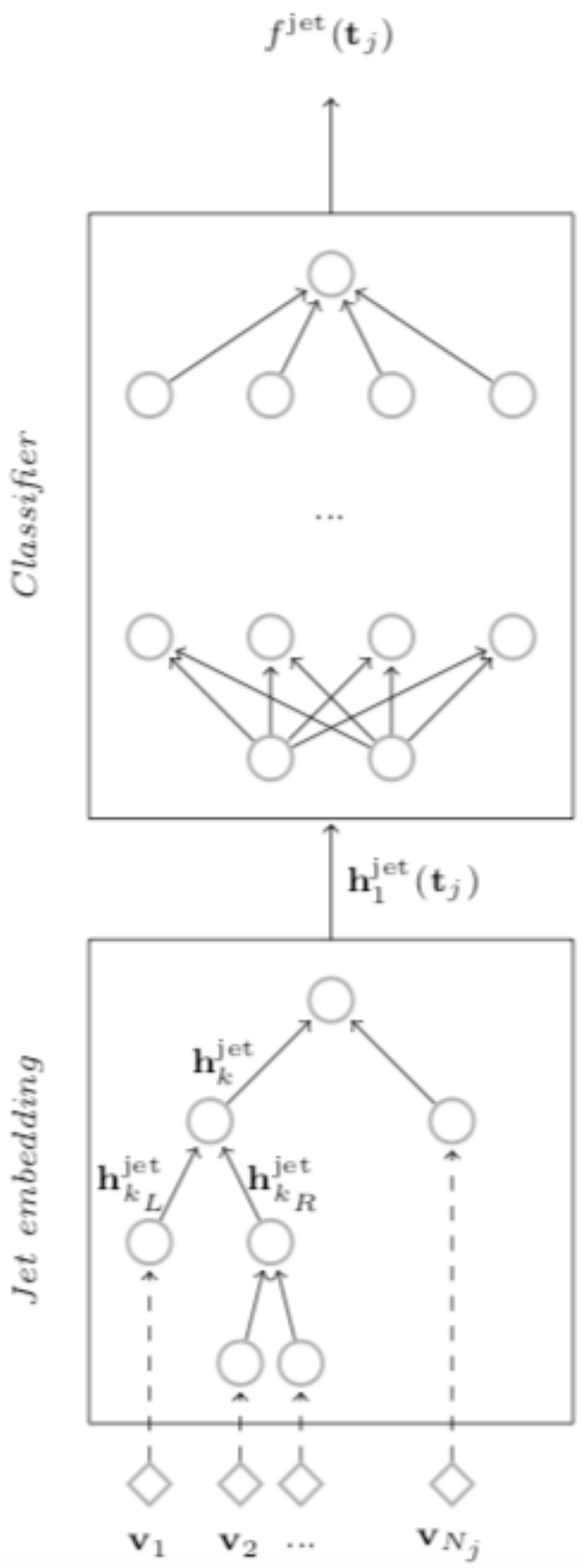


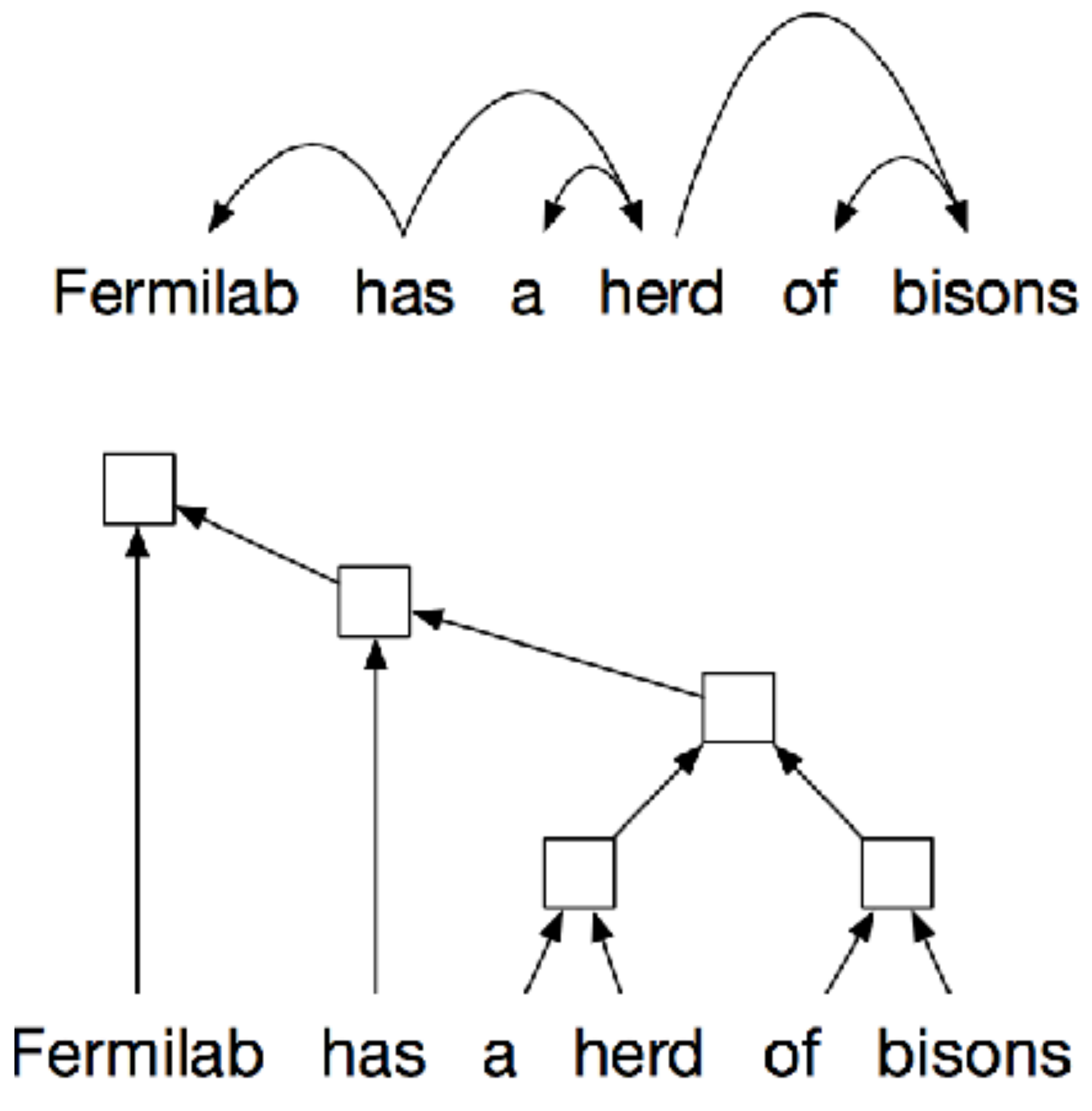


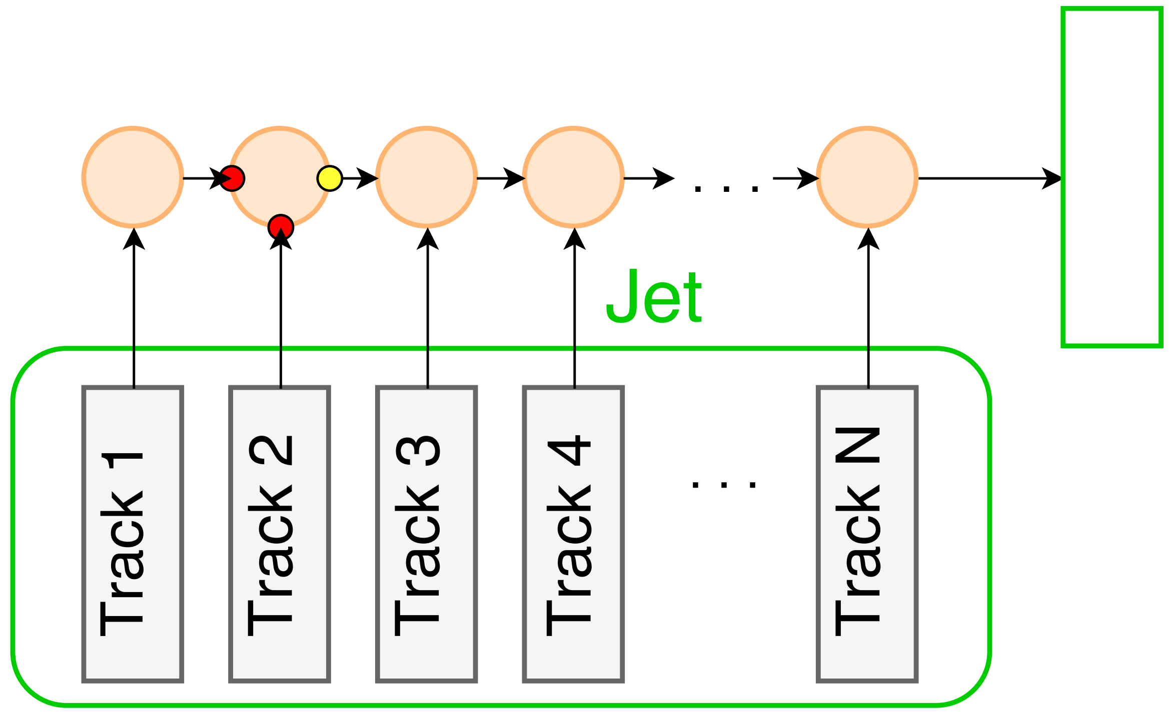
arXiv: 1701.08784

DEEP LEARNING TAGGING

- One can also represent a jet as a list of particles, ordered by QCD laws
  - Similar to words arranged in a sentence
  - Can use language processing techniques to tag a jet
- Deep learning offer a few opportunities (recurrent NN, recursive NN,…)
- Advantages:
  - No need to bin the image → can exploit the full angular resolution (e.g., tracking)
  - Very convenient for PF jets and track jets

