ACAT 2019

ACAT Trip Report

Lorenzo Moneta

10-15 March 2019 Saas-Fee

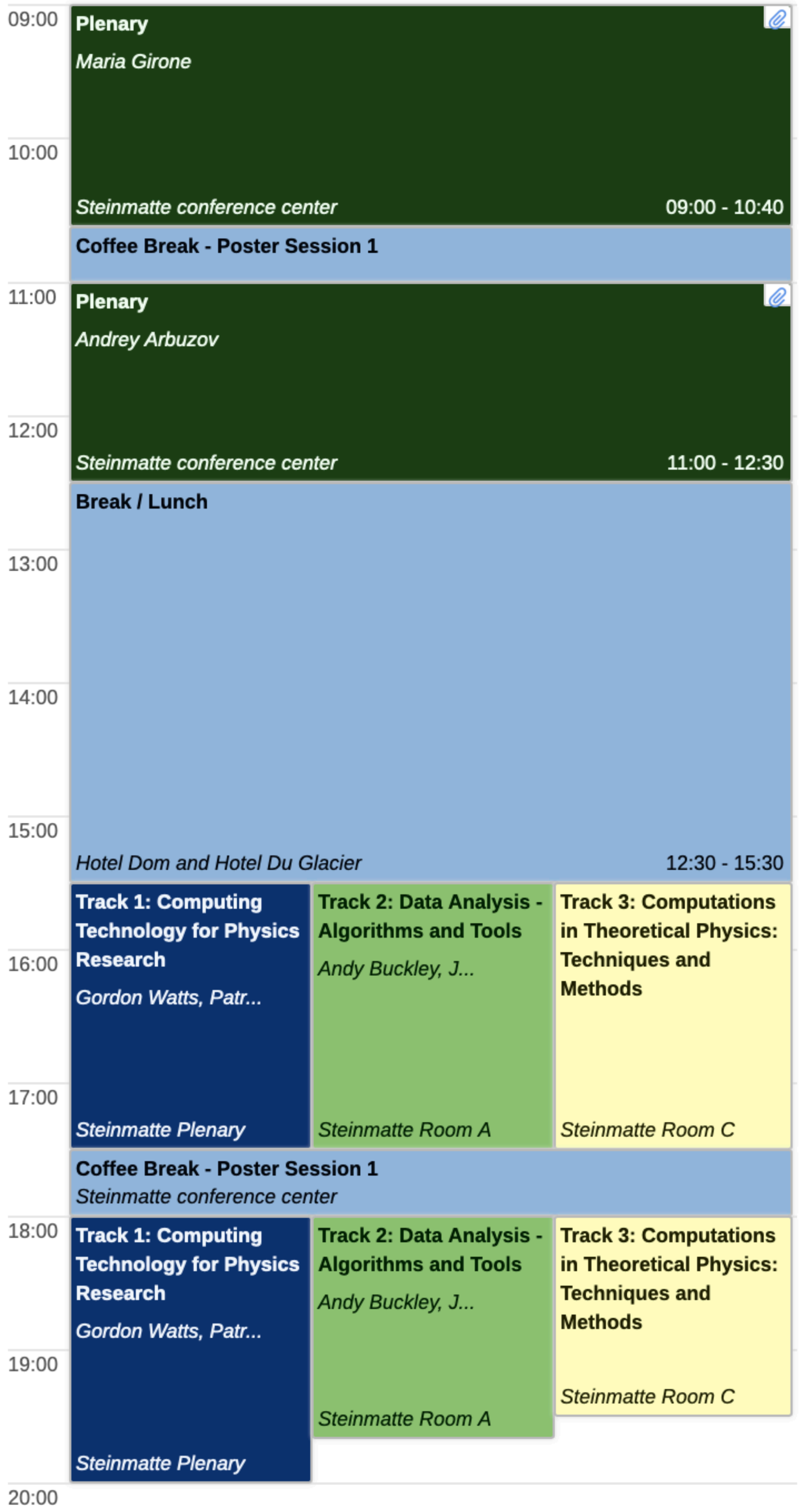
Scientific Program

- Plenaries: 35
- Parallels: 3 Sessions with 112 Presentations
  - Track 1: Computing Technology for Physics Research
  - Track 2: Data Analysis, Algorithms and Tools
  - Track 3: Computations in Theoretical Physics: Techniques and Methods
- Posters: 74
- 2 Panel discussions:
  - Diversity
  - Machine Learning
- see https://indico.cern.ch/event/708041/

Venue

Program

- Morning: plenary presentations
- Long lunch break: -> skiing,….
- Afternoon : Parallel sessions



Participants

- A large number of participants: 190
- One of the most attended ACAT workshops

ACAT History

- ACAT: *Advanced Computing and Analysis techniques for Physics Research*
- Before 2000 called AIHENP: *Artificial Intelligence in High Energy and Nuclear Physics*.
- New projects presented:
  - WWW (1992)
  - ROOT (1996)
- and also we had:
  - TMVA (2007)
  - RooStats (2010)

