Jet substructure, machine learning, and a neutrino connection

Nhan Tran
Fermilab

Neutrino Division Seminar Series
April 12, 2018

Things that I am thinking about, or have thought about, that I think maybe would be relevant to the things that you are thinking about or have thought about

Nhan Tran
Fermilab

Neutrino Division Seminar Series
April 12, 2018

Things that I am thinking about, or have thought about, that I think maybe would be relevant to the things that you are thinking about or have thought about

Disclaimer #1:

Based on my (very) limited knowledge of neutrino experiments, I will try to draw some parallels between what we do at LHC and what I think are important to neutrino experiments. Assumptions may be misguided.

Neutrino Division Seminar Series

Disclaimer #2:

I'll cover a lot of different topics, some only superficially due to time and lack of expert knowledge, but also to give you a broad view of what people are thinking about. There are lots of experts, many who sit on the 10/11th floors. Hopefully this can be the start of interesting dialogues.

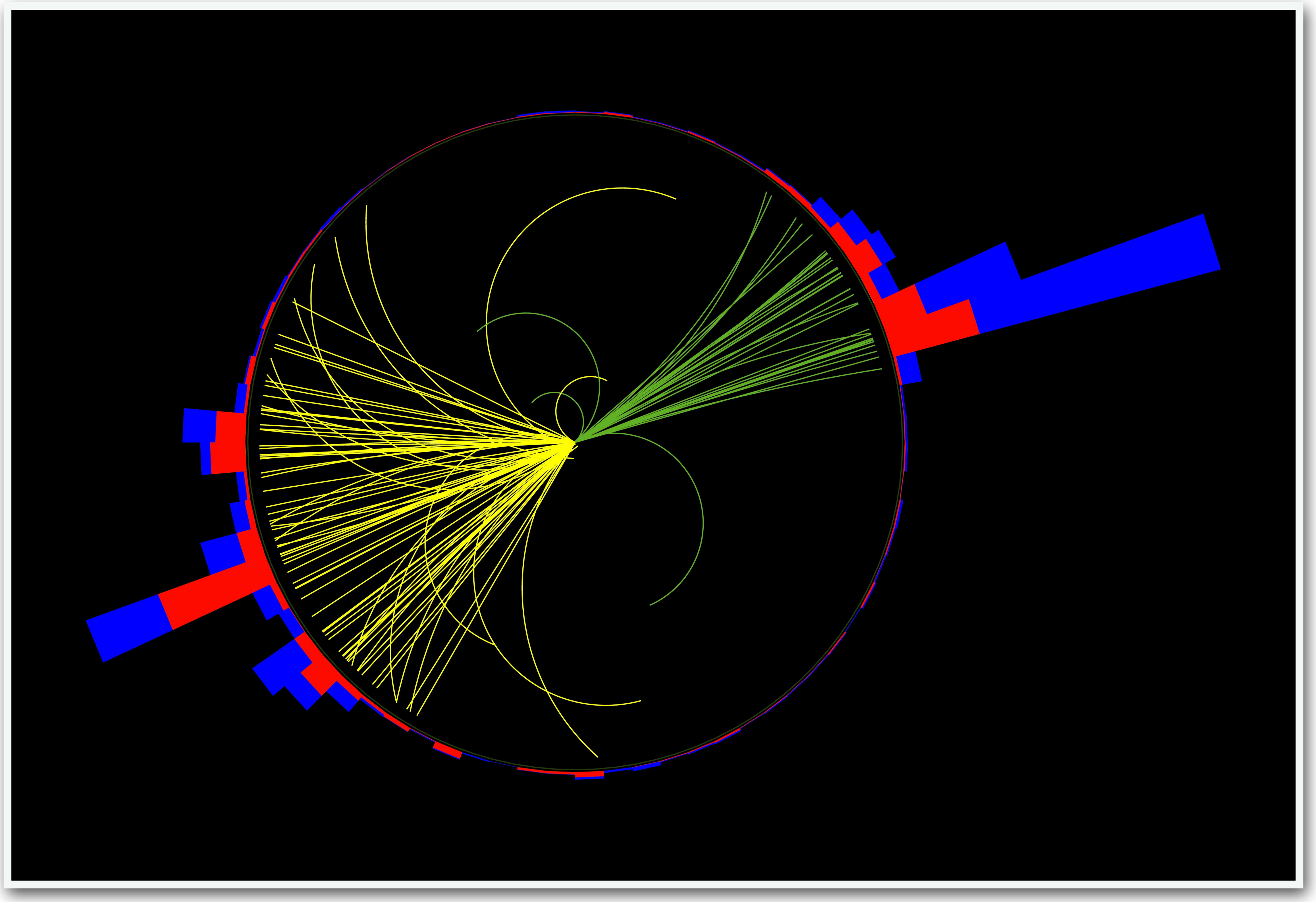
Outline

the task of jet substructure

rise of the machines

the fast and the furious

the task of jet substructure

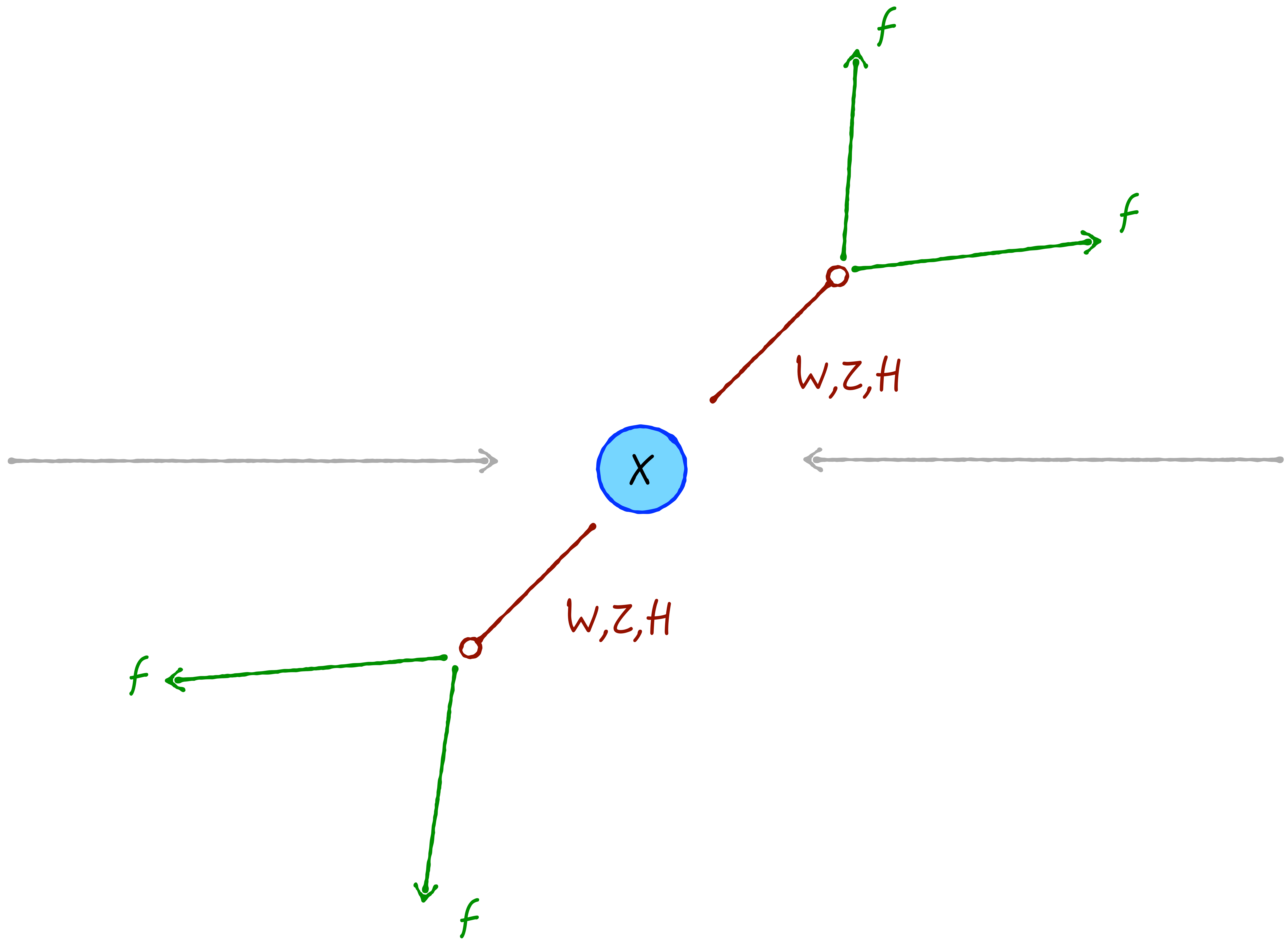


PROLOGUE

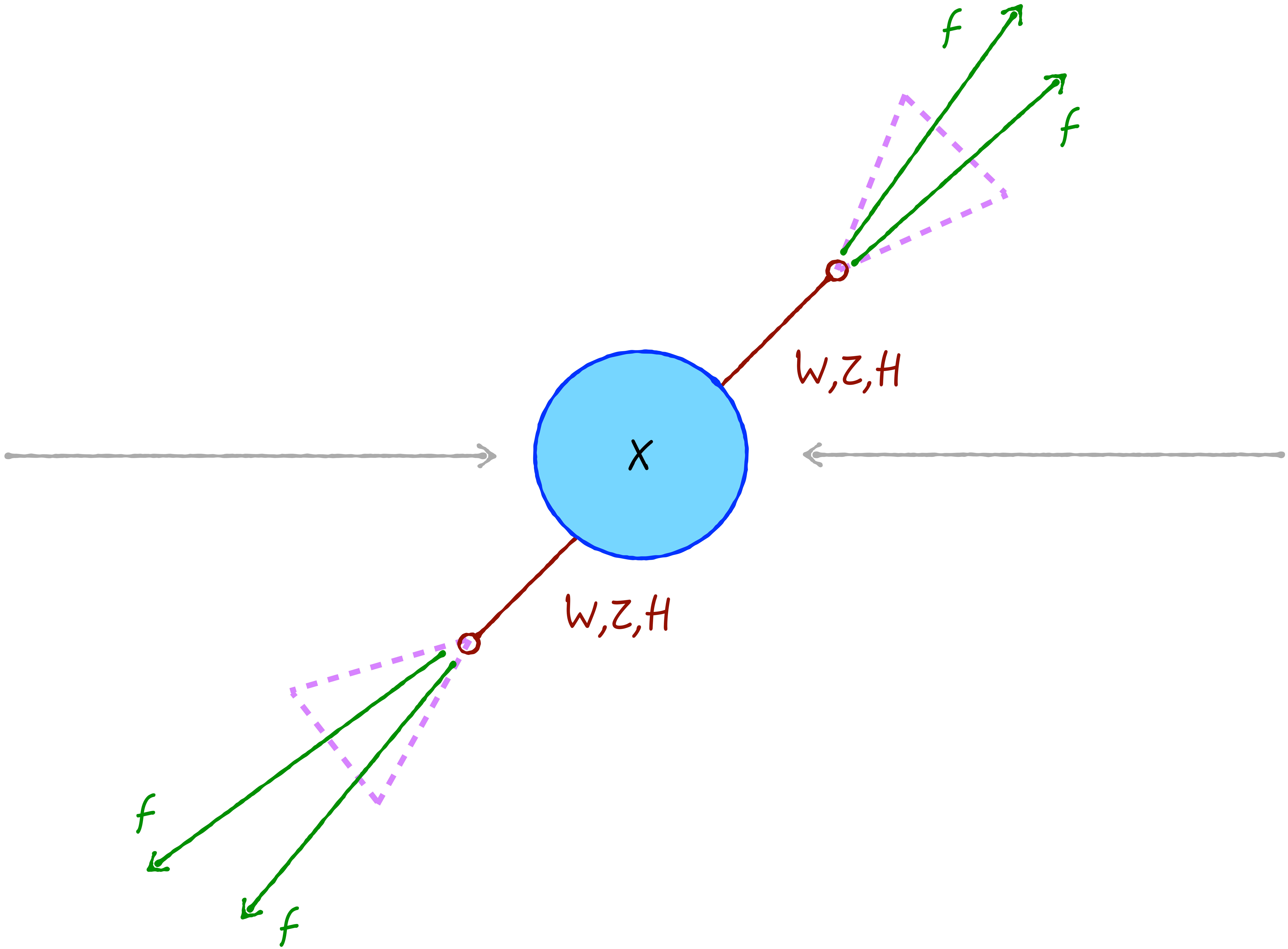
CMS & ATLAS:
A very broad and significant physics program

LHC era in a nutshell:
More energy
More luminosity

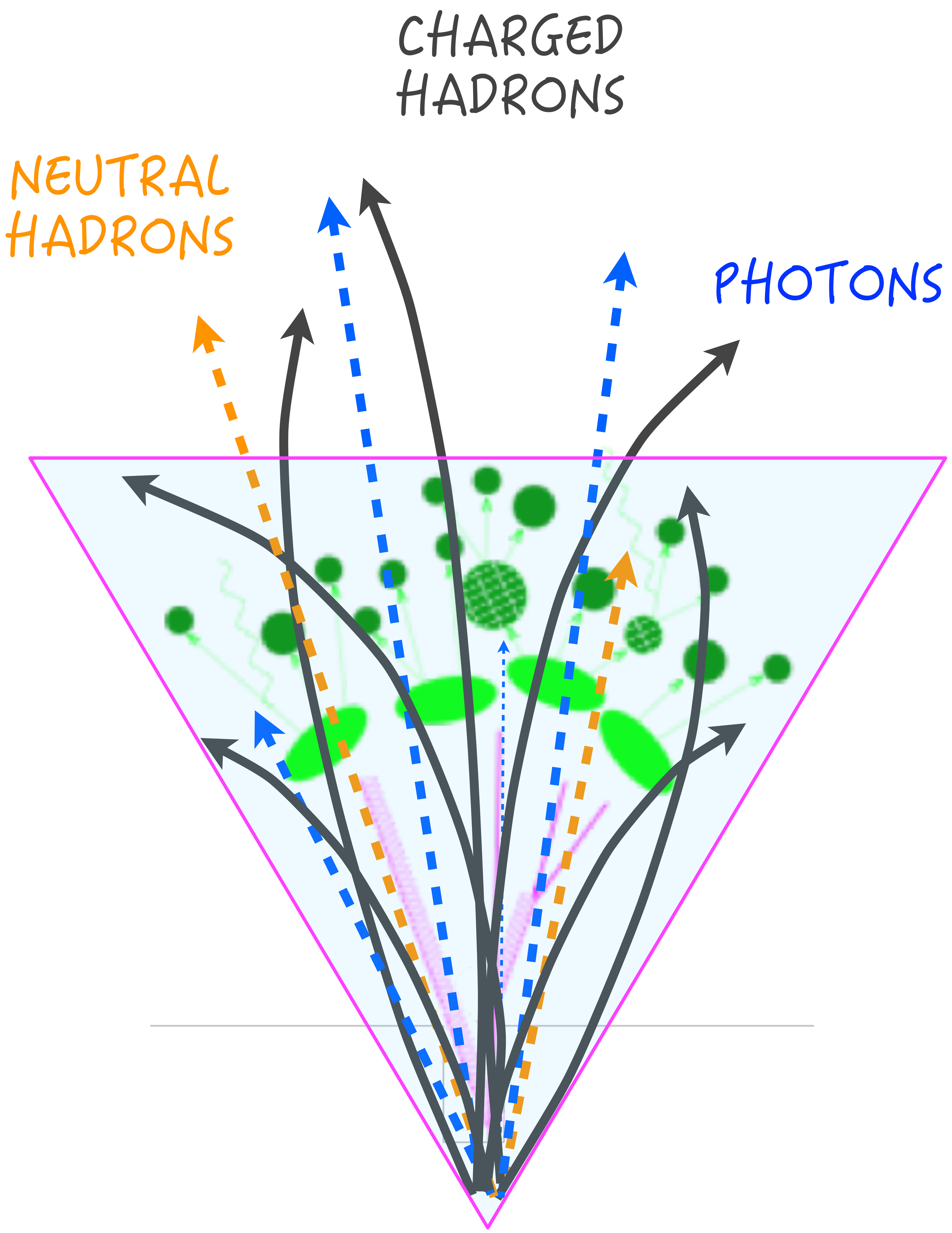
More energy

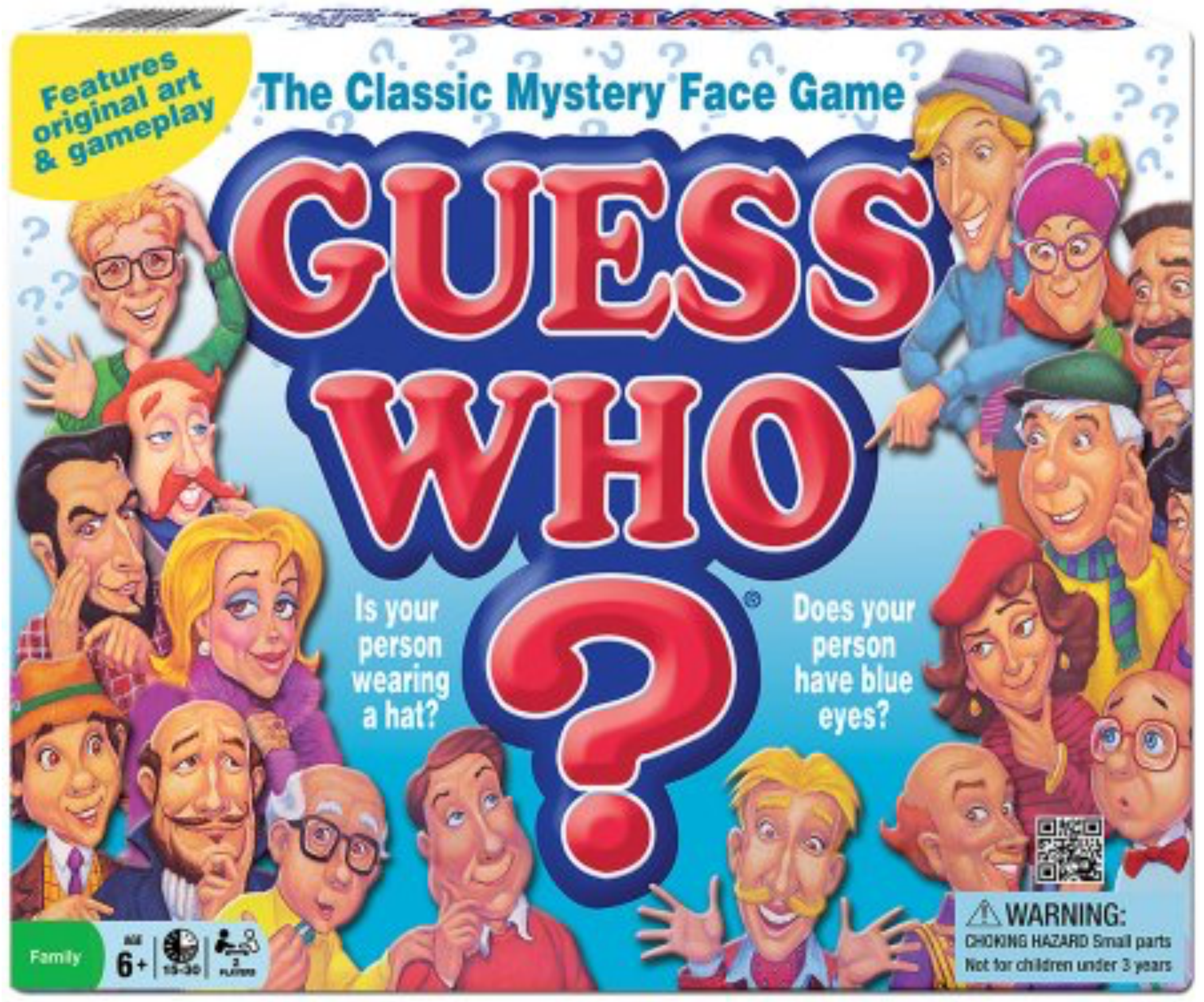


More energy

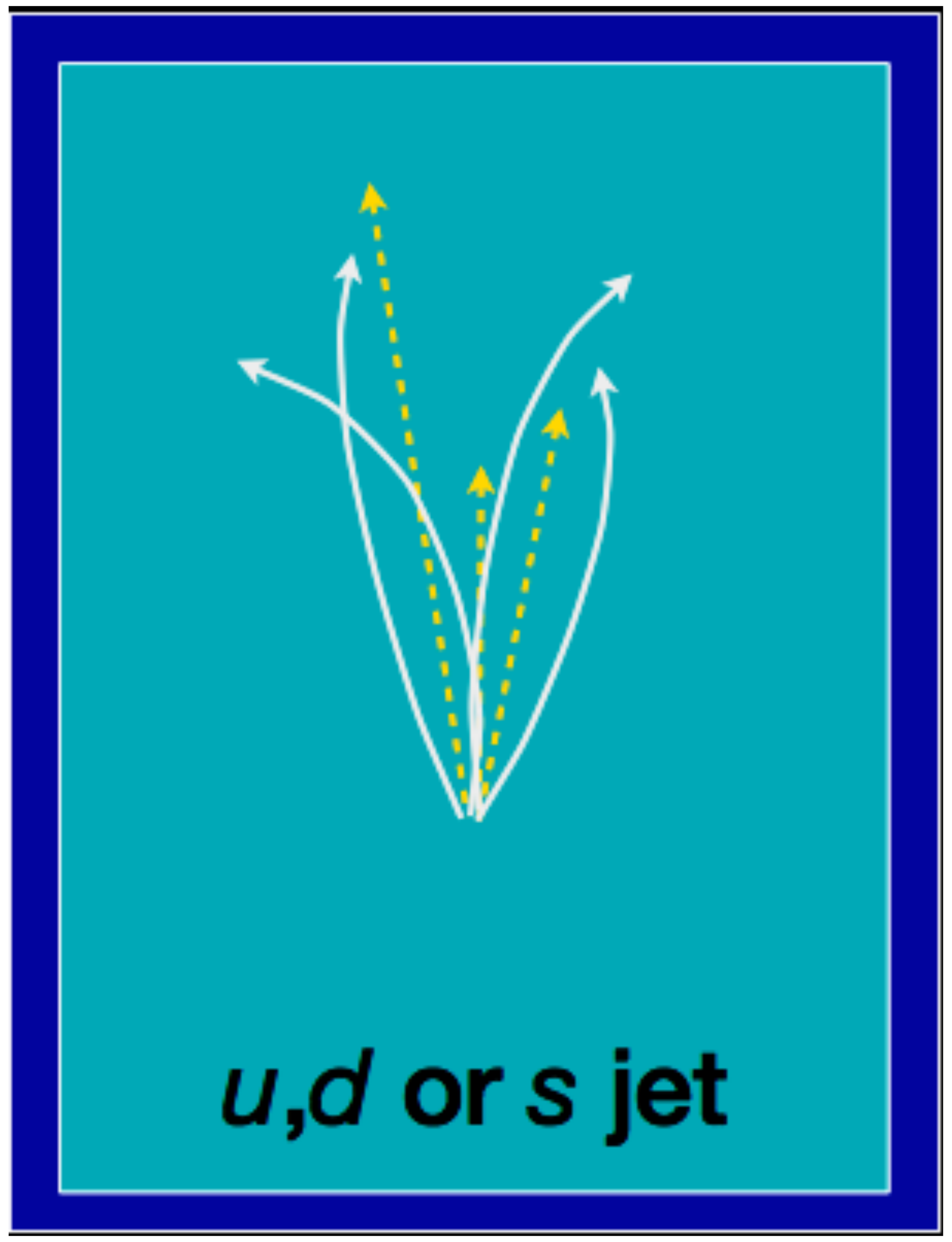


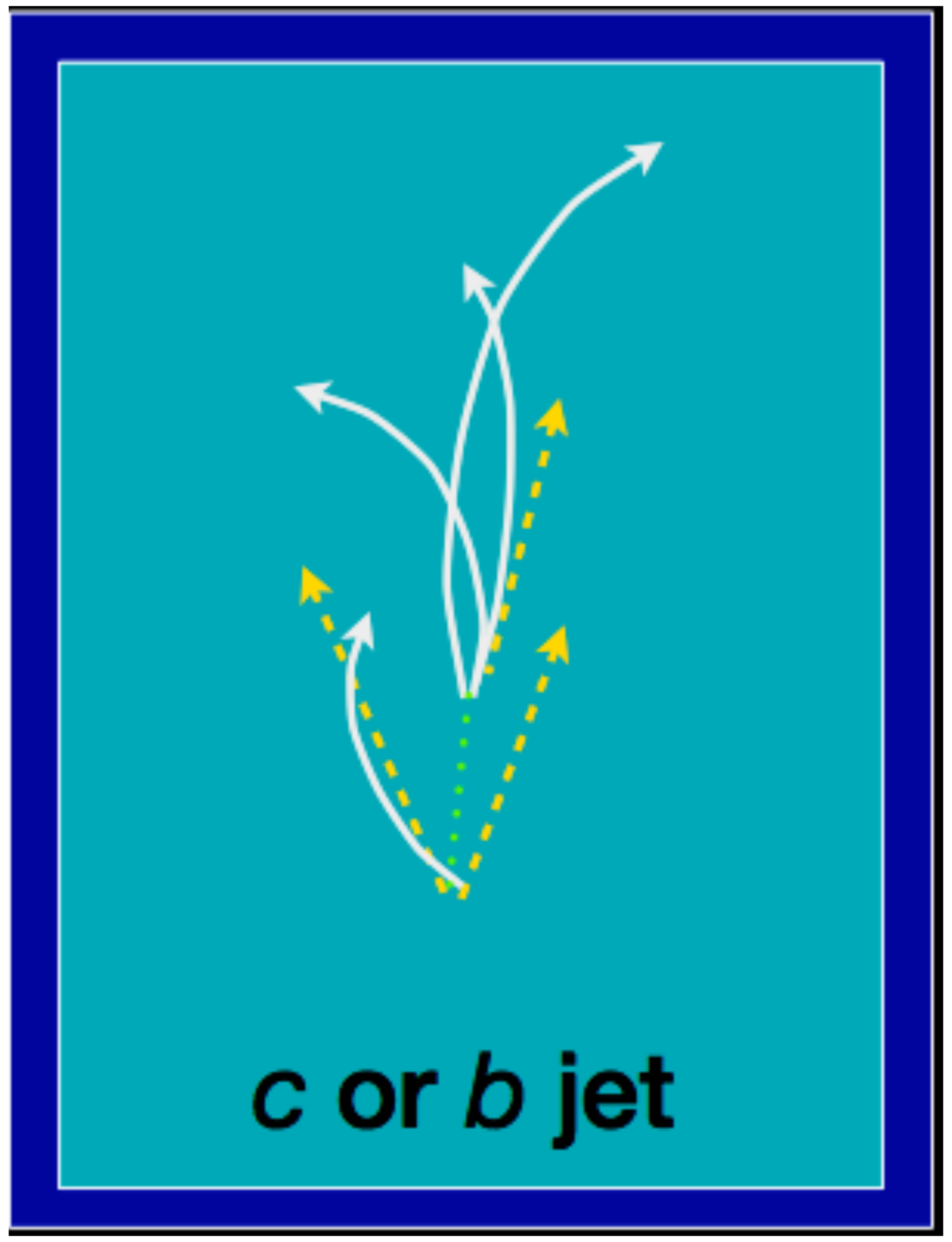
More energy


CHARGED HADRONS
NEUTRAL HADRONS
PHOTONS


Features original art & gameplay
The Classic Mystery Face Game
GUESS WHO ?
Is your person wearing a hat?
Does your person have blue eyes?
WARNING:

MORE ENERGY


u,d or s jet


c or b jet


gluon jet


pileup jet

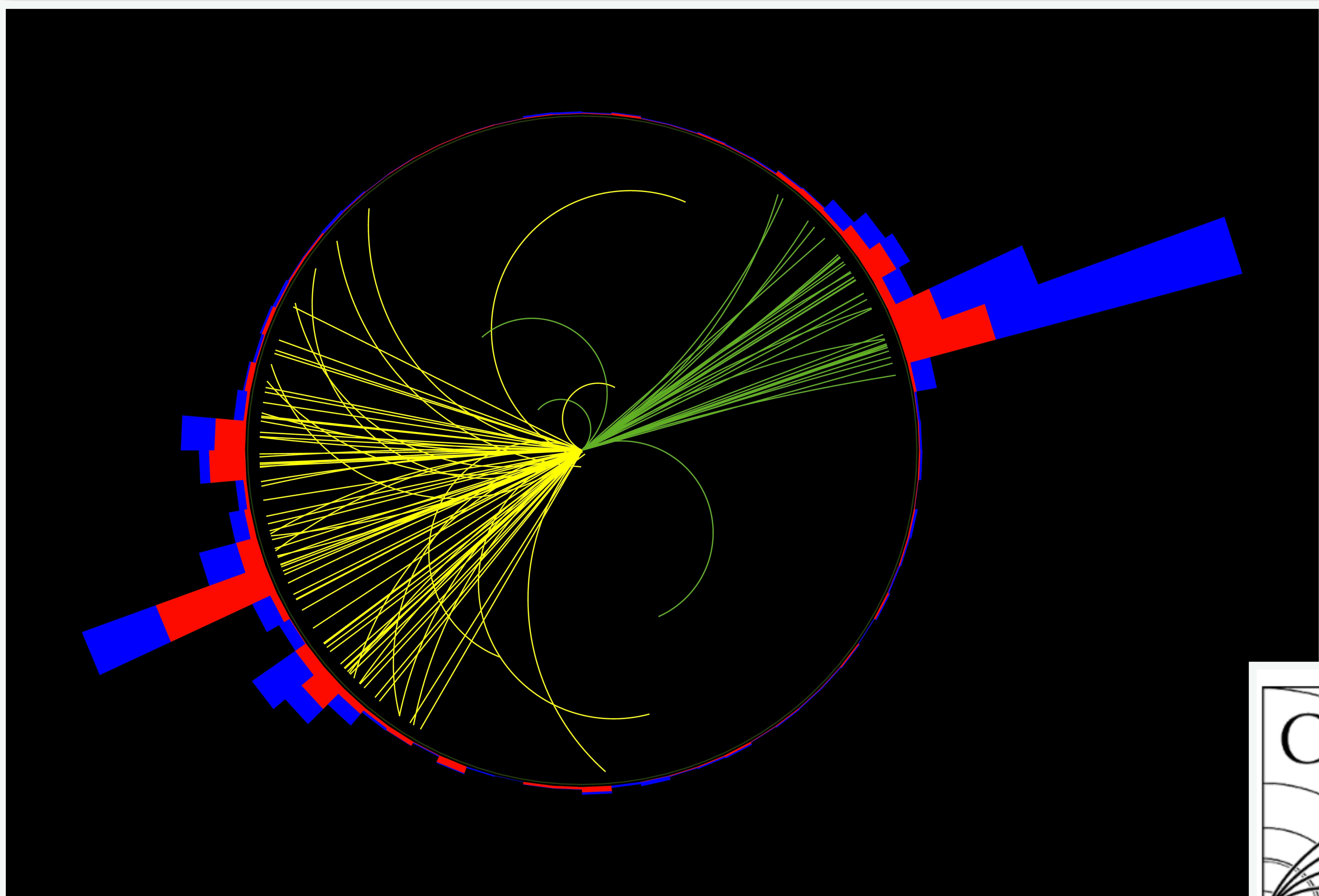

W or Z jet

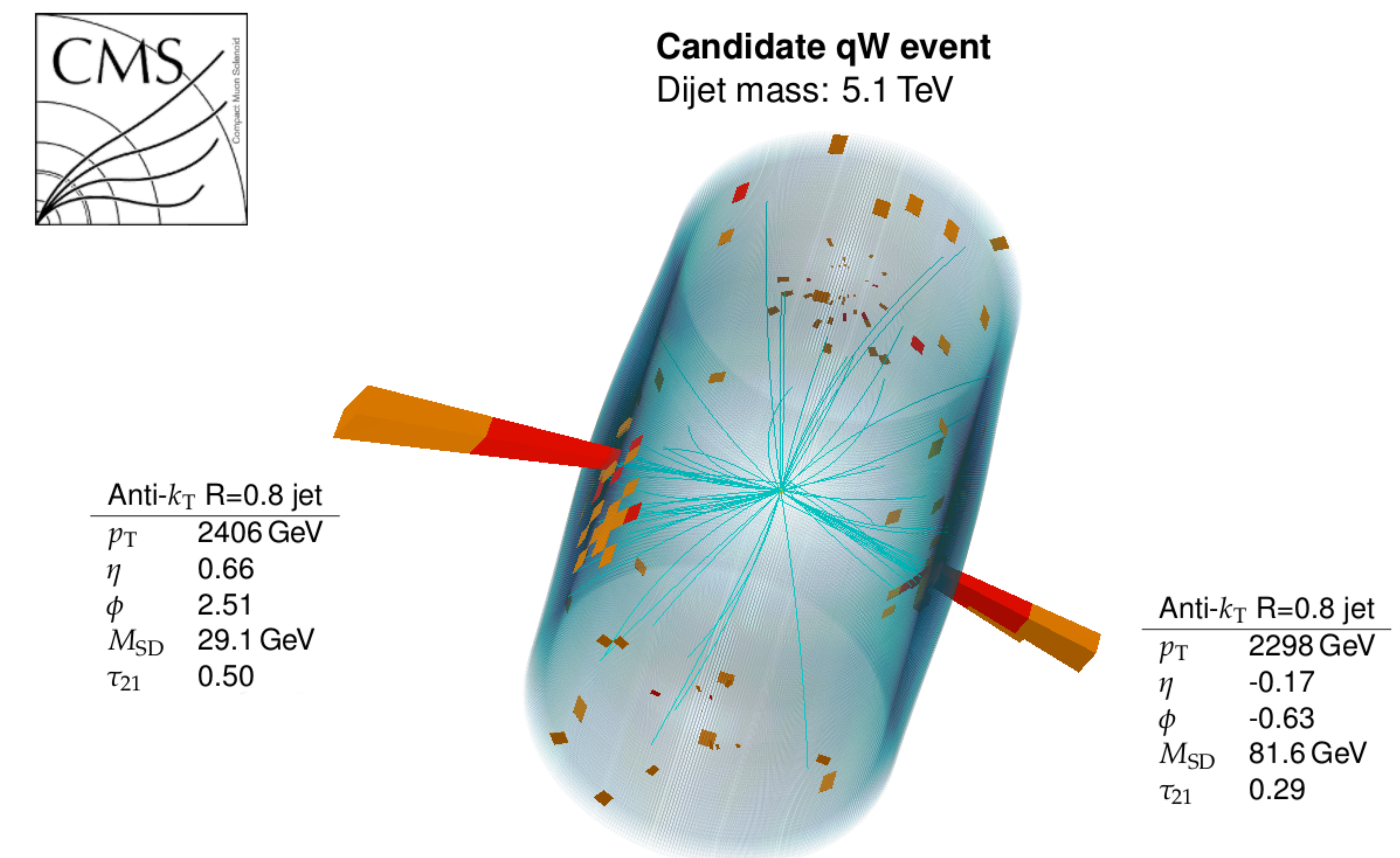

Higgs jet

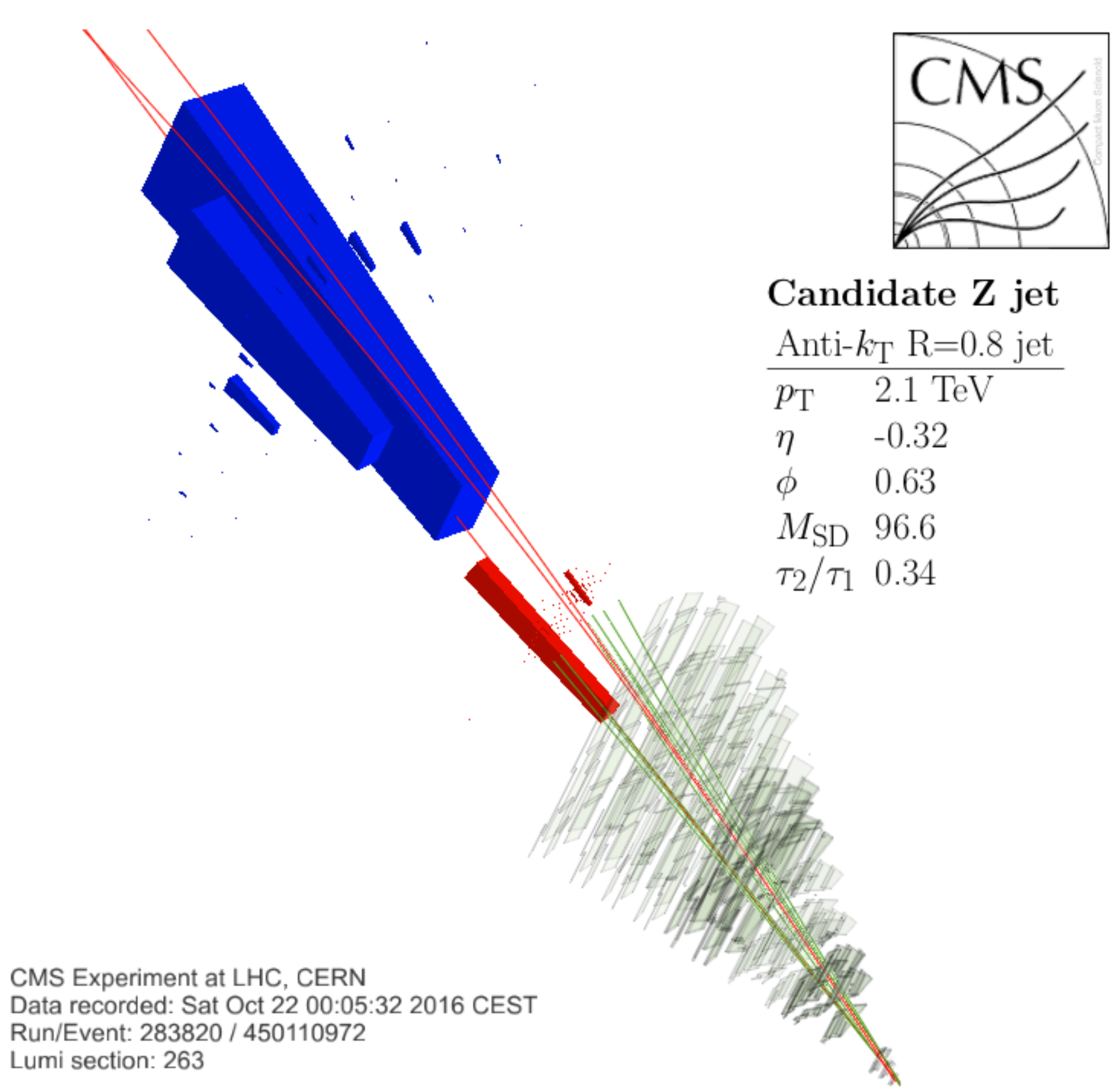

top jet

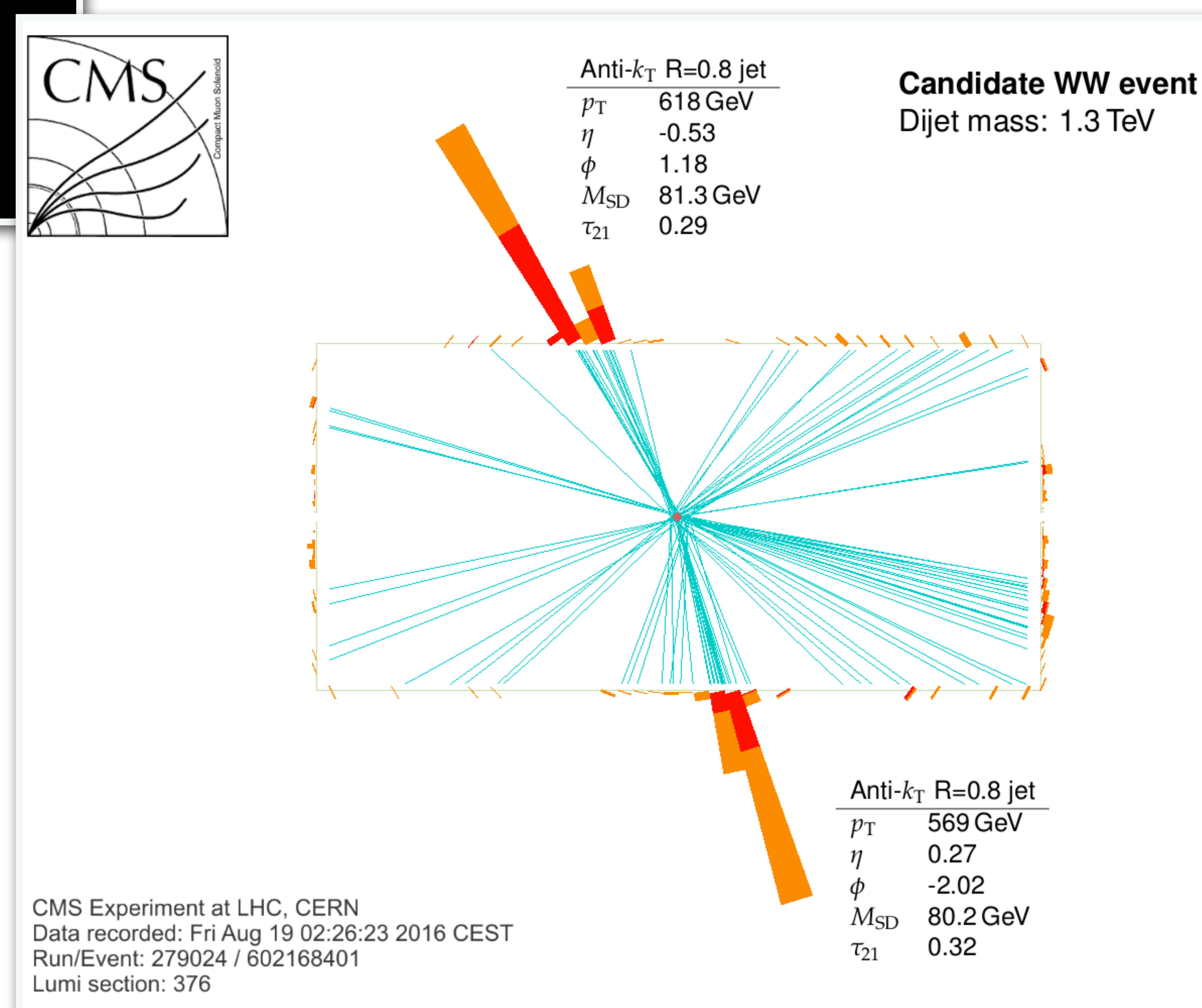

?

PICTURES

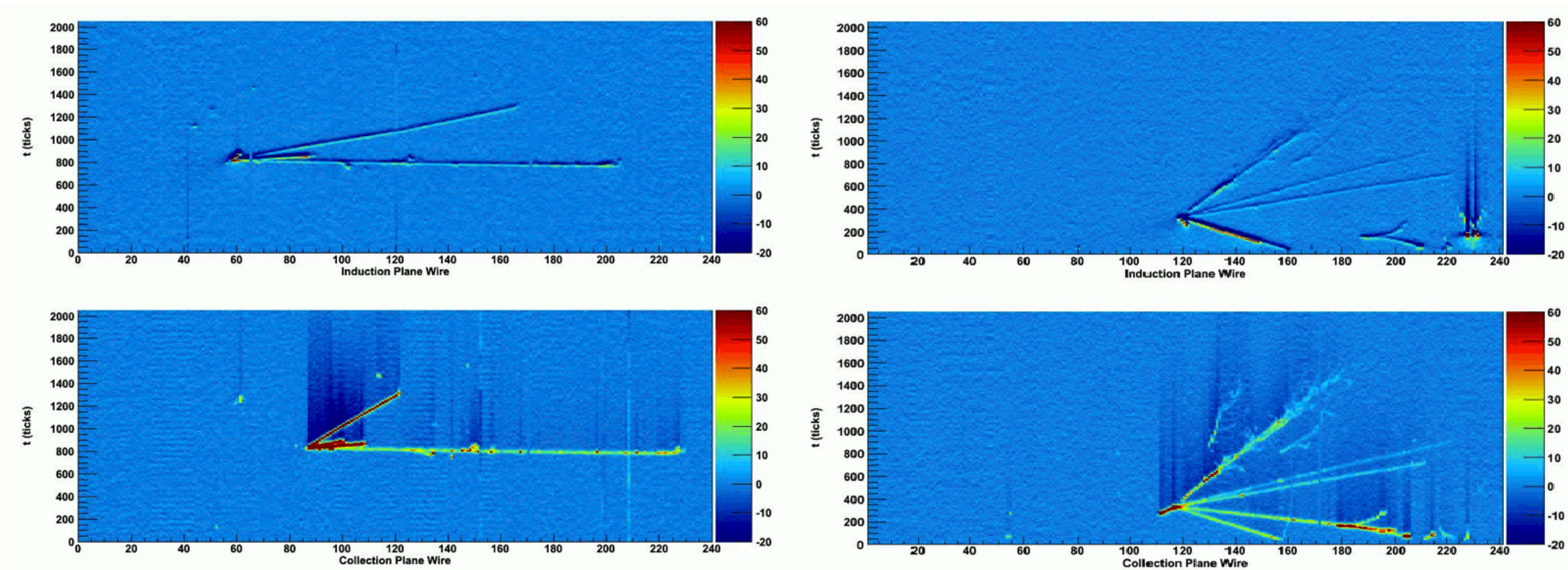



CMS
Candidate qW event
Dijet mass: 5.1 TeV
Anti-k_T R=0.8 jet
p_T 2406 GeV
η 0.66
ϕ 2.51
M_{SD} 29.1 GeV
τ_{21} 0.50
Anti-k_T R=0.8 jet
p_T 2298 GeV
η -0.17
ϕ -0.63
M_{SD} 81.6 GeV
τ_{21} 0.29


CMS
Candidate Z jet
Anti-k_T R=0.8 jet
p_T 2.1 TeV
η -0.32
ϕ 0.63
M_{SD} 96.6
τ_2/τ_1 0.34
CMS Experiment at LHC, CERN
Data recorded: Sat Oct 22 00:05:32 2016 CEST
Run/Event: 283820 / 450110972
Lumi section: 263


CMS
Compact Muon Solenoid
Candidate WW event
Dijet mass: 1.3 TeV
Anti-k_T R=0.8 jet
p_T 618 GeV
η -0.53
ϕ 1.18
M_{SD} 81.3 GeV
τ_{21} 0.29
Anti-k_T R=0.8 jet
p_T 569 GeV
η 0.27
ϕ -2.02
M_{SD} 80.2 GeV
τ_{21} 0.32
CMS Experiment at LHC, CERN
Data recorded: Fri Aug 19 02:26:23 2016 CEST
Run/Event: 279024 / 602168401
Lumi section: 376

OTHER SIMILAR-LOOKING PICTURES



IDENTIFYING BOOSTED JETS

pT,Y,ϕ + tracking

mass

4-vector sum of jet constituents

highly sensitive to soft QCD and pileup; **grooming** can be used to mitigate these dependencies

substructure

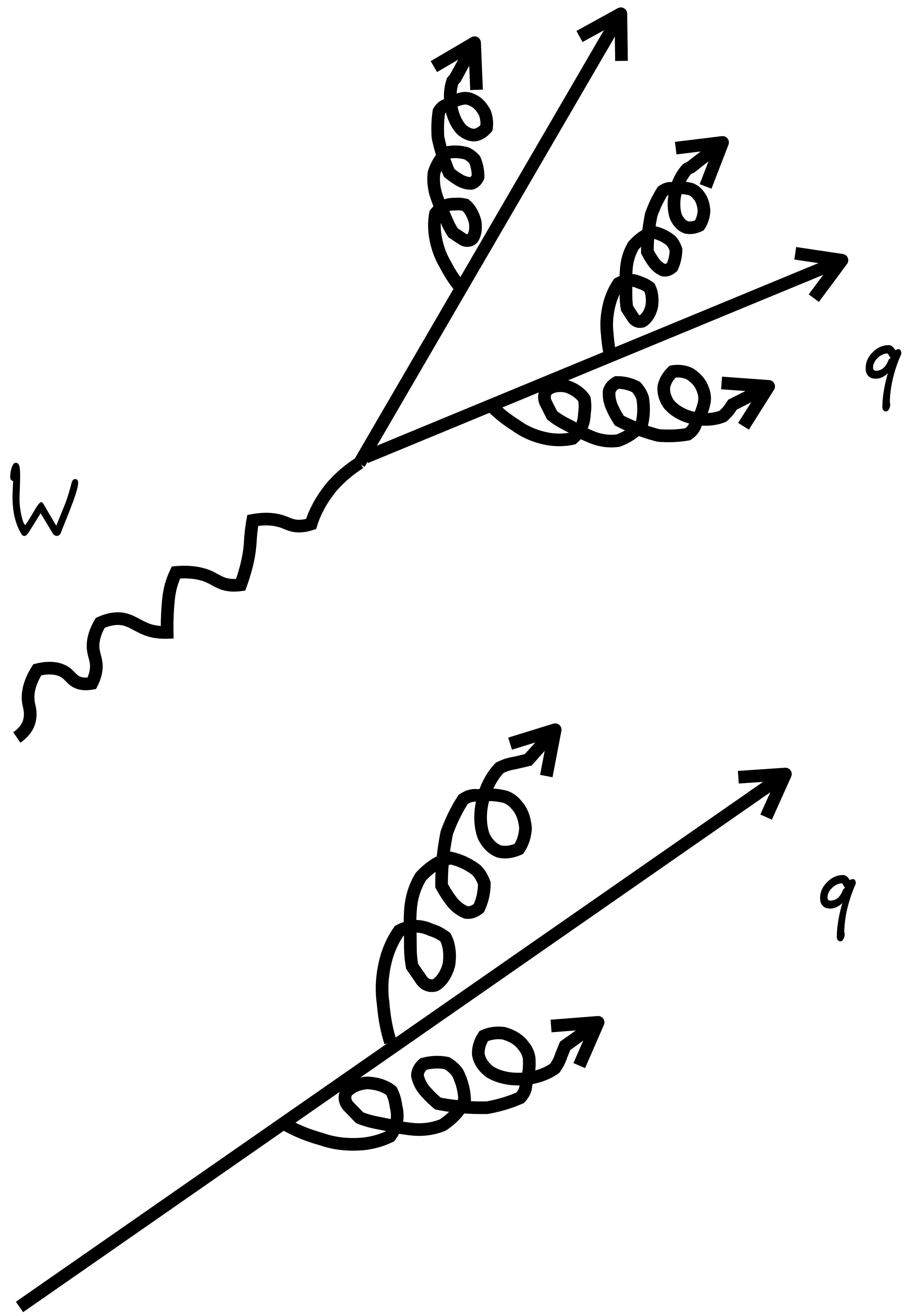
several classes: declustering/reclustering, generalized jet shapes and energy flow, statistical interpretation (Qjets), jet charge

algorithms

some combination of cuts on mass, shapes, tracking

most typical in **top tagging**

JET MASS



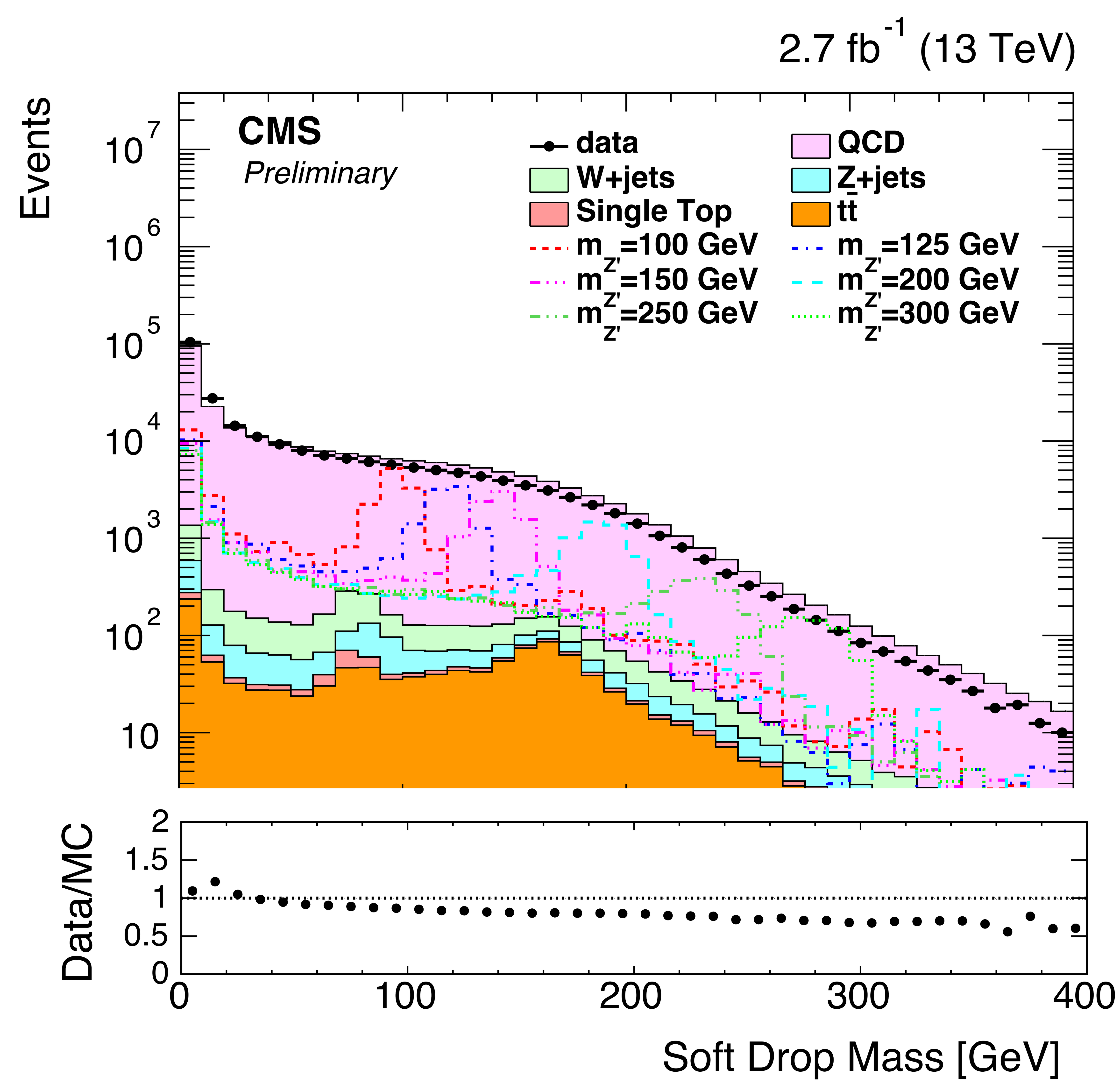
$$\boxed{\langle M^2 \rangle \simeq \left.\begin{array}{ll} \text{quarks:} & 0.16 \\ \text{gluons:} & 0.37 \end{array}\right\} \times \alpha_s p_t^2 R^2}$$

mass is the most useful jet shape observable

at parton level, these are pretty easy to tell apart

But jet mass is a perturbative quantity

And it's tough to model!

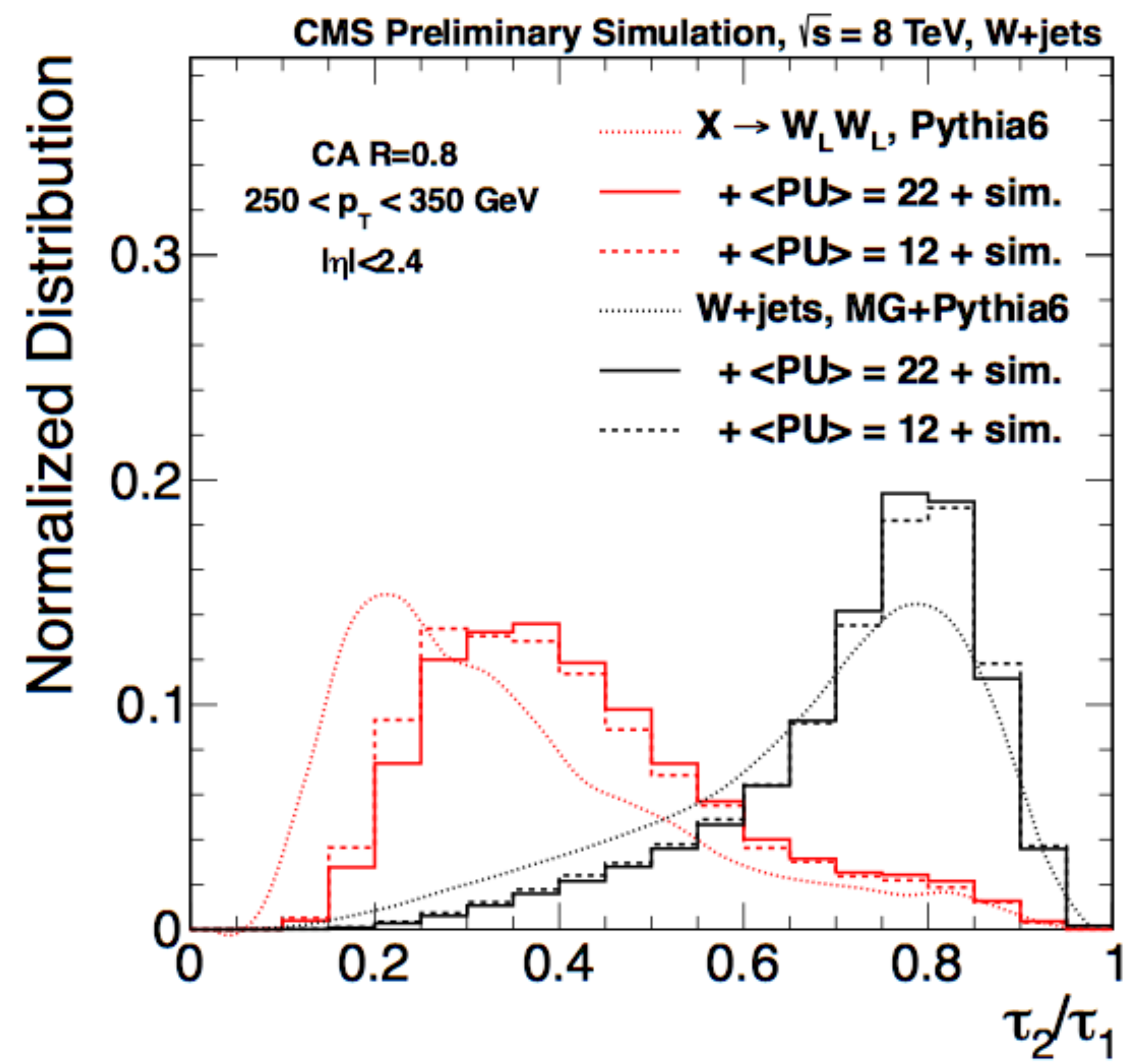


JET SHAPES, A TASTE

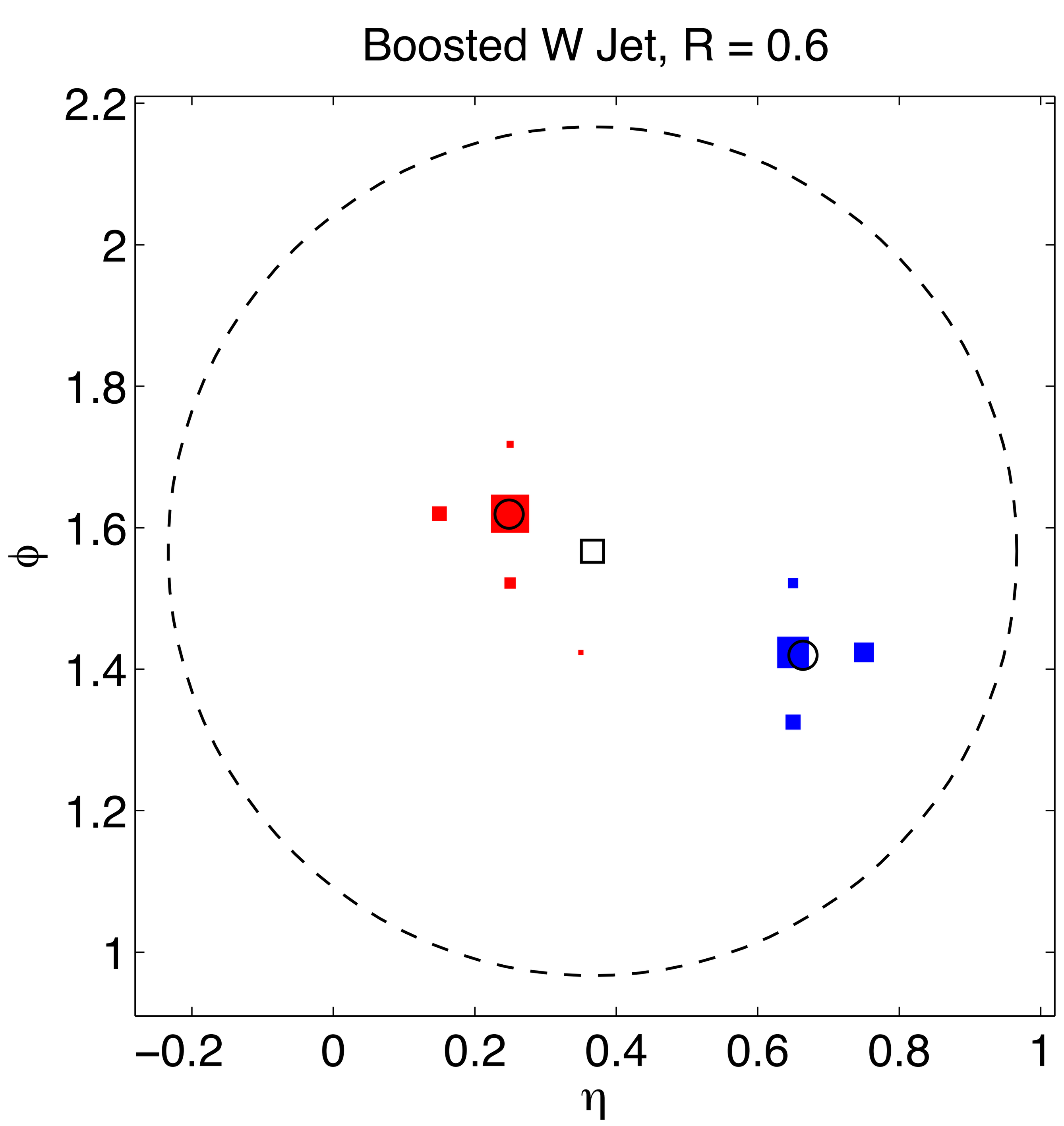
"Prongy-ness"

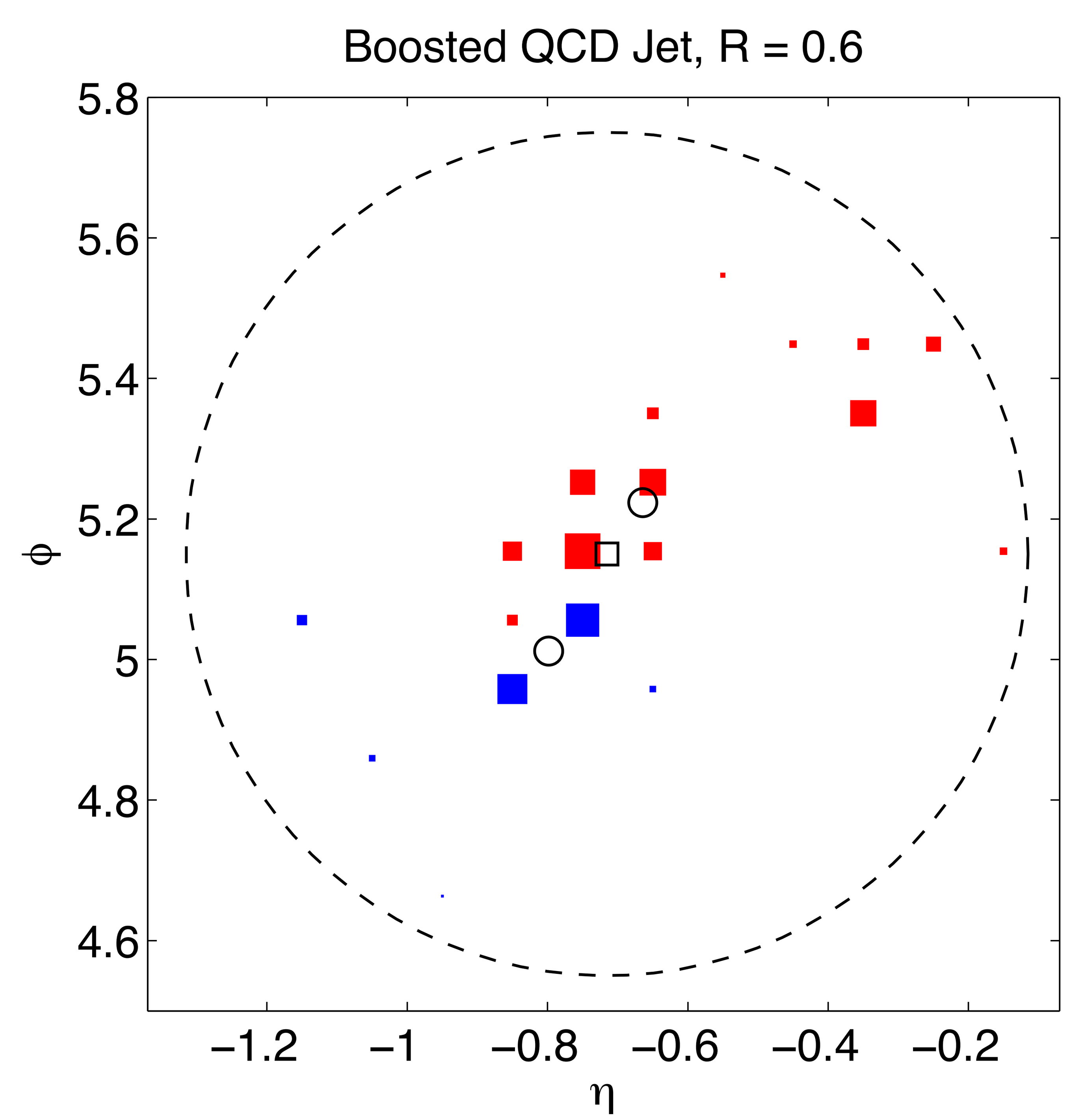
N-subjettiness: a measure of how consistent a jet is with having N subjets, τ_N

$$\tau_N = \frac{1}{d_0} \sum_k p_{T,k} \min\{\Delta R_{1,k}, \Delta R_{2,k}, \cdots, \Delta R_{N,k}\}$$



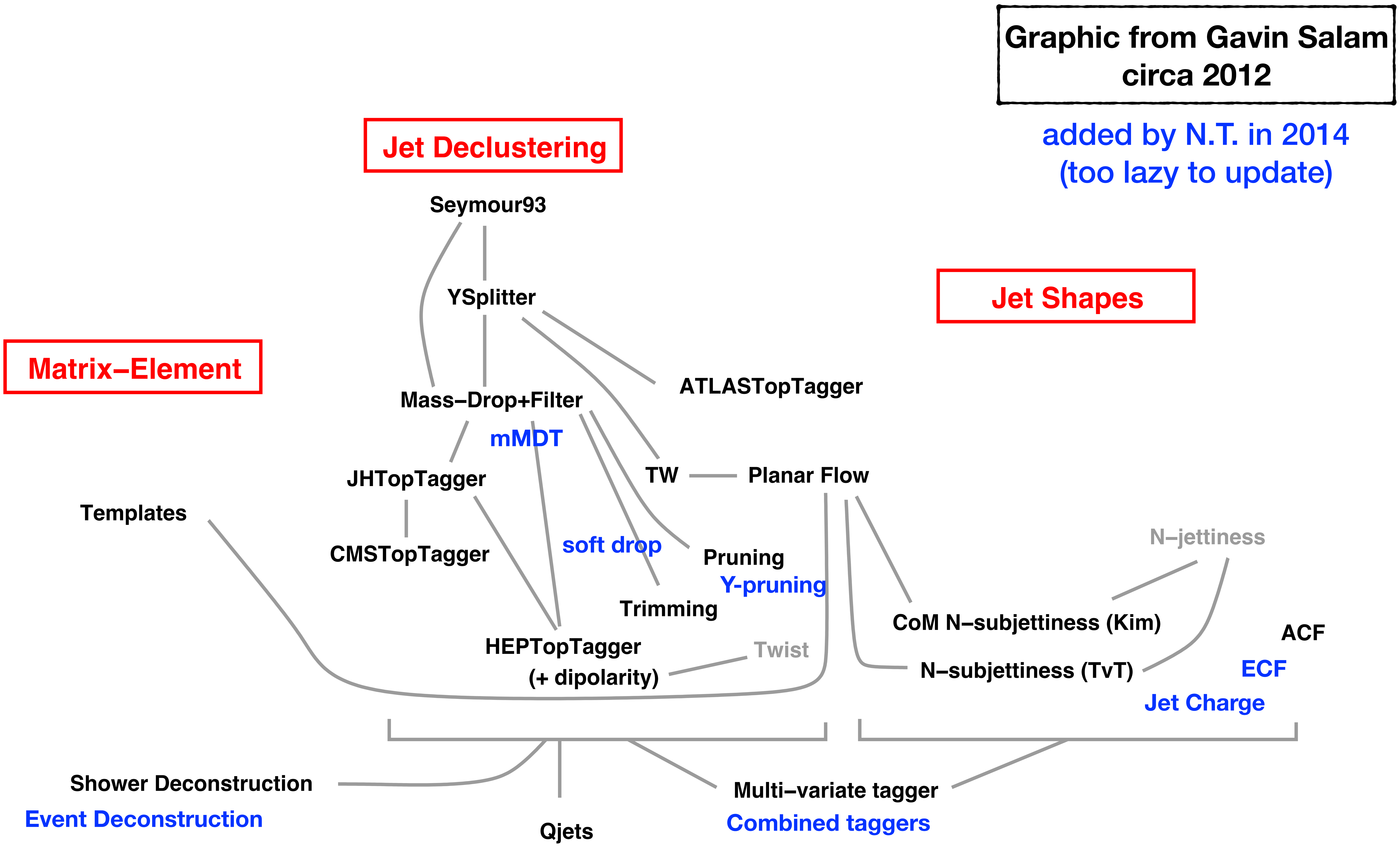
Ratios are typically used:
τ_2/τ_1 for separating **W jets** from **quark and gluon jets**





Tons of variables to measure jet radiation profiles for different tasks
Could you use variables like this to separate DIS,Res,QE?