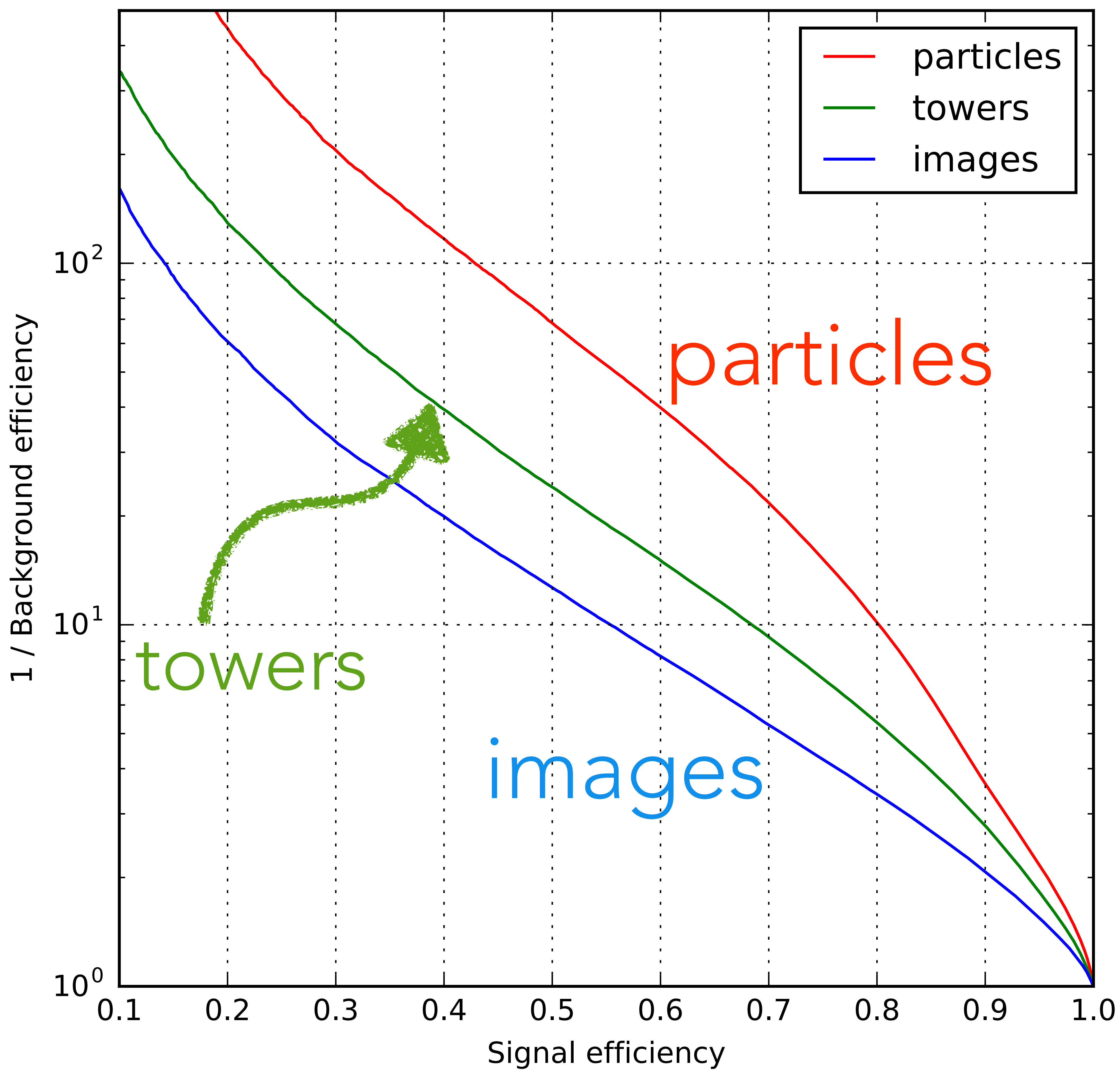






DEEP LEARNING TAGGING

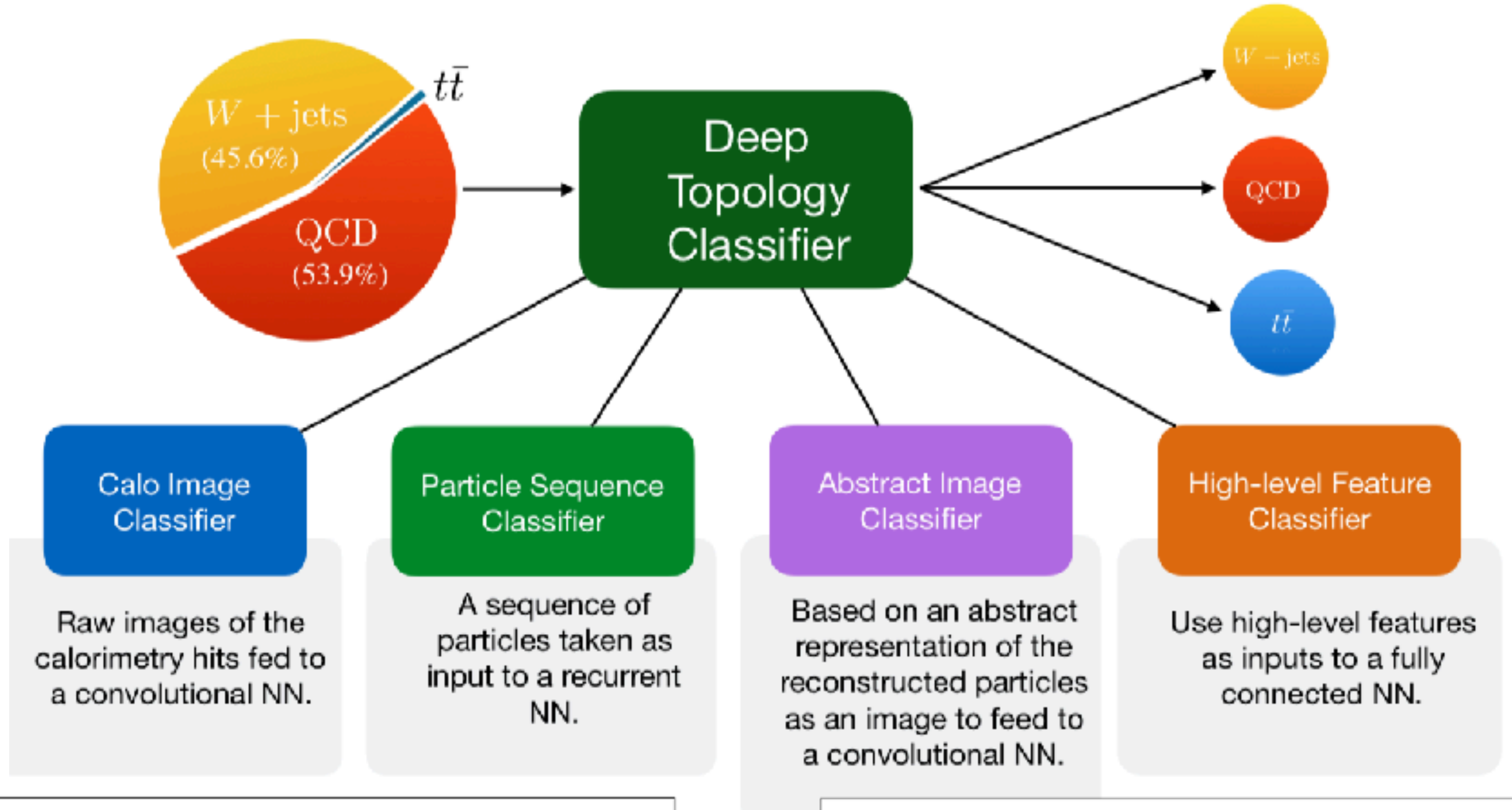
- *One can also represent a jet as a list of particles, ordered by QCD laws*
  - *Similar to words arranged in a sentence*
  - *Can use language processing techniques to tag a jet*
- *Deep learning offer a few opportunities (recurrent NN, recursive NN,…)*
- *Advantages:*
  - *No need to bin the image → can exploit the full angular resolution (e.g., tracking)*
  - *Very convenient for PF jets and track jets*

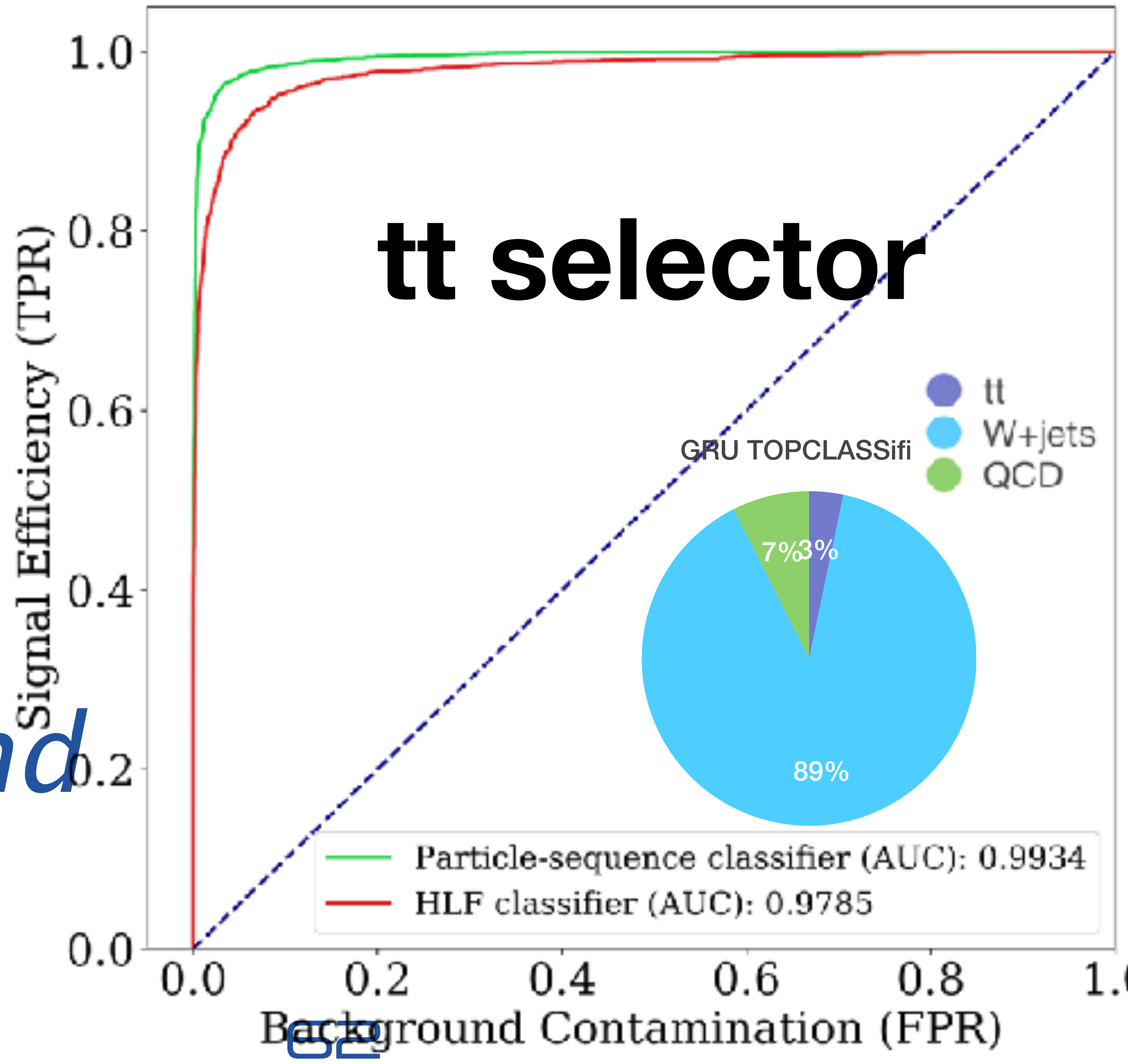


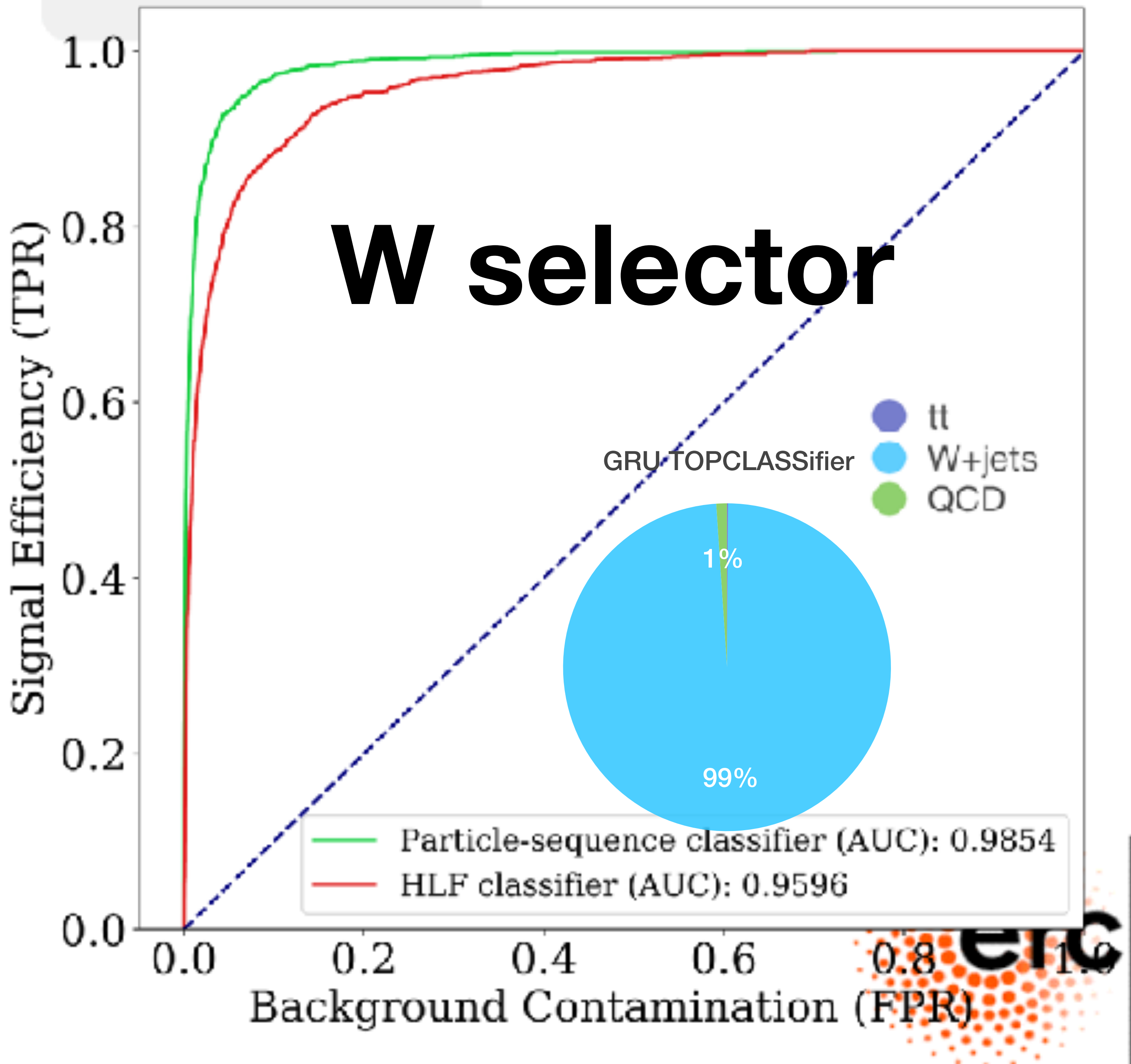
European Research Council

EVENT TAGGING WITH RNN

- *One can push this approach beyond jet, building a topology tagger for the full event*
- *Tested with simulated events, as a way to implement a more efficient trigger strategy*
- *Could have impact on the way we process and analyze data in the future*







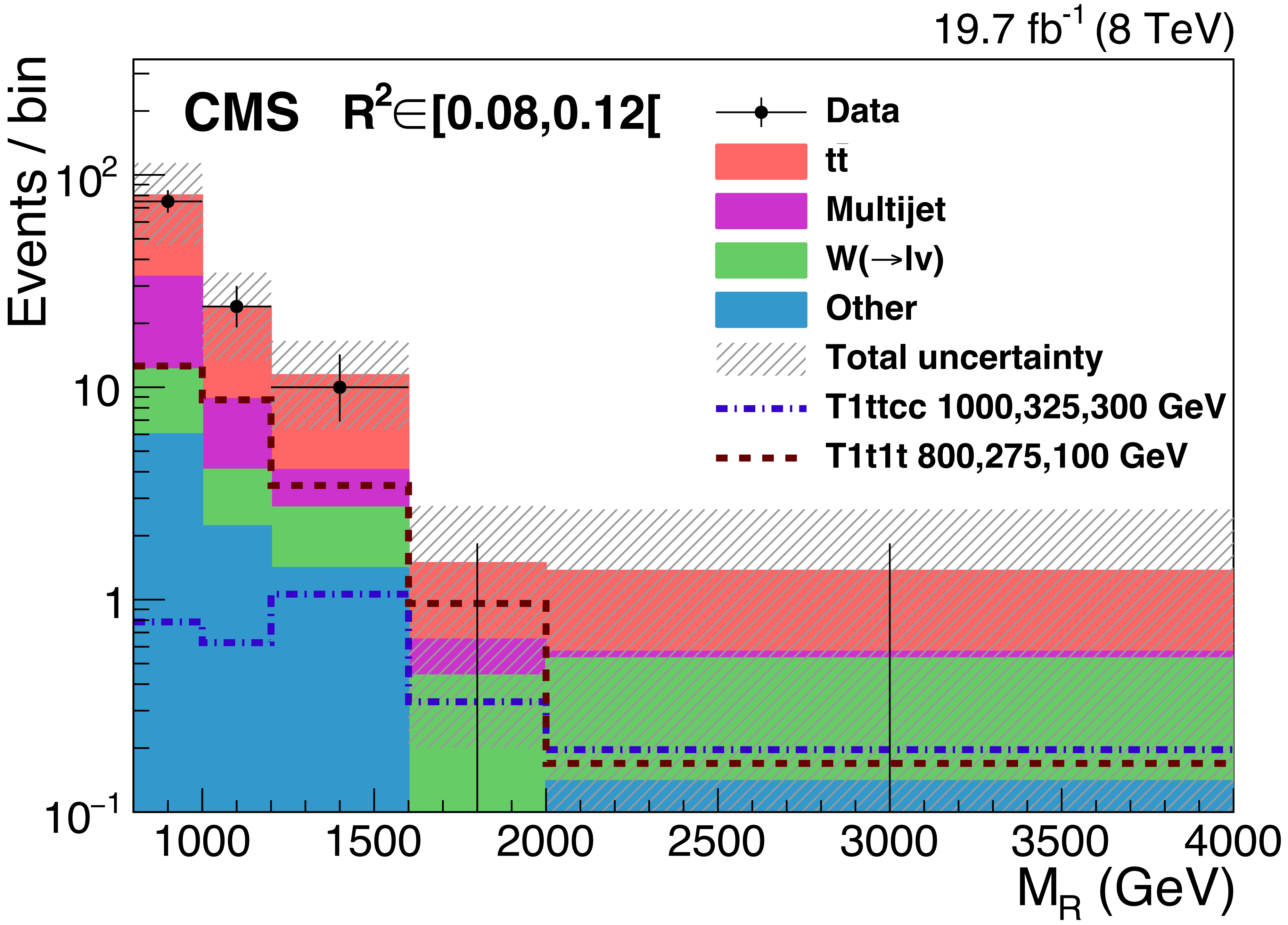
European Research Council

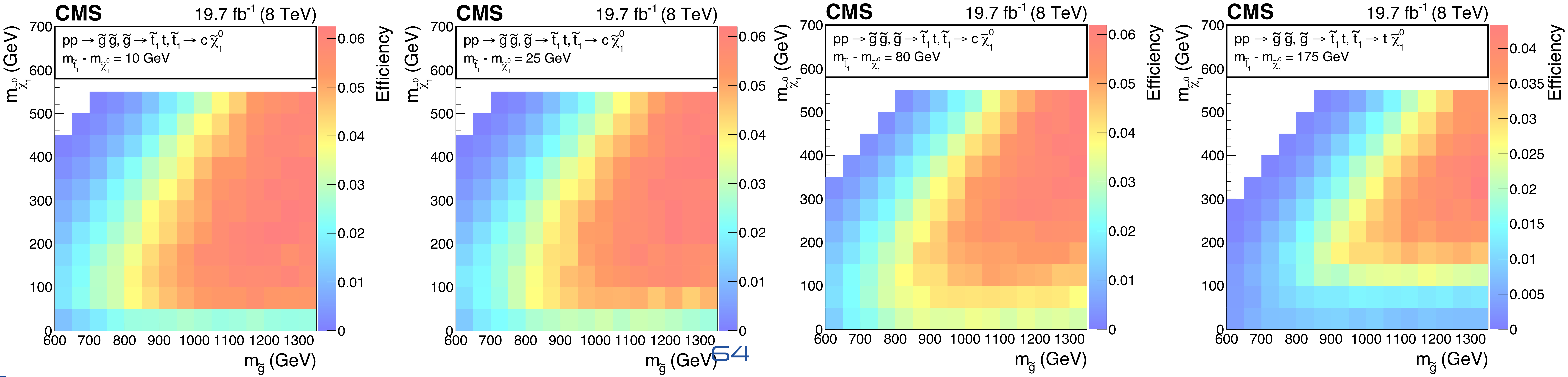
)