	Workshop	place	Date	Proceedings
1	AIHENP 1990	Lyon (France)	March 19-24 1990	some contributions: CDS, Inspire, Table of Content
2	AIHENP 1992	La Londe Les Maures (France)	Jan. 13-18 1992	some contributions: CDS, Inspire, Content
3	AIHENP 1993	Oberammergau (Germany)	Oct. 4-8 1993	some contributions: CDS, Inspire, Content
4	AIHENP 1995	Pisa (Italy)	April 3-8 1995	Some contribCDS1,CDS2 Inspire, Content
5	AIHENP 1996	Lausanne (EPFL-UNIL) (Switzerland)	Sept. 2-6 1996	NIM A389
6	AIHENP 1999	Heraklion (Crete, Greece)	April 12-16, 1999	Some Contributions: CDS, Inspire
7	ACAT 2000	Chicago (FERMILAB) (USA)	Oct. 16-20, 2000	AIP 583
8	ACAT 2002	Moscow (MSU)(Russia)	June 24-28, 2002	NIM A502
9	ACAT 2003	Tsukuba (KEK) (Japan)	Dec. 1-5,2003	NIM A534
10	ACAT 2005	Zeuthen (DESY) (Germany)	May 22-27, 2005	NIM A559
11	ACAT 2007	Amsterdam (NIKHEF) (The Netherlands)	April 23-27, 2007	POS
12	ACAT 2008	Erice (Italy)	Nov. 3-7, 2008	POS
13	ACAT 2010	Jaipur (India)	Feb. 22-27, 2010	POS
14	ACAT 2011	Uxbridge (UK)	Sept. 5-9, 2011	IOP
15	ACAT 2013	Beijing (China)	May 16-21 2013	IOP
16	ACAT 2014	Prague (Czech Republic)	Sept. 1-5, 2014	IOP
17	ACAT 2016	Valparaiso (Chile)	Jan. 18-22, 2016	IOP
18	ACAT 2017	Seattle (USA)	Aug. 21-25, 2017	IOP
19	ACAT 2019	Saas Fee (Switzerland)	Mar.. 11-15, 2019	

- Founder of ACAT conference series
- Chair of ACAT IAC

Denis Perret-Gallix
(1949-2018)

Track 1

Computing Technology for Physics Research

- 39 presentations
- Several experiment presentations
  - CMS, LHCb and ATLAS
  - presenting strategy to cope with missing computing resources for HL-LHC

SFT Presentations

- Several SFT presentations at the conference (all in Track 1)
  - from ROOT team:
    - *Lorenzo*: Data Analysis with ROOT
    - *Iliana:* new ROOT graphics language
    - *Stephan*: New PyROOT
    - *Axel*: New C++ 17 and 20
    - *Yuka*: C++ Modules
  - from Simulation
    - *Andrei*: Performance Results of GeantV
    - *Soon*: Vectorization of Random Number Generators
- Posters:
  - *Stephan*: Making RooFit ready for Run3 and beyond(Stephan)
  - *Oxsana*: ROOT I/O compression algorithm ; Evolution of ROOT package management
  - *Simone*: CernVM-FS Container Image Integration

- Example: Performance of Vectorized RNG (Soon Yung Jun)

Preliminary Performance: AVX - Vc [7] Backend

- The average CPU time [ms] for generating 10M random numbers
  - std::rand() = (93.24 ± 0.03) ms

Generator	MRG32k3a	Threefry	Philox
Scalar	137.24 ± 0.05	74.13 ± 0.08	54.59 ± 0.03
Vector(AVX)	57.59 ± 0.02	43.06 ± 0.17	76.90 ± 0.22
CUDA Backend	0.45 ± 0.05	12.12 ± 0.02	12.19 ± 0.01
Curand [8]	0.51 ± 0.01	N/A	0.67 ± 0.05

  - Intel(R) Xeon(R) CPU E5-2650 v2 @ 2.60GHz (Ivy Bridge)
  - NVidia Tesla K40M-12GB (2880 cores)
  - The word size (W) and round (R) used for Random-123 were W4x32_R20 for Threefry and W4x32_R10 for Philox
- VecMath/Rng also supports vectorized PDFs
  - Gauss, Gamma, Poisson, etc.
  - Vector gain $\sim 3-4$ on Intel(R) Core(TM) i7-4510U CPU
  - For details, see the poster by Oscar Chaparro (IPN, Mexico)