ANALYTIC/ALGORITHMIC CLASSIFICATION



apologies for omitted taggers, arguable links, etc.

ANALYTIC/ALGORITHMIC CLASSIFICATION

Jet substructure is an interesting field because it attracted a lot of interest from theory — SCET vs. perturbative QCD.

This resulted in an interesting blend of tagging algorithms and observables calculable by theory.

Very interesting from an "information theory" point-of-view.

A lot of physical underpinning and ways to think about the information content of a jet.

We'll come back to this when discussing ML, but you can start to ask:

What is the machine learning?

THE TASK OF JET SUBSTRUCTURE

Identify interesting highly-boosted, highly energetic objects

Complicated correlated multi-body final states

A broad range of very interesting physics!

SM, Higgs, Exotics, Susy,…

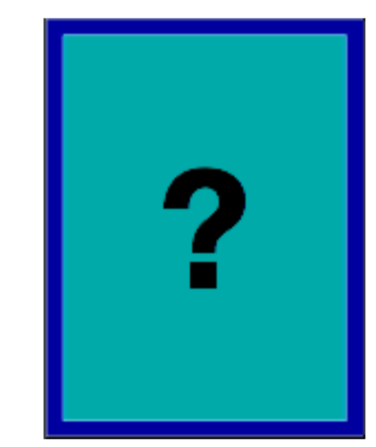
Imperfect MC modeling

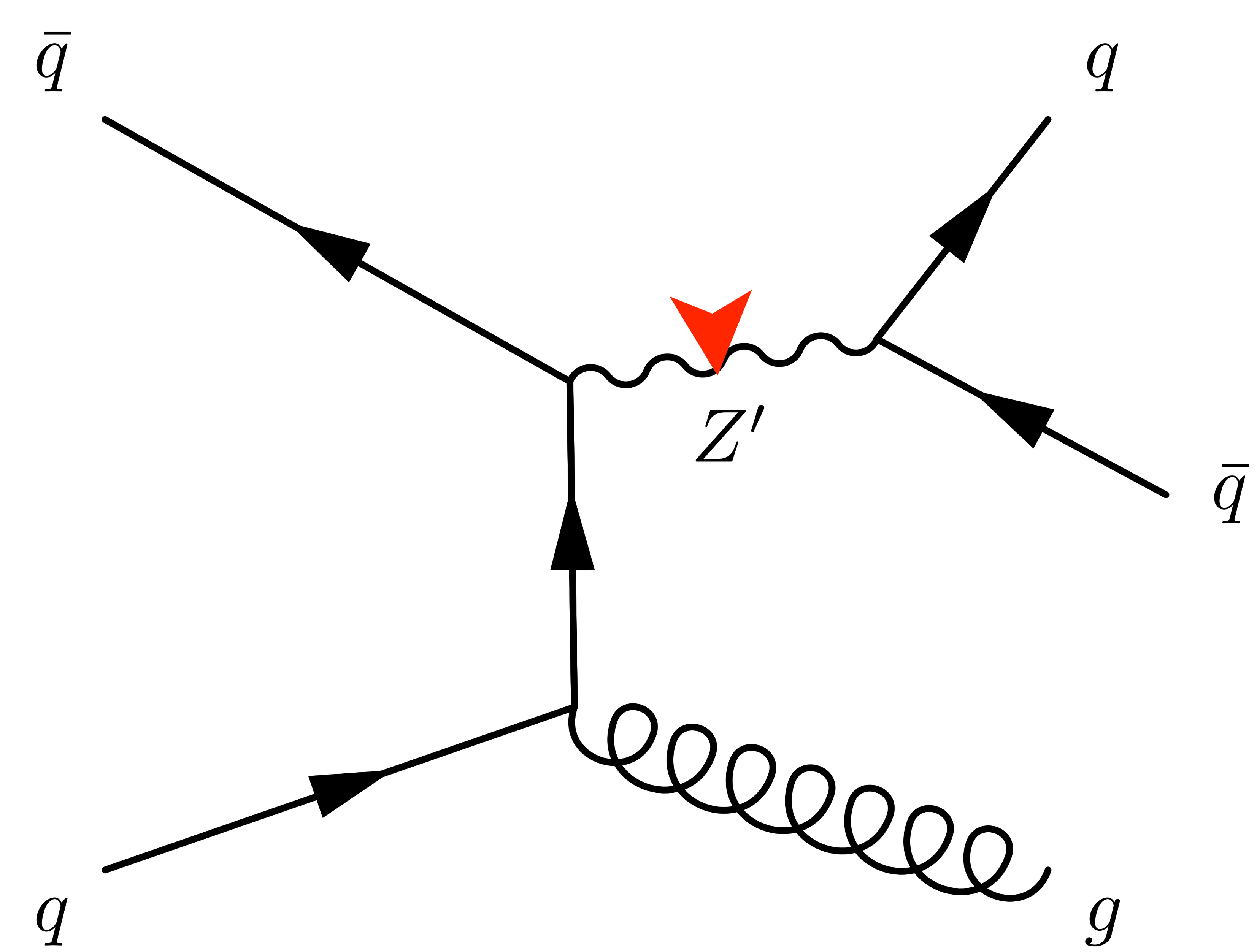
Strategies for background, signal, and related systematic uncertainty estimation?

New particles in substructure?

Generic features

Training away new physics?





STANDARD CANDLES AND CORRELATIONS

How do we perform measurements and searches on things that are not well-modeled?

No one may admit it now, but in the early days of jet substructure people thought you would never be able to understand the structure of QCD well-enough to employ these methods

Signals: find standard candles in the standard model and extrapolate with generous modeling uncertainties

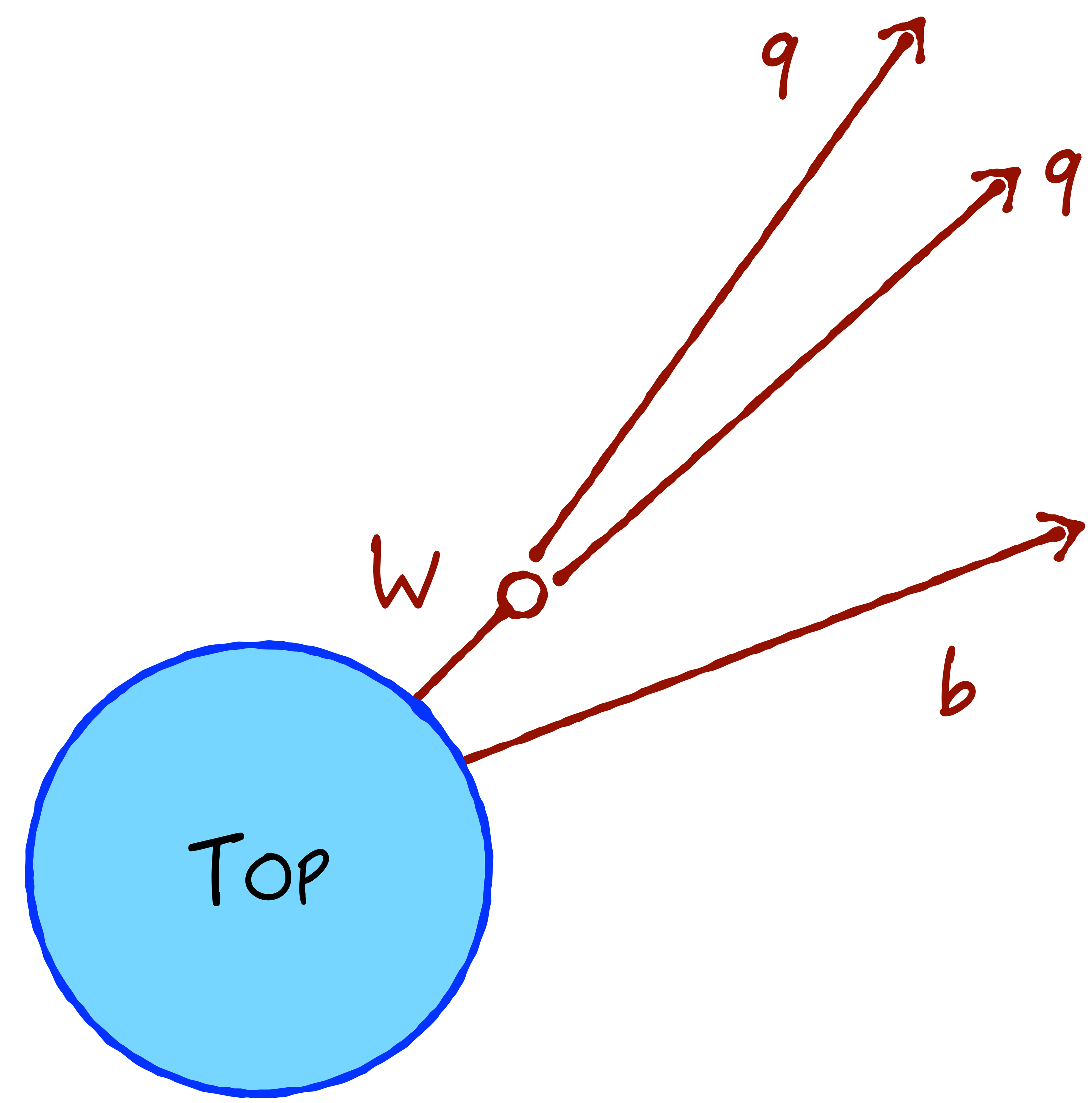
Backgrounds: build orthogonal signal-depleted, background-rich control regions to study and estimate background; requires a good understand of ***correlations*** between observables

Systematics come from standard candle, extrapolation uncertainties, and sideband fits

STANDARD CANDLES

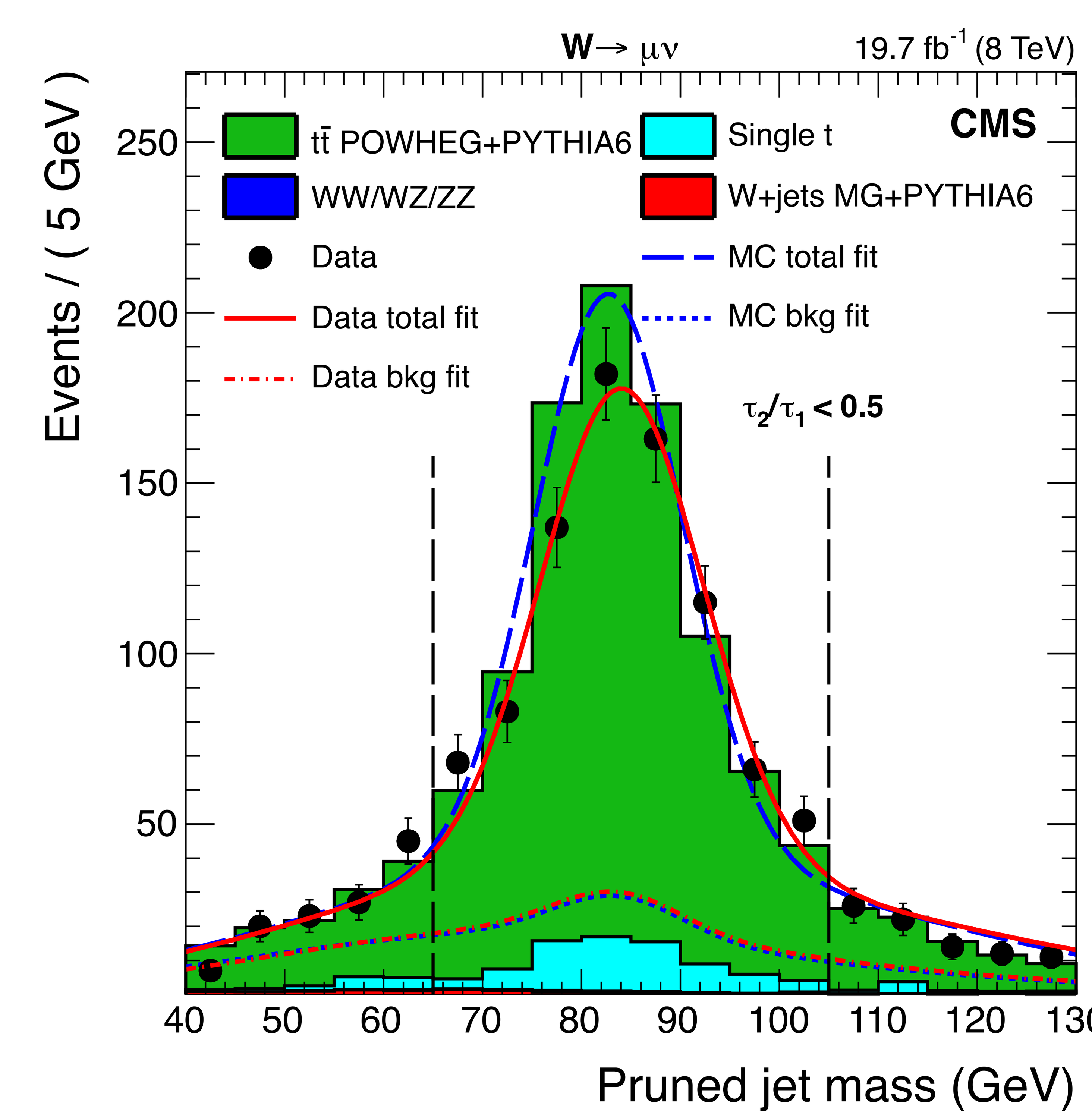
How do we perform measurements on things that are not well-modeled?

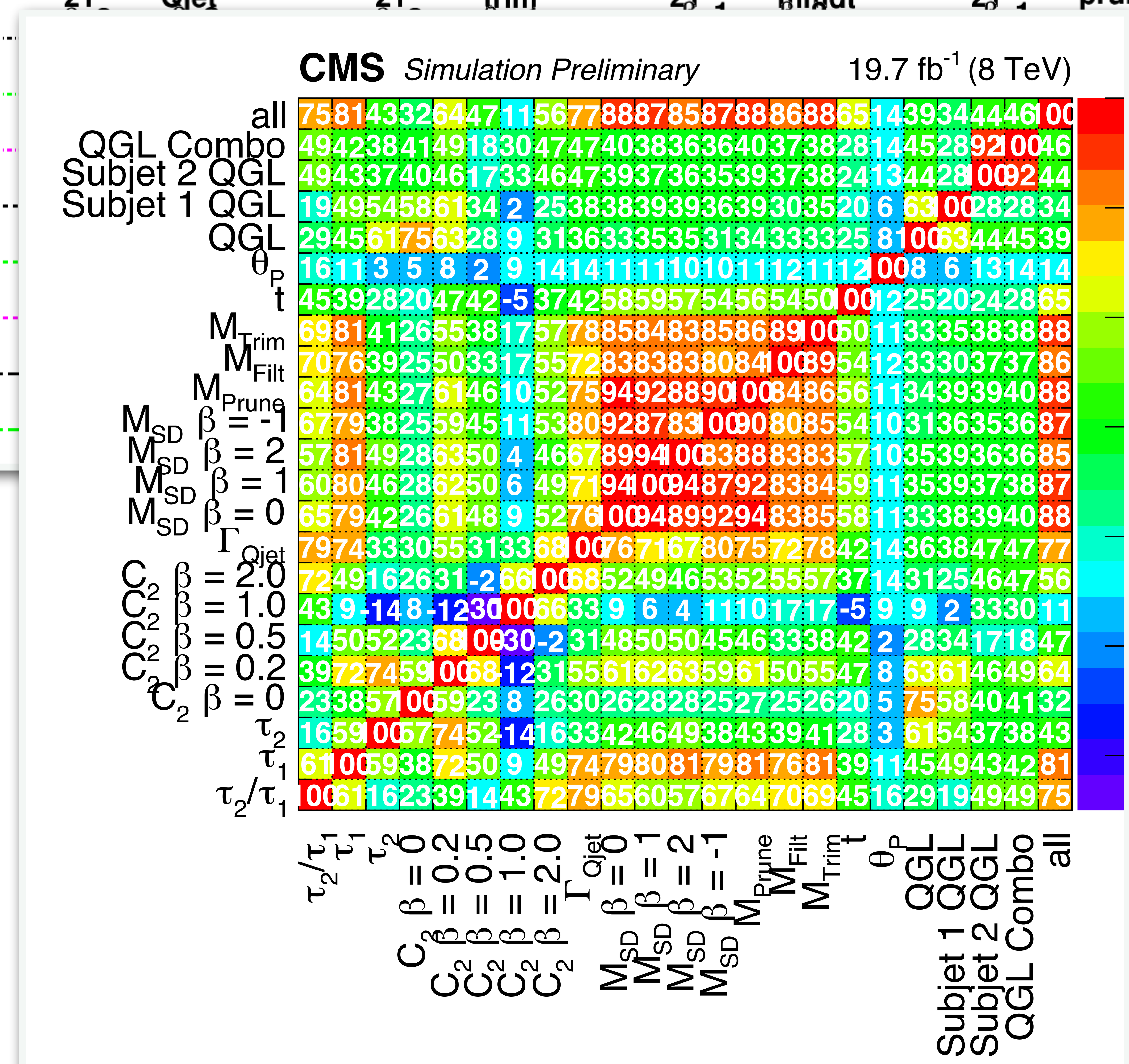
Signals: find standard candles in the standard model and extrapolate with generous modeling uncertainties



Top quarks provide both W and top jet standard candles.

More subtle: using gluon → bbar as a standard candle for understanding H(bb)





Towards an Understanding of the Correlations in Jet Substructure

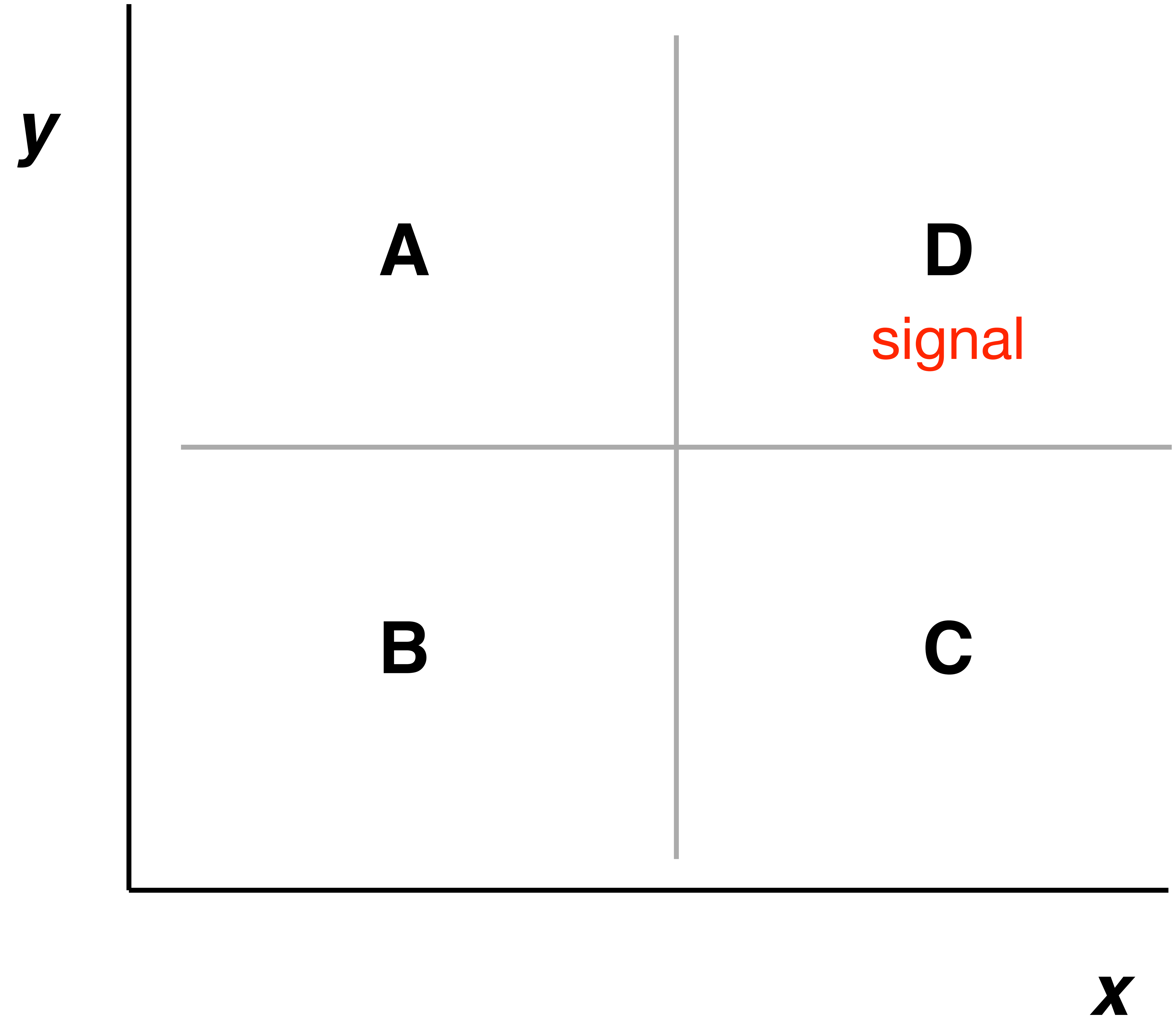
Report of BOOST2013, hosted by the University of Arizona, 12^{th}-16^{th} of August 2013.

D. Adams[1], A. Arce[2], L. Asquith[3], M. Backovic[4], T. Barillari[5], P. Berta[6], D. Bertolini[7], A. Buckley[8], J. Butterworth[9], R. C. Camacho Toro[10], J. Caudron[11], Y.-T. Chien[12], J. Cogan[13], B. Cooper[9], D. Curtin[14], C. Debenedetti[15], J. Dolen[16], M. Eklund[17], S. El Hedri[11], S. D. Ellis[18], T. Embry[17], D. Ferencek[19], J. Ferrando[8], S. Fleischmann[20], M. Freytsis[21], M. Giulini[22], Z. Han[23], D. Hare[24], P. Harris[25], A. Hinzmann[26], R. Hoing[27], A. Hornig[12], M. Jankowiak[28], K. Johns[17], G. Kasieczka[29], R. Kogler[27], W. Lampl[17], A. J. Larkoski[30], C. Lee[12], R. Leone[17], P. Loch[17], D. Lopez Mateos[21], H. K. Lou[31], M. Low[32], P. Maksimovic[33], I. Marchesini[27], S. Marzani[30], L. Masetti[11], R. McCarthy[34], S. Menke[5], D. W. Miller[32], K. Mishra[24], B. Nachman[13], P. Nef[13], F. T. O'Grady[17], A. Ovcharova[35], A. Picazio[10], C. Pollard[8], B. Potter-Landua[25], C. Potter[25], S. Rappoccio[16], J. Rojo[36], J. Rutherfoord[17], G. P. Salam[25,37], R. M. Schabinger[38], A. Schwartzman[13], M. D. Schwartz[21], B. Shuve[39], P. Sinervo[40], D. Soper[23], D. E. Sosa Corral[22], M. Spannowsky[41], E.. Strauss[13], M. Swiatlowski[13], J. Thaler[30], C. Thomas[25], E. Thompson[42], N. V. Tran[24], J. Tseng[36], E. Usai[27], L. Valery[43], J. Veatch[17], M. Vos[44], W. Waalewijn[45], J. Wacker[46], and C. Young[25]

DECORRELATION IS AN OLD IDEA

A familiar example:
ABCD method

If observables x and y are uncorrelated, then
D = C * A / B



especially nice when background not well-modeled

“data-driven”

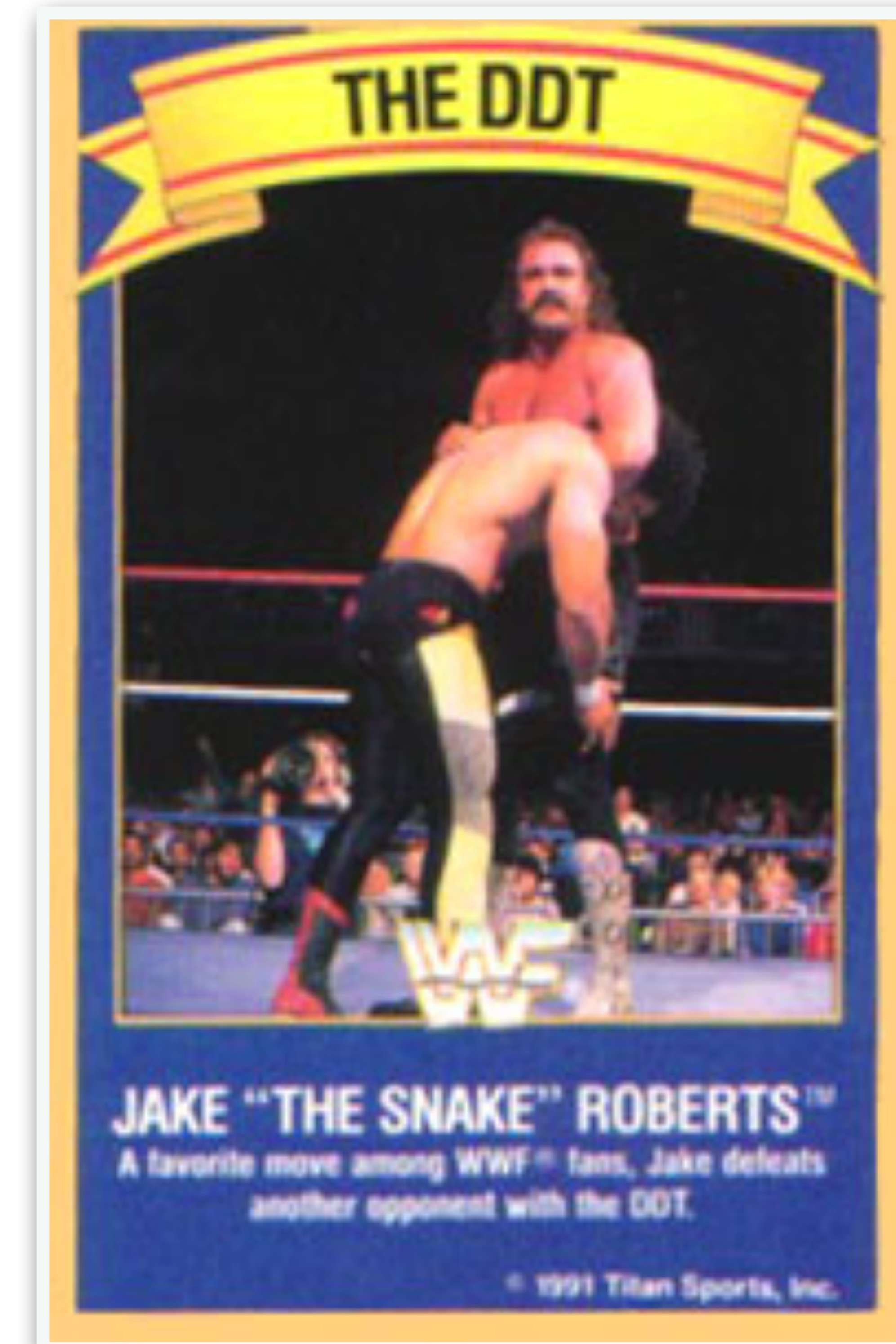
Many variants: ABCD, ABCDEF, Alphabet, alpha, rhalphabet…

DE-CORRELATIONS

Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure

James Dolen,[a] **Philip Harris,**[b] **Simone Marzani,**[a] **Salvatore Rappoccio,**[a] **and Nhan Tran**[c]

[a] *University at Buffalo, The State University of New York, Buffalo, NY 14260-1500, USA*
[b] *CERN, European Organization for Nuclear Research, Geneva, Switzerland*
[c] *Fermi National Accelerator Laboratory (FNAL), Batavia, IL 60510, USA*

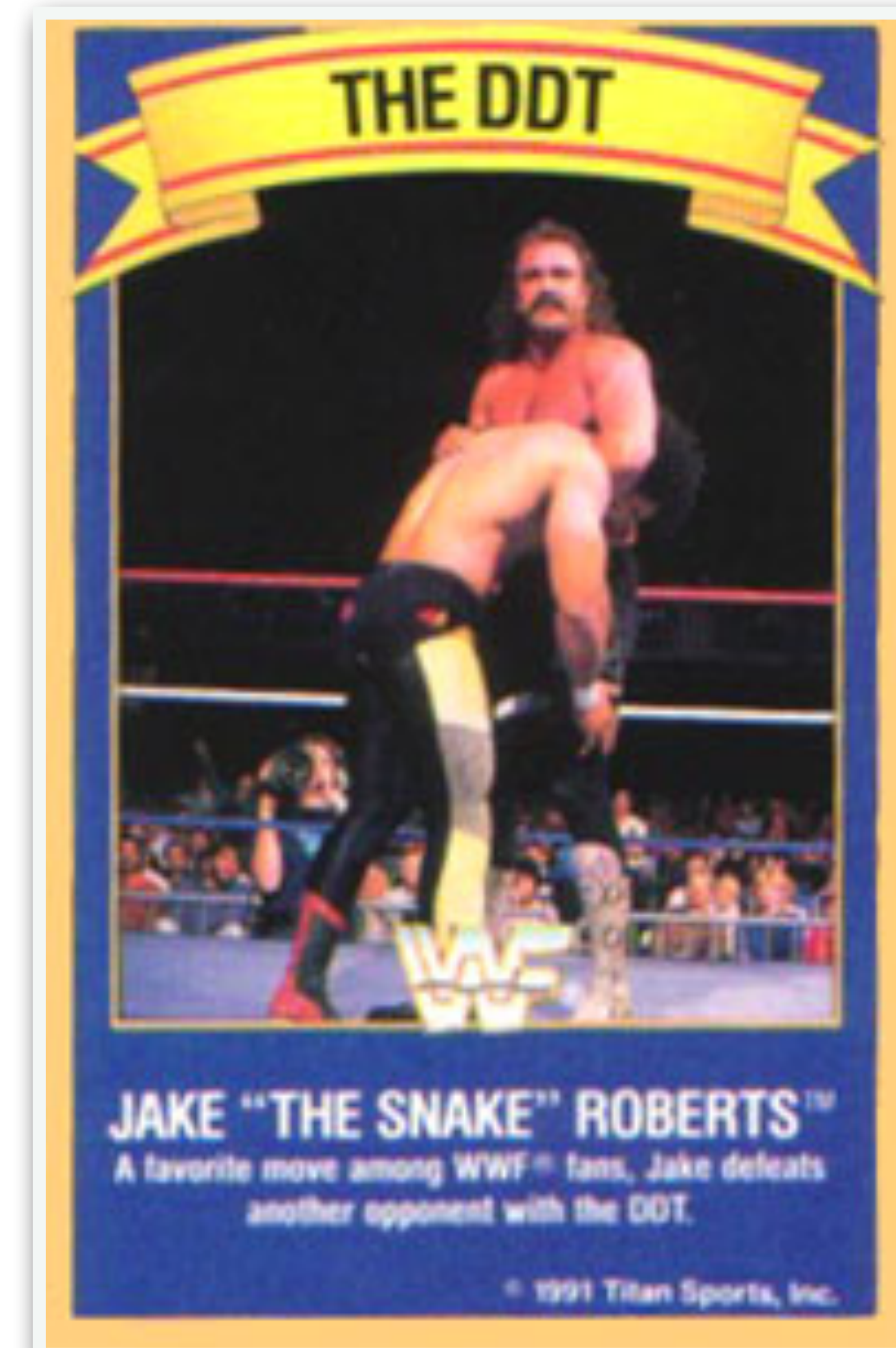


De-correlations

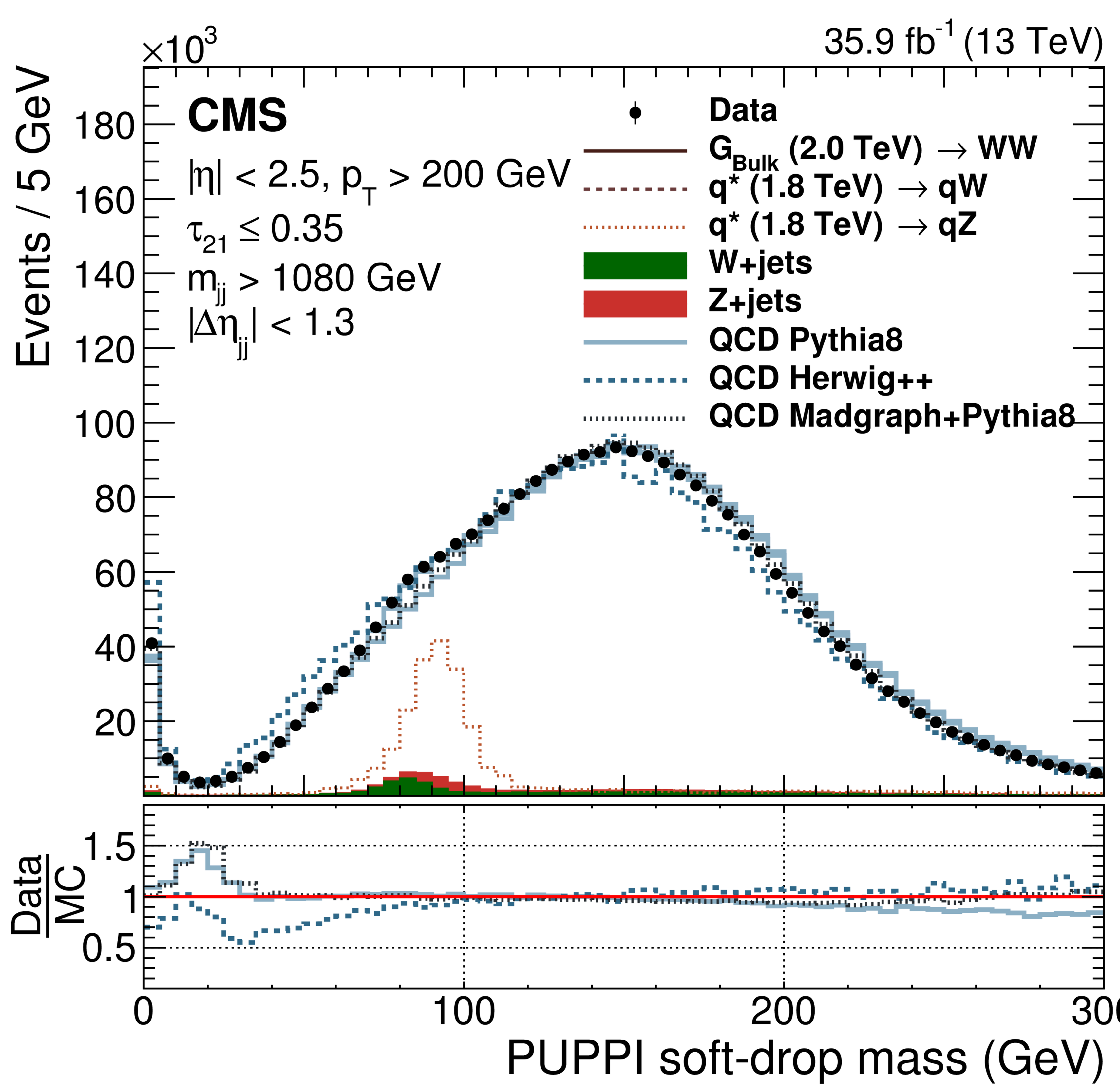
Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure

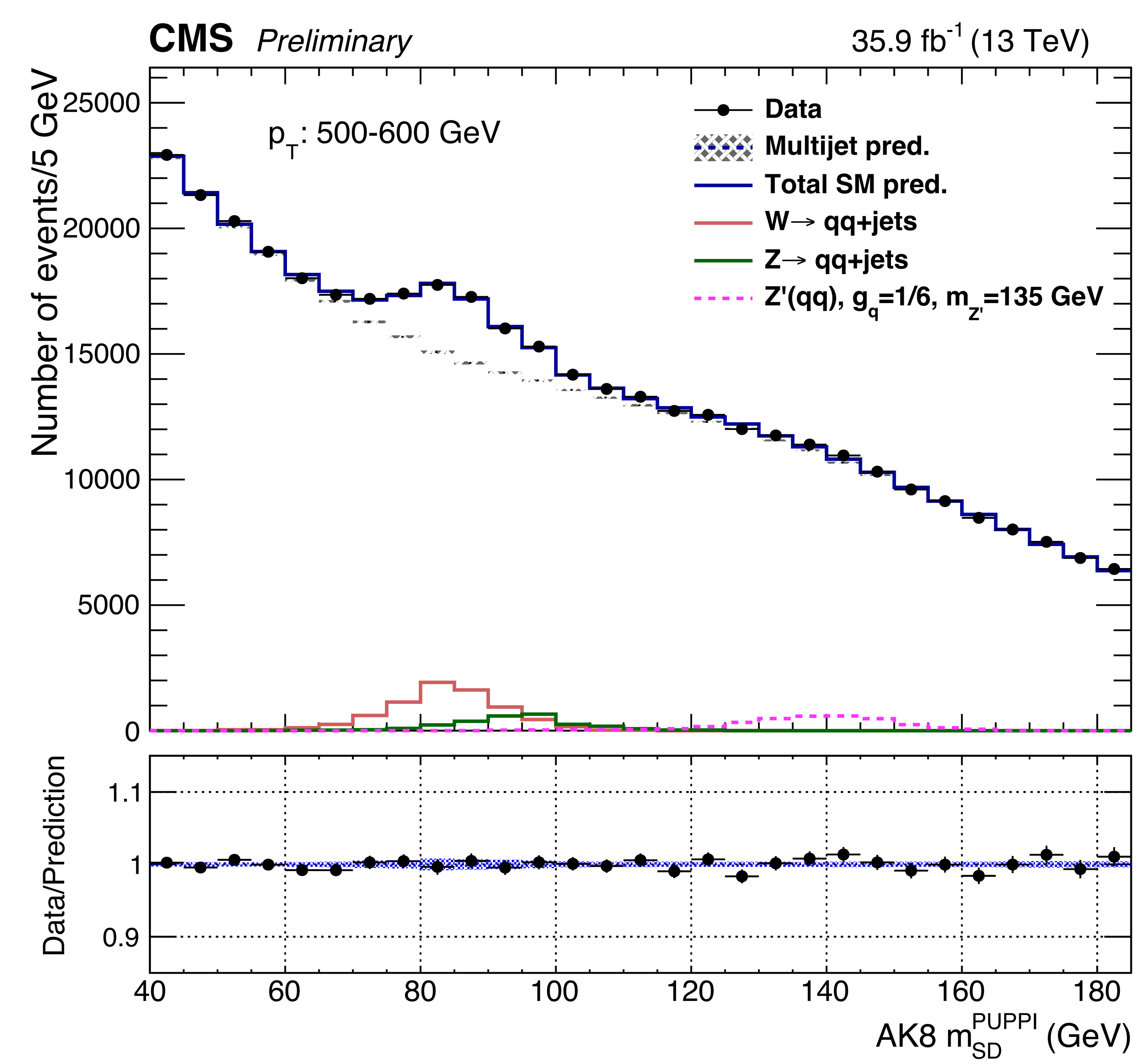
James Dolen,[a] Philip Harris,[b] Simone Marzani,[a] Salvatore Rappoccio,[a] and Nhan Tran[c]

[a] *University at Buffalo, The State University of New York, Buffalo, NY 14260-1500, USA*
[b] *CERN, European Organization for Nuclear Research, Geneva, Switzerland*
[c] *Fermi National Accelerator Laboratory (FNAL), Batavia, IL 60510, USA*



Where is the W peak?