Parenthesis closed

BOOSTED-JETS SUSY SEARCH

- *Regime already probed in Run I (update ongoing)*
- *Will become more relevant with HL-LHC and higher-energy colliders (if any)*





SUMMARY OF EPISODE II

- *When R-parity is postulated, the lightest SUSY particle becomes stable (a natural dark matter candidate)*
- *Dark matter cannot be detected*
- *But LHC can probe dark matter production using balance on transverse plane*
    - *direct production, when high-pT jet/photon/etc is radiated*
    - *in cascade, from the production of other SUSY partners*
- *Several new methods proposed since LHC started*
- *A large part of the parameter space was explored, particularly in the context of Natural SUSY*
- *Now looking at the corners of the parameter space, where experimental conditions are more complicated*
- *Deep learning (e.g., for jet tagging) will help us to deal with this*

European Research Council

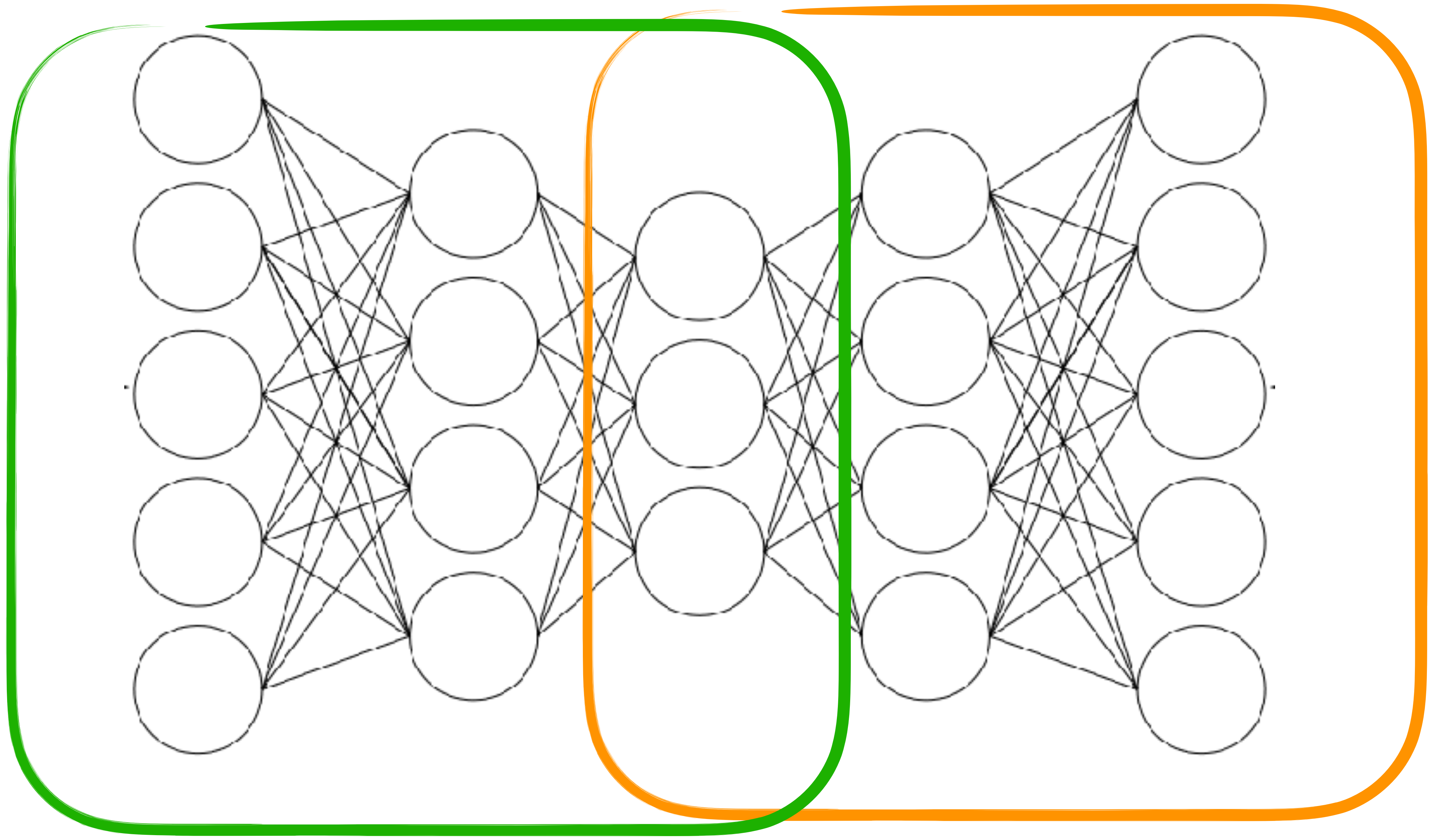
BACKUP SLIDES

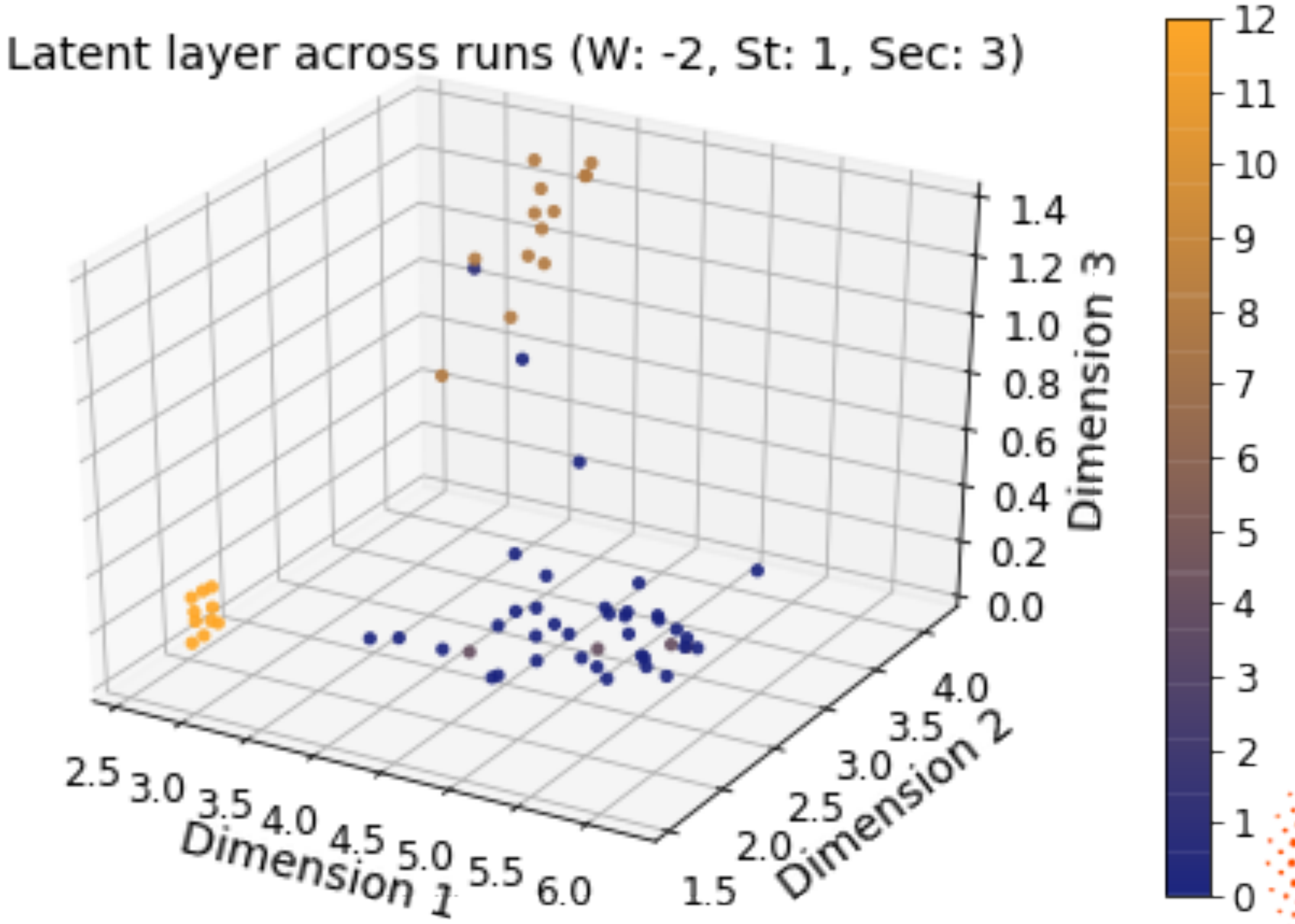
Search for New Physics @LHC

- We don't know what we are searching for
- Our theory-inspired fully-supervised hypothesis test approach not guaranteed to work
- We know our background
- use (ALSO) a semi-supervised approach:
  - <u>train on known physics</u>
  - <u>look for anomalies</u>
- Tried on the same setup used for trigger cleanup

Autoencoders

- *Autoencoders are a useful tool for anomaly detection*
- *They consist of two parts:*
  - ***encoder****: compressing high-dimensional input on a lower-dimension latent space*
  - ***decoder****: decompressing the point in the latent space back to the original representation*
- *Trainable with just "normal" data*
- *Anomalies = failed compression-decompression*