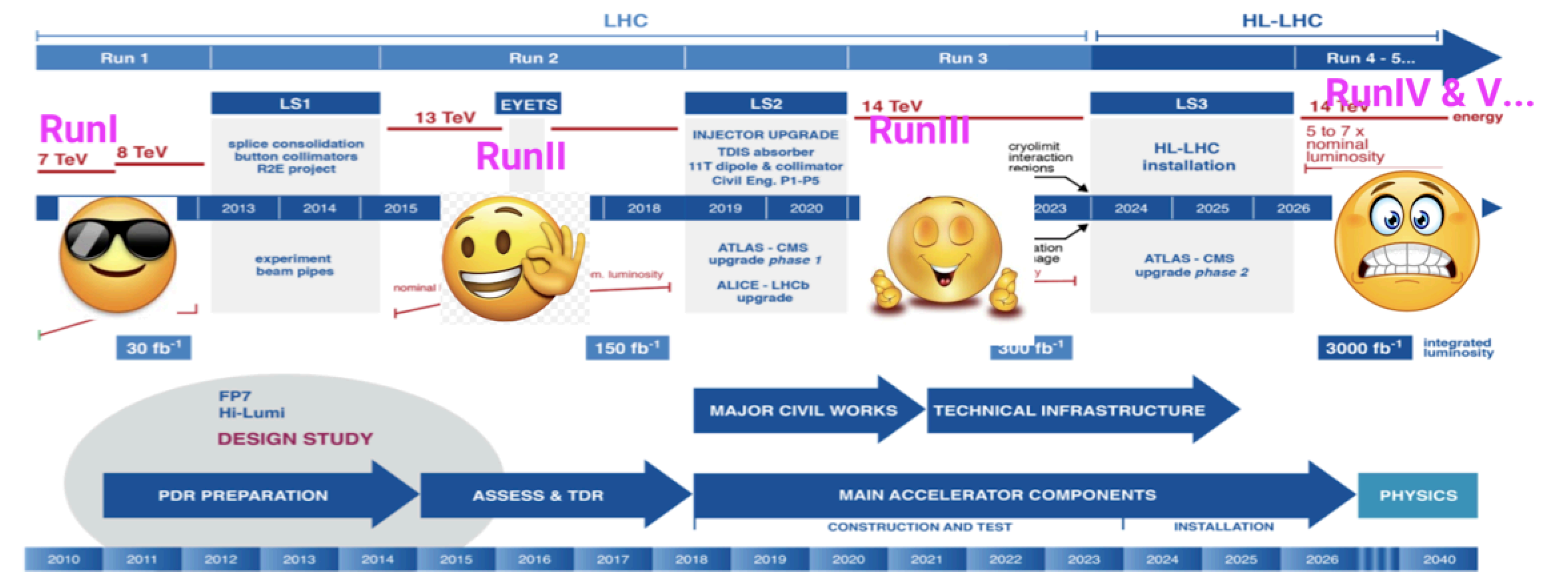
CMS Software

T.Boccali (CMS)

CMS Software and Offline preparation for future runs

So where is the problem?

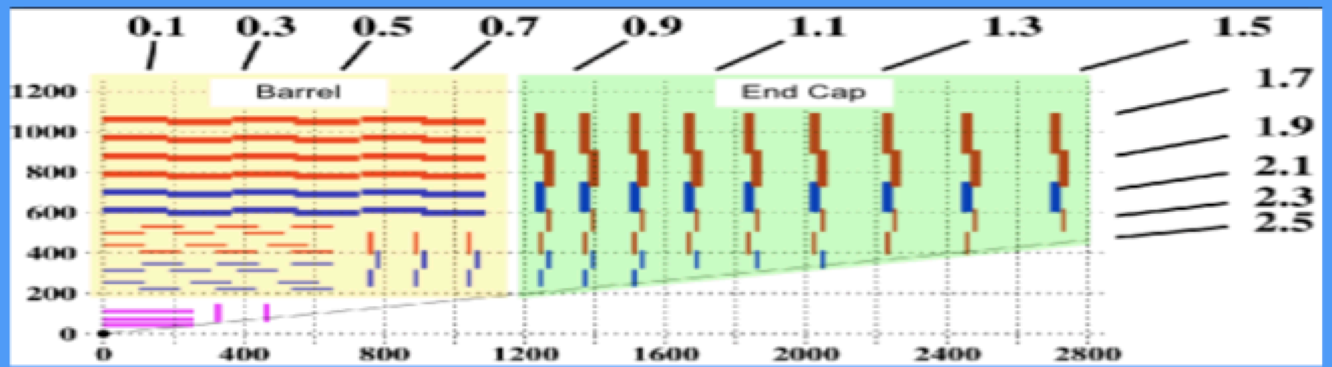
- **2009-2012 (RunI):** resources somehow overprovisioned, luxury mode
- **2013-2014 (LS1)**: Funding Agencies imposed a "flat funding", which means ~20% increase/y thanks to Moore's law (and friends)
- **2015-2018 (RunII)**: resources more and more constrained, Moore's law starting to be excessively optimistic
- **2019-2020 (LS2)**: virtually no increase in resources granted
- **2021-2023 (RunIII)**: in principle not incredibly different from RunII, but LHC is willing to surprise us
- **2024-2025 (LS3)**: no increases?
- **2026+** : the **LHC Phase II, the problem!**