De-correlations

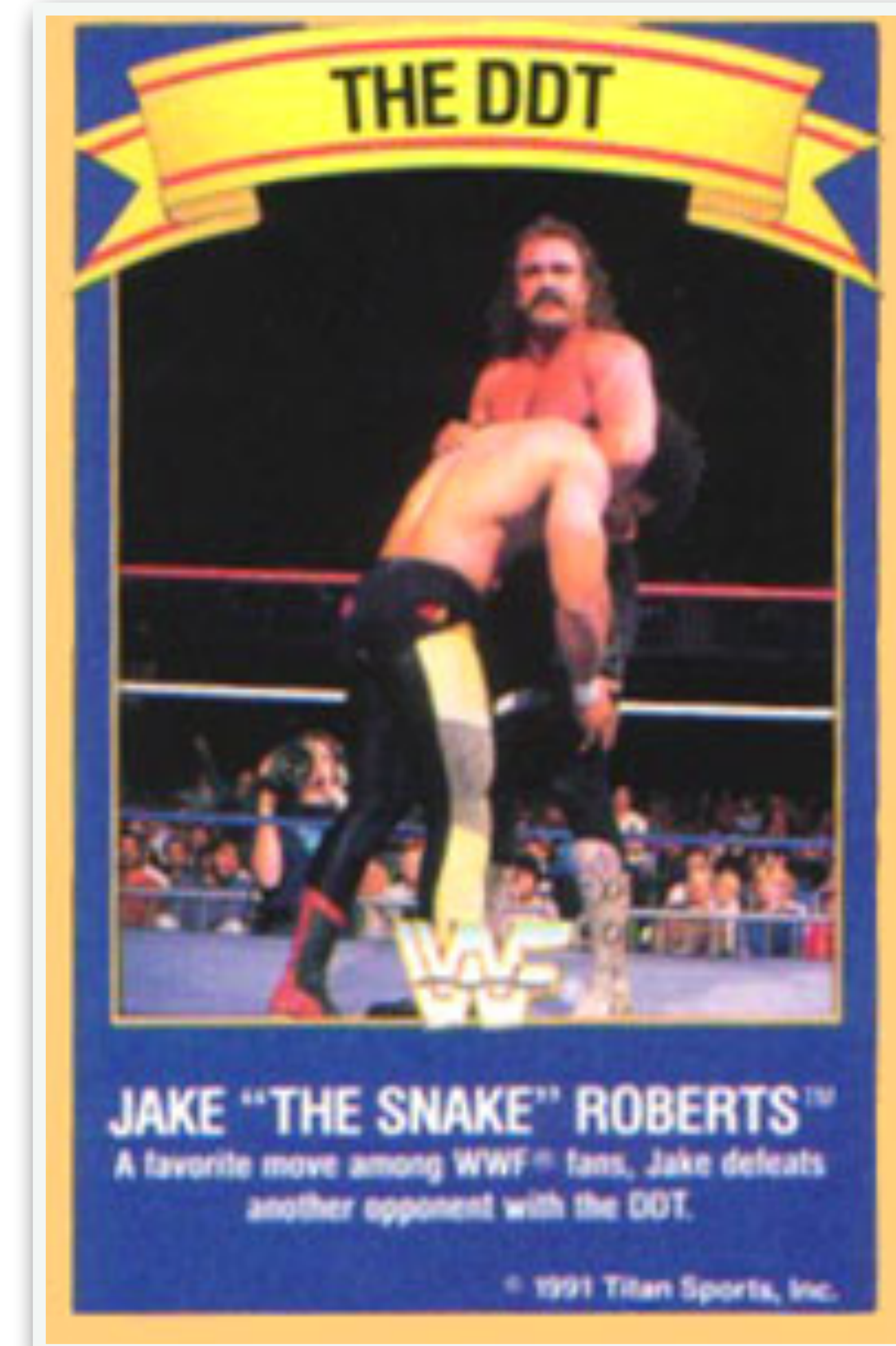
Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure

James Dolen,[a] Philip Harris,[b] Simone Marzani,[a] Salvatore Rappoccio,[a] and Nhan Tran[c]

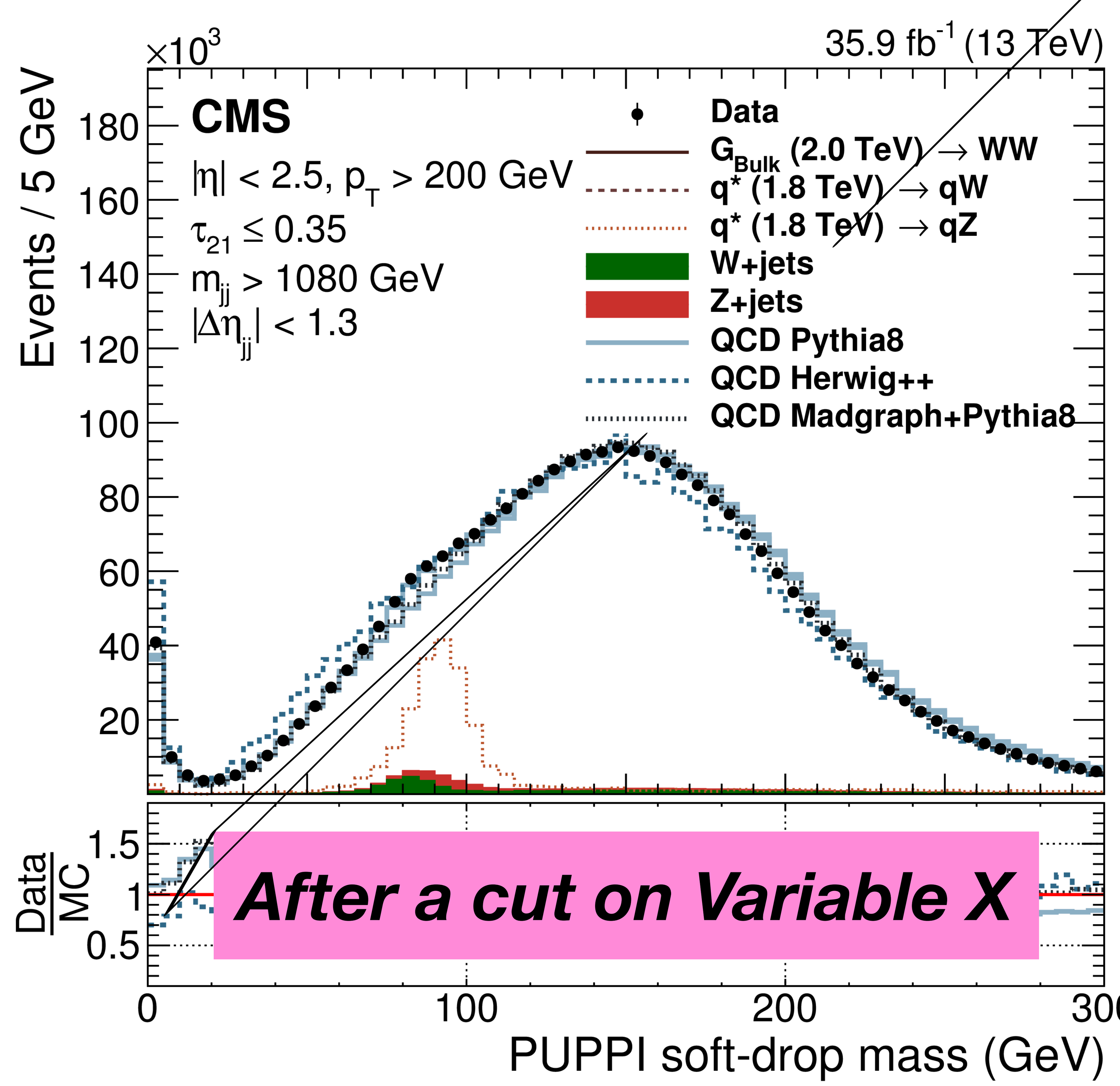
[a] *University at Buffalo, The State University of New York, Buffalo, NY 14260-1500, USA*

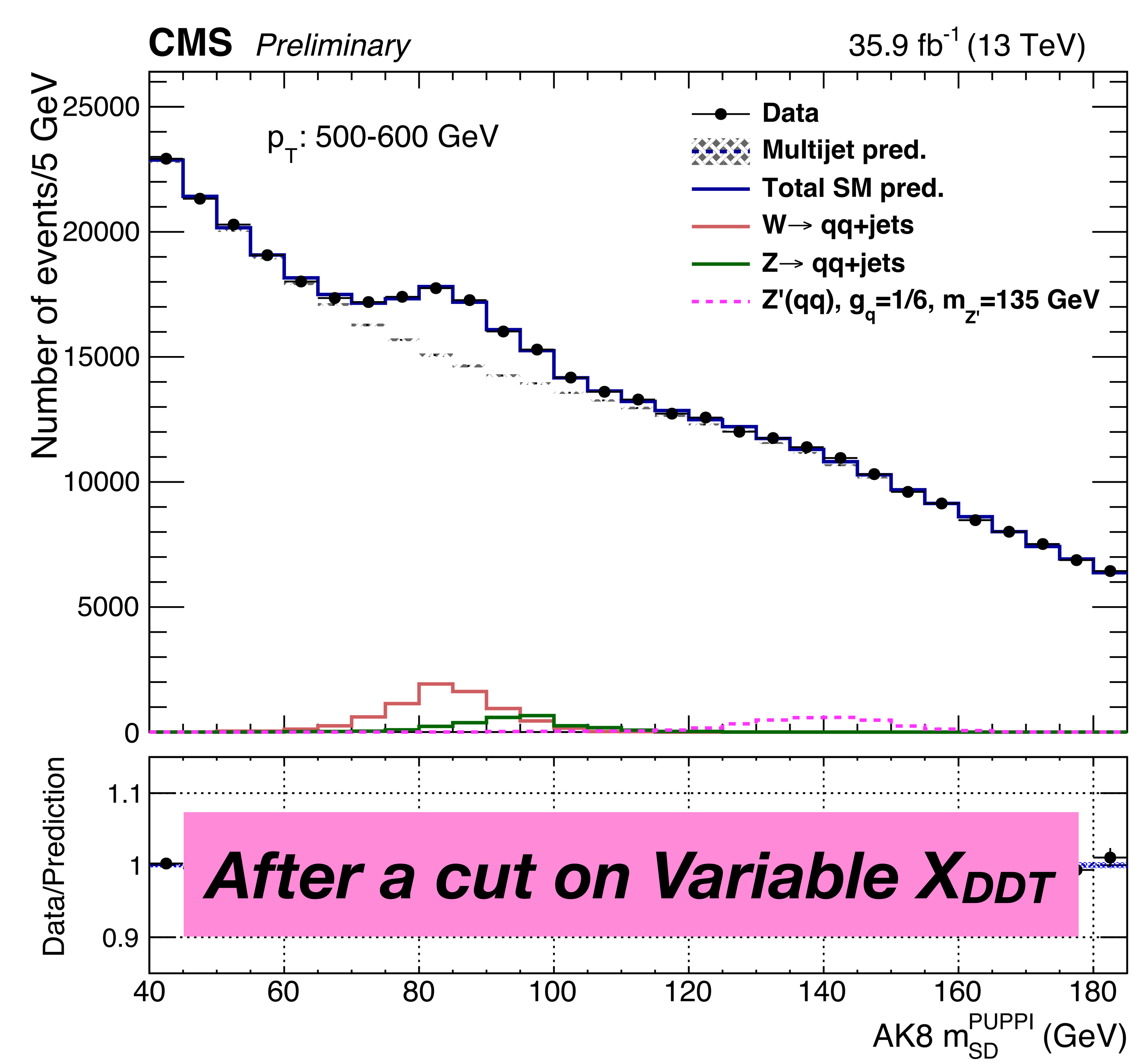
[b] *CERN, European Organization for Nuclear Research, Geneva, Switzerland*

[c] *Fermi National Accelerator Laboratory (FNAL), Batavia, IL 60510, USA*



Where is the W peak?



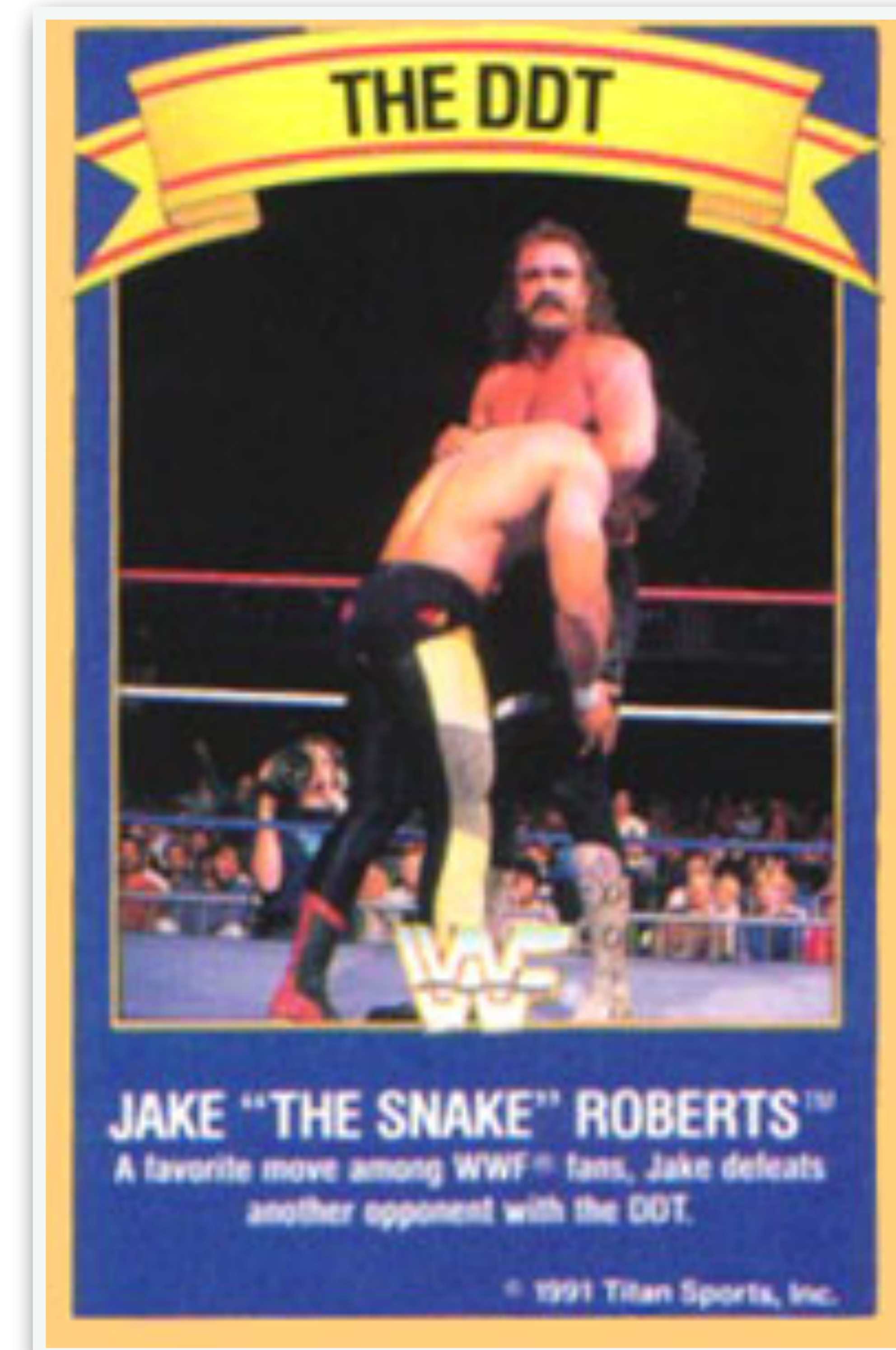


DE-CORRELATIONS

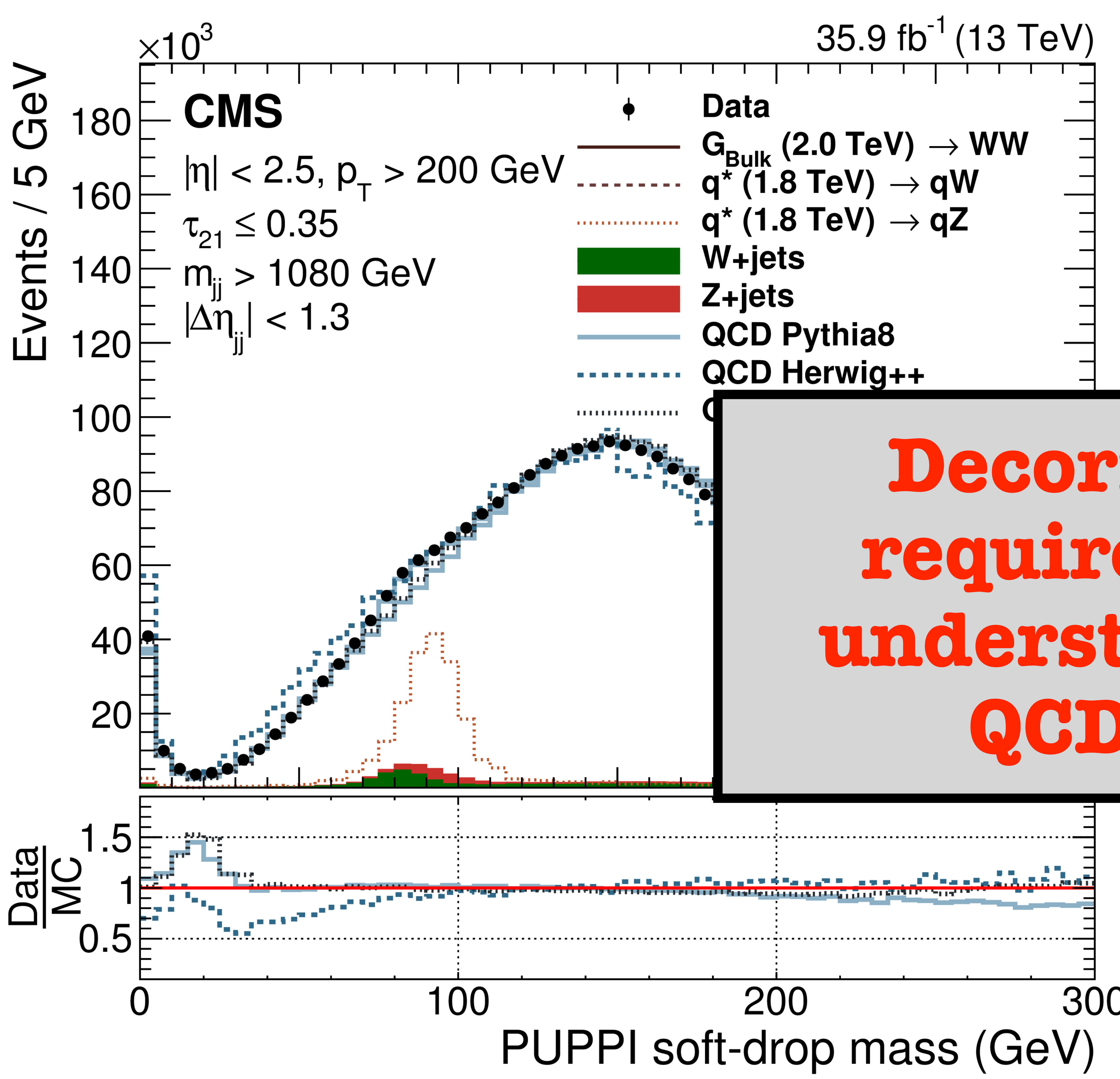
Thinking outside the ROCs: Designing Decorrelated Taggers (DDT) for jet substructure

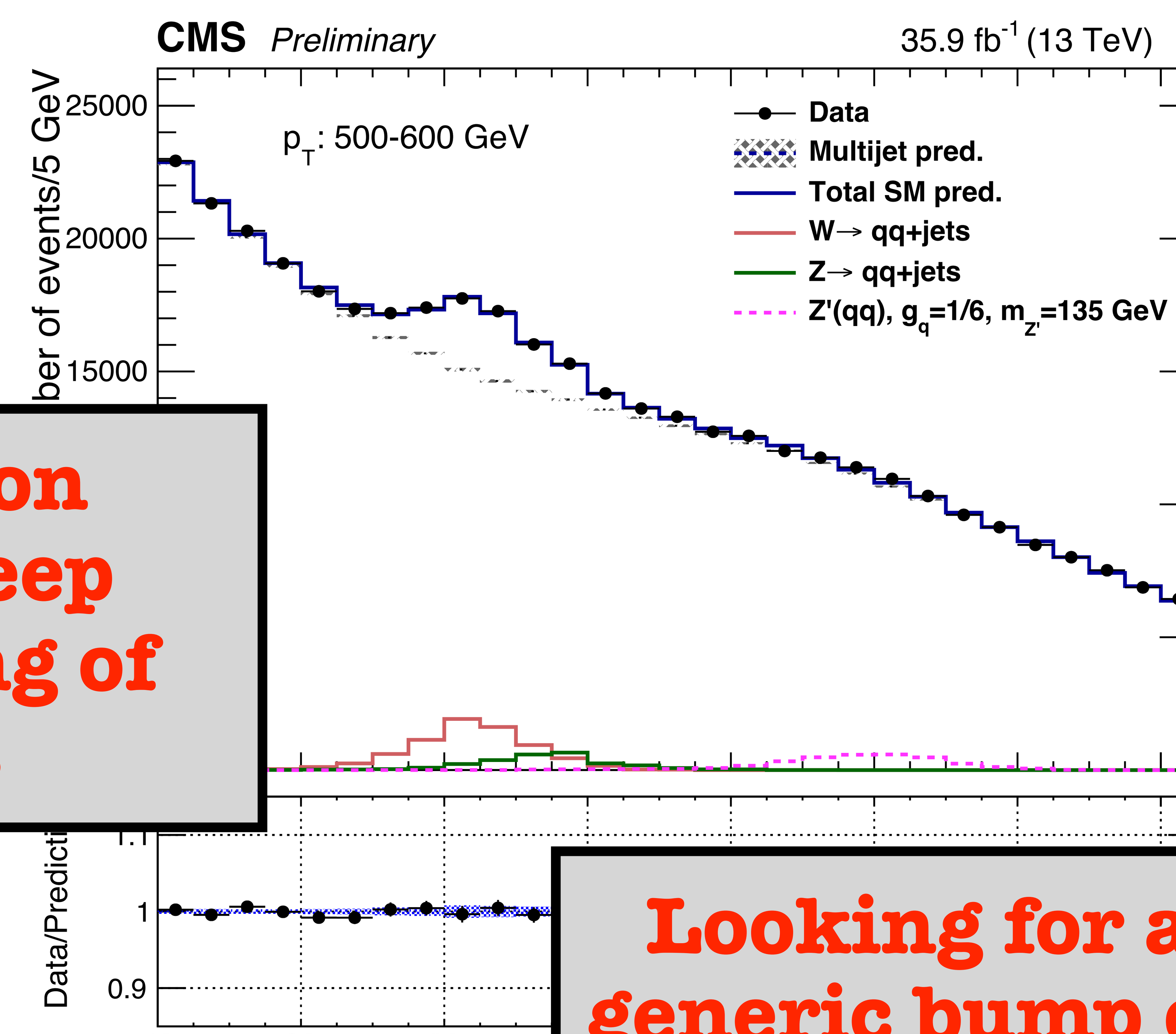
James Dolen,[a] Philip Harris,[b] Simone Marzani,[a] Salvatore Rappoccio,[a] and Nhan Tran[c]

[a] *University at Buffalo, The State University of New York, Buffalo, NY 14260-1500, USA*
[b] *CERN, European Organization for Nuclear Research, Geneva, Switzerland*
[c] *Fermi National Accelerator Laboratory (FNAL), Batavia, IL 60510, USA*



Where is the W peak?

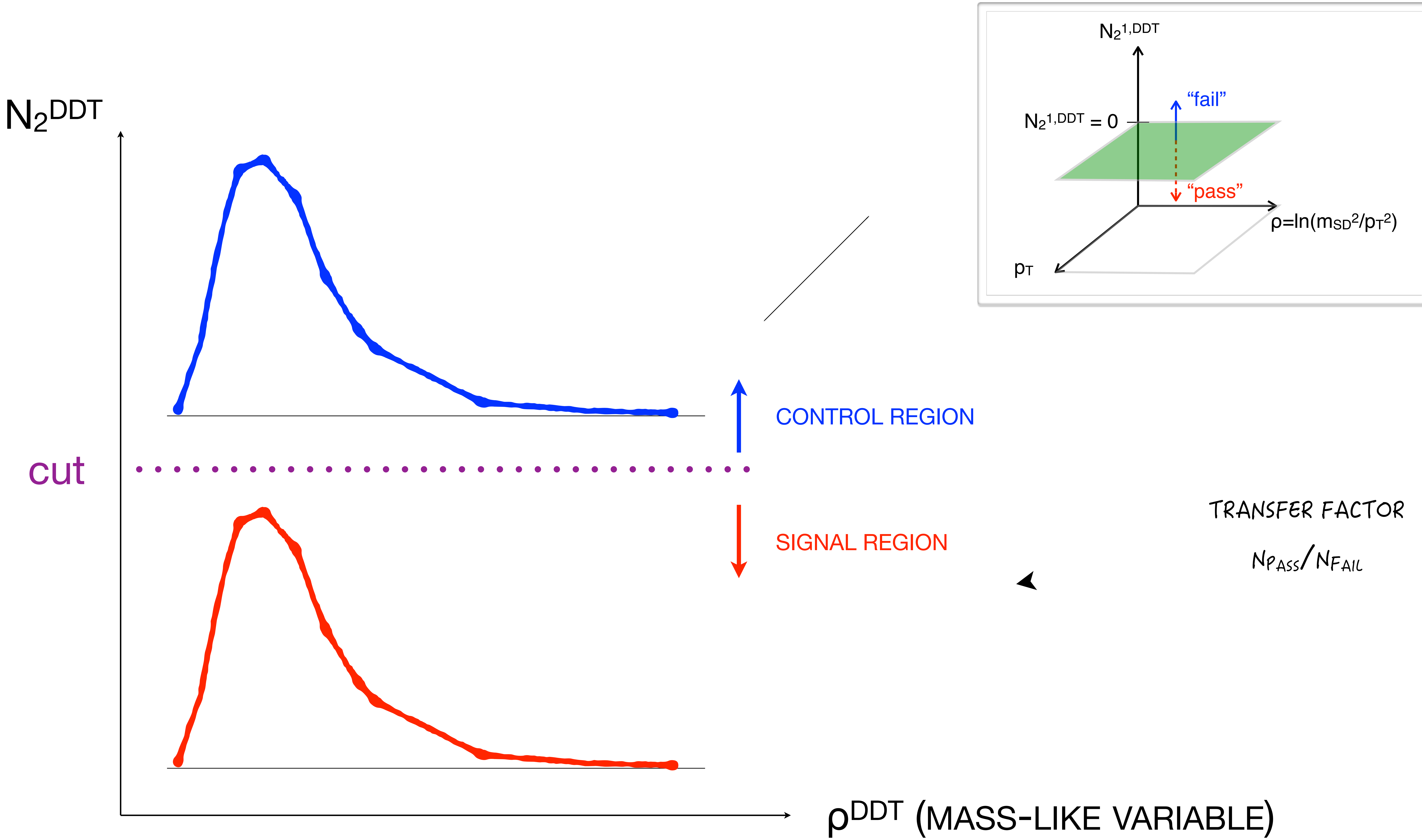




Decorrelation requires a deep understanding of QCD mass

Looking for a Z', a generic bump on this mass spectrum

DATA-DRIVEN SCHEMES



JET SUBSTRUCTURE MINI-SUMMARY

Task: classify different type of jets based on radiation patterns and secondary vertex information

Many “expert features” invented to discriminate between various types of jets

Dealing with modeling challenges:

- Standard Candles give a handle on signal systematics
- Sidebands and control regions for data-driven backgrounds; requires a detailed understanding of feature correlations

rise of the machines

RISE OF THE MACHINES

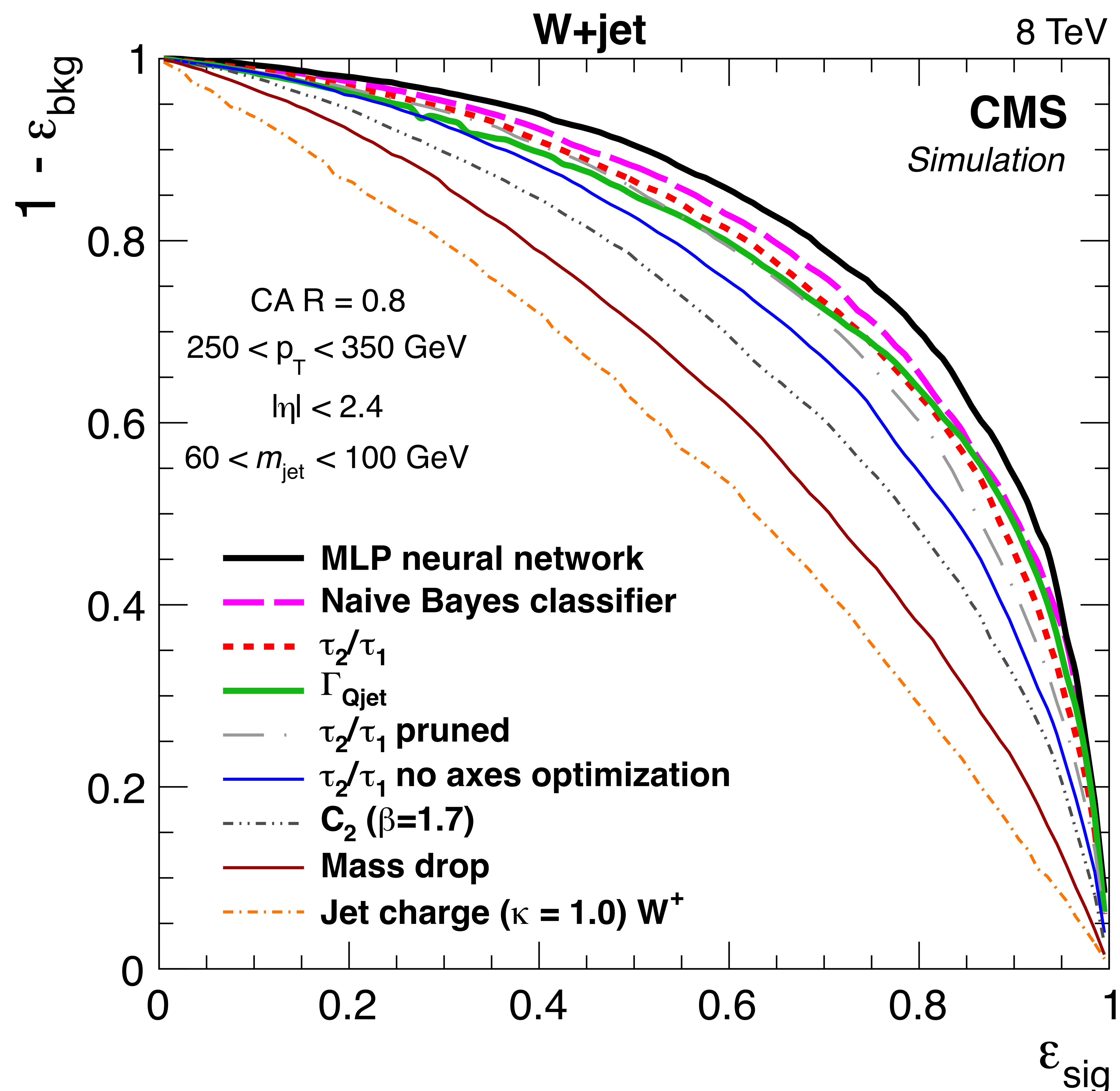
An obvious thing to do:

Put all the expert features into a multivariate classifier

Observe an improvement in classifier performance, O(10-20%) increase in background rejection

“Simple” machine learning, typically shallow networks or BDTs

2013!