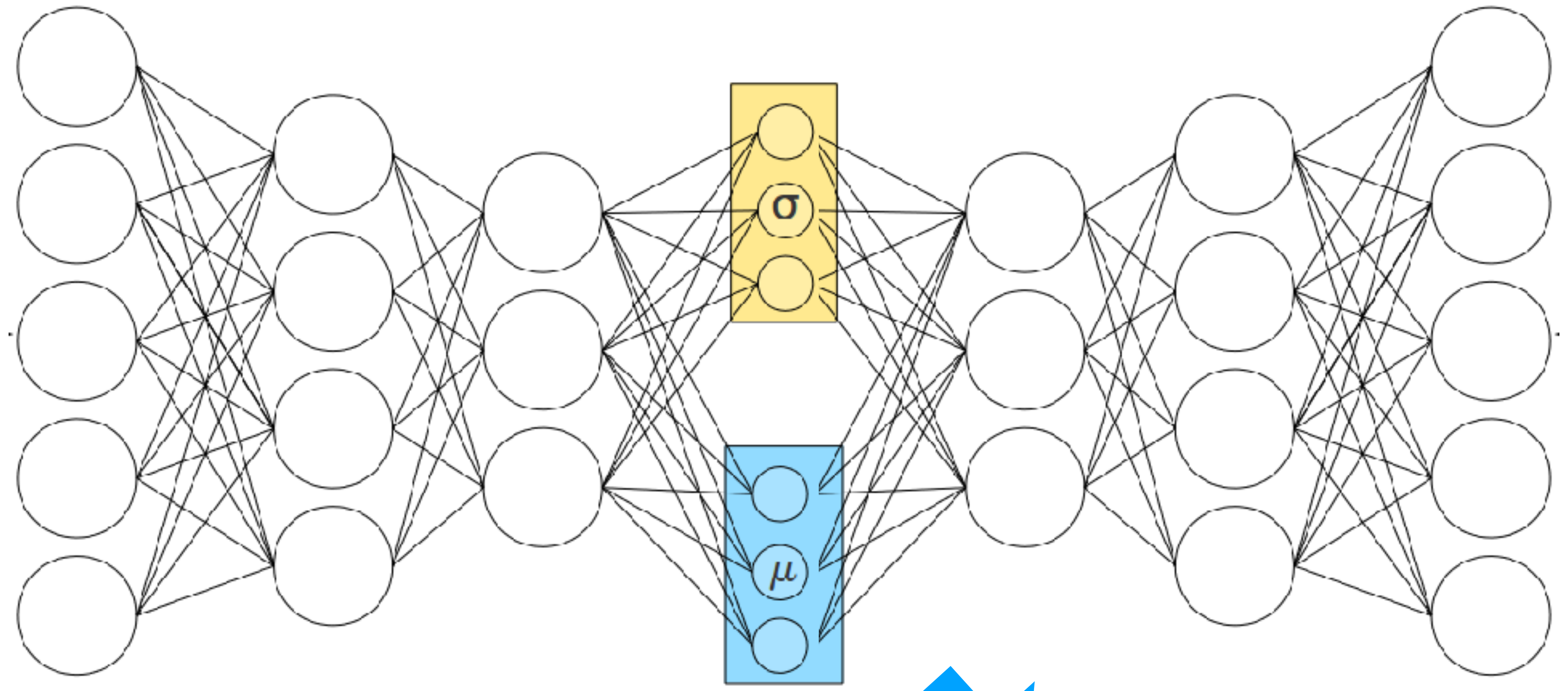


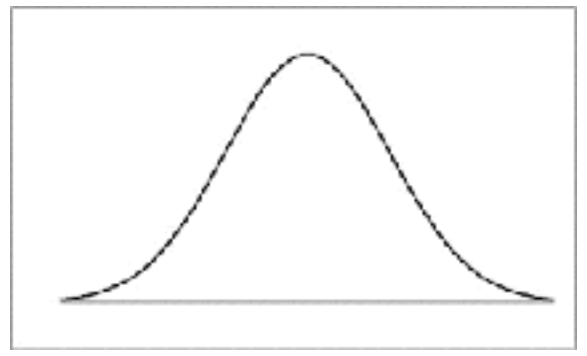


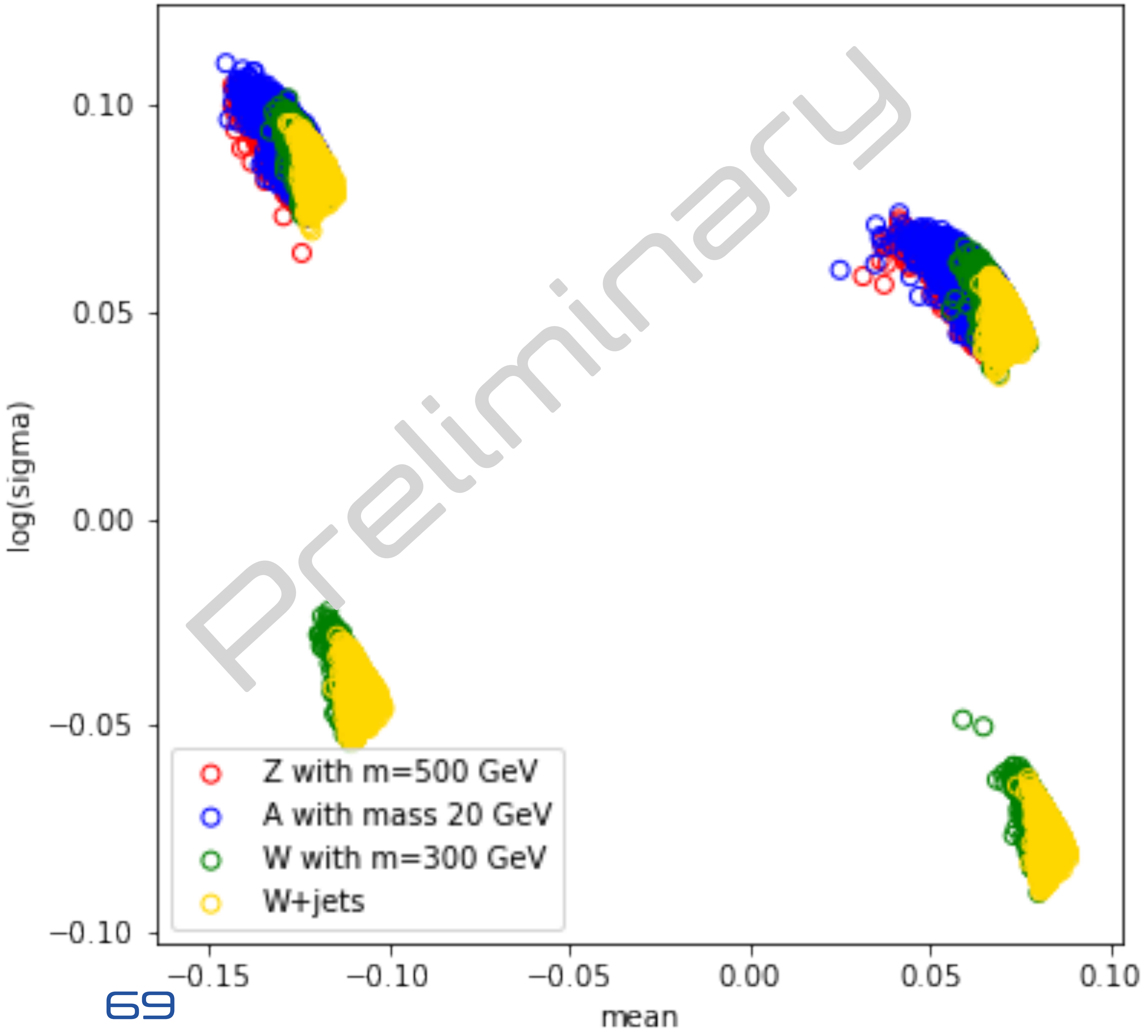
European Research Council

Variational Autoencoders

- Variational autoencoders are a special kind of encoders
- in VAE, the inner nodes represent parameters of a smooth function
  - easier assessment of anomaly by mean of distance between functions
  - can also be used to generate events

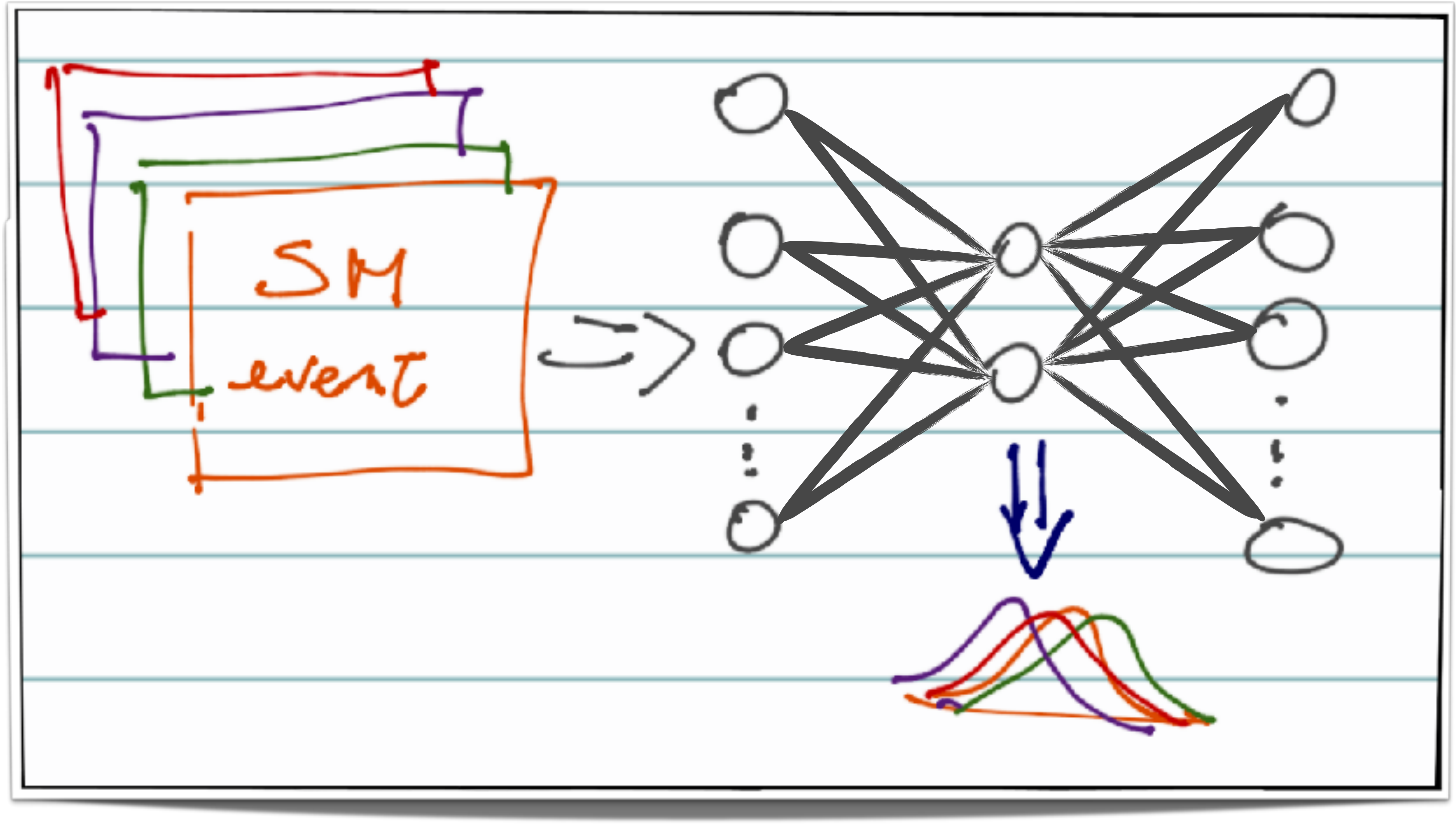


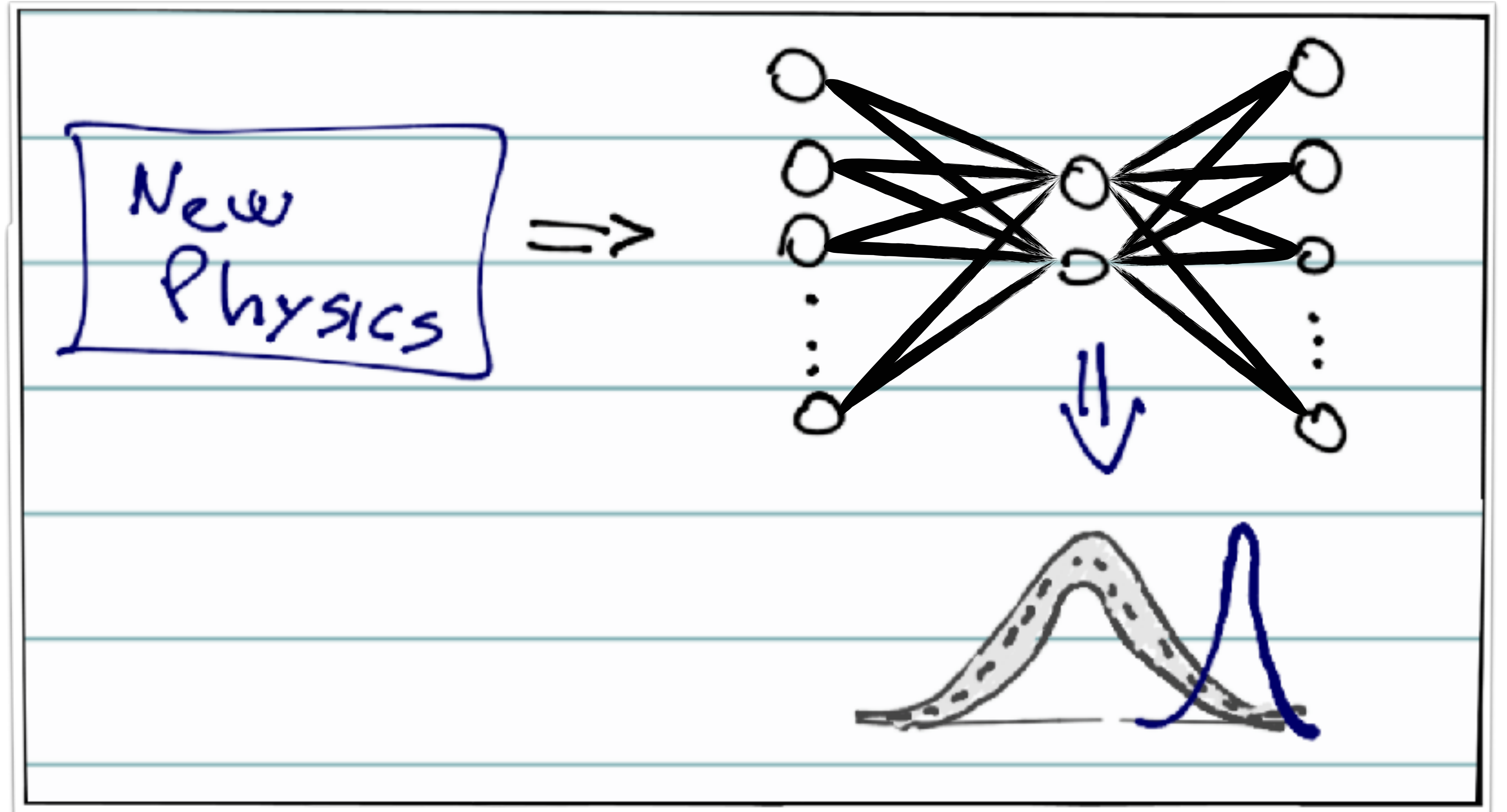




Quantifying anomaly

- Processing SM events (Monte Carlo simulation) through autoencoders, one derive an ensemble of distributions
- An observed SM event would be instances of this ensemble
- A observed new physics event might instead give a far-away/ broader Gaussian
- We could then use a metric of function distance to quantify anomaly