CMS Software

- HL-LHC Computing problem

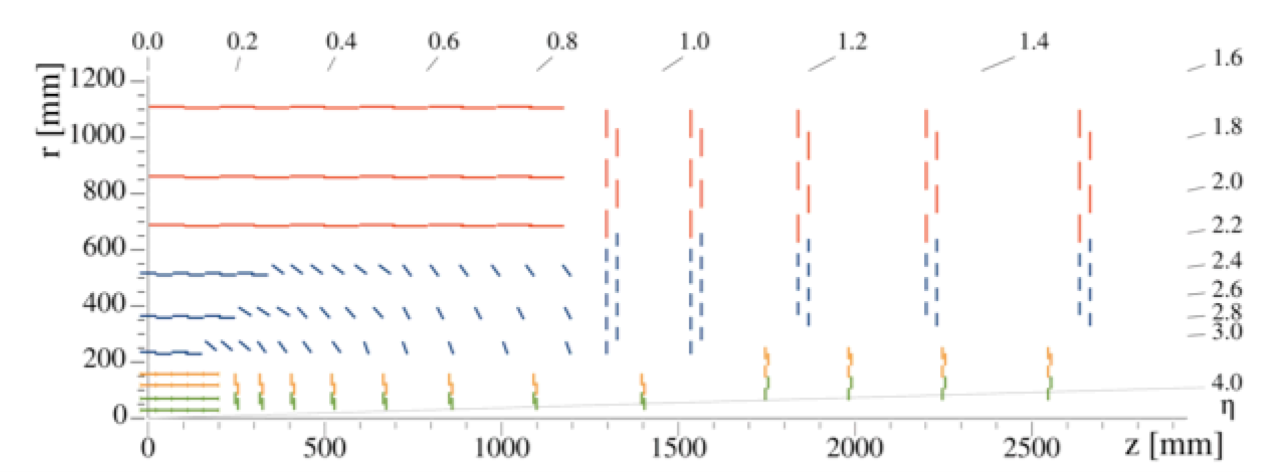
RunIII → RunIV

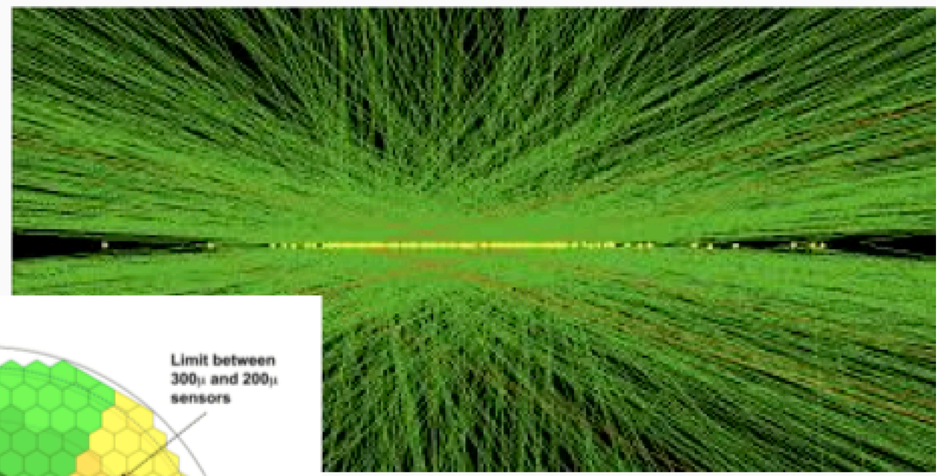


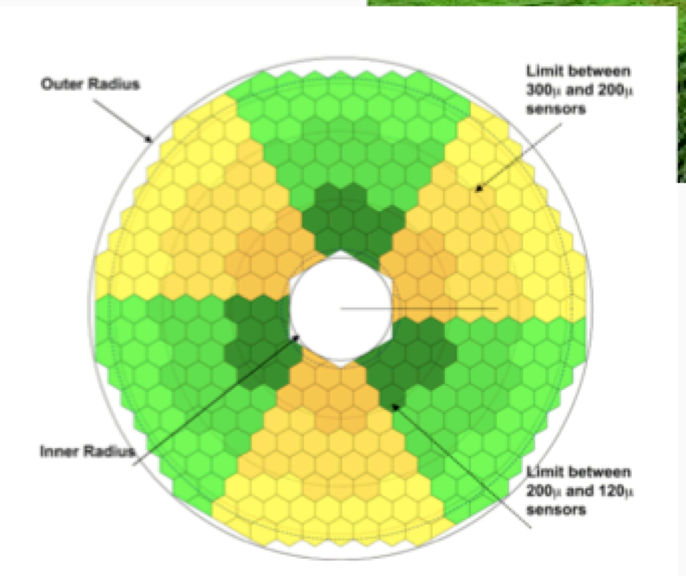
- Current modelling expects up to **7.5e34 cm^{-2}s^{-1} flat luminosity during fills**
- **PU = <PU> = 200**
- CMS with an upgraded detector:
  - Many more silicon tracker measurements
  - Completely revamped forward calorimetry
  - **→ many more channels, much larger expected algorithmic complexity**
- Physics requirements currently suggesting 5-10x increased trigger rate

→ on paper, easy to get factors 50-100x more resources needed with respect to RunII!

(see **An analytics driven computing model for HL-LHC** here at ACAT)