Machine learning in jet substructure

Deep learning taking off in recent years…

arXiv:1709.04464

Jet Substructure at the Large Hadron Collider:
A Review of Recent Advances in Theory and Machine Learning

Andrew J. Larkoski*
Physics Department, Reed College, Portland, OR 97202, USA

Ian Moult†
Berkeley Center for Theoretical Physics, University of California, Berkeley, CA 94720, USA and Theoretical Physics Group, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

Benjamin Nachman‡
Physics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA
(Dated: September 15, 2017)

Incomplete collection of approaches:

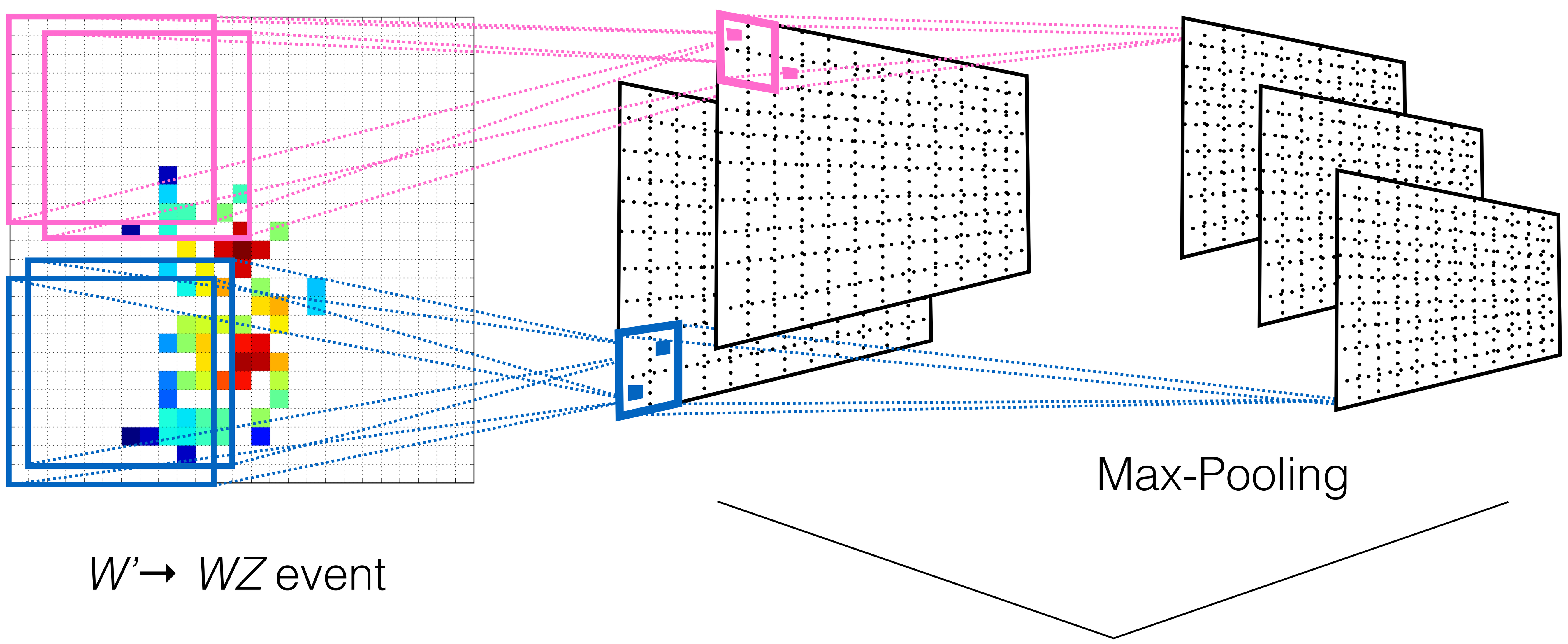
CNN, RNN, Unsupervised, GANs, Regression

Partial list, focused on recent results

[472] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, JHEP **07**, 069 (2016), arXiv:1511.05190 [hep-ph].
[473] ATLAS Collaboration, *Identification of Hadronically-Decaying W Bosons and Top Quarks Using High-Level Features as Input to Boosted Decision Trees and Deep Neural Networks in ATLAS at $\sqrt{s}$ = 13 TeV*, ATLAS PUB Note ATL-PHYS-PUB-2017-004 (2017).
[474] ATLAS Collaboration, *Optimisation and performance studies of the ATLAS b-tagging algorithms for the 2017-18 LHC run*, ATLAS PUB Note ATL-PHYS-PUB-2017-013 (2017).
[475] ATLAS Collaboration, ATLAS-CONF-2017-064 (2017).
[476] J. Cogan, M. Kagan, E. Strauss, and A. Schwartzman, JHEP **02**, 118 (2015), arXiv:1407.5675 [hep-ph].
[477] L. G. Almeida, M. Backović, M. Cliche, S. J. Lee, and M. Perelstein, JHEP **07**, 086 (2015), arXiv:1501.05968 [hep-ph].
[478] P. Baldi, K. Bauer, C. Eng, P. Sadowski, and D. Whiteson, Phys. Rev. **D93**, 094034 (2016), arXiv:1603.09349 [hep-ex].
[479] J. Barnard, E. N. Dawe, M. J. Dolan, and N. Rajcic, Phys. Rev. **D95**, 014018 (2017), arXiv:1609.00607 [hep-ph].
[480] G. Kasieczka, T. Plehn, M. Russell, and T. Schell, (2017), arXiv:1701.08784 [hep-ph].
[481] L. de Oliveira, M. Paganini, and B. Nachman, (2017), arXiv:1701.05927 [stat.ML].
[482] ATLAS Collaboration, ATL-PHYS-PUB-2017-017 (2017).
[483] CMS Collaboration, *New Developments for Jet Substructure Reconstruction in CMS*, CMS Detector Performance Summary (2017).
[484] CMS Collaboration, *Heavy flavor identification at CMS with deep neural networks*, CMS Detector Performance Summary CMS-DP-2017-005 (2017).
[485] CMS Collaboration, *CMS Phase 1 heavy flavour identification performance and developments*, CMS Detector Performance Summary CMS-DP-2017-013 (2017).
[486] J. Pearkes, W. Fedorko, A. Lister, and C. Gay, (2017), arXiv:1704.02124 [hep-ex].
[487] A. Butter, G. Kasieczka, T. Plehn, and M. Russell, (2017), arXiv:1707.08966 [hep-ph].
[488] K. Datta and A. Larkoski, JHEP **06**, 073 (2017), arXiv:1704.08249 [hep-ph].
[489] J. A. Aguilar-Saavedra, J. H. Collins, and R. K. Mishra, (2017), arXiv:1709.01087 [hep-ph].
[490] D. Guest, J. Collado, P. Baldi, S.-C. Hsu, G. Urban, and D. Whiteson, Phys. Rev. **D94**, 112002 (2016), arXiv:1607.08633 [hep-ex].
[491] ATLAS Collaboration, *Identification of Jets Containing b-Hadrons with Recurrent Neural Networks at the ATLAS Experiment*, ATLAS PUB Note ATL-PHYS-PUB-2017-003 (2017).
[492] G. Louppe, K. Cho, C. Becot, and K. Cranmer, (2017), arXiv:1702.00748 [hep-ph].
[493] L. M. Dery, B. Nachman, F. Rubbo, and A. Schwartzman, JHEP **05**, 145 (2017), arXiv:1702.00414 [hep-ph].
[494] T. Cohen, M. Freytsis, and B. Ostdiek, (2017), arXiv:1706.09451 [hep-ph].
[495] E. M. Metodiev, B. Nachman, and J. Thaler, (2017), arXiv:1708.02949 [hep-ph].
[496] G. Louppe, M. Kagan, and K. Cranmer, (2016), arXiv:1611.01046 [stat.ME].
[497] I. J. Goodfellow *et al.*, (2014), 1406.2661.
[498] S. Agostinelli *et al.* (GEANT4), Nucl. Instrum. Meth. **A506**, 250 (2003).
[499] ATLAS Collaboration, Eur. Phys. J. **C77**, 490 (2017), arXiv:1603.02934 [hep-ex].
[500] M. Paganini, L. de Oliveira, and B. Nachman, (2017), arXiv:1705.02355 [hep-ex].
[501] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, (2017), arXiv:1707.08600 [hep-ph].
[502] M. Cacciari, G. P. Salam, and G. Soyez, Eur. Phys. J. **C75**, 59 (2015), arXiv:1407.0408 [hep-ph].
[503] D. Bertolini, P. Harris, M. Low, and N. Tran, JHEP **10**, 059 (2014), arXiv:1407.6013 [hep-ph].
[504] D. Krohn, M. D. Schwartz, M. Low, and L.-T. Wang, Phys. Rev. **D90**, 065020 (2014), arXiv:1309.4777 [hep-ph].

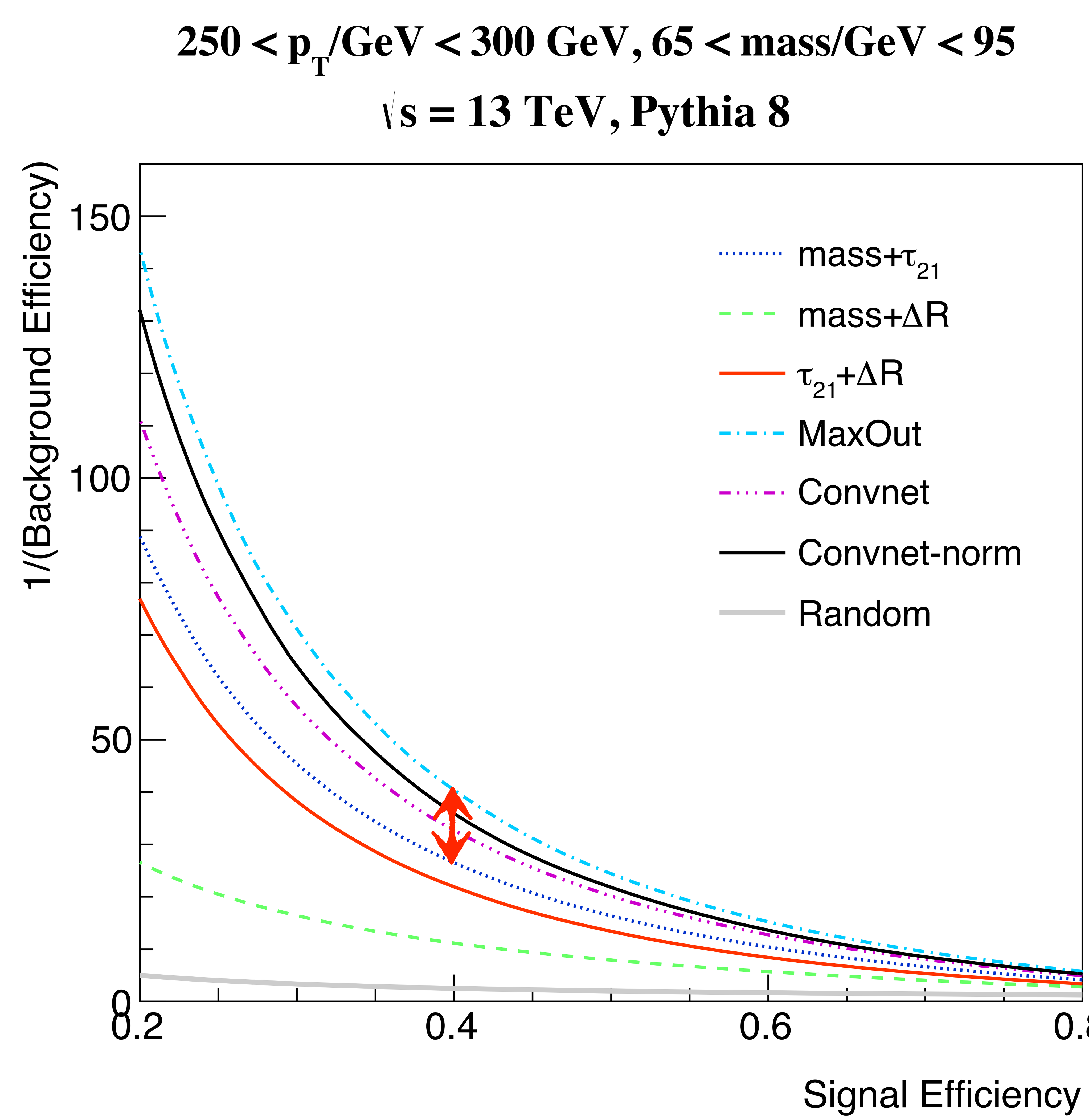
JET IMAGES

Oliveira, Kagan, Mackey, Machman, Schwartzman
arXiv:1511.05190



An early study of convolution neural networks in jet substructure

Special care in discussing preprocessing of information

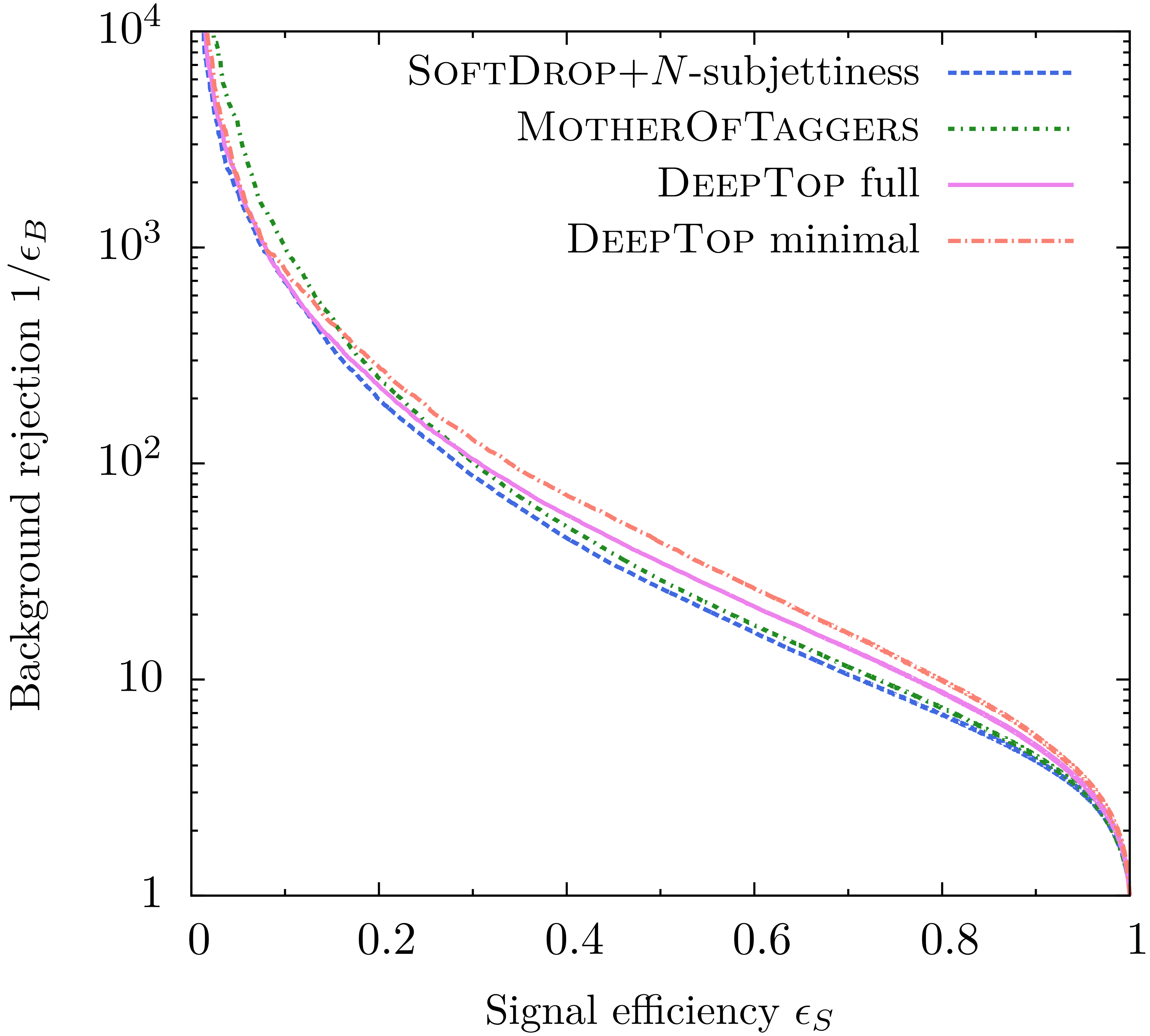


Comparison of performance of expert features vs. CNN performance

Performance gain over 2-variable expert features

ANOTHER CNN STUDY

Kasieczka, Plehn, Russel, Schell
arXiv:1701.08784



2 expert features

9 expert features

CNN, w/preprocessing

CNN, minimal preprocessing

A SAMPLING

A selection of some other ideas I wanted to highlight

QCD-Aware Recursive Neural Networks for Jet Physics

Gilles Louppe,[1] Kyunghyun Cho,[1] Cyril Becot,[1] and Kyle Cranmer[1]
[1] *New York University*

arXiv:1702.00748

Variable length inputs for jet substructure

Deep-learned Top Tagging with a Lorentz Layer

Anja Butter[1], Gregor Kasieczka[2], Tilman Plehn[1], and Michael Russell[1,3]

1 Institut für Theoretische Physik, Universität Heidelberg, Germany
2 Institute for Particle Physics, ETH Zürich, Switzerland
3 School of Physics and Astronomy, University of Glasgow, Scotland
plehn@uni-heidelberg.de

arXiv:1707.08966

A cute paper that had the network learn the Minkowski metric

Classification without labels: Learning from mixed samples in high energy physics

arXiv:1708.02949

Eric M. Metodiev,[a] Benjamin Nachman,[b] and Jesse Thaler[a]

Weakly supervised learning (CWoLa)

How Much Information is in a Jet?

Kaustuv Datta and Andrew Larkoski

arXiv:1704.08249

Building a complete analytic basis for the jet

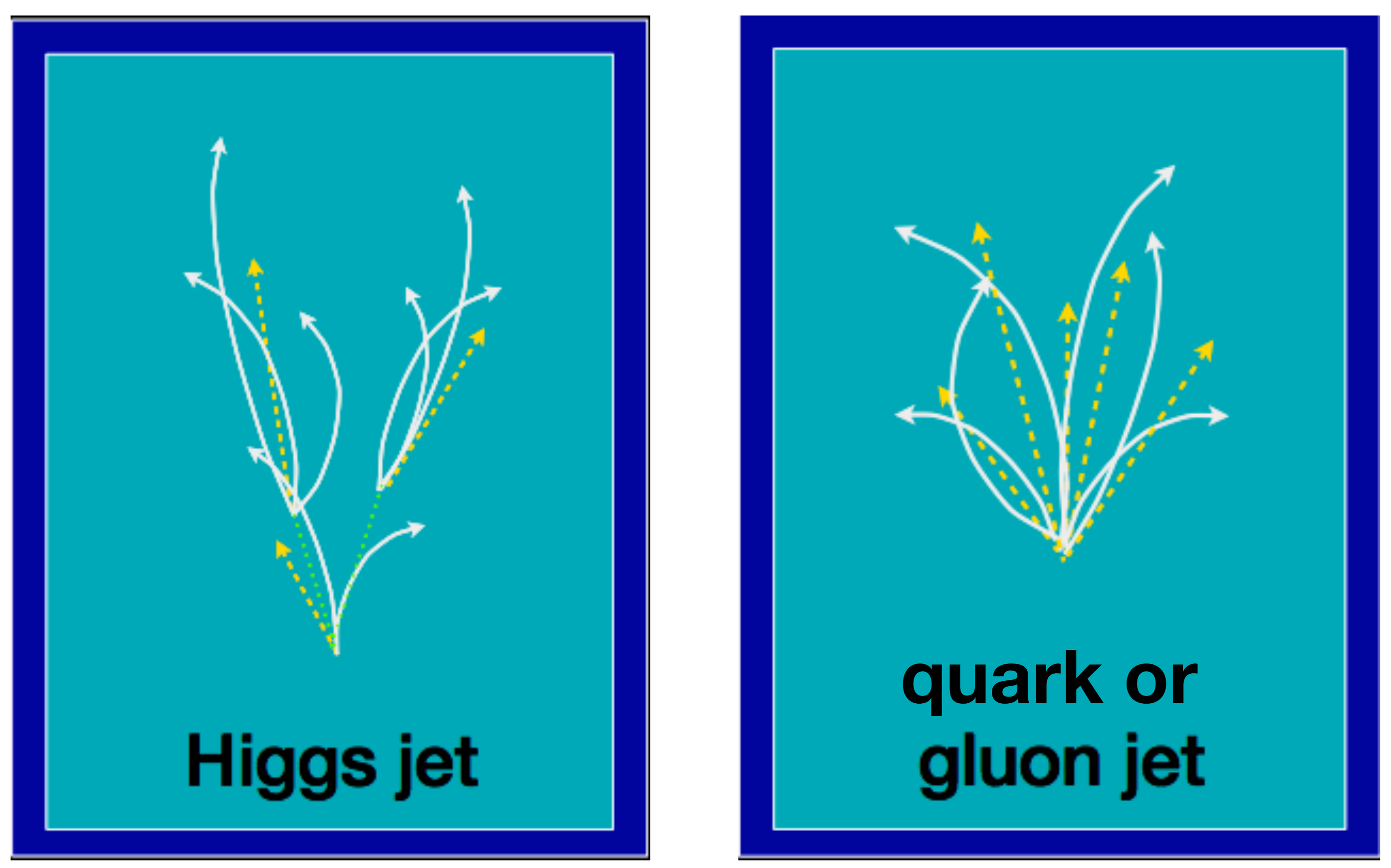
Decorrelated Jet Substructure Tagging using Adversarial Neural Networks

arXiv:1703.03507: Shimmin, Sadowski, Baldi, Weik, Whiteson, Goul, Sogaard

Decorrelations! (see next slide)

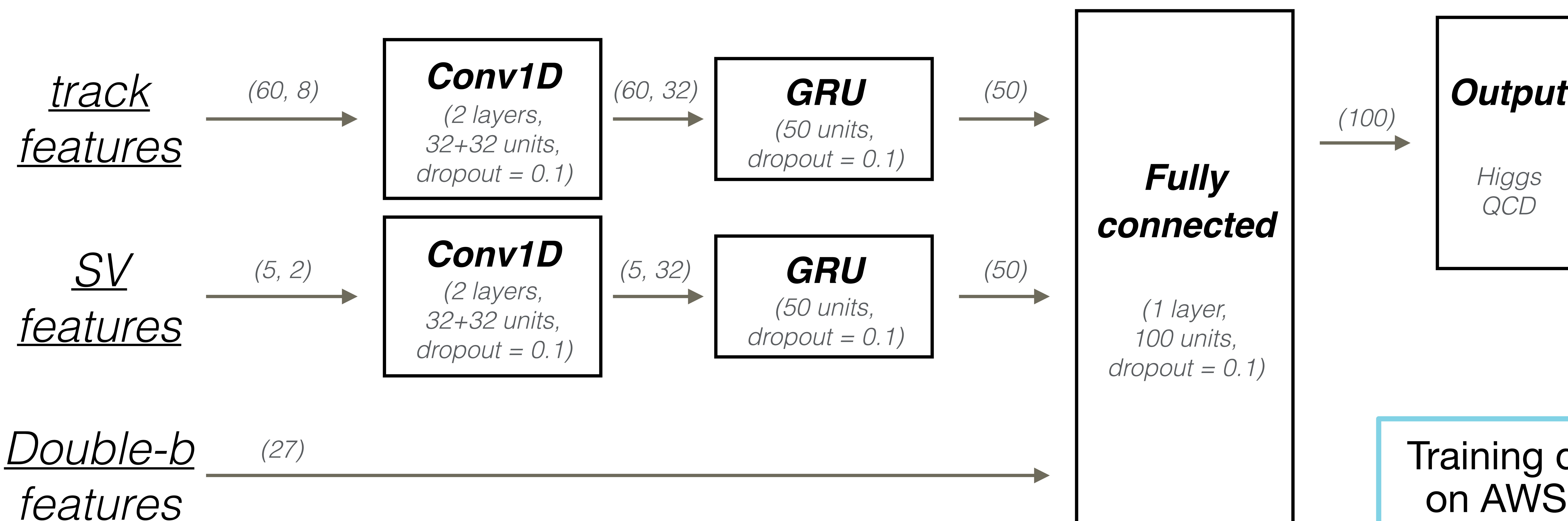
Javier Duarte (FNAL),
Caterina Vernieri (FNAL)

A concrete example to think about: Higgs tagging



Train a network to distinguish a Higgs(bb) jet from a quark/gluon jet using secondary vertex information

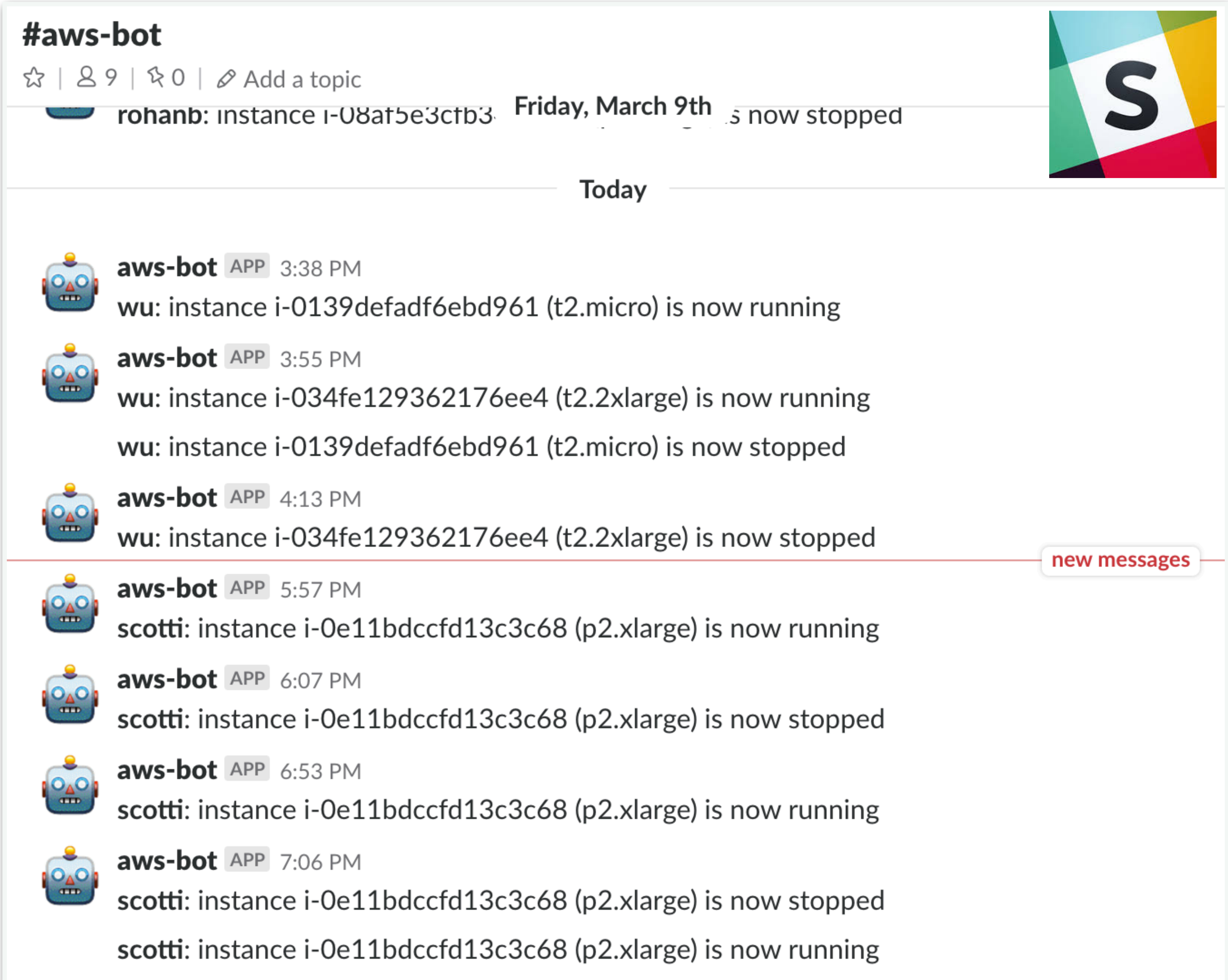
On the face of it, this is *not* a substructure tagger, only using tracking information