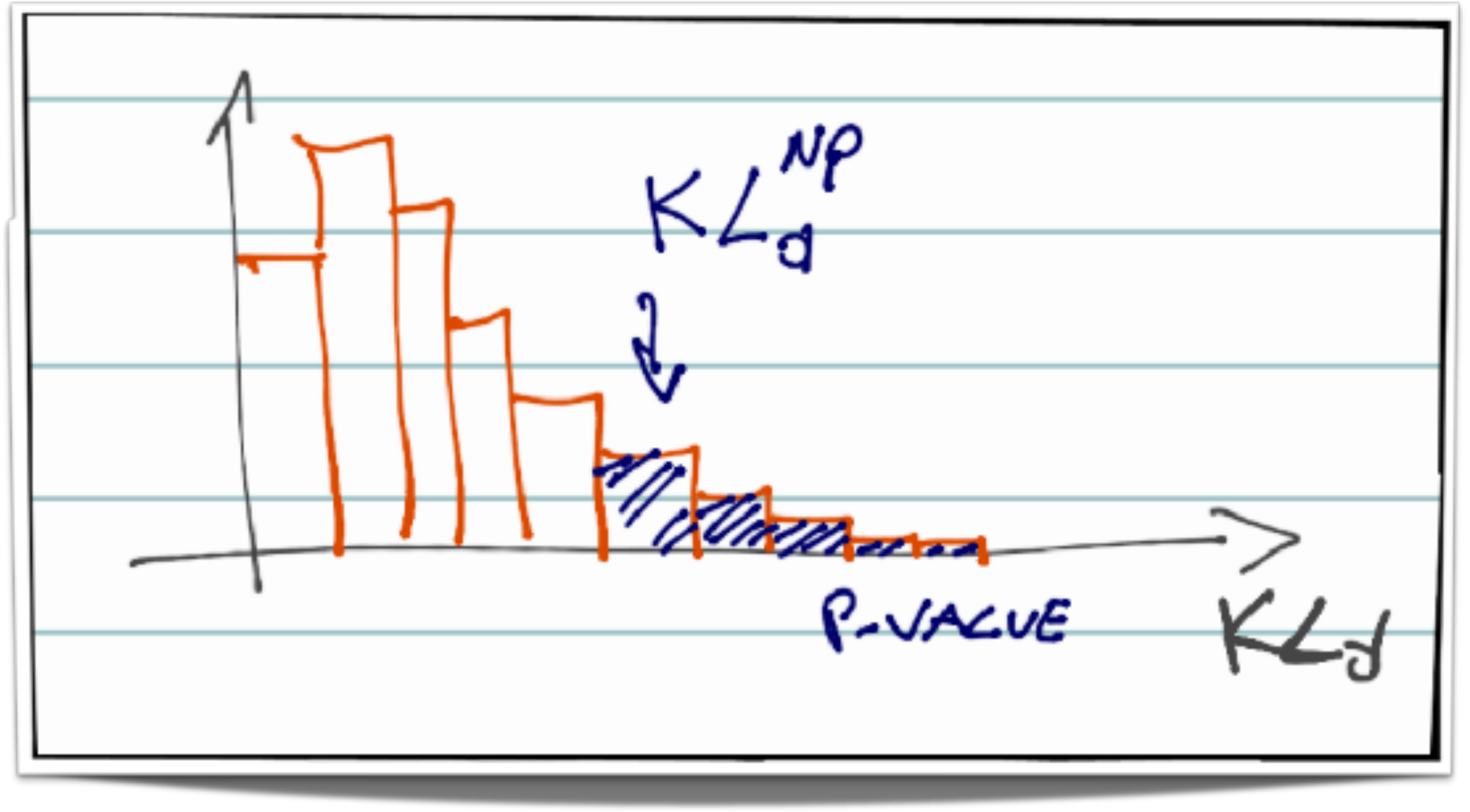


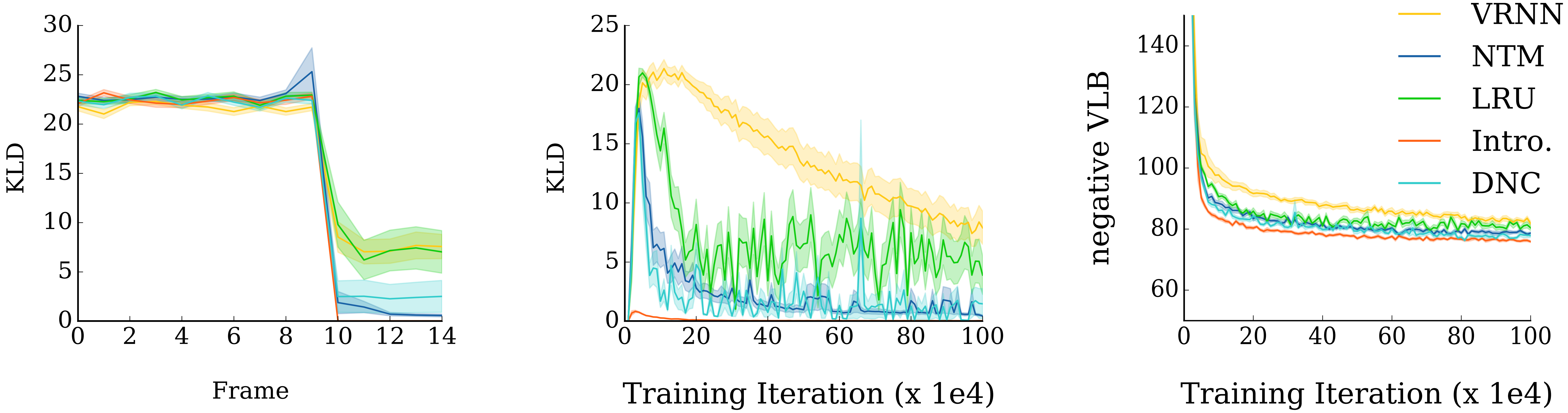


Quantifying anomaly

- *Information theory tells us to use KL divergence*
- *This is embedded in the loss function used for training (KL regularization)*
- *KL loss = # of extra bits needed to encode p in q. Measures p being anomalous or not*

$$KL^i_d = \int dx\, P^i(x) \log\left(\frac{P^i(x)}{q(x)}\right)$$

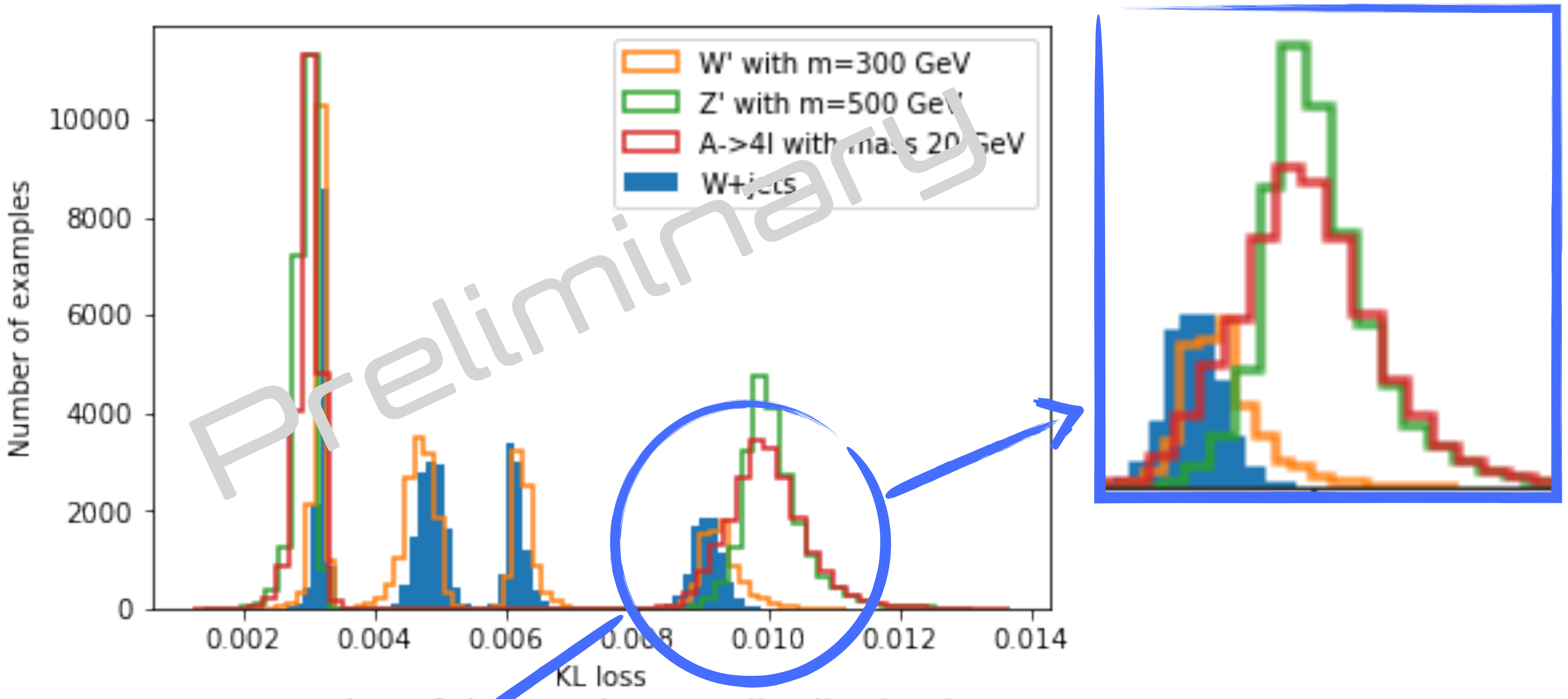


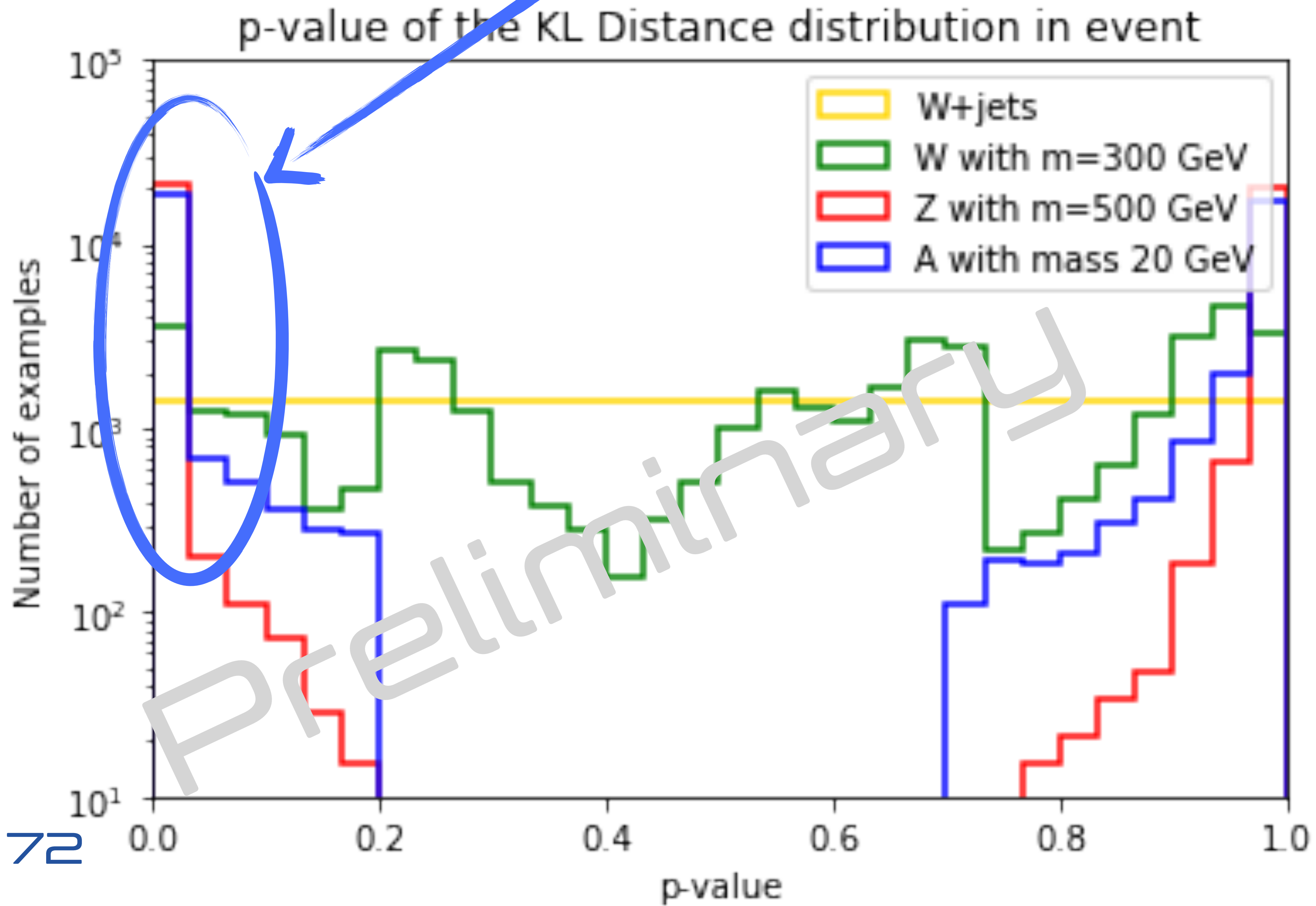


https://arxiv.org/pdf/1702.04649.pdf

VAE for Anomaly detection

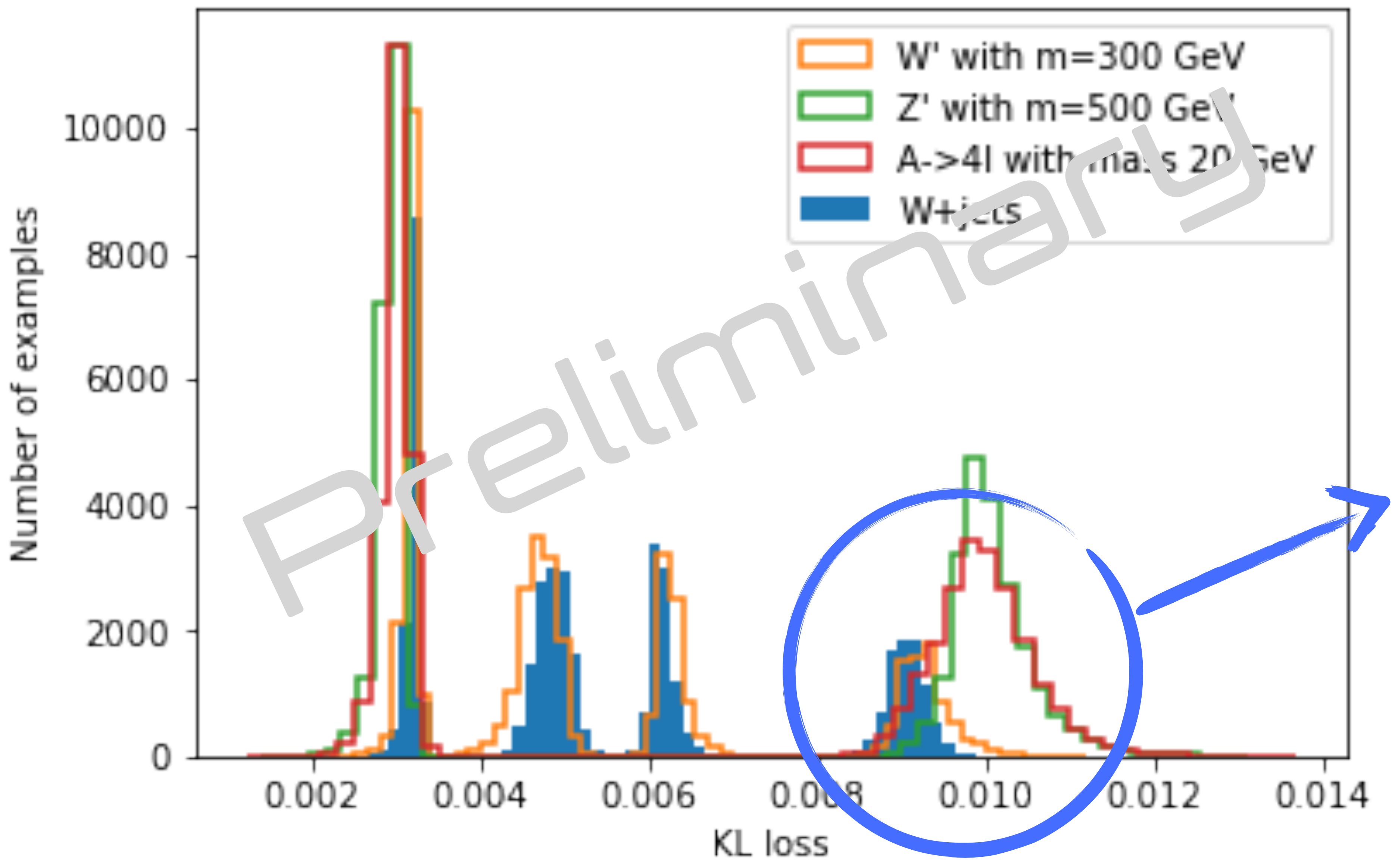
- *Experiment: consider a sample of W+jets events, selected by single-lepton trigger*
- *Train a VAE on them*
- *Use the encoder to compute the KL loss for*
  - *W+jets events*
  - *A W' with m = 300 GeV*
  - *A Z' with m = 500 GeV*
  - *A light A-> 4ℓ with m_A = 10 GeV*