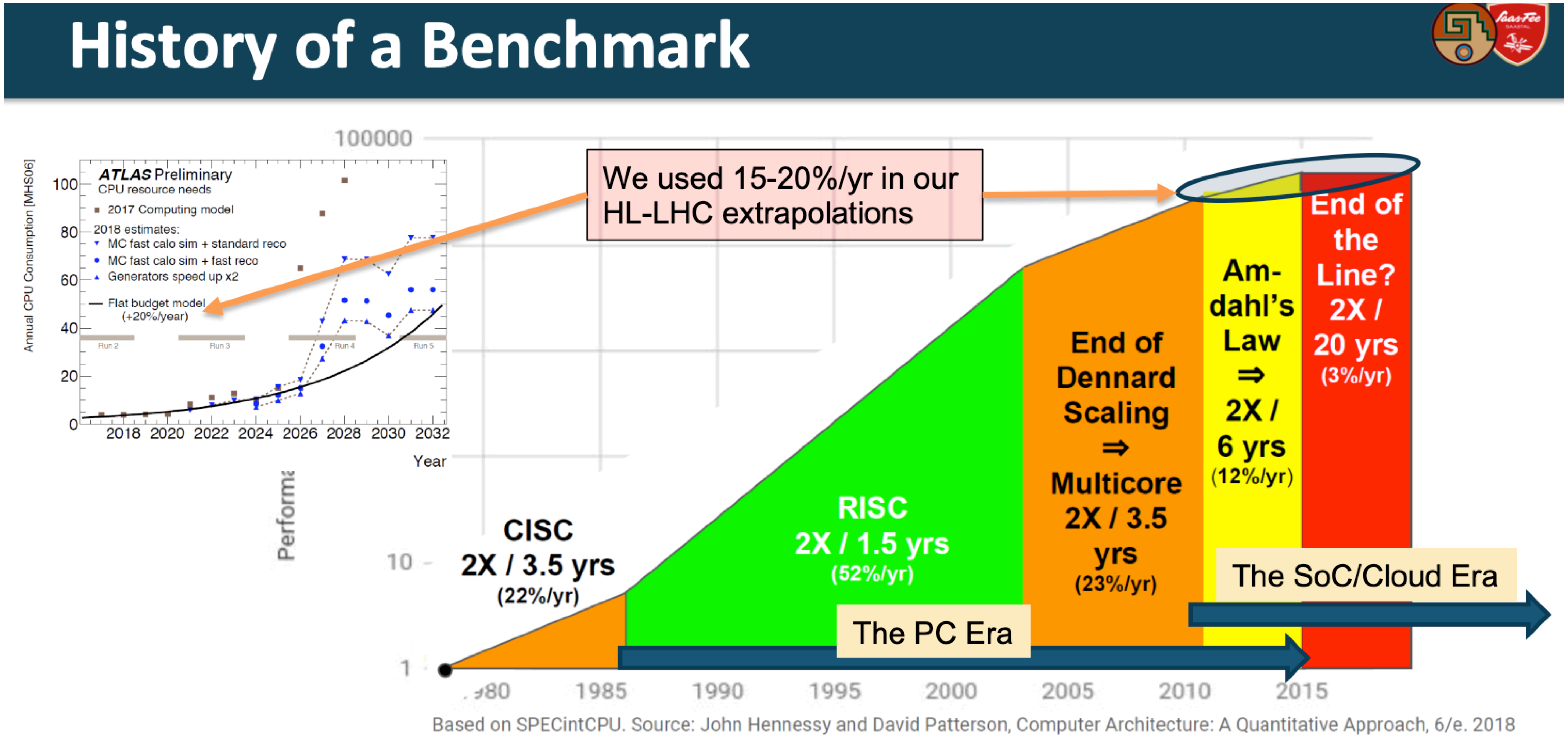


6

ACAT 2019

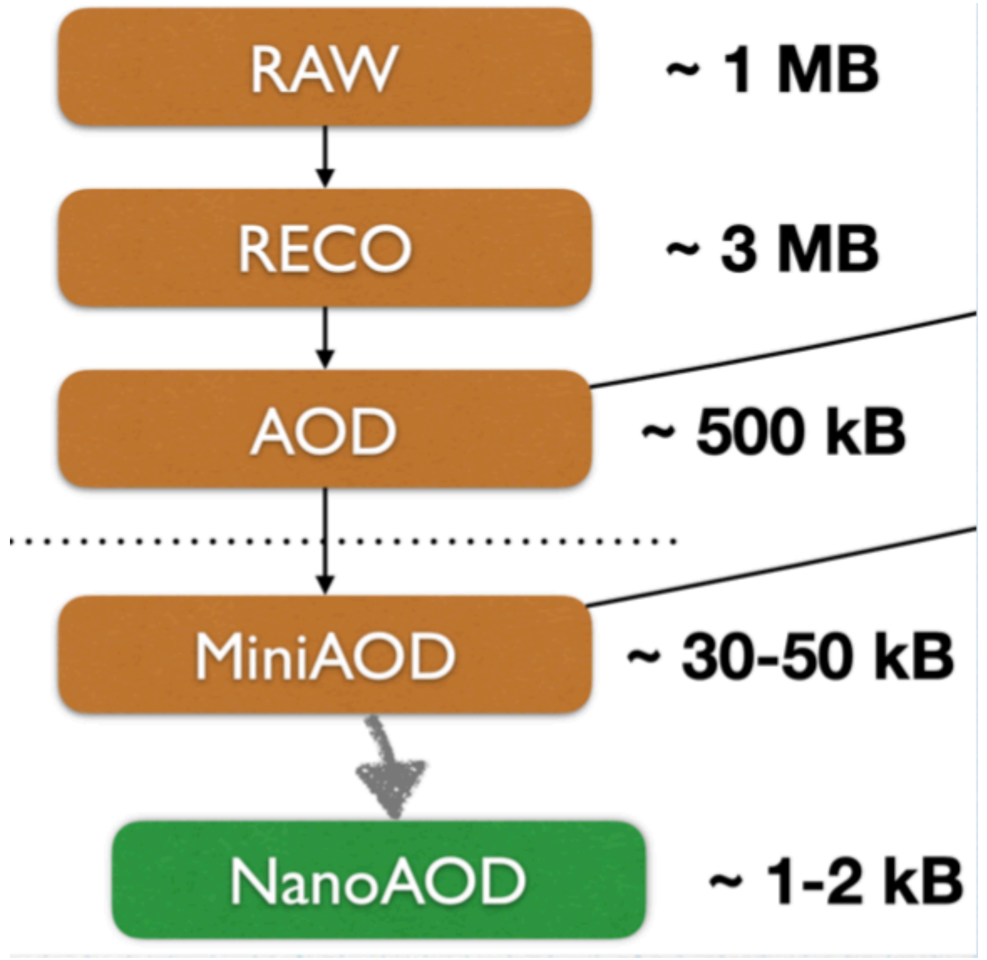
Future Computing Architectures

- P. Calafuria: Computer Architecture in HL-LHC



Possibile Solutions

- CMS try to exploit heterogeneous computing
  - GPU for calorimeter reconstruction at event filter level
  - FPGA (ML inference on FPGA)
  - New compact data format for analyse (NanoAOD)
    - expect 50% of analysis can profit from it



- LHCb
  - use full software trigger at 30MHz (collision rates) to reduce rate to ~ 1kHz
  - 5 times more collisions in the coming Run3

- Use ML algorithms as much as possible and take advantage of deployment on accelerated devices