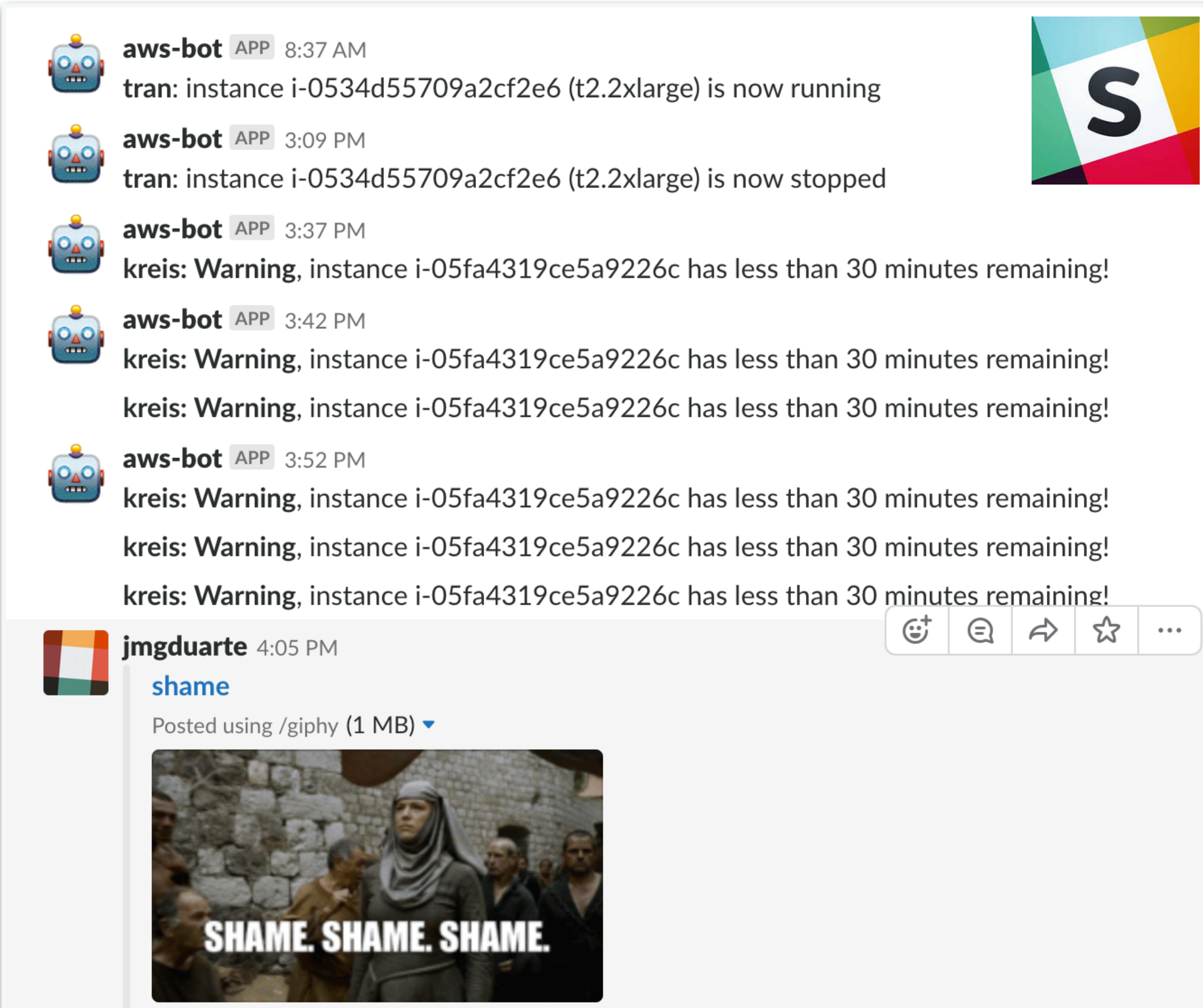
Training done on AWS P2 GPU resources

GPU MONITORING

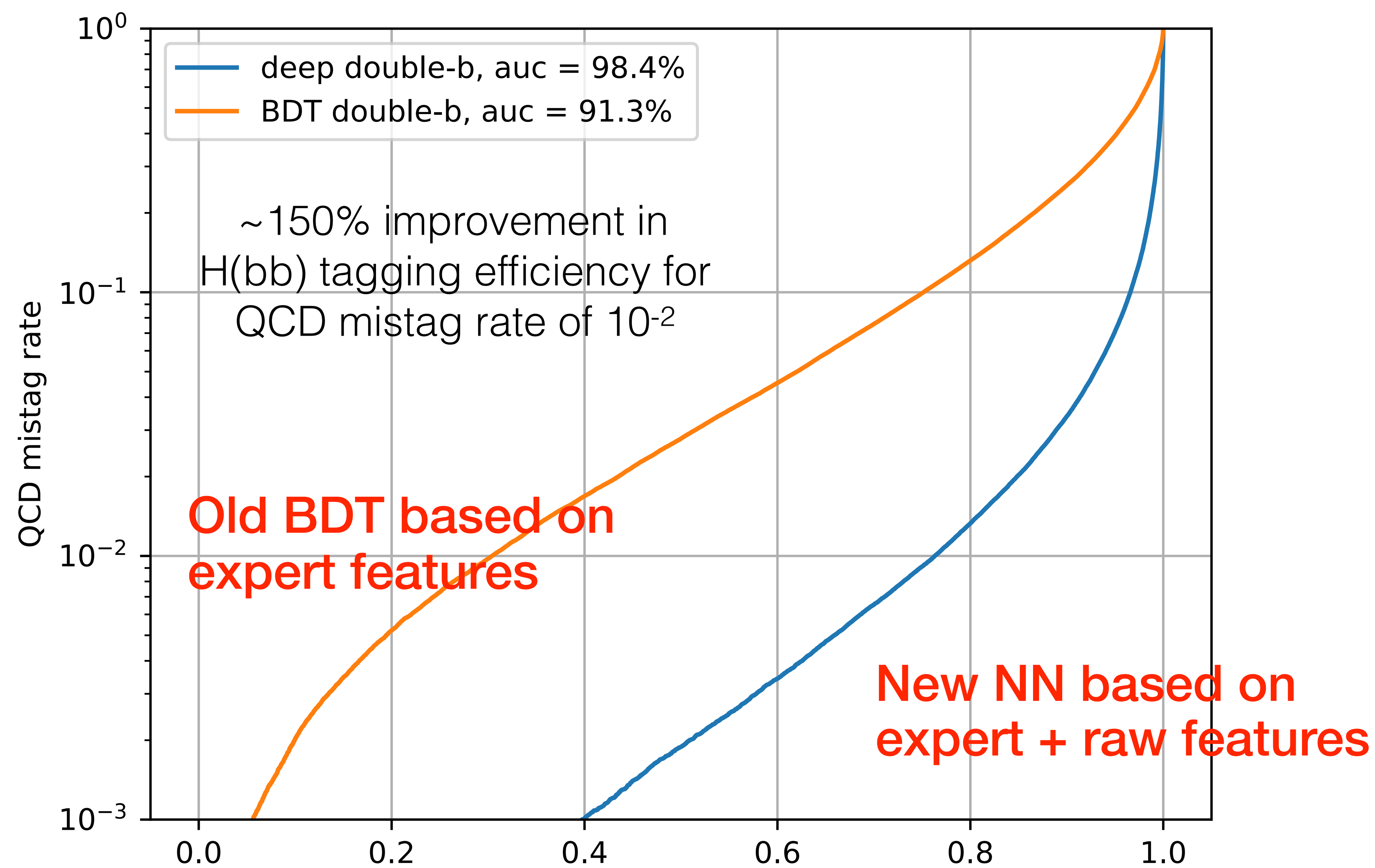
Burt Holzmann, FNAL



GPU MONITORING



EXAMPLE: ML AND CORRELATIONS

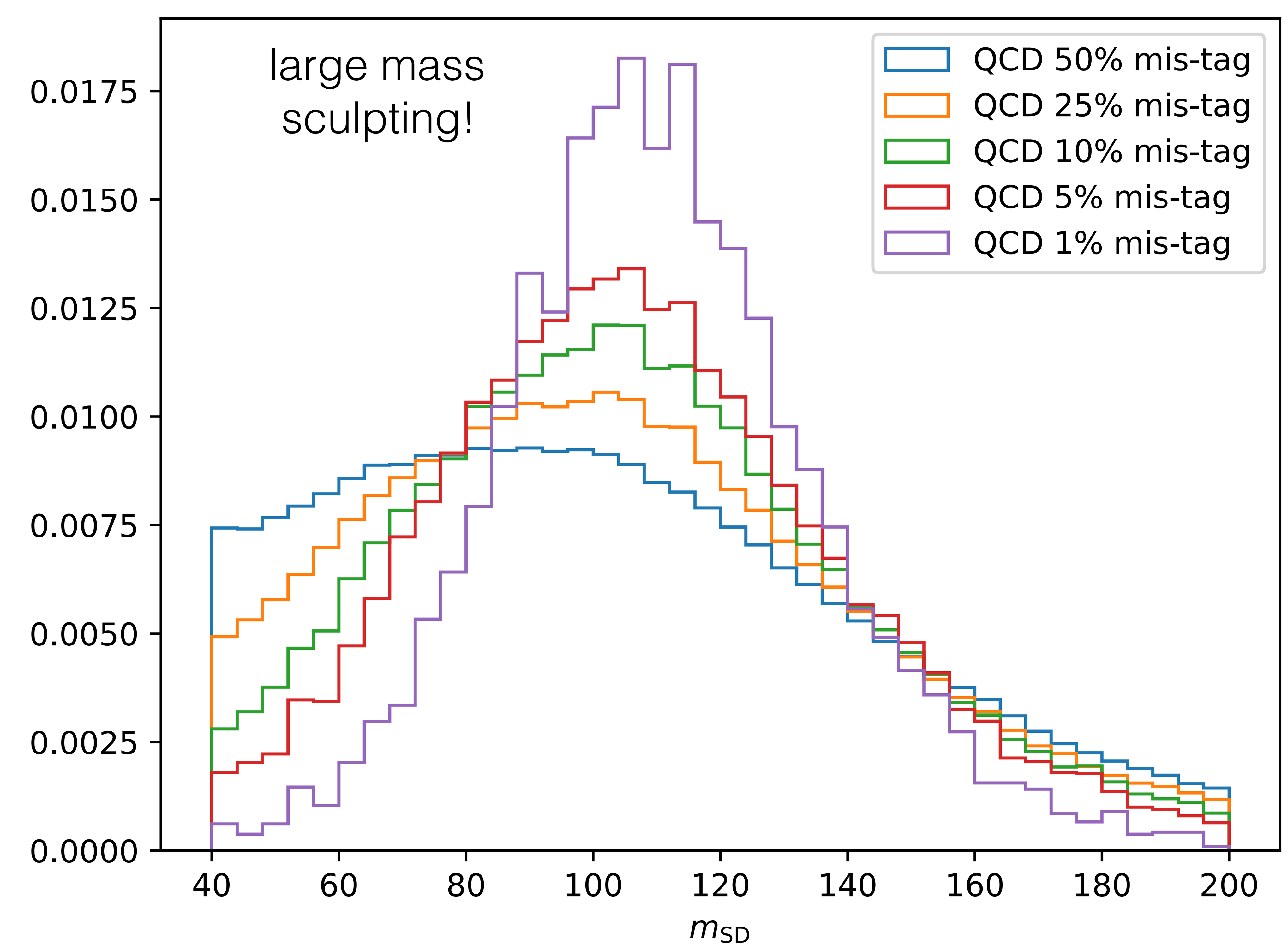


Big improvement in the performance coming from a new neural network!

But…umm…

Crud, it looks like the tagger learned the Higgs jet mass!

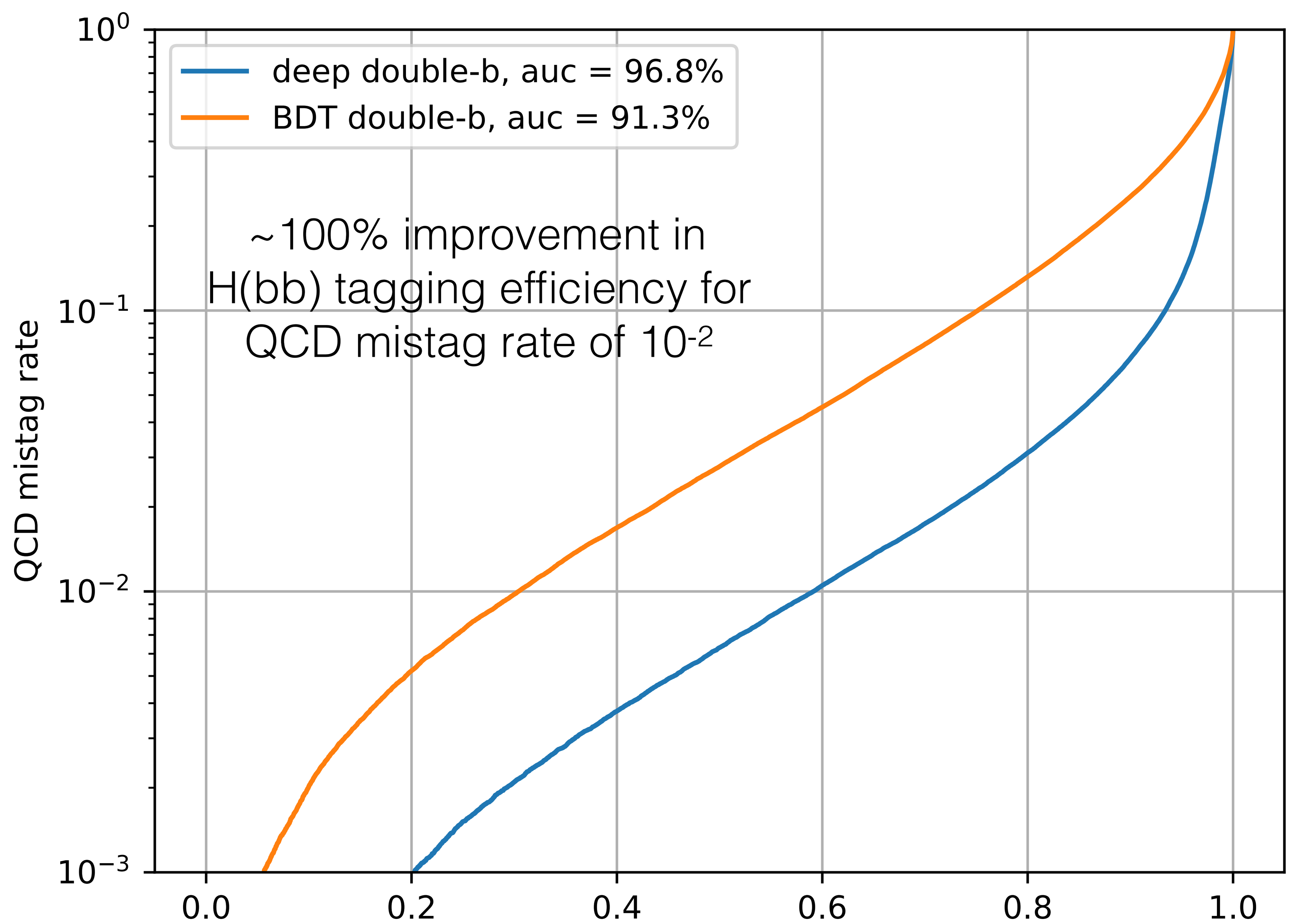
QCD background is sculpted to look like the Higgs jet signal

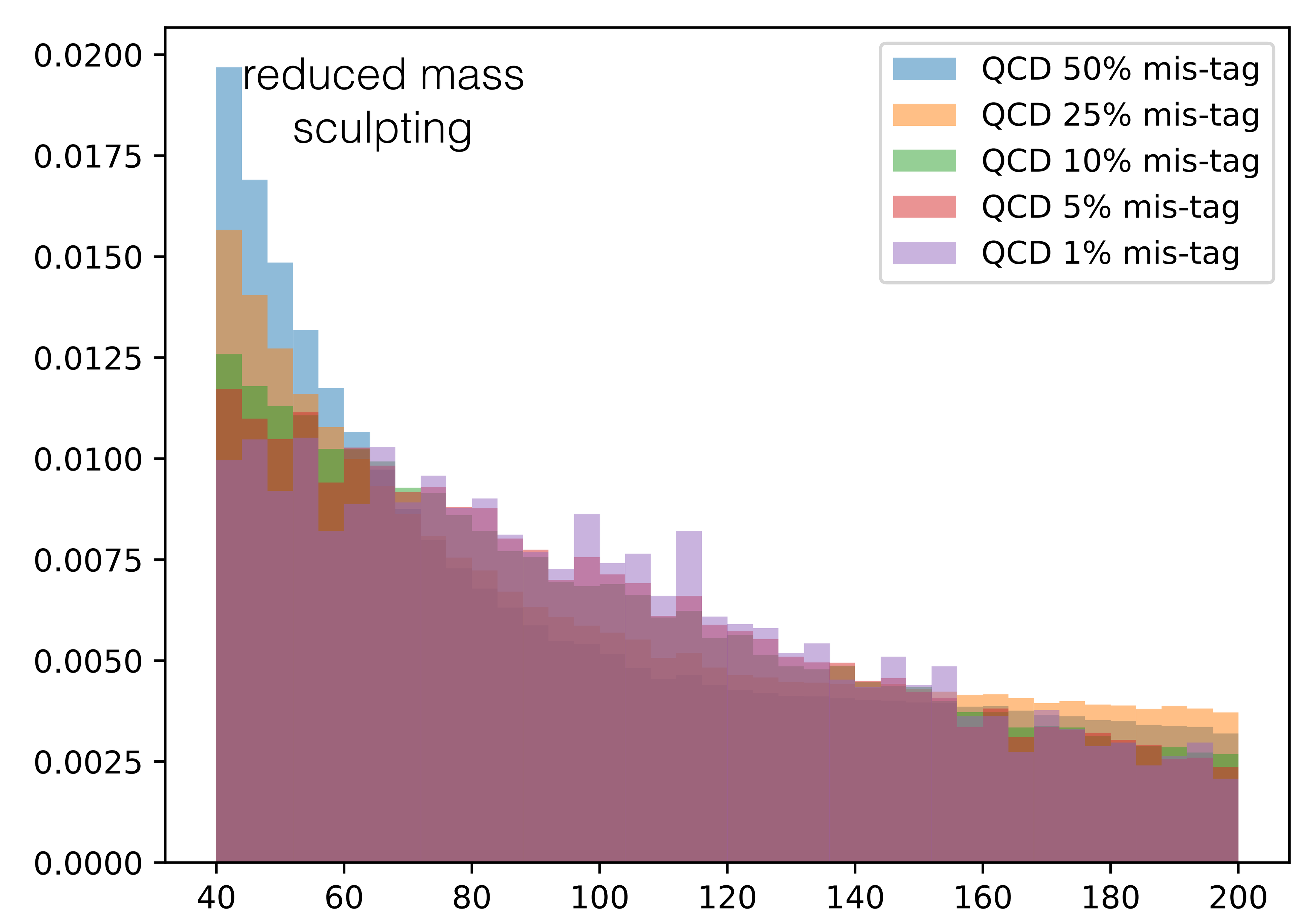


EXAMPLE: ML AND CORRELATIONS

How to decorrelate the secondary vertex tagger from the mass in some machine learning algorithm?

- Some early work into principle component analysis
- Pointed to a first paper in physics using Adversarial Networks
- Another approach here using modified loss functions (J. Duarte et al)
  - Categorical cross-entropy loss function with additional mass-binned Kullback-Leibler term for mass sculpting





Lessons learned

Did my ML algorithm learn too much?

- Learning specific modeling differences in the MC
- Sculpting backgrounds to look like signal
- Throwing away interesting anomalous signals

Did my ML algorithm learn enough?

- Did I give it enough information to learn all the physics?

What did my ML algorithm learn?

Always a tricky discussion

Lessons Learned

Did my ML algorithm learn too much?

Learning specific modeling differences in the MC
Sculpting backgrounds to look like signal
Throwing away interesting anomalous signals

Did my ML algorithm learn enough?

Did I give it enough information to learn all the physics?

What did my ML algorithm learn?

Always a tricky discussion

Is my NN sculpting the neutrino energy?
Other hidden dependencies?

Is granularity sufficient for the image?
What about other orthogonal info, like time?

Throwing out any exotic interesting physics signatures?

LESSONS LEARNED

Did my ML algorithm learn too much?

Learning specific modeling differences in the MC
Sculpting backgrounds to look like signal
Throwing away interesting anomalous signals

Did my ML algorithm learn enough?

Did I give it enough information to learn all the physics?

What did my ML algorithm learn?

Always a tricky discussion

My personal take away:

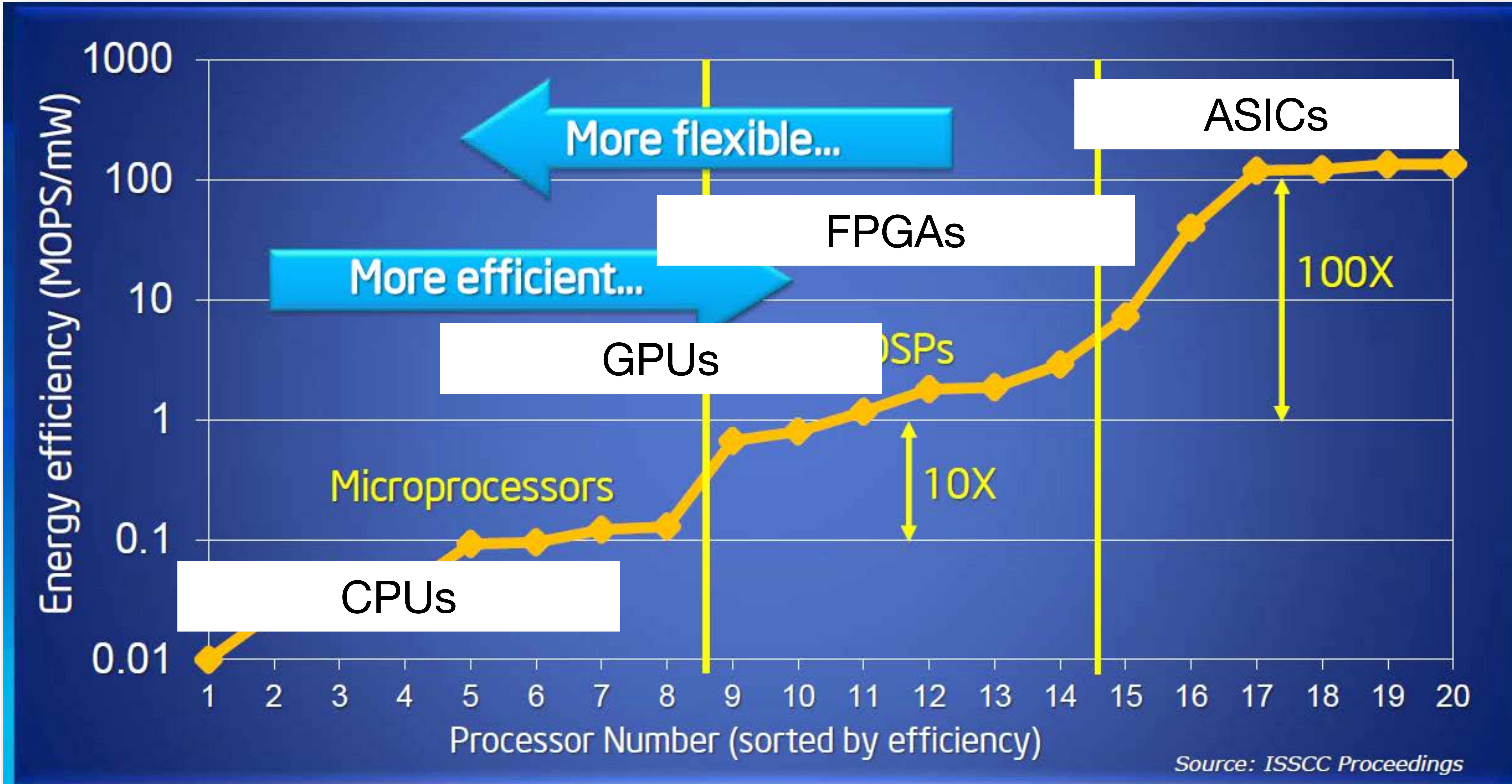
Provided it's well-understood,
we should use the best (performance & speed) algorithm

A good way to develop understanding is to have a suite of performant expert features (to understand complete information content and correlations); we are physicists after all

the fast and the furious

(the reason I personally like machine learning)

COMPUTING ARCHITECTURES



Source: Bob Broderson, Berkeley Wireless group

* GPUs still best option for training
* FPGAs generally much more power efficient

The trigger @ ATLAS and CMS

Goal:
reduce event rate from 40MHz to 500 Hz

How: multi-tier system

custom hardware (“L1”)

latency, O(μs)

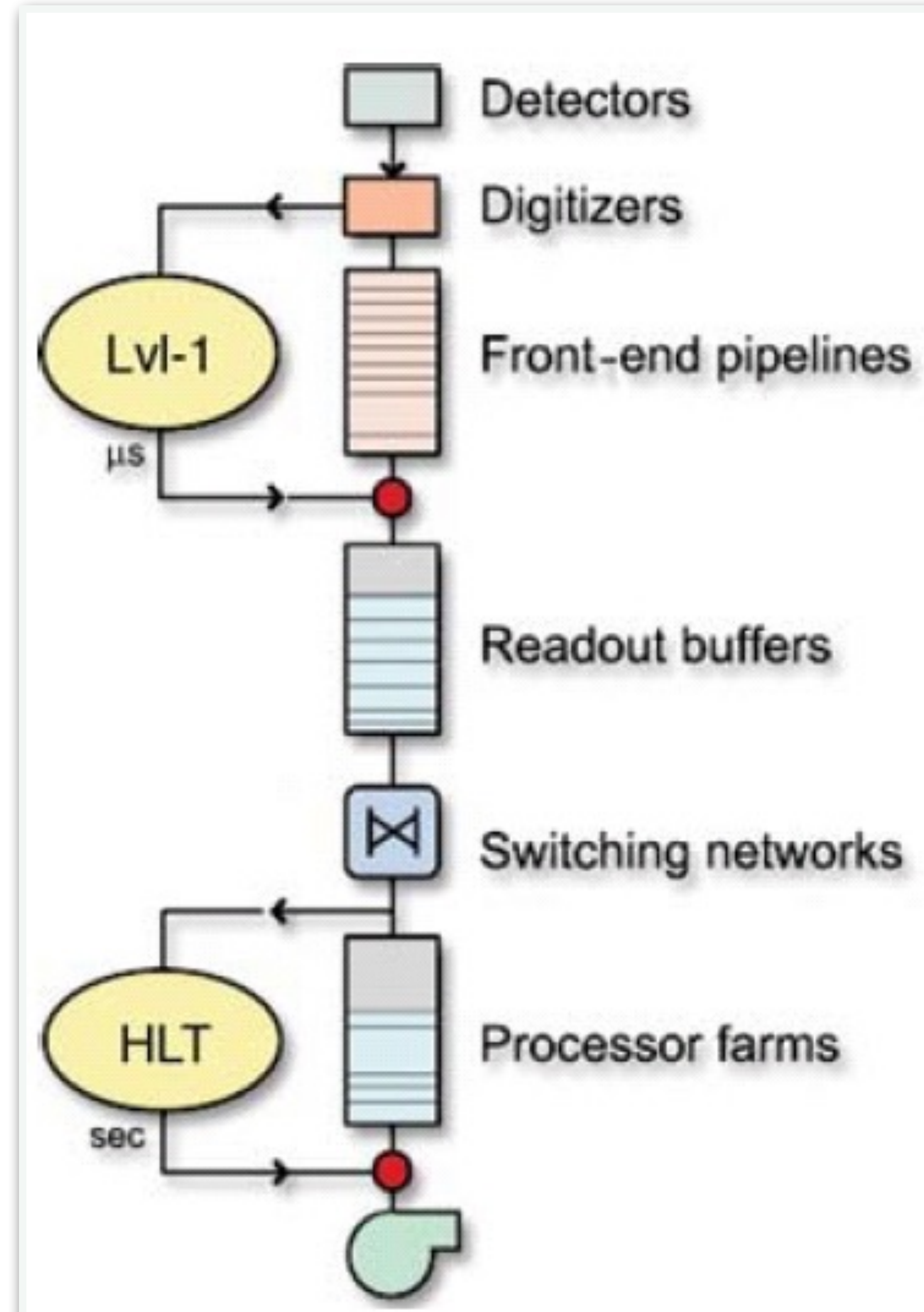
rate in/out: 40 MHz / 100 KHz

computing farm (“HLT”)

latency, O(100 ms)

rate in/out: 100 KHz / 500 Hz

n.b. all numbers approximate



For HL-LHC upgrade: latency and output rates go up ~5

THE TRIGGER @ ATLAS AND CMS

Goal:
reduce event rate from 40MHz to 500 Hz

How: multi-tier system

custom hardware (“L1”)

latency, O(μs)

rate in/out: 40 MHz / 100 KHz

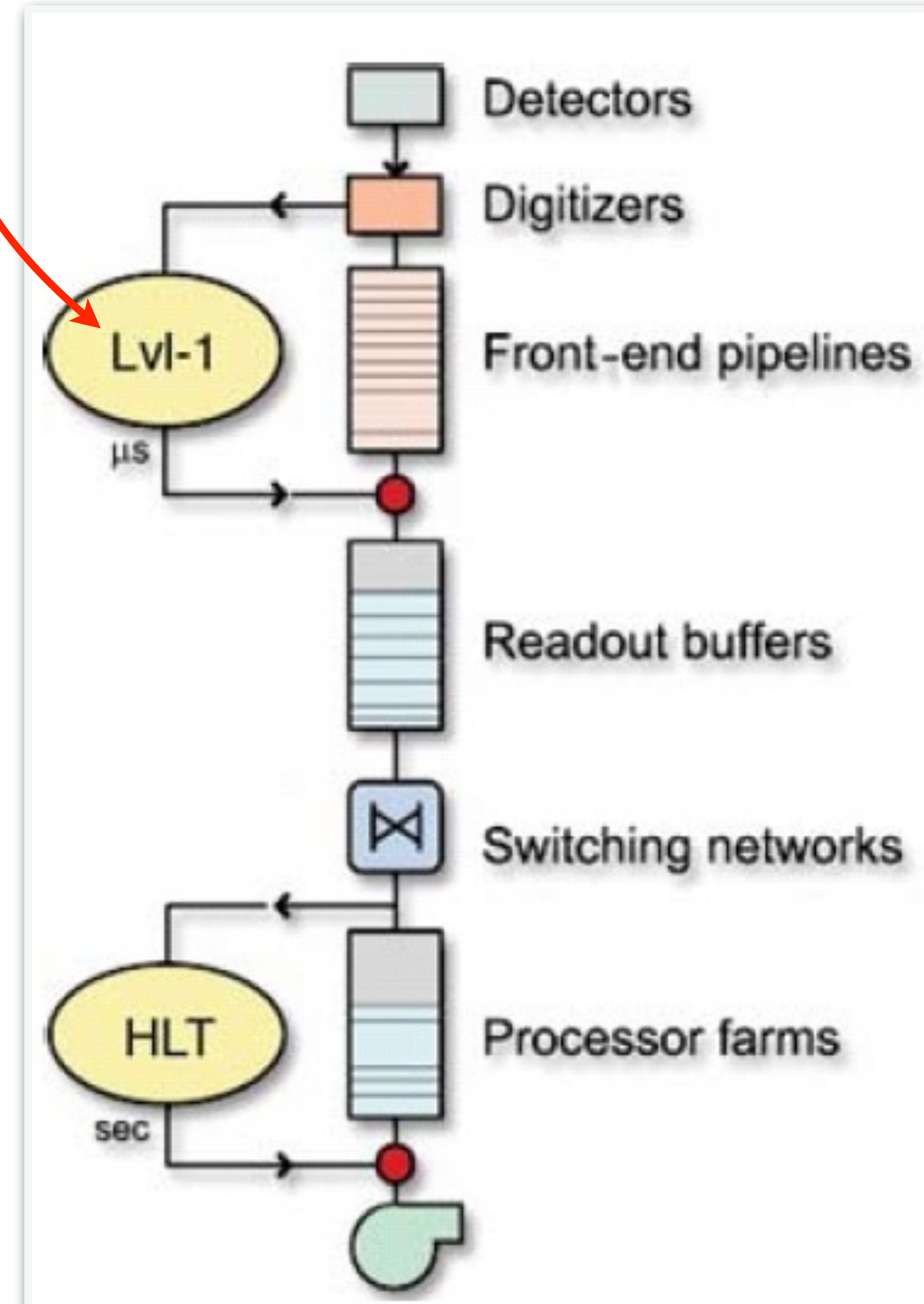
computing farm (“HLT”)

latency, O(100 ms)

rate in/out: 100 KHz / 500 Hz

n.b. all numbers approximate

Latencies necessitate all-FPGA design!

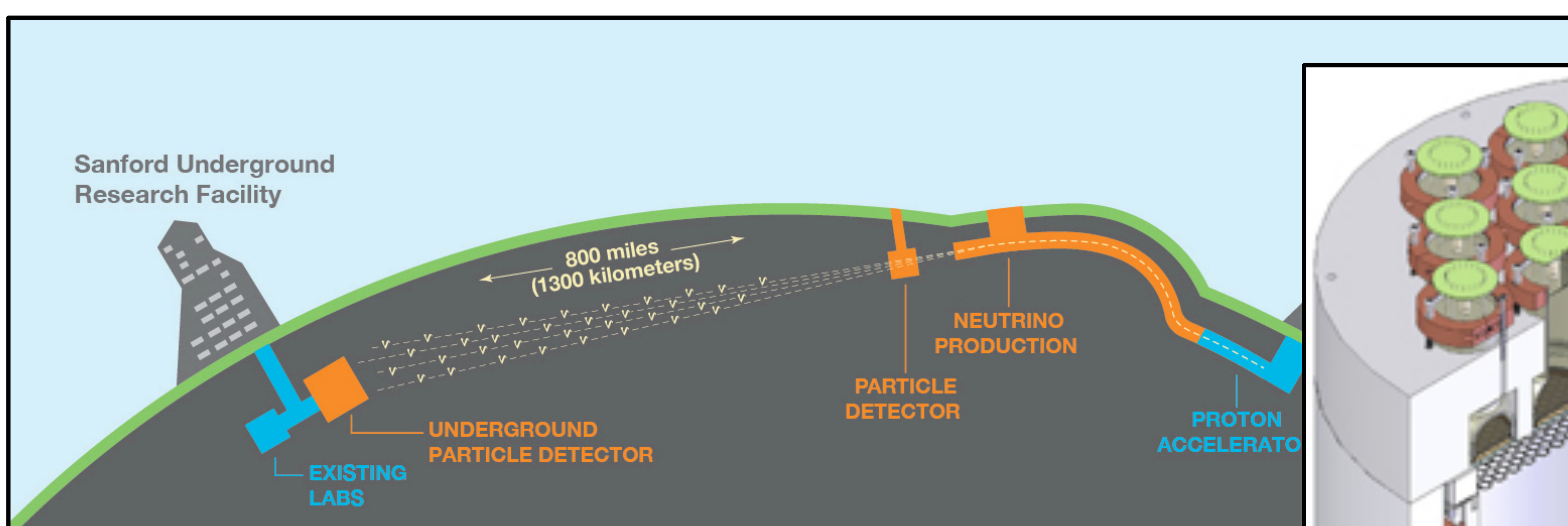


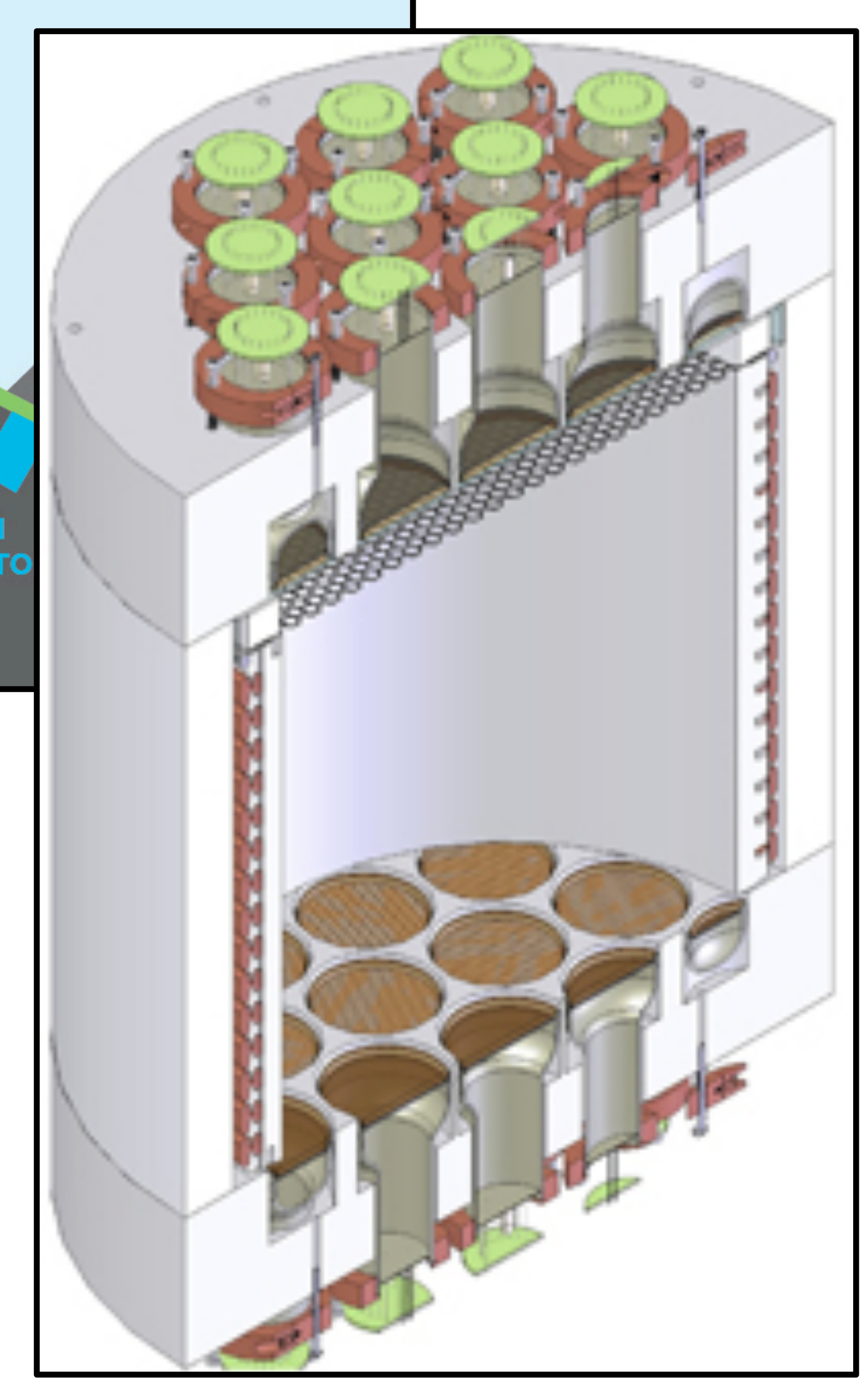
For HL-LHC upgrade: latency and output rates go up ~5