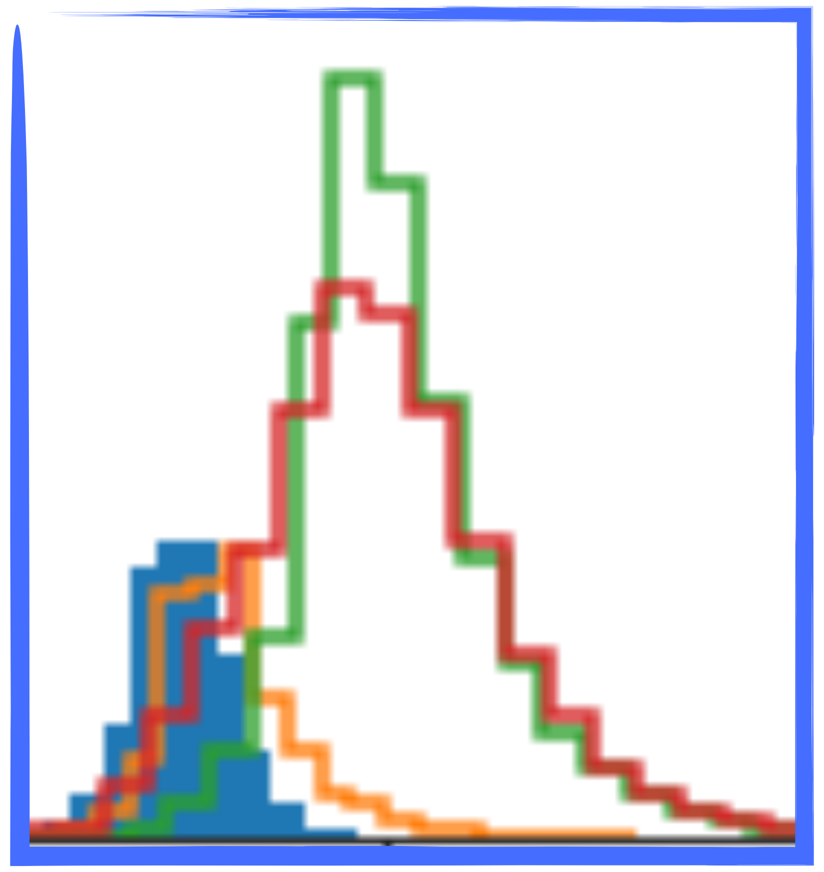




VAE for Anomaly detection

- *Experiment: consider a sample of W+jets events, selected by single-lepton trigger*
- *Train a VAE on them*
- *Use the encoder to compute the KL loss for*
    - *W+jets events*
    - *A W' with m = 300 GeV*
    - *A Z' with m = 500 GeV*
    - *A light A-> 4ℓ with m_A = 10 GeV*



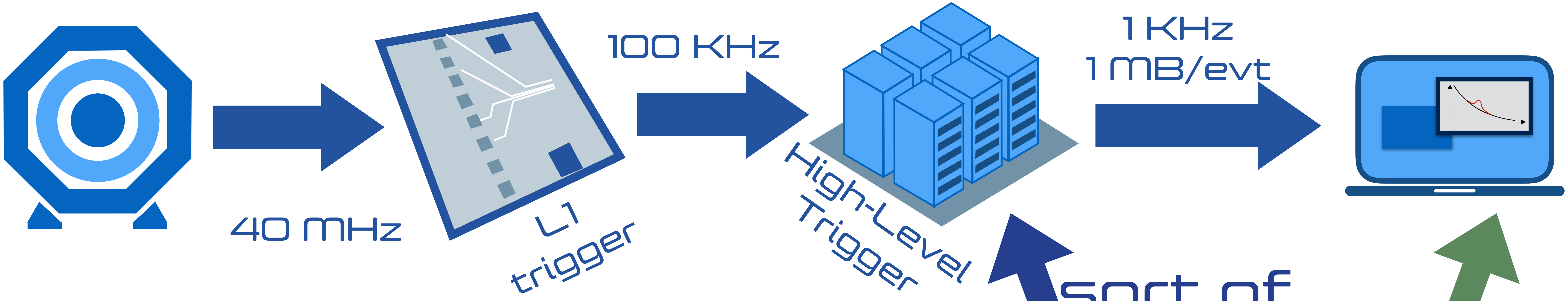


```
Fraction of events with p-value<0.00001
Wprime: 0.000612
Zprime: 0.065976
A->4mu: 0.069271

Fraction of events with p-value<0.001
Wprime: 0.013600
Zprime: 0.390918
A->4mu: 0.324071
Fraction of events with p-value<0.01
Wprime: 0.039953
Zprime: 0.476635
A->4mu: 0.412282
```


European Research Council

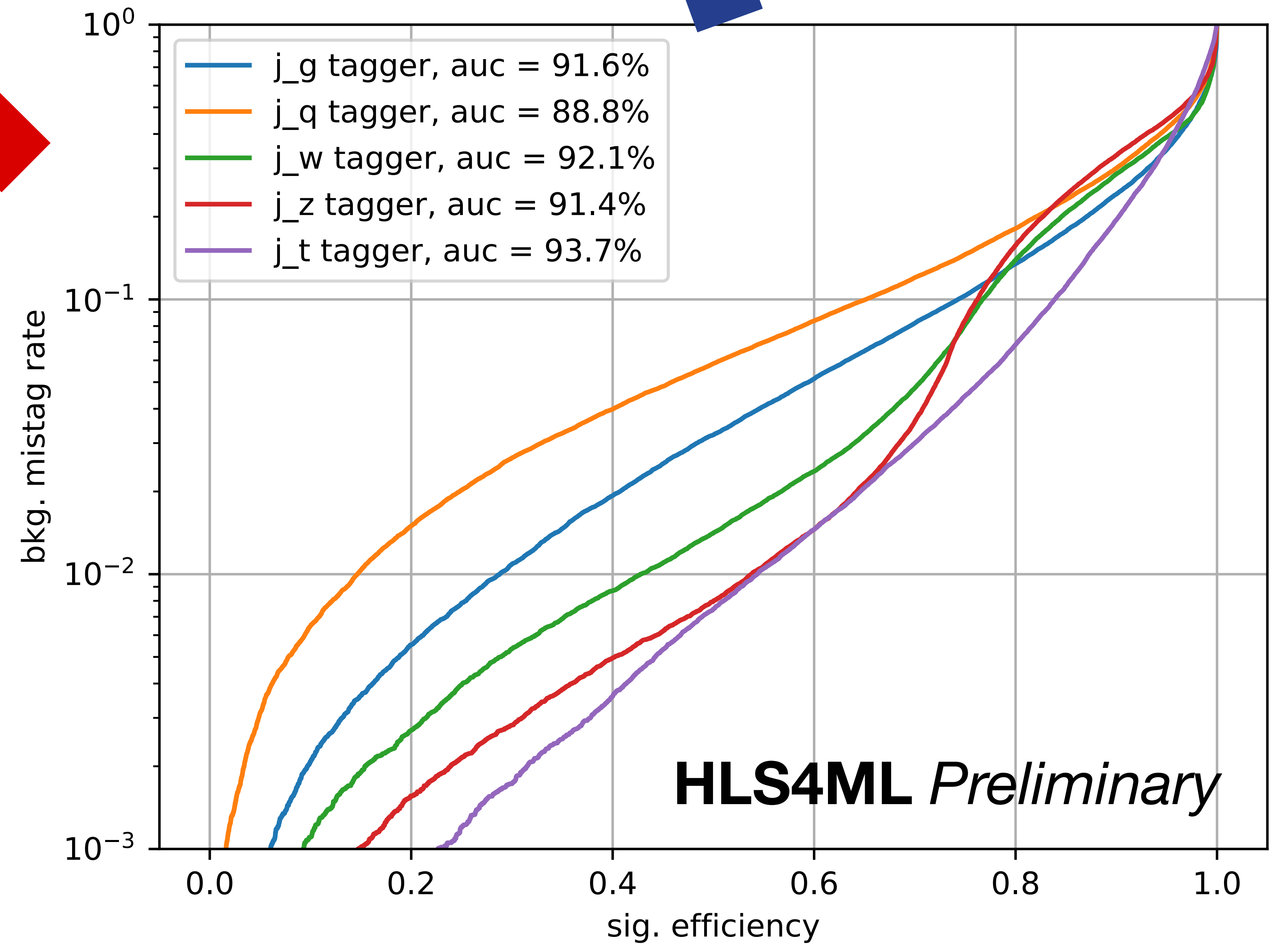
The LHC Big Data Problem

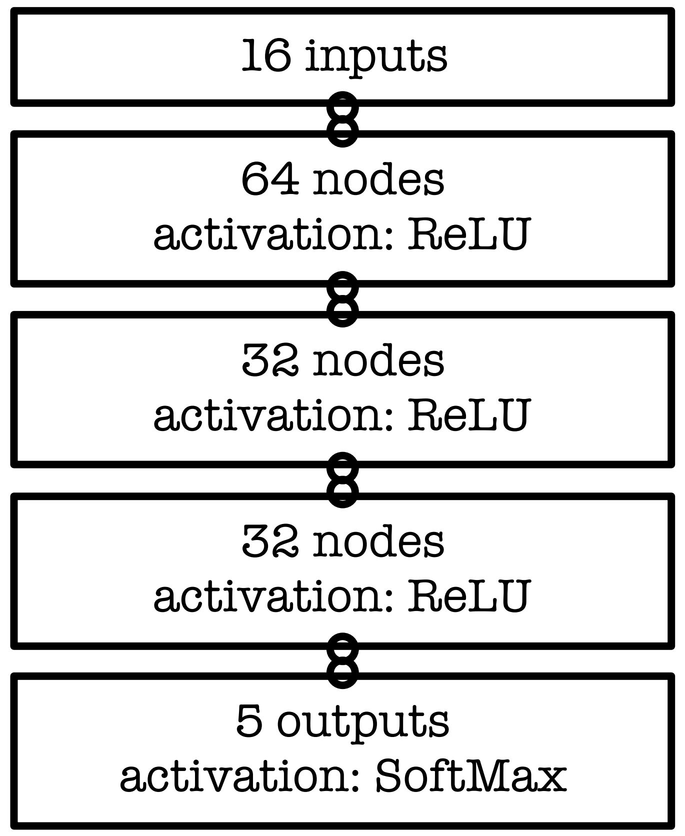


sort of OK

Easy

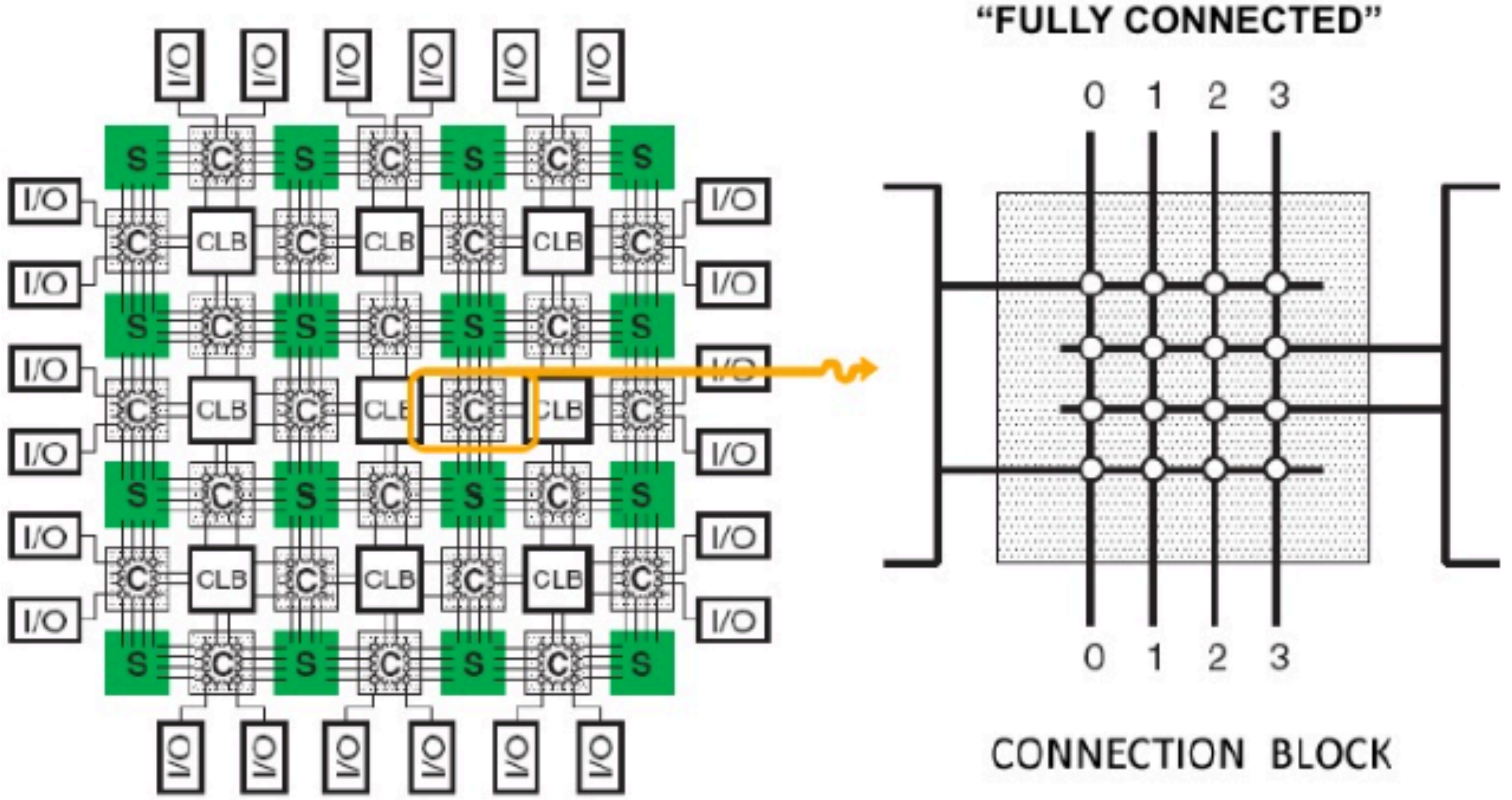
Complicated: need to be fast (10 ms) and with very small resources