Track 2

- Data Analysis - Algorithms and Tools
- 36 talks, 39 posters
- A lot of Machine Learning related presentations

Tracking, online (Monday)	Simulation and Event reconstruction (Tuesday)
Statistics, Uncertainty (Wednesday)	New Physics, Cosmic (Thursday)

Session 2 : Tracking

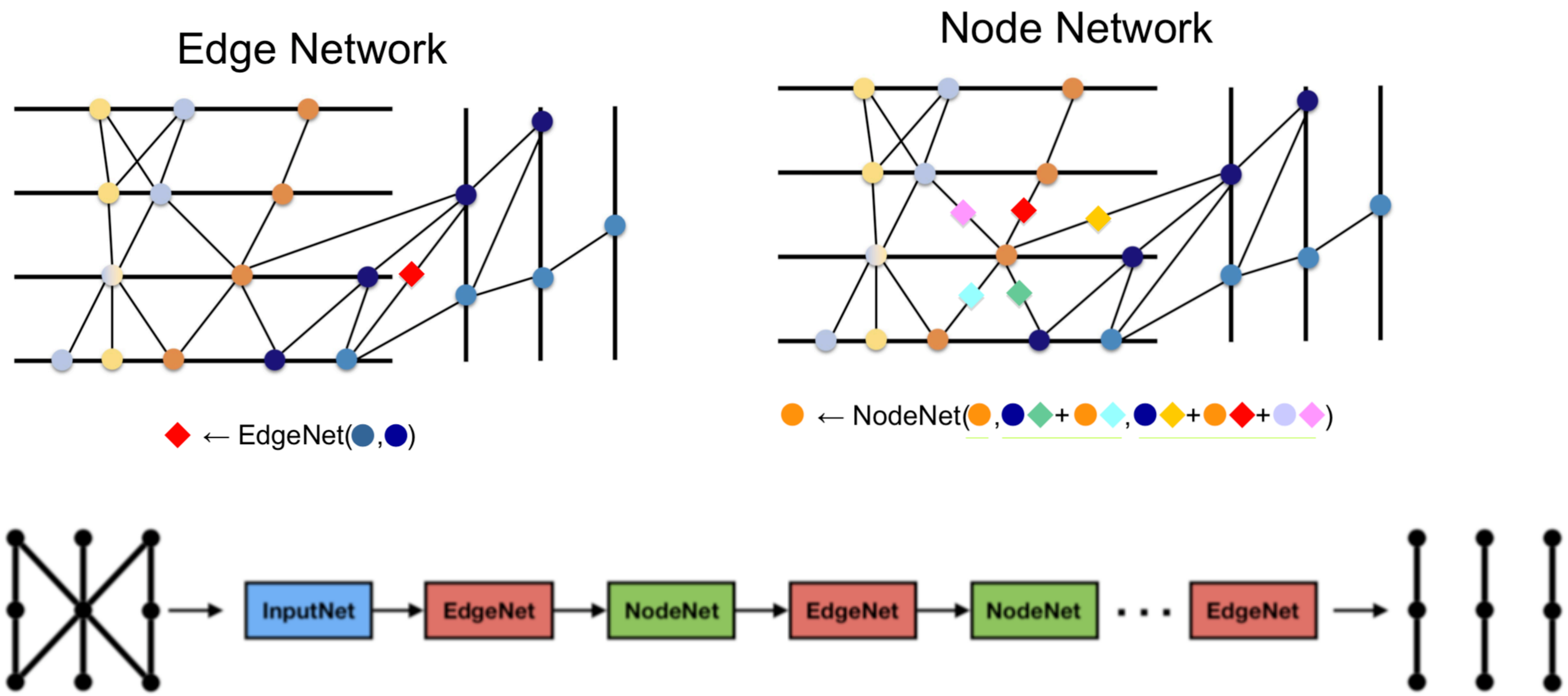
- Tracking session
  - very rich
  - presentations on traditional methods
    - paralleled Kalman filter
    - 3D track finder for Belle2 trigger
  - Machine Learning methods:
    - tracking with Graphics NN
    - Quantum annealing tracking
  - FPGA
    - Deep Learning on FPGA for trigger
    - Real time trading on FPGA for LHCb run 4