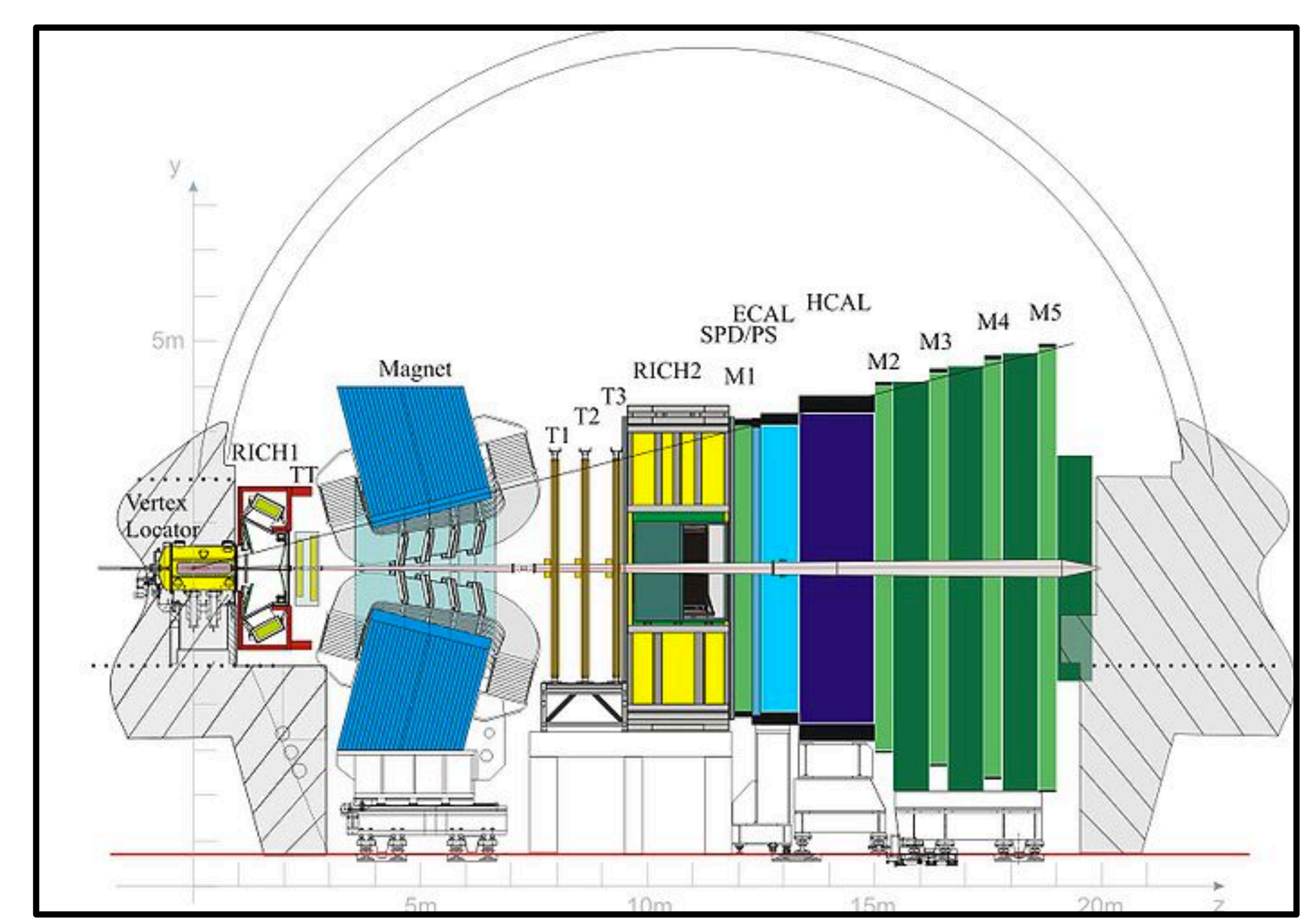
MORE OPPORTUNITIES

In the era of big science, more sophisticated triggers and DAQ systems are required

- Even in traditional “low” rate experiments
- Other LHC applications, like LHCb, and ATLAS/CMS HLT and cosmic and intensity frontier experiments







MACHINE LEARNING & FPGAs

Machine learning algorithms are ubiquitous in HEP

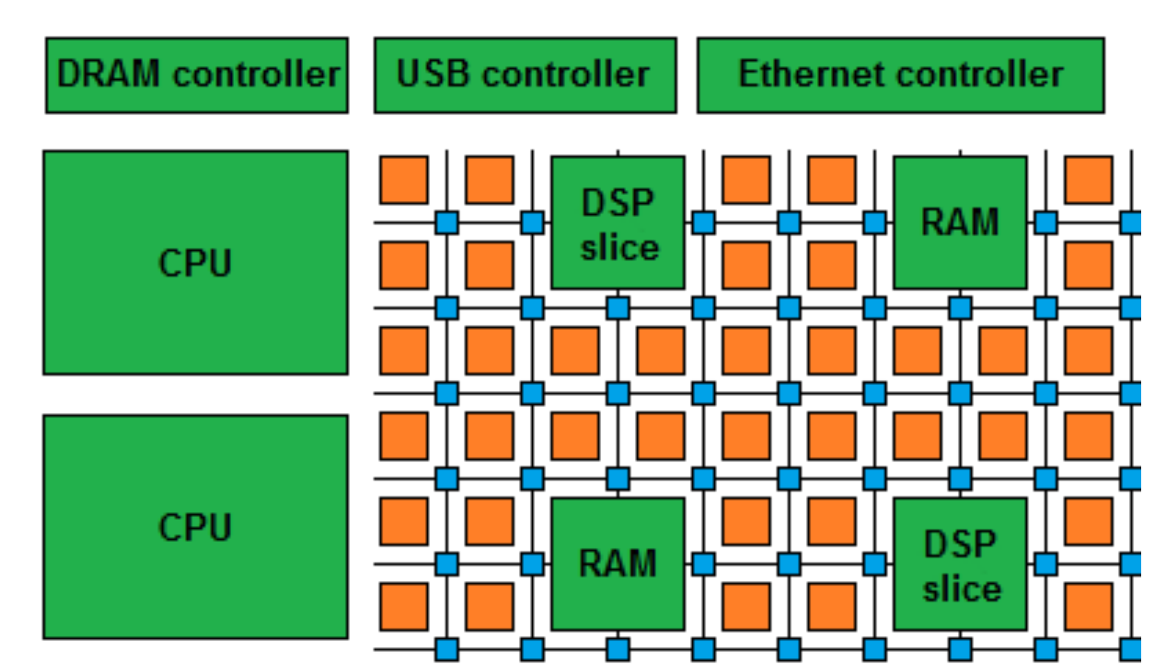
FPGA usage broad across HEP experiments

- Centered on DAQ and trigger development
- Some early adaptions of ML techniques in trigger [1]
- FPGA development becoming more accessible
  - **High Level Synthesis, OpenCL**

FPGA interest in industry is growing

Programmable hardware with structures that maps nicely onto ML architectures

FPGA
"programmable hardware"

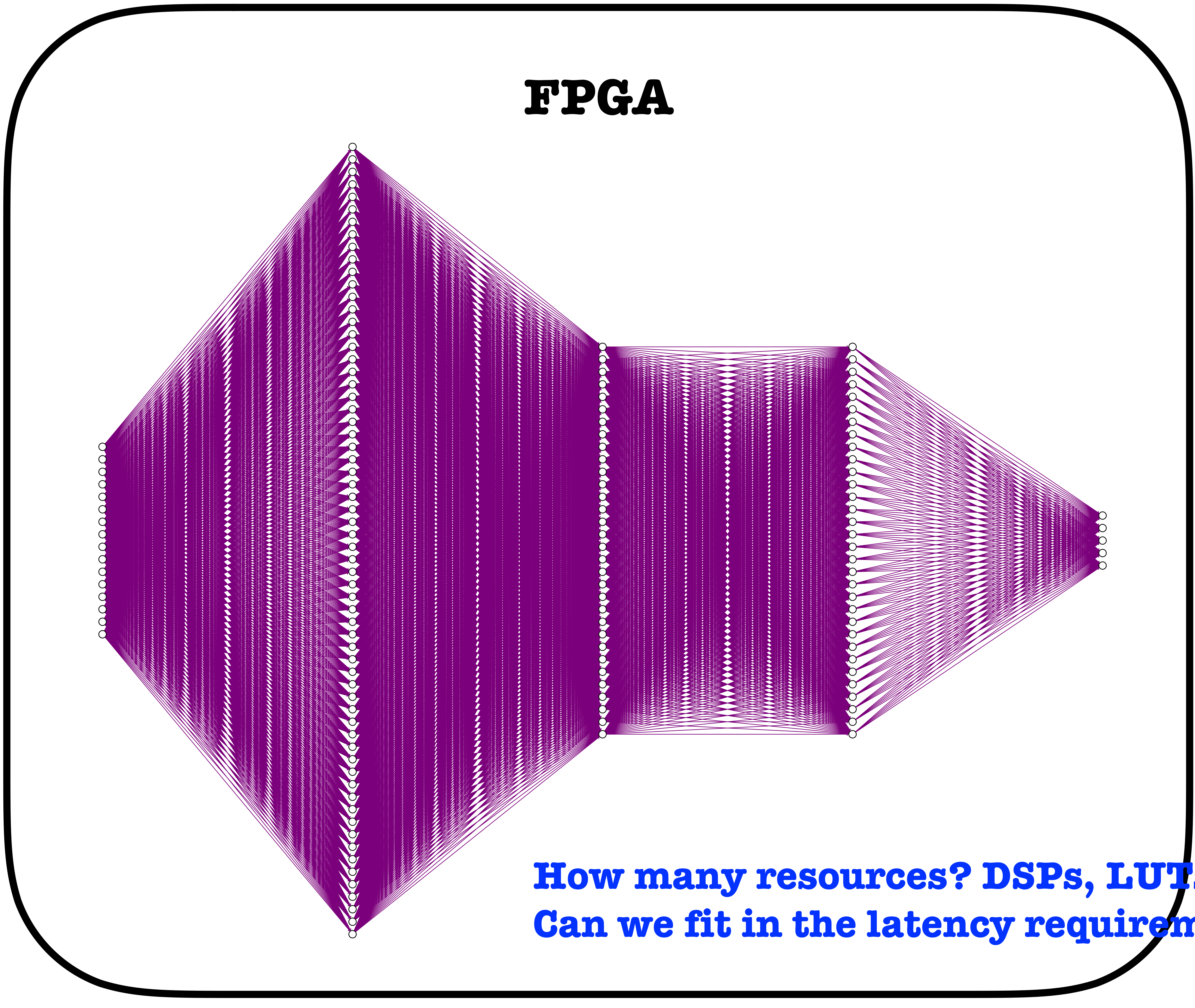


Modern FPGA: lots of hard, not-field-programmable gates

DSPs (multiply-accumulate, etc.)
Flip Flops (registers/distributed memory)
LUTs (logic)
Block RAMs (memories)

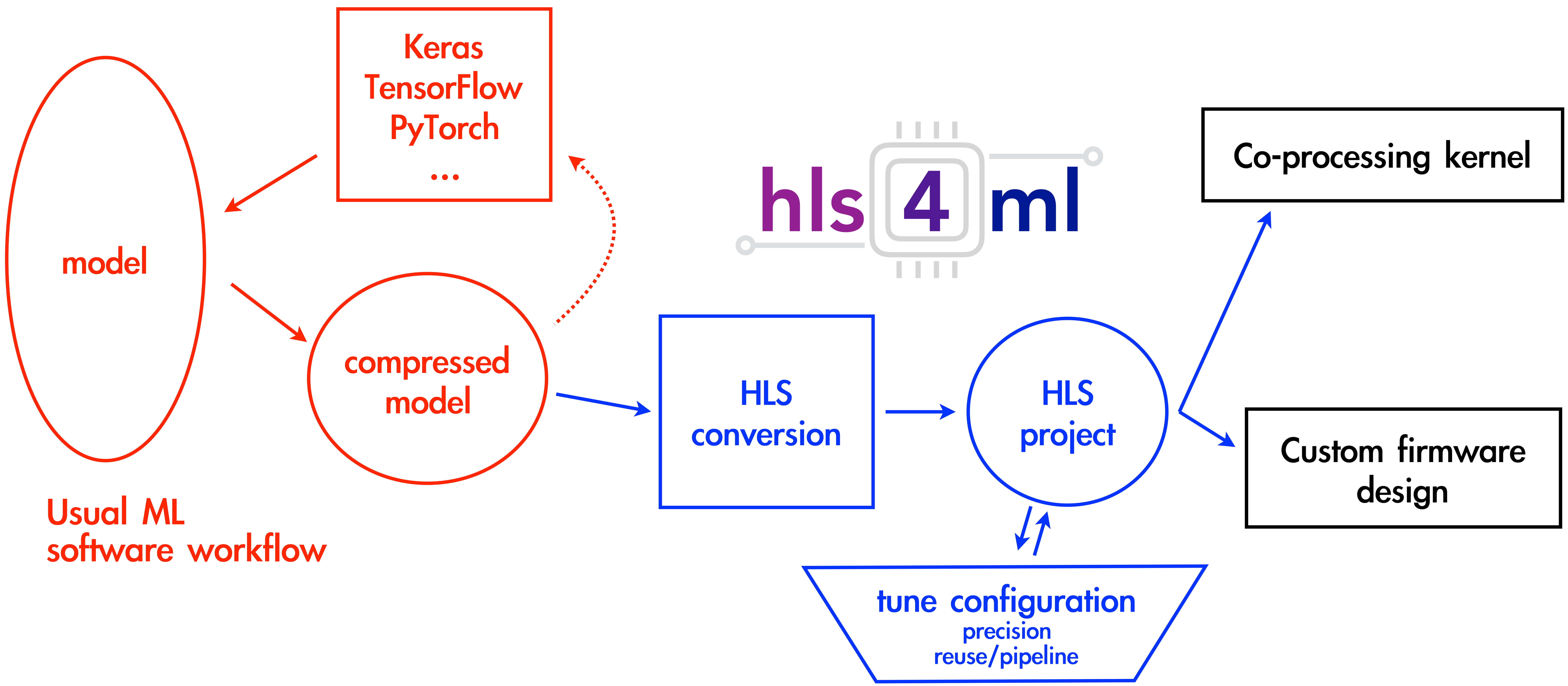
[1] Carnes et al., https://indico.cern.ch/event/567550/contributions/2629686/

ML IN FPGAs?



How many resources? DSPs, LUTs, FFs?
Can we fit in the latency requirements?

PROJECT OVERVIEW

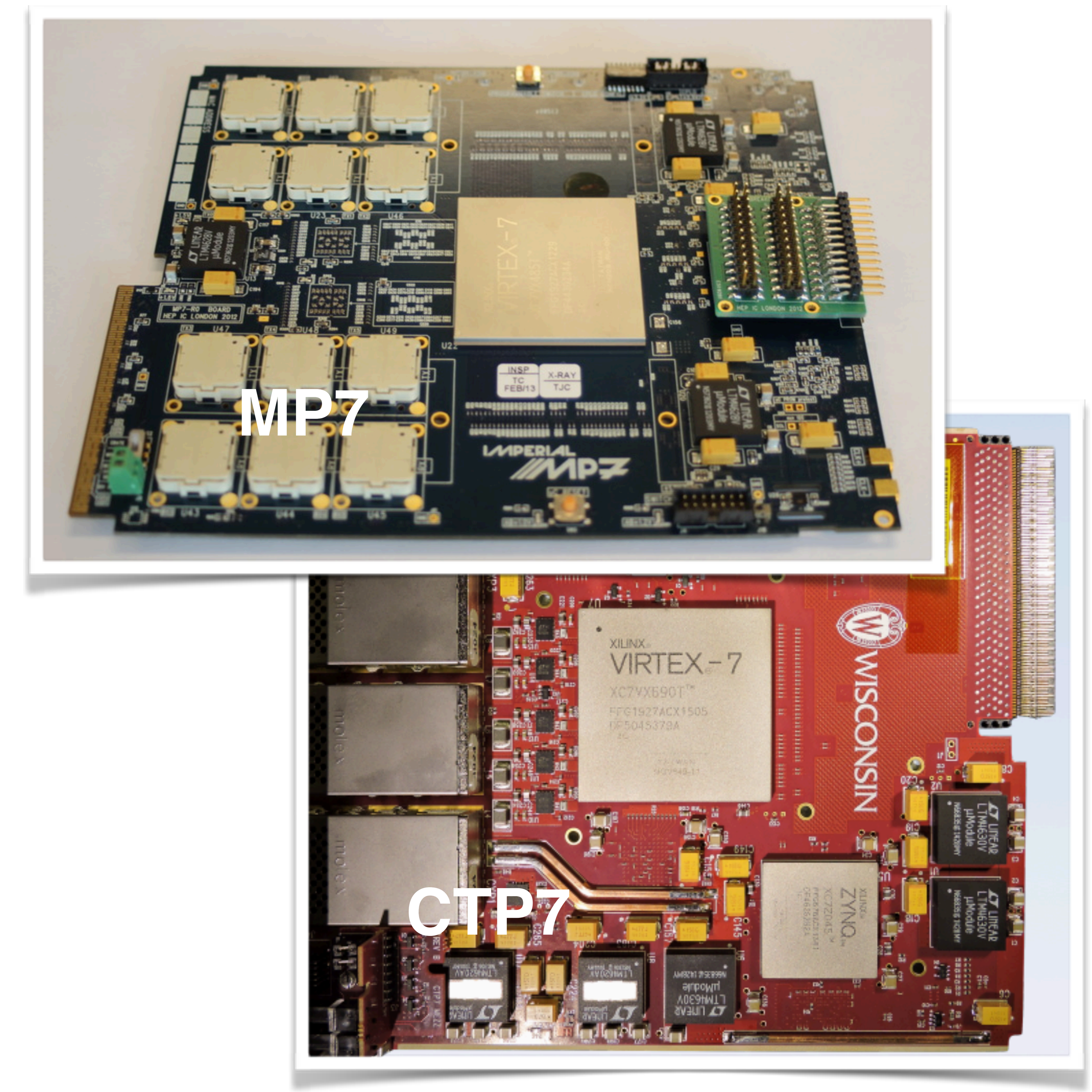


SNEAK PEAK

Fast inference of deep neural networks in FPGAs for particle physics

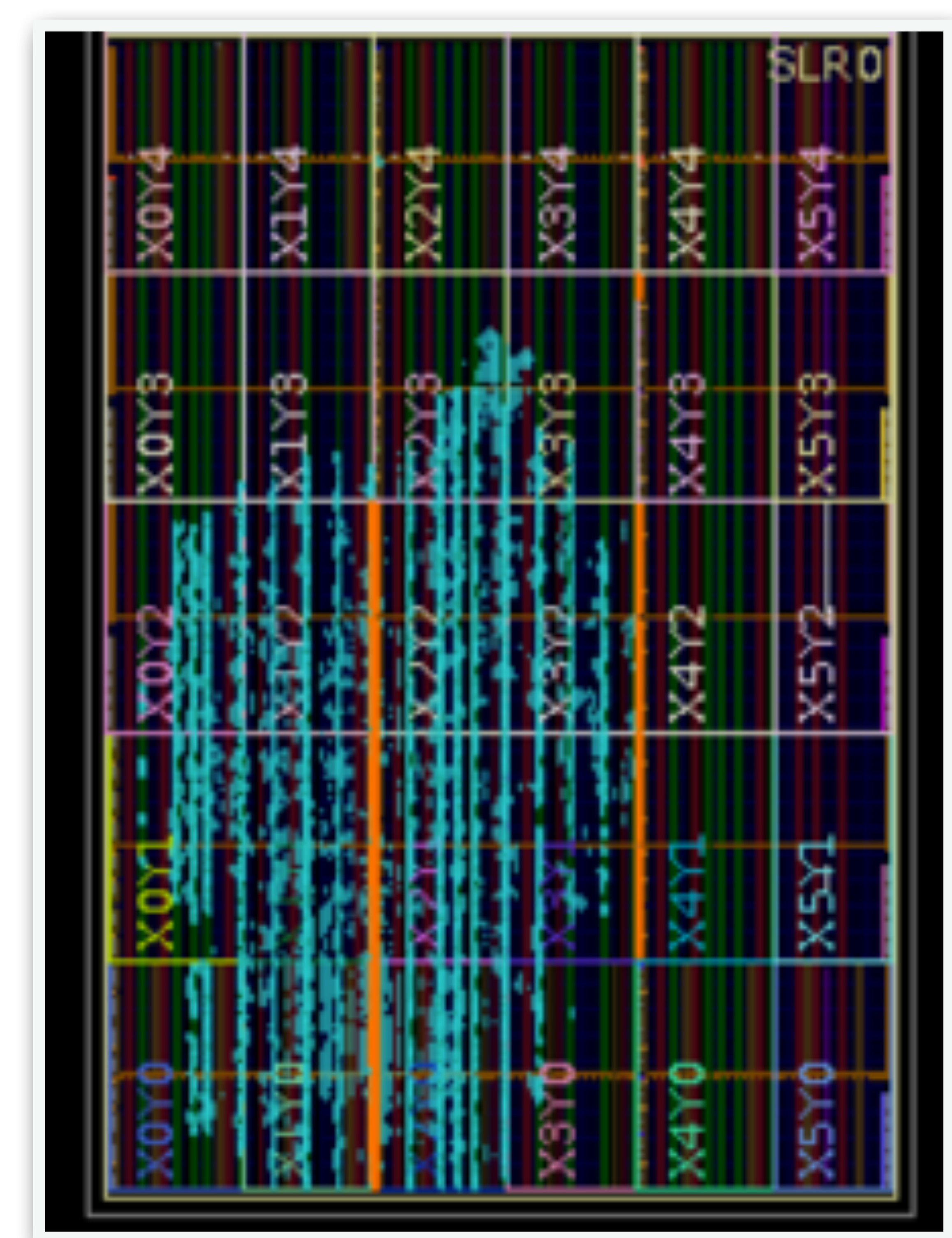
Javier Duarte[a], Song Han[b,c], Philip Harris[c], Sergo Jindariani[a], Edward Kreinar[d], Benjamin Kreis[a], Jennifer Ngadiuba[e], Maurizio Pierini[e], Nhan Tran[a], Zhenbin Wu[f]

[a] *Fermi National Accelerator Laboratory, Batavia, IL 60510, USA*
[b] *Stanford University, Menlo Park, CA 94025, USA*
[c] *Massachusetts Institute of Technology, Cambridge, MA 02139, USA*
[d] *HawkEye360, Herndon, VA 20170, USA*
[e] *European Center for Nuclear Research, Geneva, Switzerland*
[f] *University of Illinois at Chicago, Chicago, IL 60607, USA*



Network	Substructure (uncompressed)	Substructure (compressed)
AUC / Expected AUC	99.68%	99.55%
Parameters	4389	1338
Compression rate	-	3.3×
DSP48E	3329	954
Logic (LUT + FF)	263,234	88,797
Latency	75 ns	75 ns

Table 2: A summary of the vital statistics and HLS resource estimates of the uncompressed and compressed jet substructure tagging model with a network precision of fixed-point `<16,6>` and fully pipelined with clock frequency of 200 MHz synthesized on a Xilinx Kintex Ultrascale+ FPGA.



We introduce a software/firmware package, `hls4ml`

Automated translation of neural networks into firmware using HLS

Case study present with jet substructure in L1 trigger

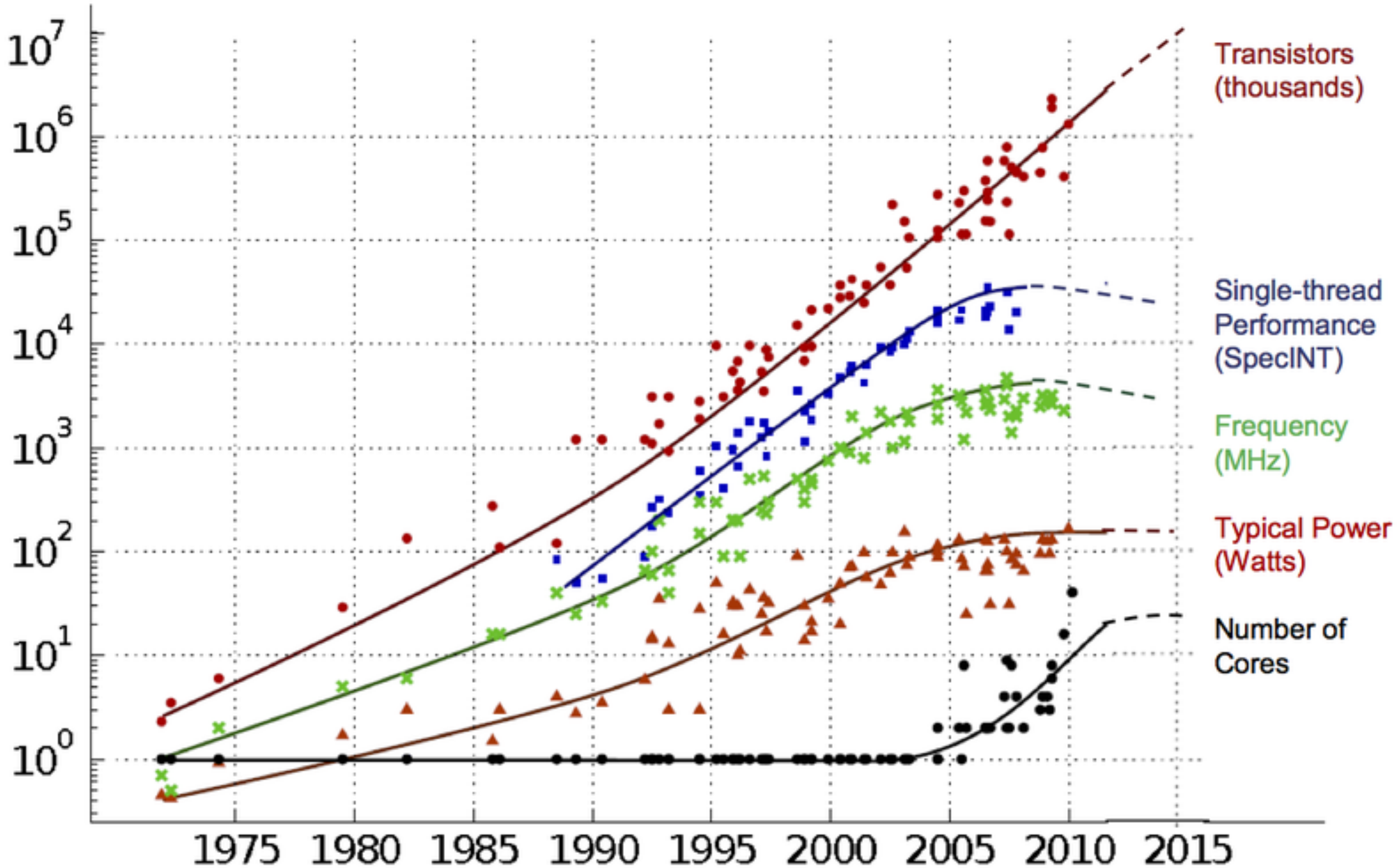
Tunable configuration for a broad range use cases

More info here:

https://hls-fpga-machine-learning.github.io/hls4ml/

Look out for research techniques seminar by
Javier Duarte (FNAL) on April 24th for many more details!

MOORE'S LAW AND DENNARD SCALING

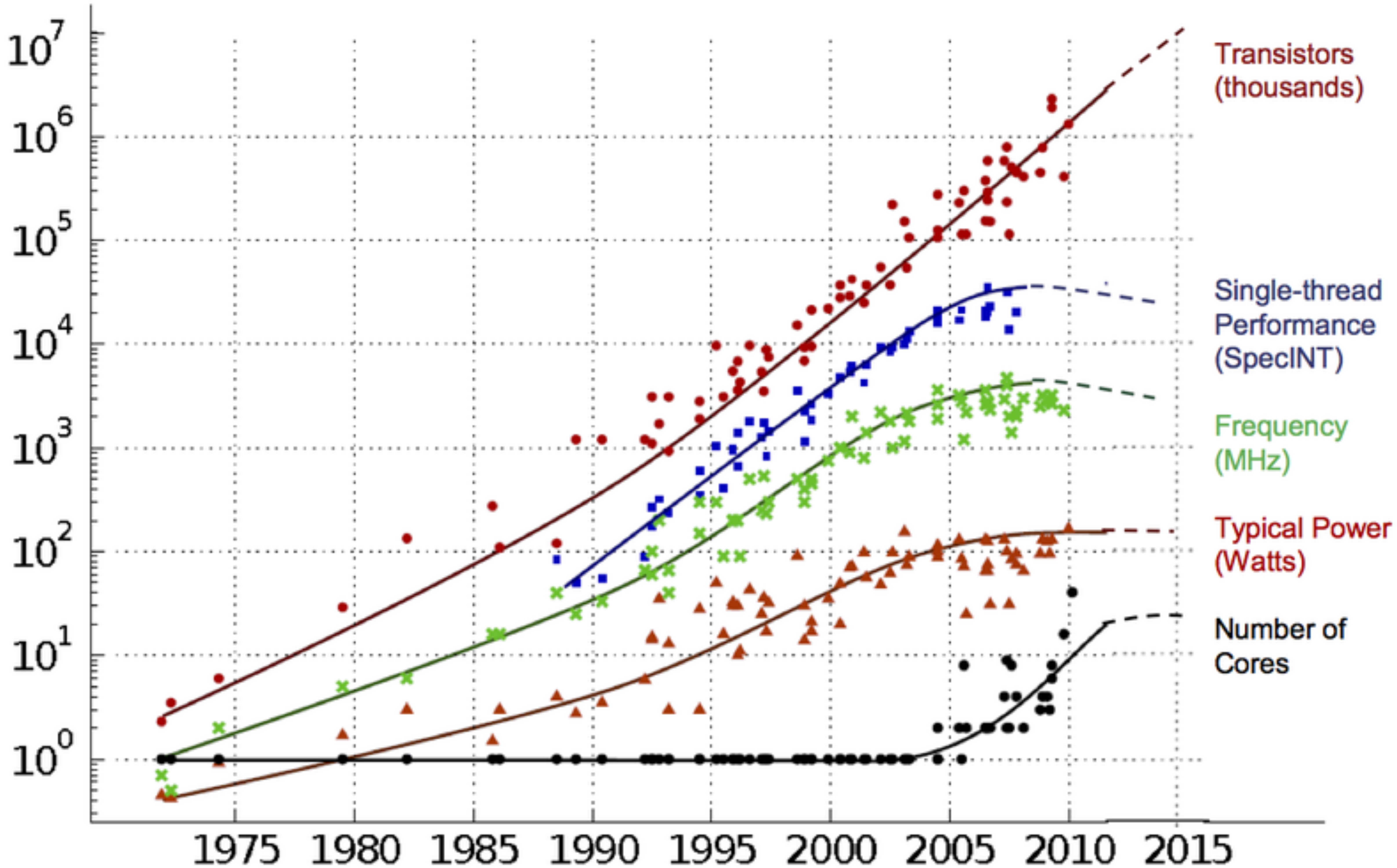


Moore's Law continues

Dennard Scaling fails

Original data collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond and C. Batten
Dotted line extrapolations by C. Moore

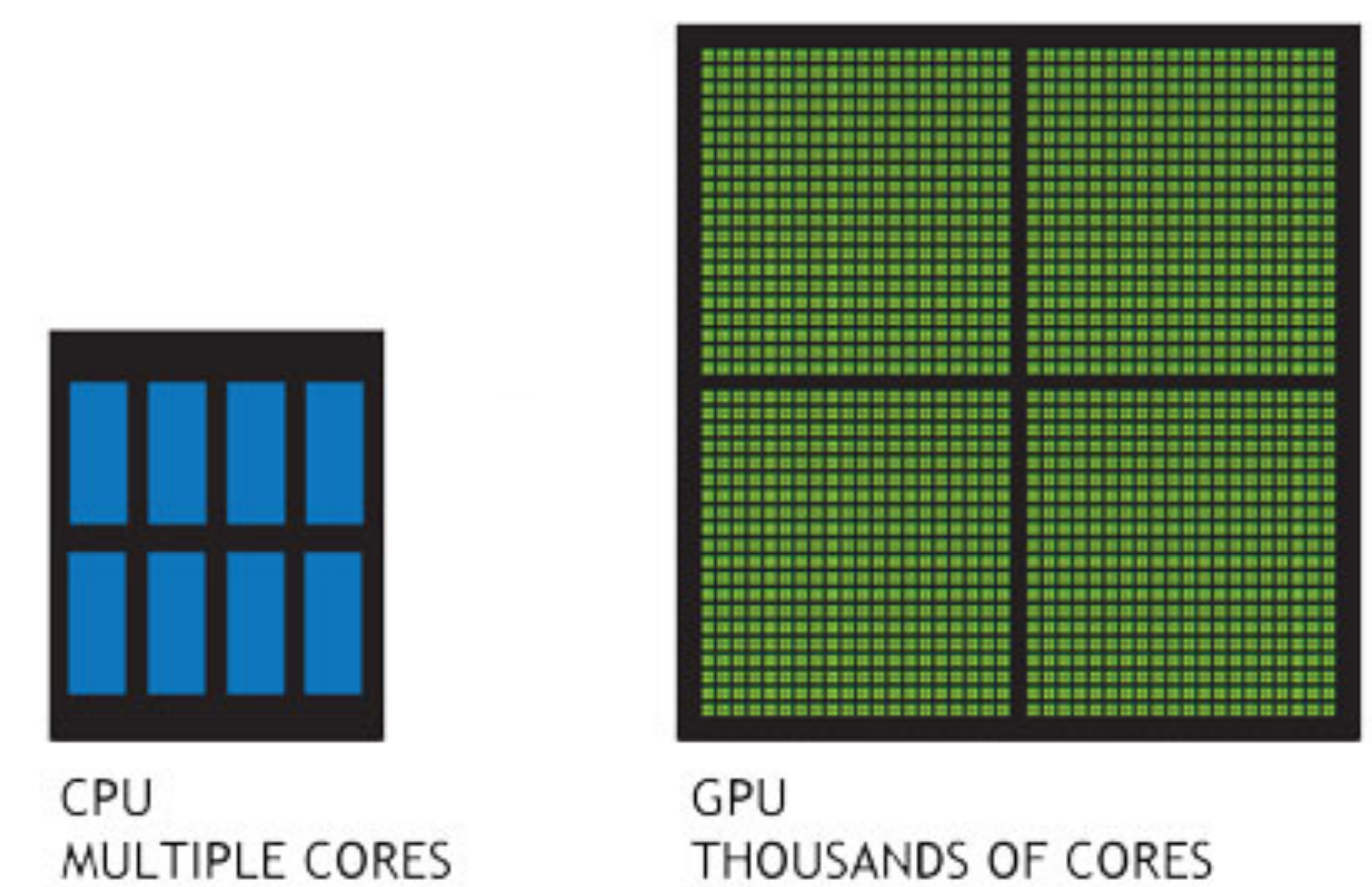
MOORE'S LAW AND DENNARD SCALING



Original data collected and plotted by M. Horowitz, F. Labonte, O. Shacham, K. Olukotun, L. Hammond and C. Batten
Dotted line extrapolations by C. Moore