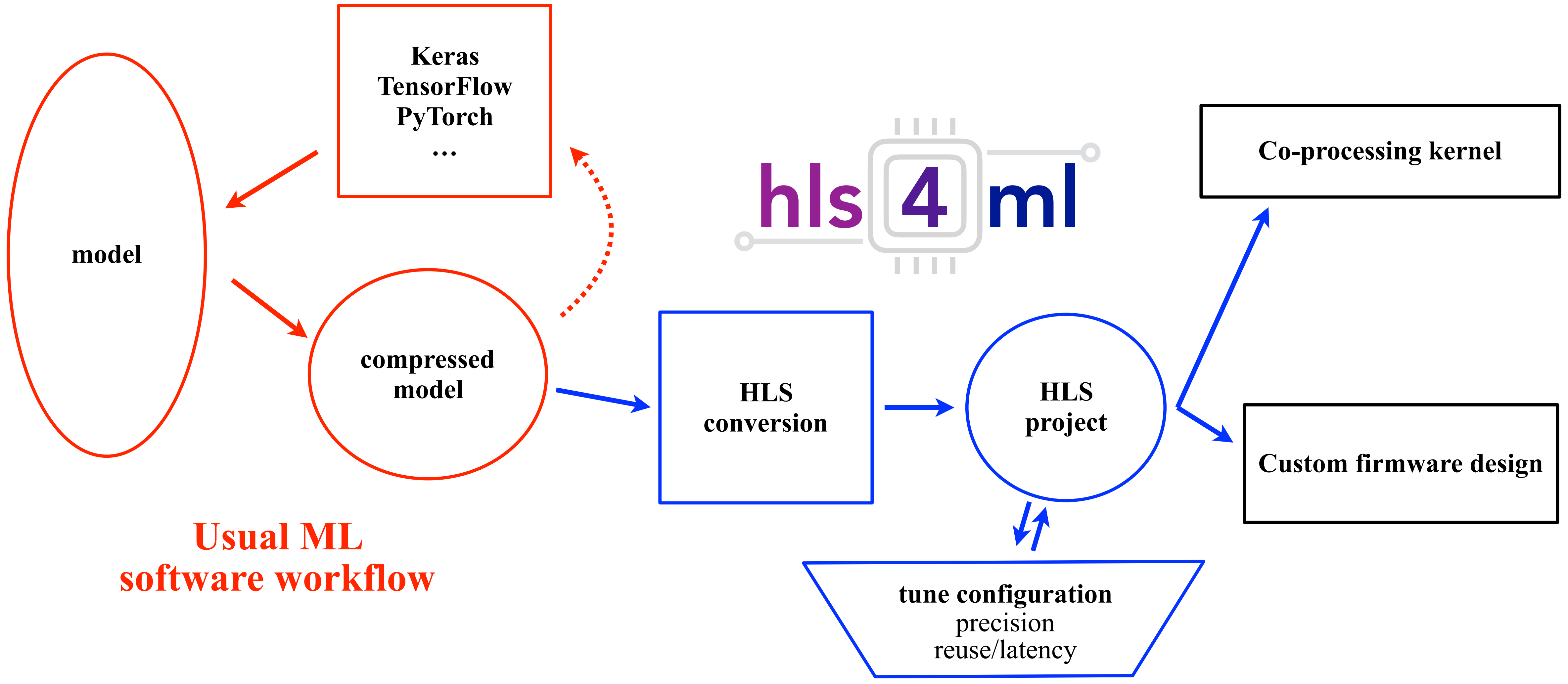


Bring DL to L1

- The L1 trigger is a complicated environment
  - decision to be taken in ~10 μsec
  - only access to local portions of the detector
  - processing on Xilinx FPGA, with limited memory resources
- Some ML already running @L1
  - CMS has BDT-based regressions coded as look-up tables
- Working to facilitate DL solutions @L1 with dedicated library



[HLS4ML: CERN/FNAL/MIT collaboration]

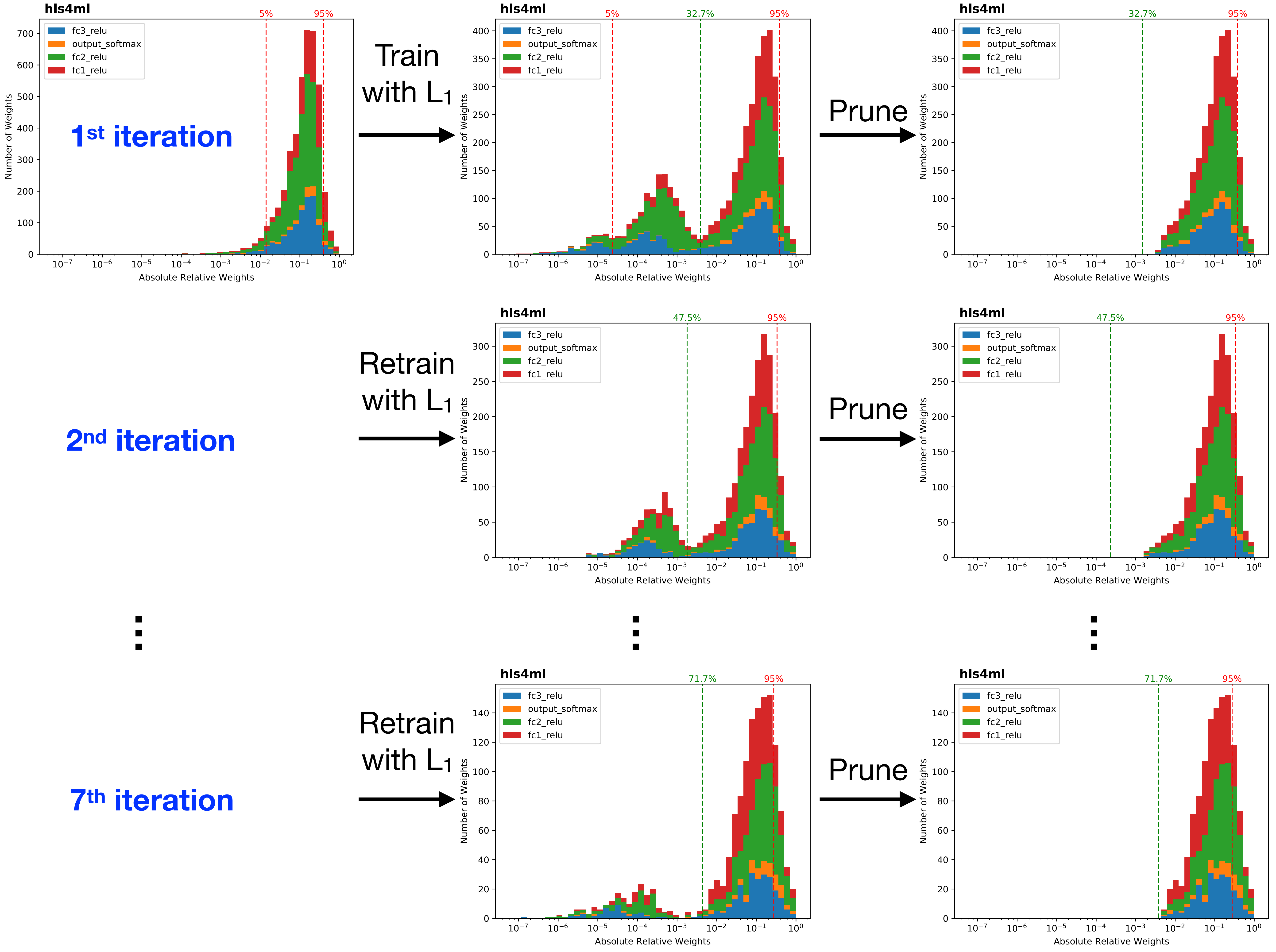


Make the model cheaper

- ***Pruning**: remove parameters that don't really contribute to performances*
  - *force parameters to be as small as possible (regularization)*

$$L_\lambda(\vec{w}) = L(\vec{w}) + \lambda||\vec{w}_1||$$

  - *Remove the small parameters*
  - *Retrain*

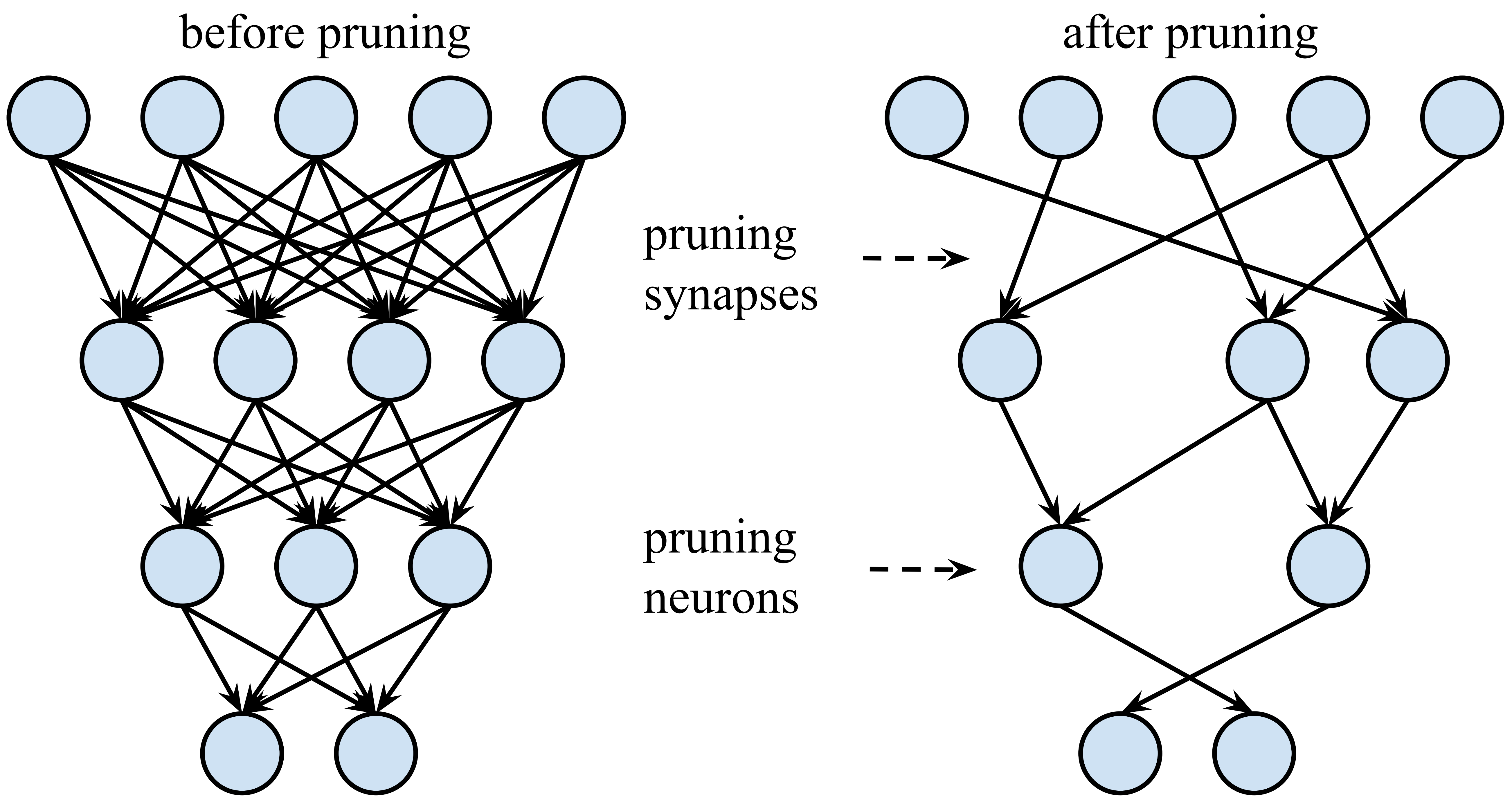


Make the model cheaper

- ***Pruning:*** *remove parameters that don't really contribute to performances*
  - *force parameters to be as small as possible (regularization)*

$$L_\lambda(\vec{w}) = L(\vec{w}) + \lambda||\vec{w}_1||$$

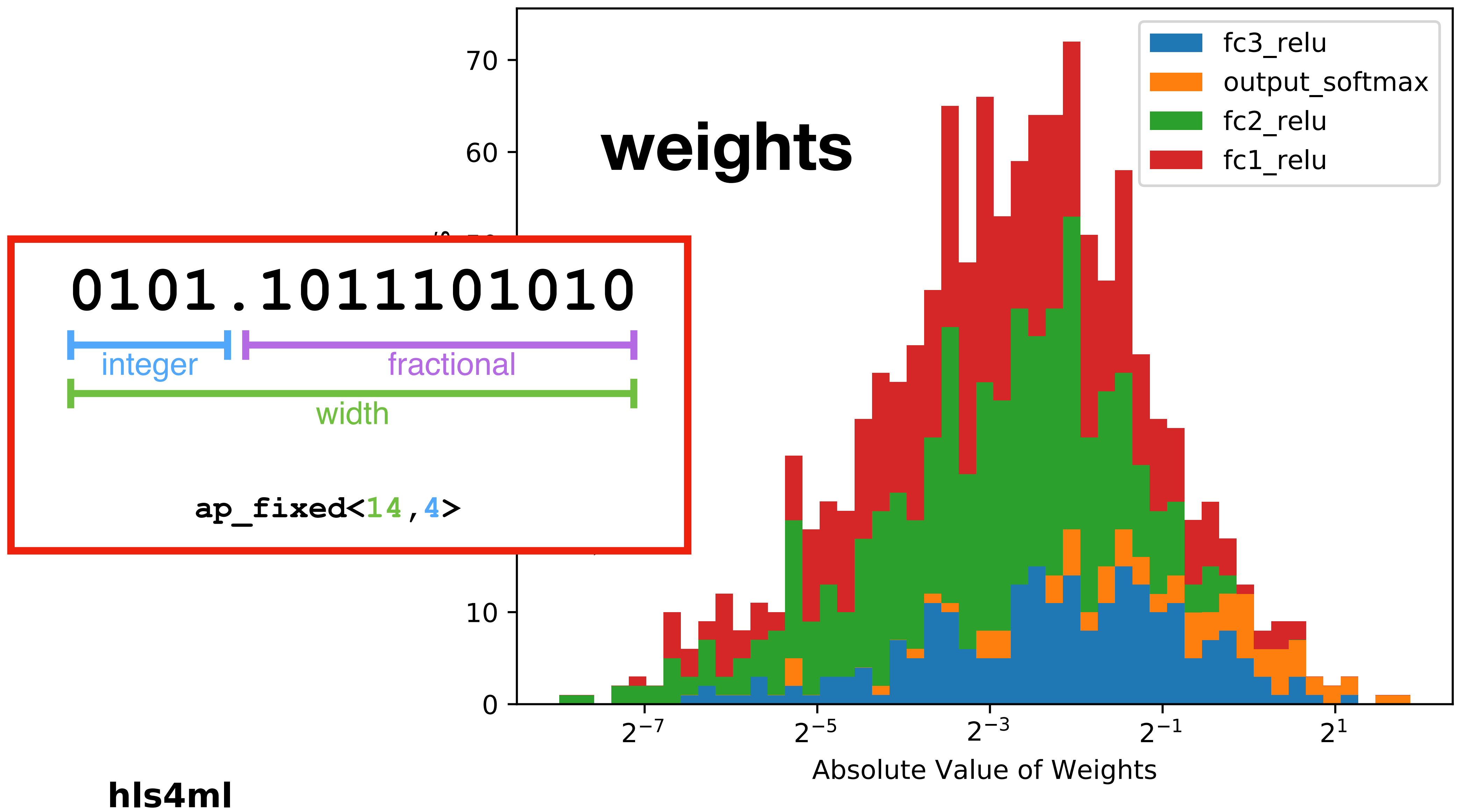
  - *Remove the small parameters*
  - *Retrain*