Tracking with Graph NN

- New technics for tracking (from HEP.TrX project, *Jean-Roch Vilmant*)
- Use novel methods of ML (Graph Neural Networks)
  - GNN seems better adapted to complex data structures of HEP
  - Seeing lately more and more examples of using GNN

Neural Networks

- **Input Network**
  - Transforms from hit features (r,φ , z) to the node latent representation (N for 8 to 128)
  - Dense : 3→...→N
- **Edge Network**
  - Predicts an edge weight from the node latent representation at both ends
  - Dense : N+N→...→1
- **Node Network**
  - Predicts a node latent representation from the current node representation, weighted sum of node latent representation from incoming edge, and weighted sum
  - Dense : N+N+N→...→N



ACAT 2019

Deep Learning on FPGA

- *Jennifer Ngadiuba*: Deep Learning on FPGA for Trigger and DAQ

With `hls4ml` package we have studied/optimized the FPGA design through:

- **compression:** reduce number of synapses or neurons

NN TRAINING

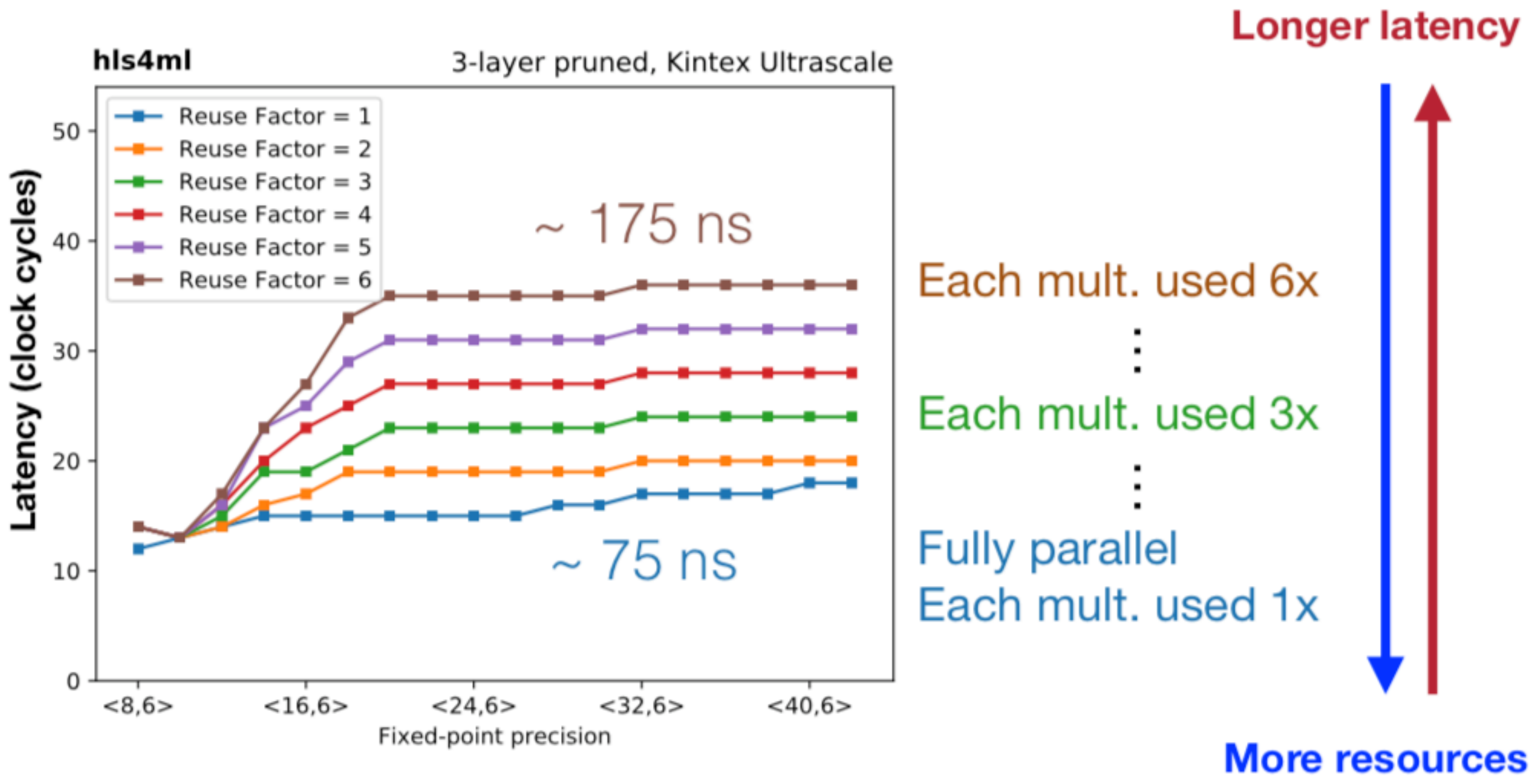
- **quantization:** reduces the precision of the calculations (inputs, weights, biases)