Moore's Law continues

Dennard Scaling fails

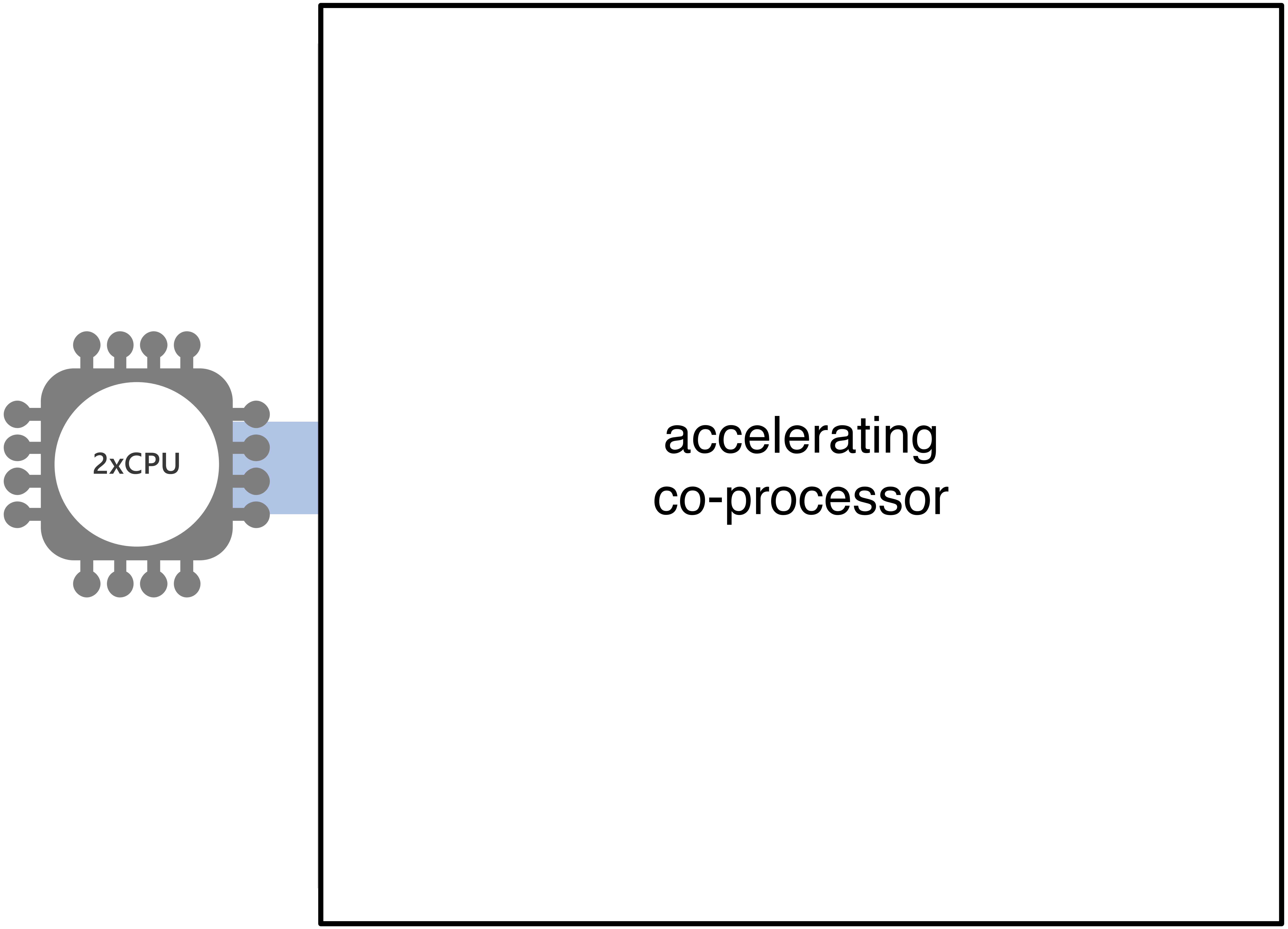


Single threaded performance not improving

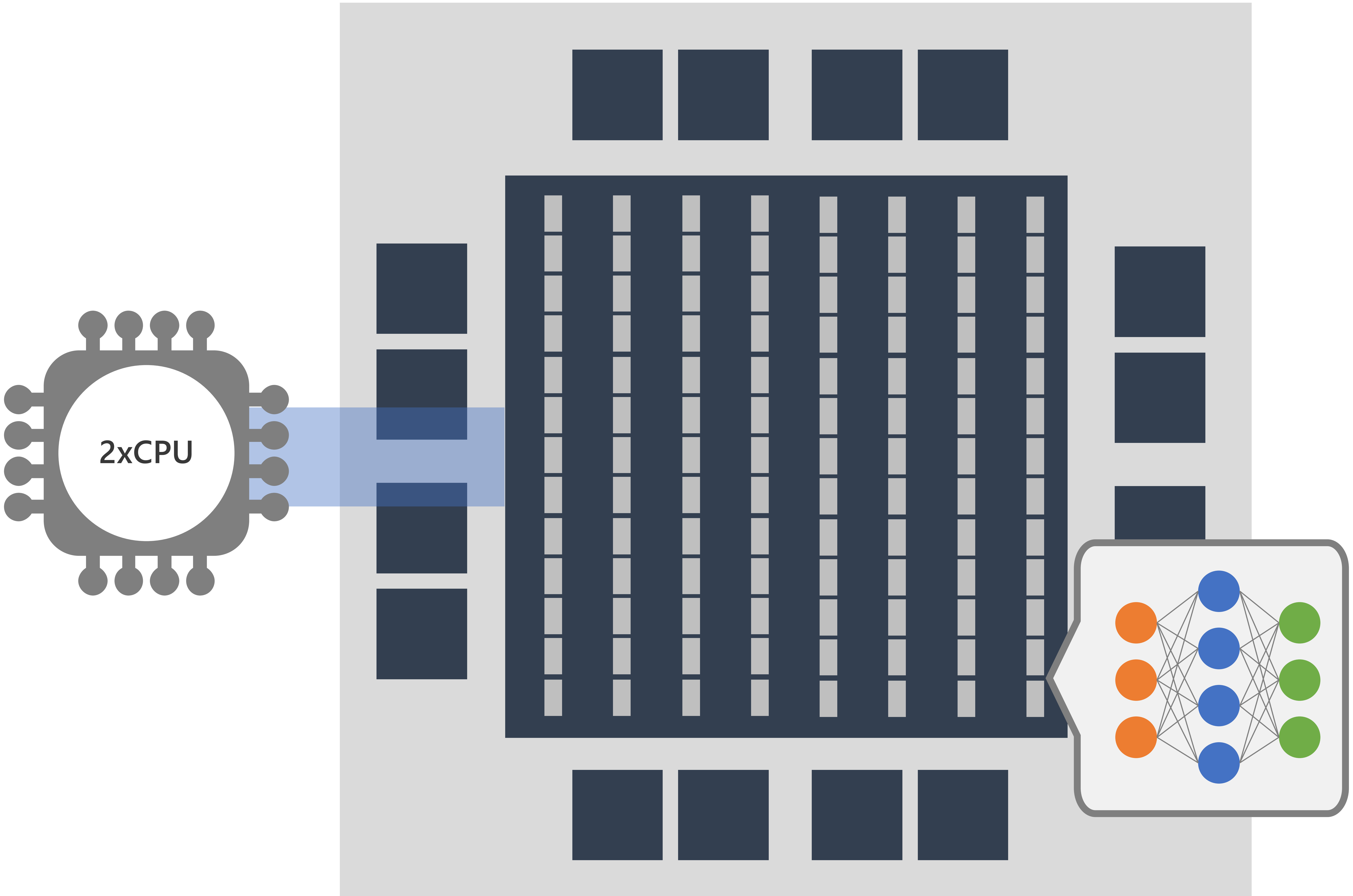
Circa ~2005: "The Era of Multicore"

→ **Today: Transition to the "Era of Specialization"?** (c.f. Doug Burger)

A NEW COMPUTING PARADIGM



A NEW COMPUTING PARADIGM



FPGA DATACENTERS!

Microsoft Catapult

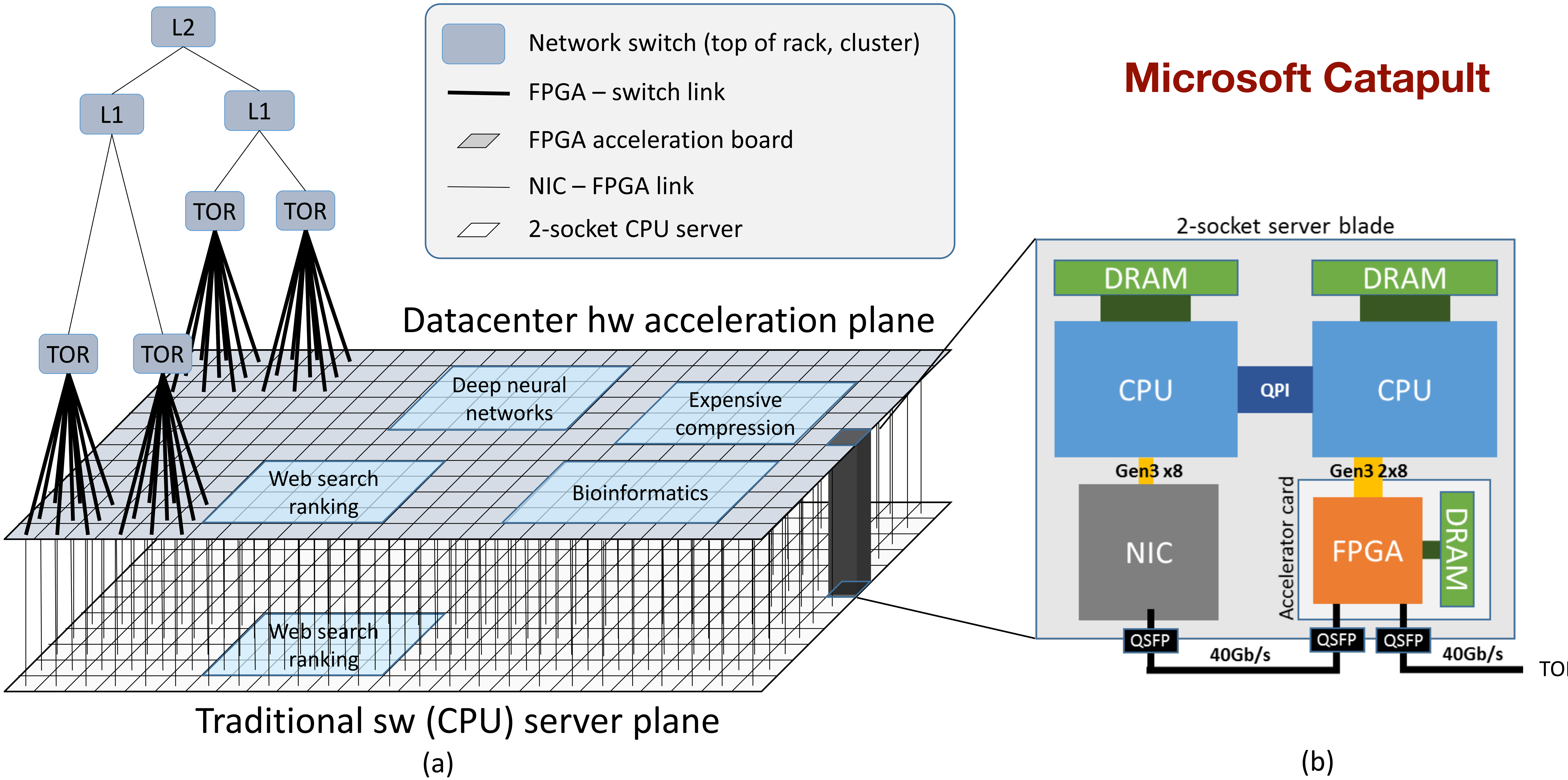
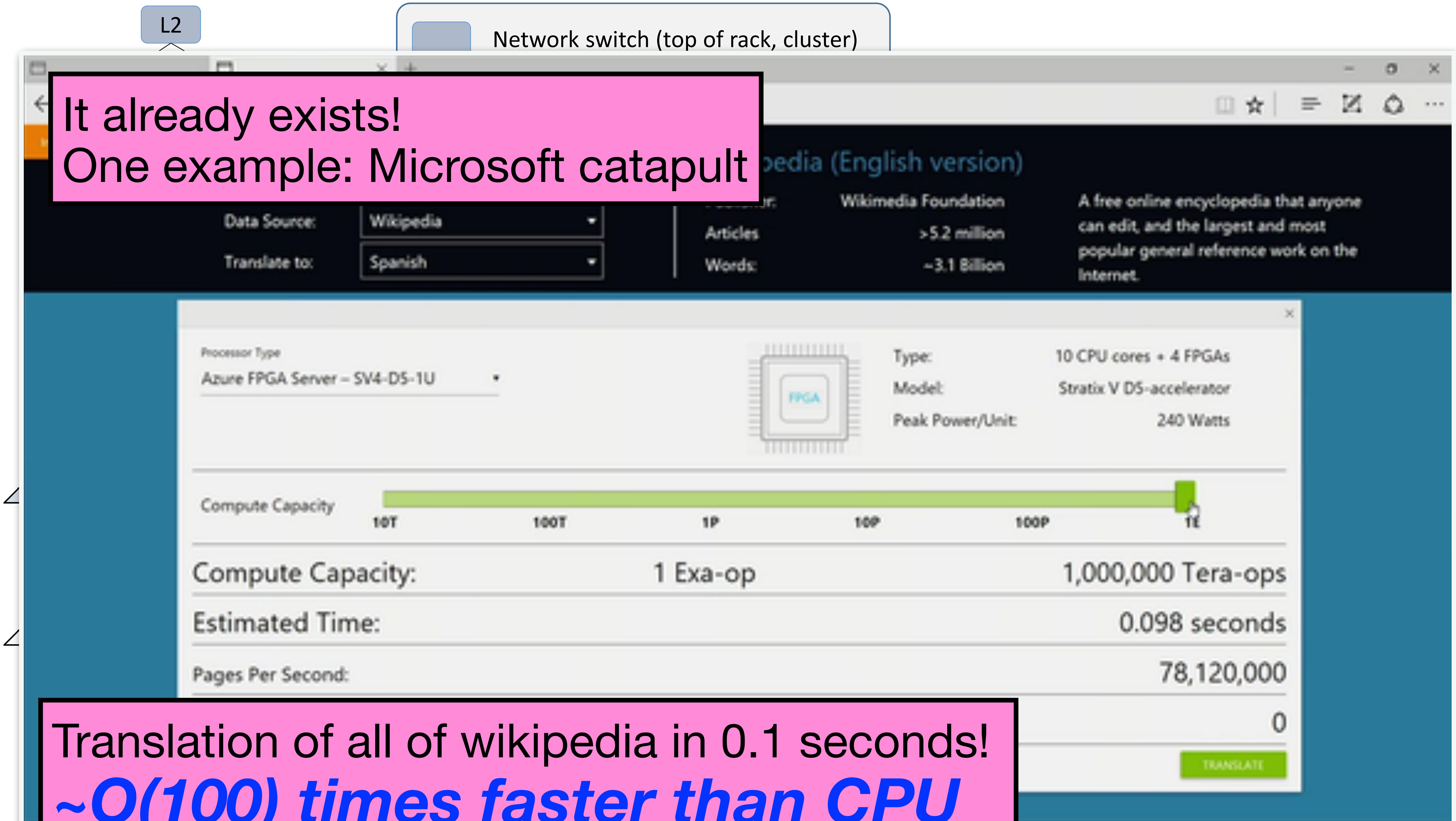


Fig. 1. (a) Decoupled Programmable Hardware Plane, (b) Server + FPGA schematic.

A. Caulfield, et al., Proceedings of the 49th Annual IEEE/ACM International Symposium on Microarchitecture (2016).
https://www.microsoft.com/en-us/research/publication/configurable-cloud-acceleration/

FPGA DATACENTERS!

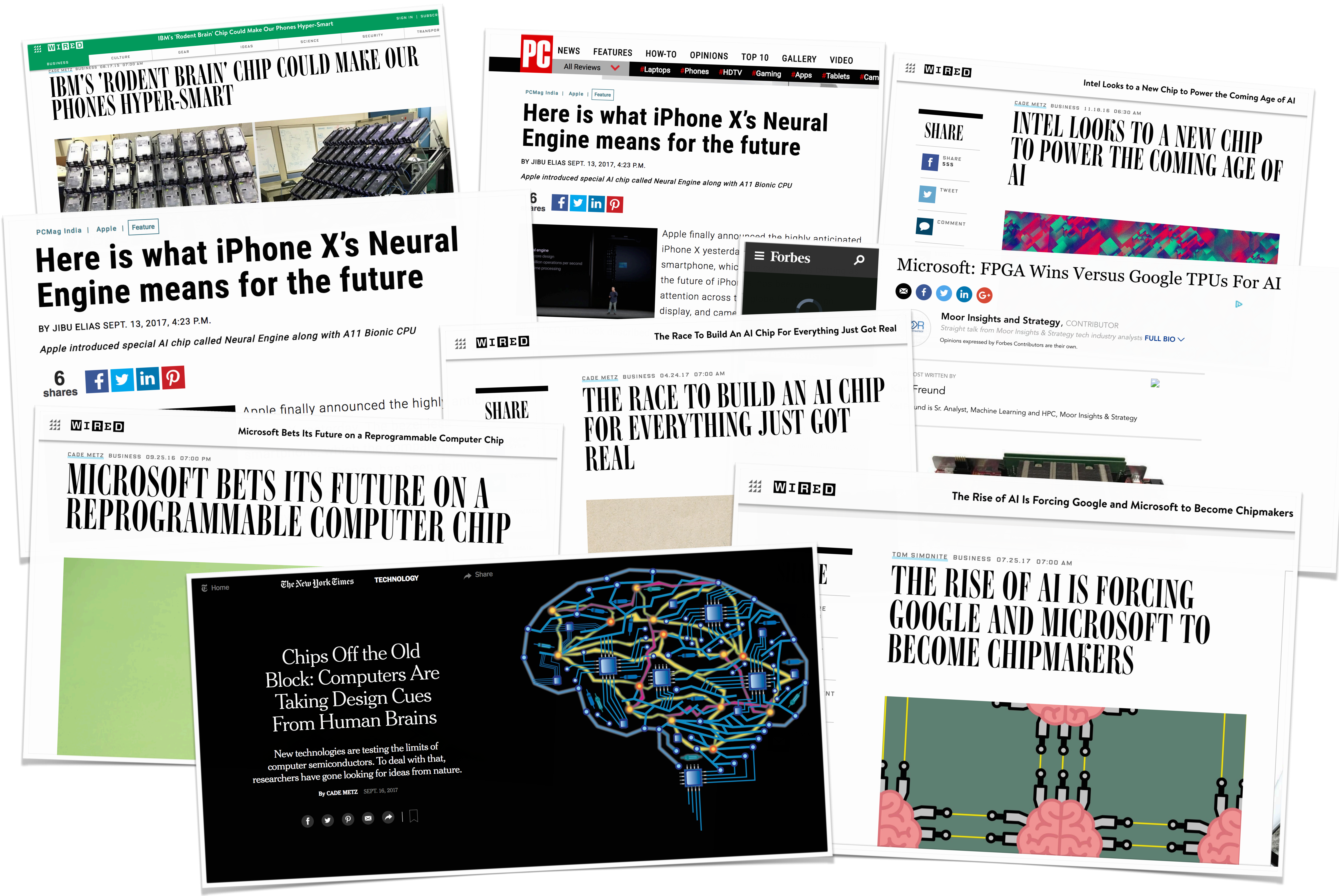
It already exists!
One example: Microsoft catapult



Translation of all of wikipedia in 0.1 seconds!
~O(100) times faster than CPU

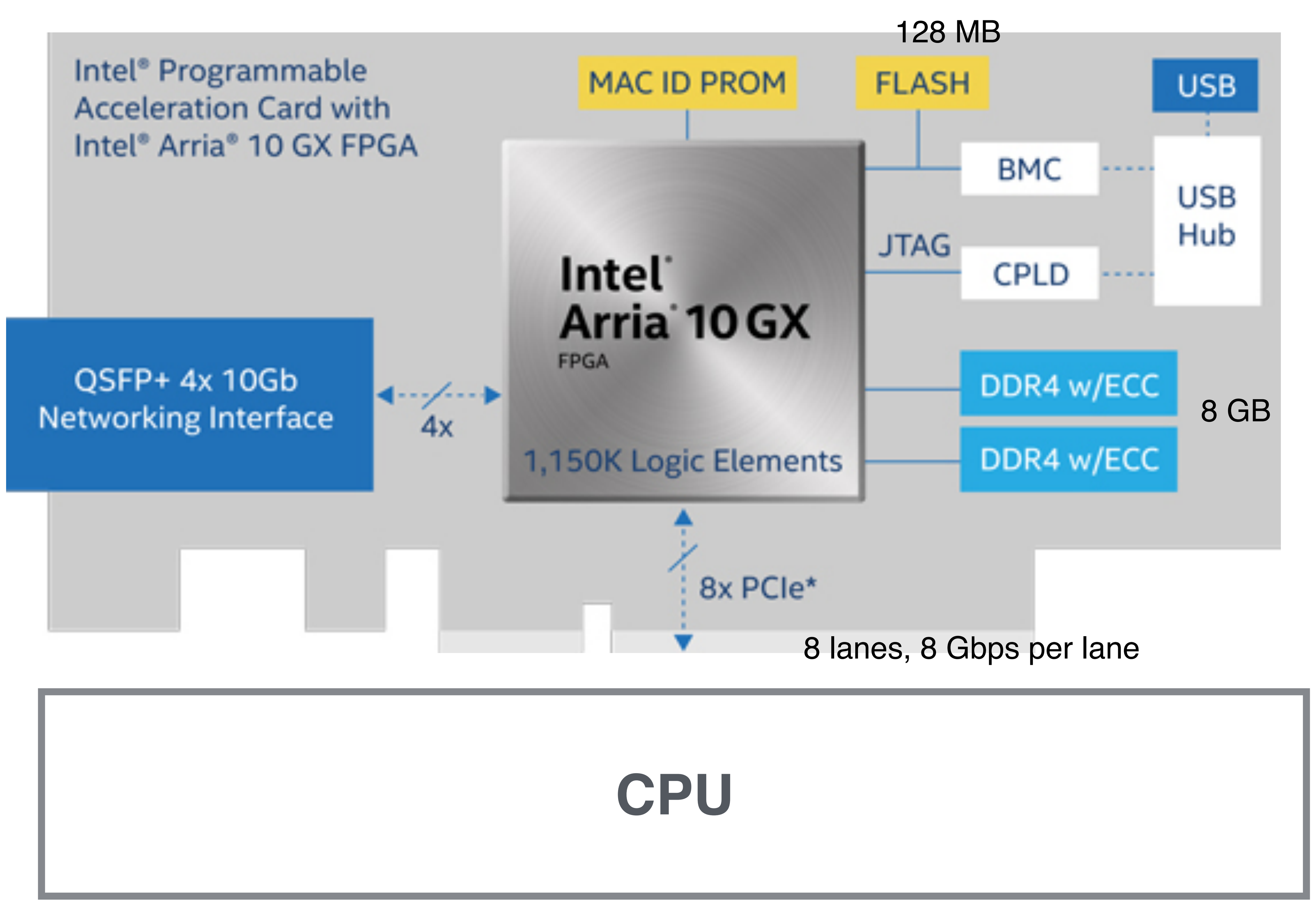
A. Caulfield, et al., Proceedings of the 49th Annual IEEE/ACM International Symposium on Microarchitecture (2016).
https://www.microsoft.com/en-us/research/publication/configurable-cloud-acceleration/

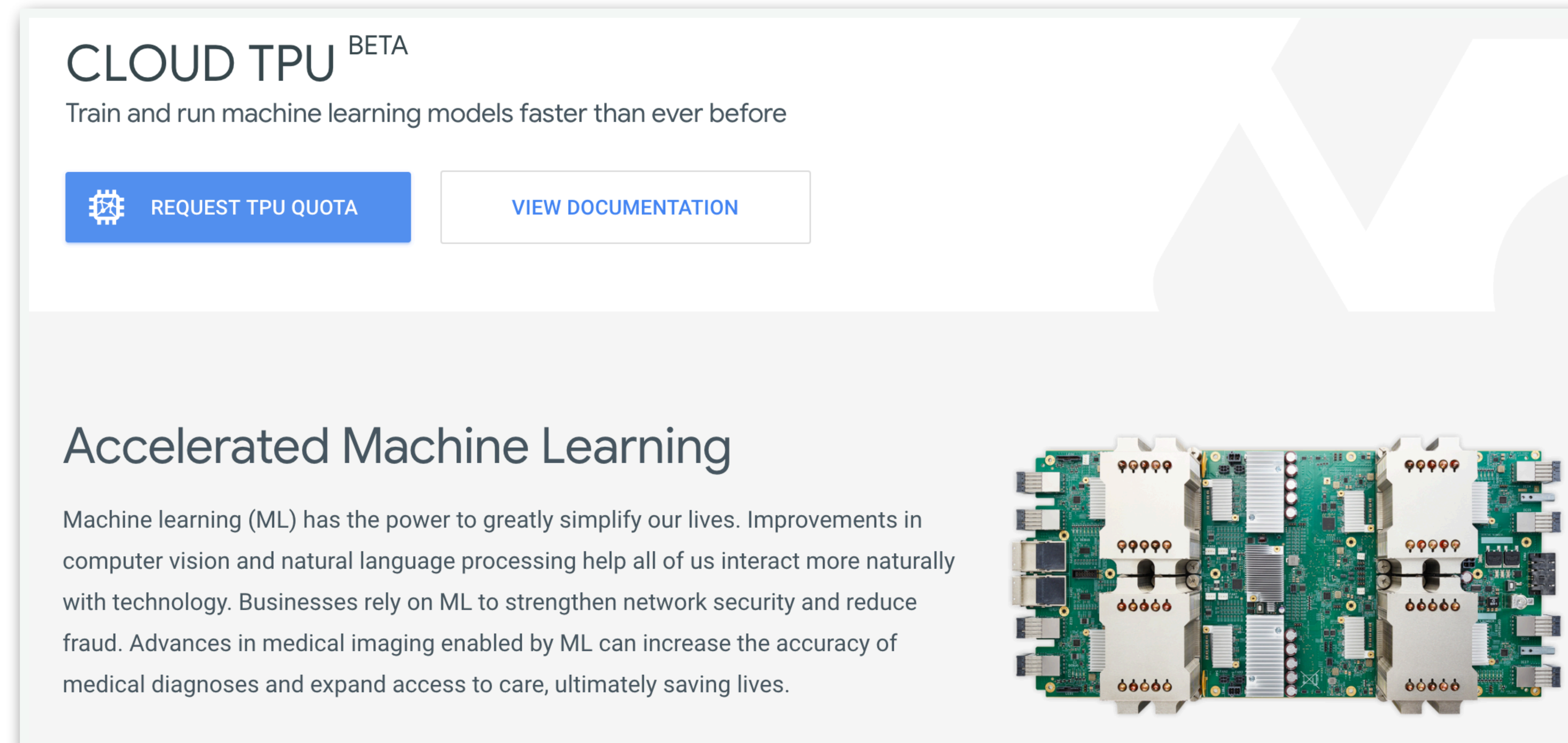
#RealTimeAI


IBM's 'Rodent Brain' Chip Could Make Our Phones Hyper-Smart
IBM'S 'RODENT BRAIN' CHIP COULD MAKE OUR PHONES HYPER-SMART
Here is what iPhone X's Neural Engine means for the future
BY JIBU ELIAS SEPT. 13, 2017, 4:23 P.M.
Apple introduced special AI chip called Neural Engine along with A11 Bionic CPU
Intel Looks to a New Chip to Power the Coming Age of AI
INTEL LOOKS TO A NEW CHIP TO POWER THE COMING AGE OF AI
Microsoft: FPGA Wins Versus Google TPUs For AI
Moor Insights and Strategy, CONTRIBUTOR
The Race To Build An AI Chip For Everything Just Got Real
THE RACE TO BUILD AN AI CHIP FOR EVERYTHING JUST GOT REAL
Microsoft Bets Its Future on a Reprogrammable Computer Chip
MICROSOFT BETS ITS FUTURE ON A REPROGRAMMABLE COMPUTER CHIP
The Rise of AI Is Forcing Google and Microsoft to Become Chipmakers
THE RISE OF AI IS FORCING GOOGLE AND MICROSOFT TO BECOME CHIPMAKERS
Chips Off the Old Block: Computers Are Taking Design Cues From Human Brains
New technologies are testing the limits of computer semiconductors. To deal with that, researchers have gone looking for ideas from nature.
By CADE METZ SEPT. 16, 2017

WHAT ABOUT ME?

Resources are available for development already!



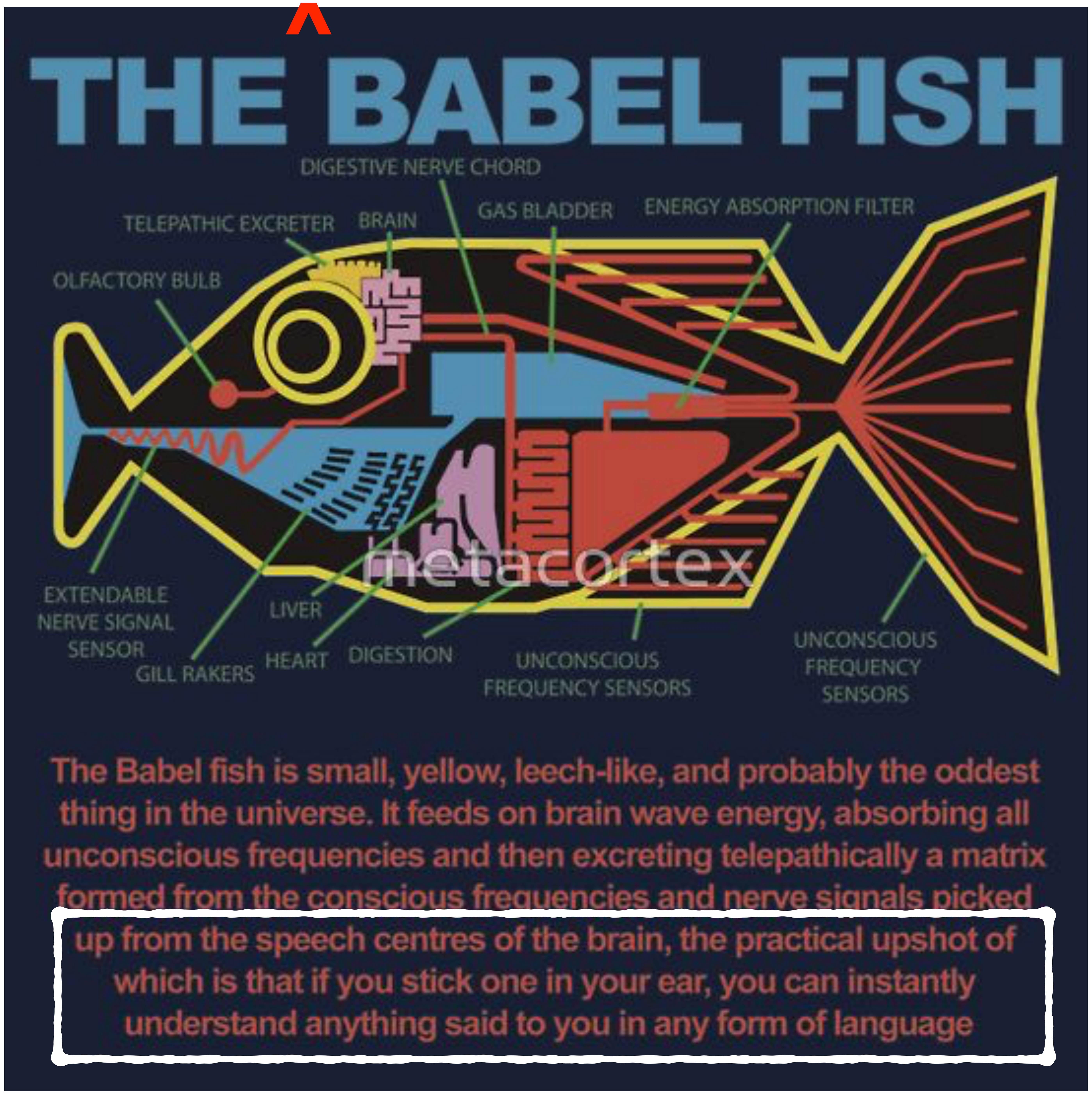


F1 FPGA Instance Types on AWS

- Up to 8 Xilinx UltraScale+ 16nm VU9P FPGA devices in a single instance
- The **f1.16xlarge** size provides:
  - 8 FPGAs, each with over 2 million customer-accessible FPGA programmable logic cells and over 5000 programmable DSP blocks
  - Each of the 8 FPGAs has 4 DDR-4 interfaces, with each interface accessing a 16GiB, 72-bit wide, ECC-protected memory

Instance Size	FPGAs	DDR-4 (GiB)	FPGA Link	FPGA Direct	vCPUs	Instance Memory (GiB)	NVMe Instance Storage (GB)
f1.2xlarge	1	4 x 16	-	-	8	122	1 x 470
f1.16xlarge	8	32 x 16	Y	Y	64	976	4 x 940

ML
^


THE BABEL FISH
DIGESTIVE NERVE CHORD
TELEPATHIC EXCRETER
BRAIN
GAS BLADDER
ENERGY ABSORPTION FILTER
OLFACTORY BULB
EXTENDABLE NERVE SIGNAL SENSOR
GILL RAKERS
LIVER
HEART
DIGESTION
UNCONSCIOUS FREQUENCY SENSORS
UNCONSCIOUS FREQUENCY SENSORS
metacortex
The Babel fish is small, yellow, leech-like, and probably the oddest thing in the universe. It feeds on brain wave energy, absorbing all unconscious frequencies and then excreting telepathically a matrix formed from the conscious frequencies and nerve signals picked up from the speech centres of the brain, the practical upshot of which is that if you stick one in your ear, you can instantly understand anything said to you in any form of language

ML Babel Fish

Large gains from hardware accelerating co-processors
Industry trending towards specialized computing paradigms

Option 1

re-write physics algorithms for new hardware

Language: OpenCL, OpenMP, HLS, …?

Hardware: FPGA, GPU

Option 2

re-cast physics problem as a machine learning problem

Language: C++, Python (TensorFlow, PyTorch,…)

Hardware: FPGA, GPU, ASIC

Why (Deep) Machine Learning?

a common *language* for solving problems
which can universally be expressed on optimized computing hardware
and follow industry trends

ML BABEL FISH

Large gains from hardware accelerating co-processors
Industry trending towards specialized computing paradigms

Option 1

re-write physics algorithms for new hardware

Language: OpenCL, OpenMP, HLS, …?

Hardware: FPGA, GPU

Option 2

re-cast physics problem as a machine learning problem

Language: C++, Python (TensorFlow, PyTorch,…)

Hardware: FPGA, GPU, ASIC

Why (Deep) Machine Learning?

a common *language* for solving problems

which can universally be expressed on optimized computing hardware

Look out for ML seminar by Andrew Putnam (Microsoft Research) on May 14/15 on FPGA datacenters and MS Catapult!

Summary

SUMMARY

Jet substructure is a rapidly developing field

- First seminal papers in 2008!
- More recently, a fast adopter of machine learning algorithms
  - A tractable and interesting problem

Jet substructure not well-modeled in MCs

- Use of SM standard candles and decorrelation techniques are key to demonstrate understanding of observables and ultimately to use in analysis

Many machine learning applications have been developed to improve jet substructure techniques

- Same challenges and principles apply to ML algorithms and physics algorithms

OUTLOOK

Why machine learning?

Performance?

O(1) improvements over physics algorithms and BDTs

The ML Babel Fish: many problems can be cast as machine learning problems

Once you cast your problem as a machine learning problem, you can use specialized hardware to accelerate your solution

Industry is not developing FPGAs and ASICs for Higgs jet substructure tagging and CCQE identification in LAr TPCs

O(100) improvements in computing and reconstruction!