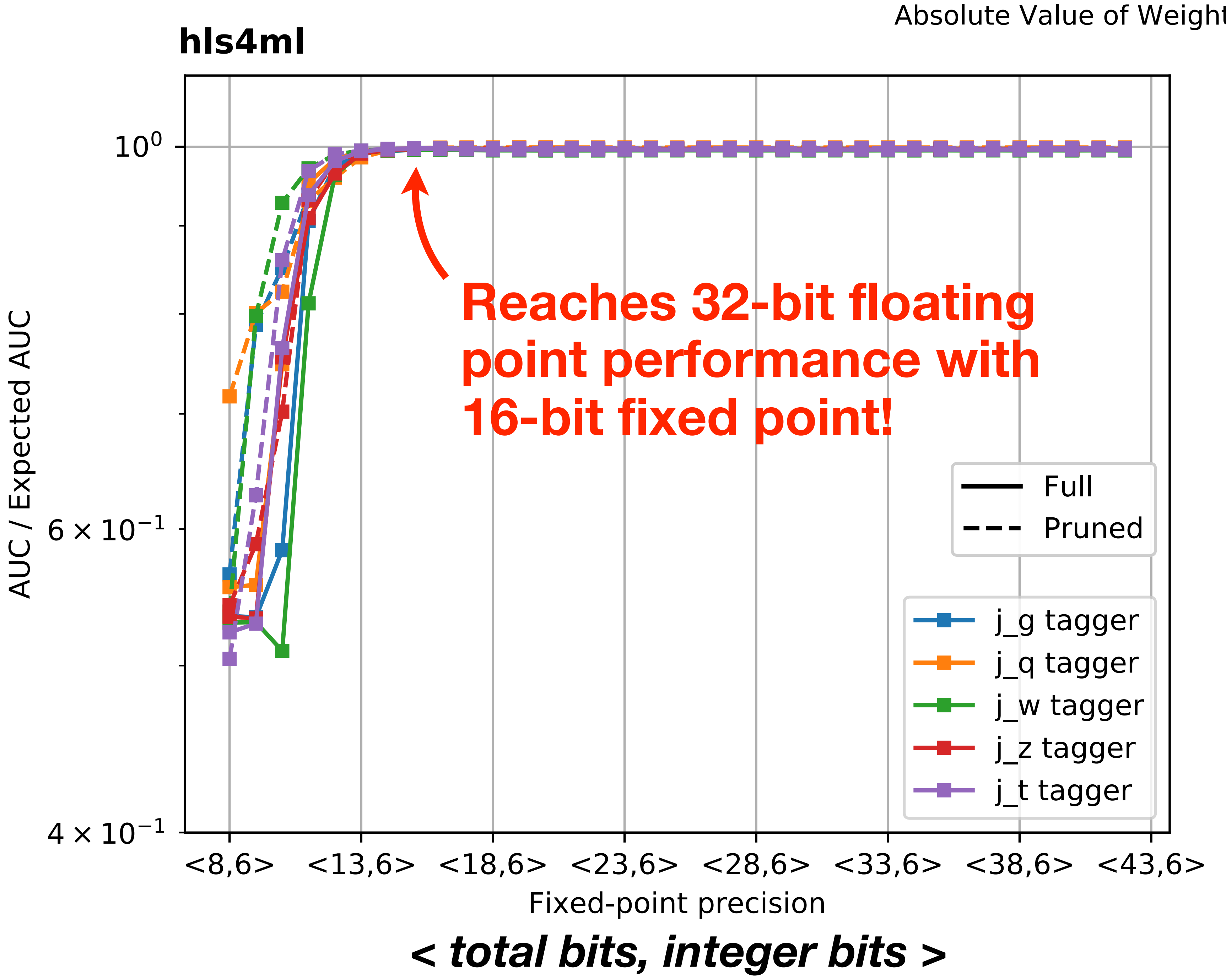


→ 70% reduction of weights and multiplications w/o performance loss

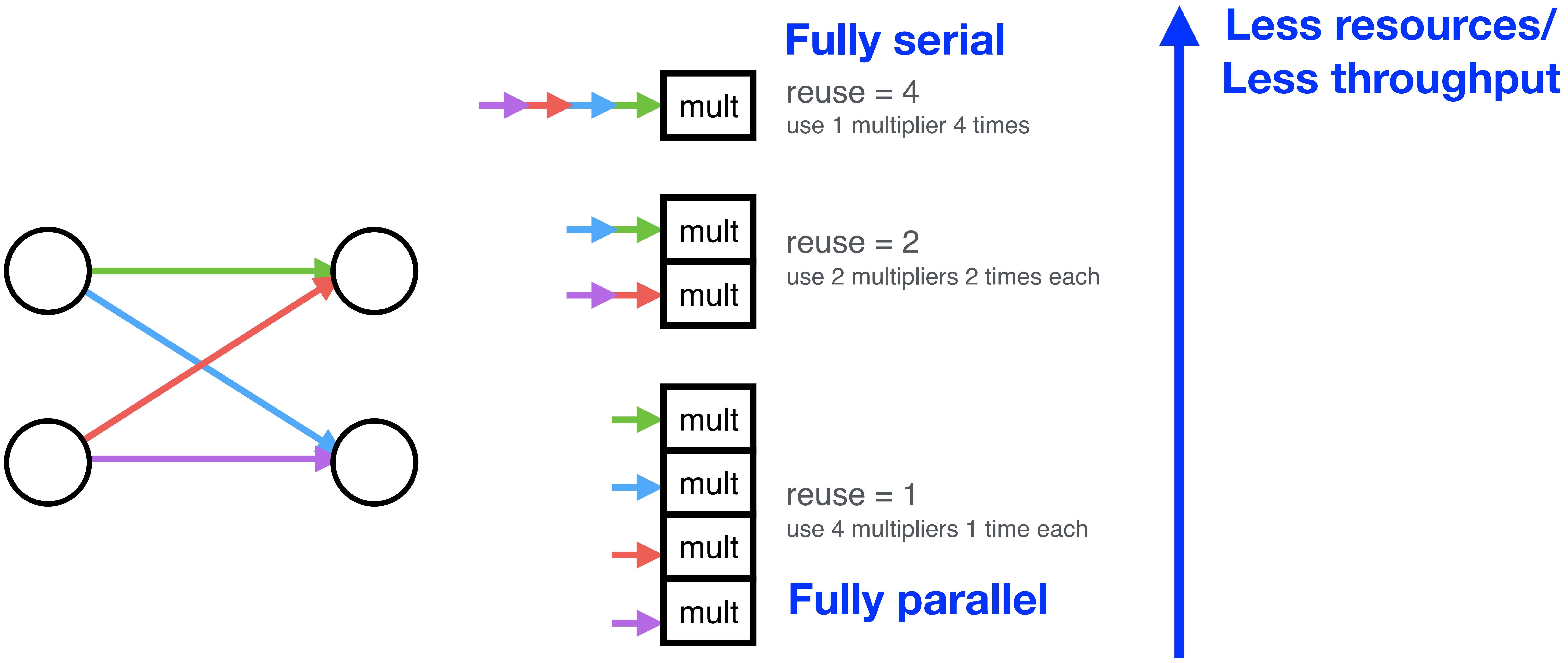
Make the model cheaper

- ***Quantization***: *reduce the number of bits used to represent numbers (i.e., reduce used memory)*
  - *models are usually trained at 64 or 32 bits*
  - *this is not necessarily needed in real life*
- *In our case, we could reduce to 16 bits w/o loosing precision*
- *Beyond that, one would have to accept some performance loss*



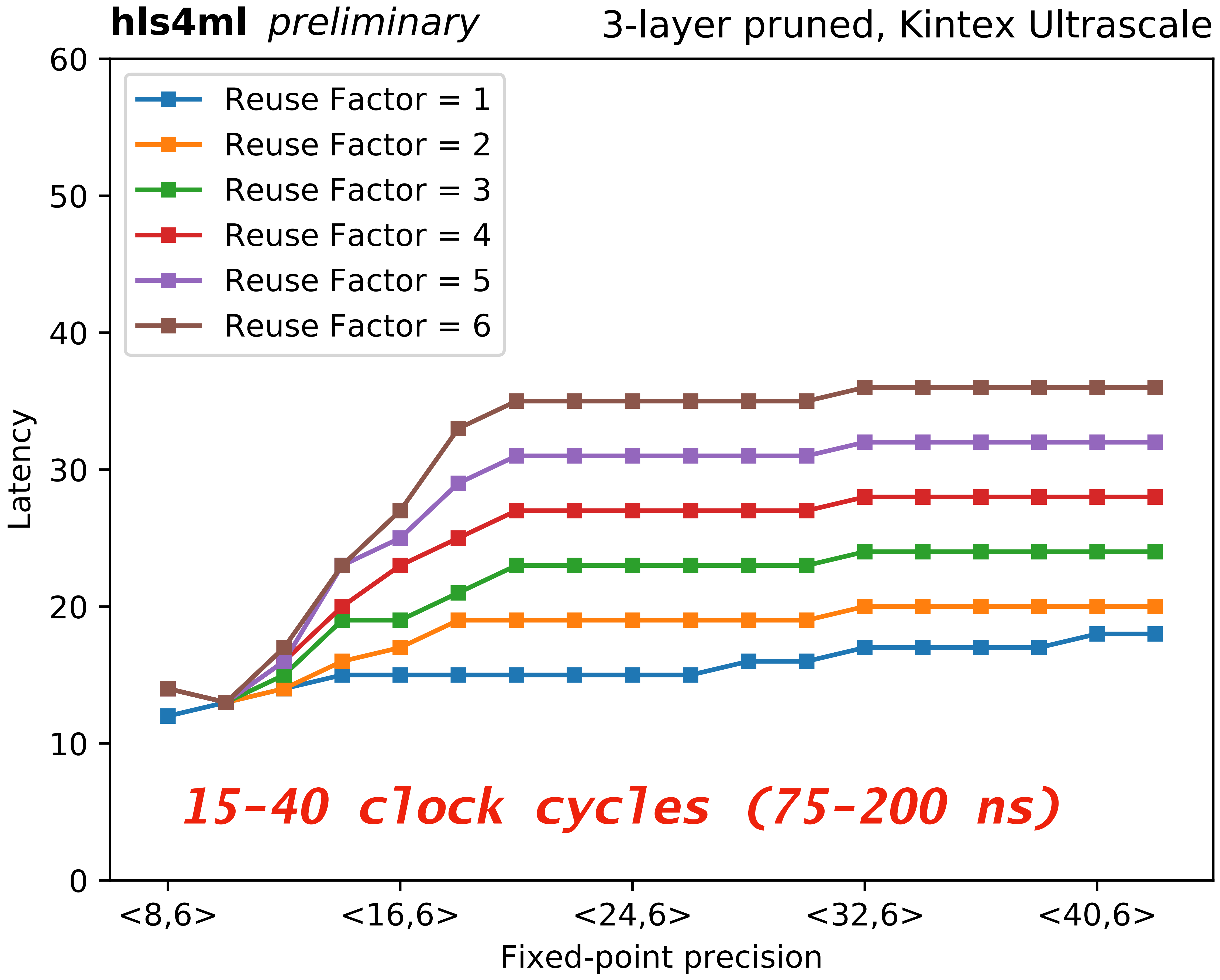


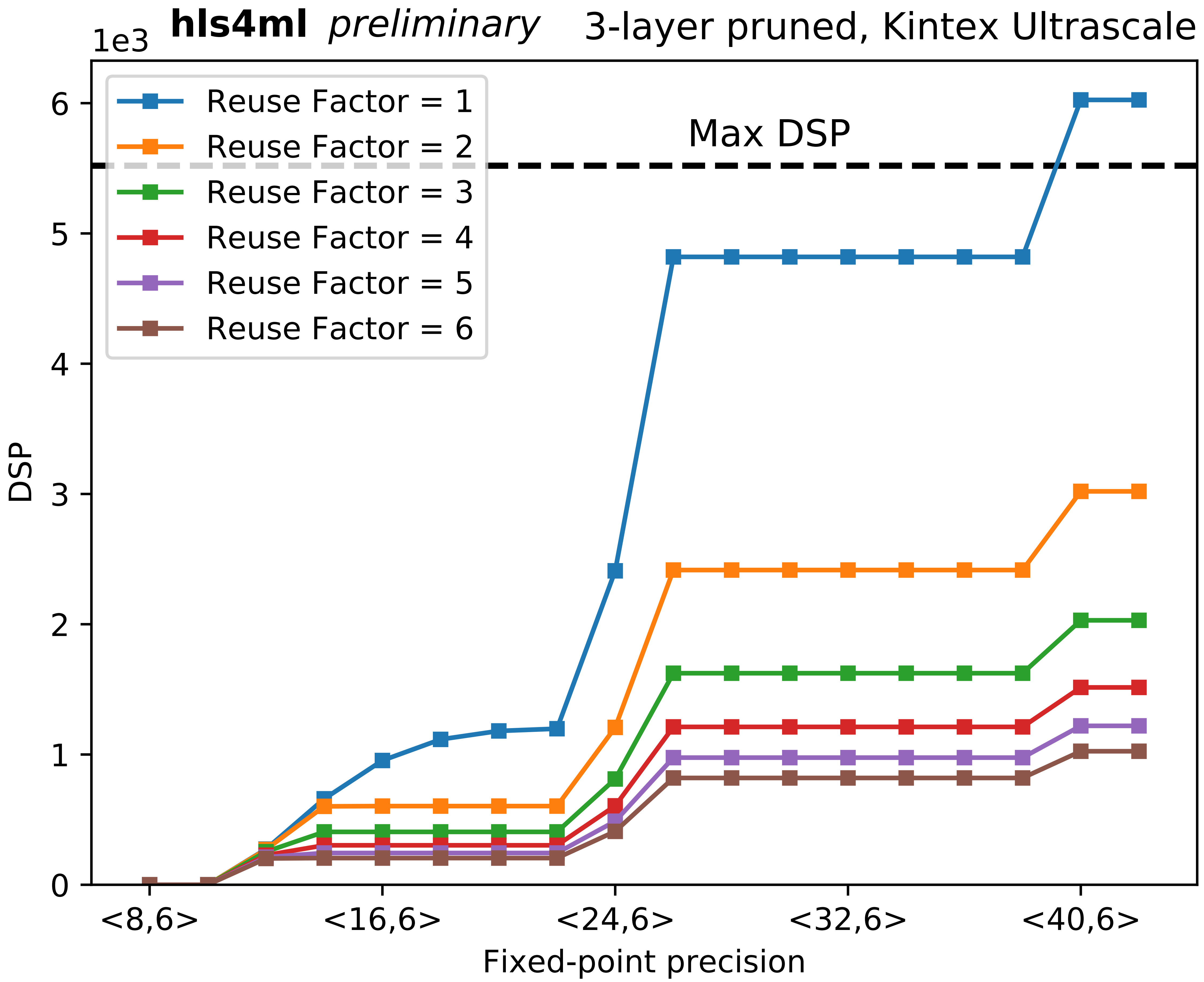
Speed vs Memory



Reuse factor: how much to parallelize operations in a hidden layer

Performances





NB: FPGA emulator over-estimates resource needs by a factor 2-4 (tested our emulation vs actual deployment)