FPGA PROJECT DESIGNING

- **parallelization**: tune how much to parallelize to make the inference faster/slower versus FPGA resources

hls4ml

Machine Learning & FPGAs



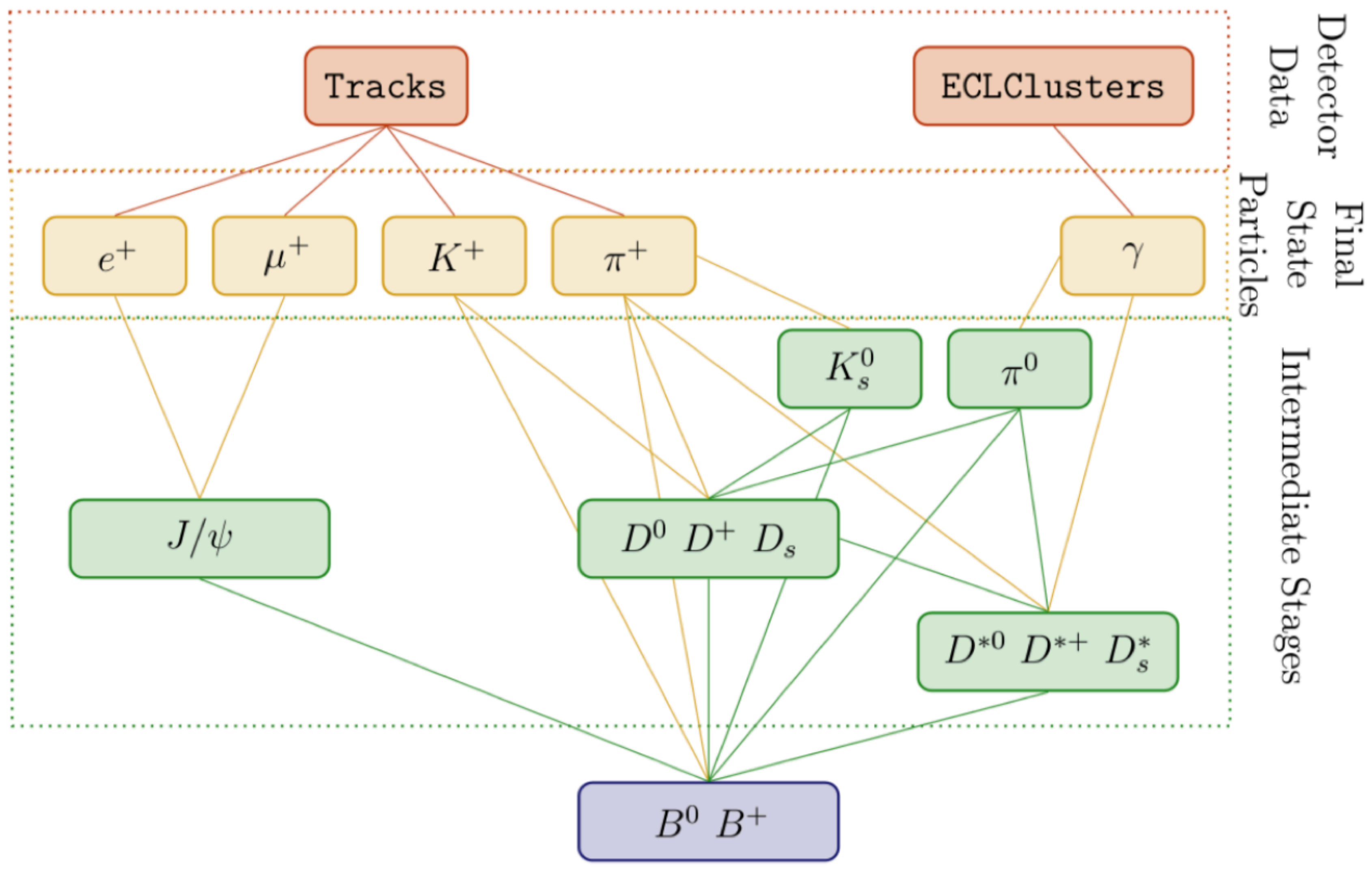
- Simulation and Event Reconstruction
  - Several presentation on Generative Adversarial Networks (GAN):
    - Recurrent GAN for generating list particles (e.g. pile-up generation)
    - GAN for LHCb RICH
    - GAN for simulation of showers of ATLAS TileCal high granularity Upgrade
  - Reinforcement learning
    - Jet grooming (iteratively removing soft radiation from jets)
    - Assignment of reconstructed object to truth-truth-level objects

- Analysis of whole events
  - Full event interpretation at Belle 2
  - Variational auto-encoders for detecting anomalous events

Full Event Interpretation at BELLE2

The Full Event Interpretation

- Utilises O(200) decay channels with classifiers (BDTs) trained for each.
- Reconstructs O(10000) unique decays chains in six stages.



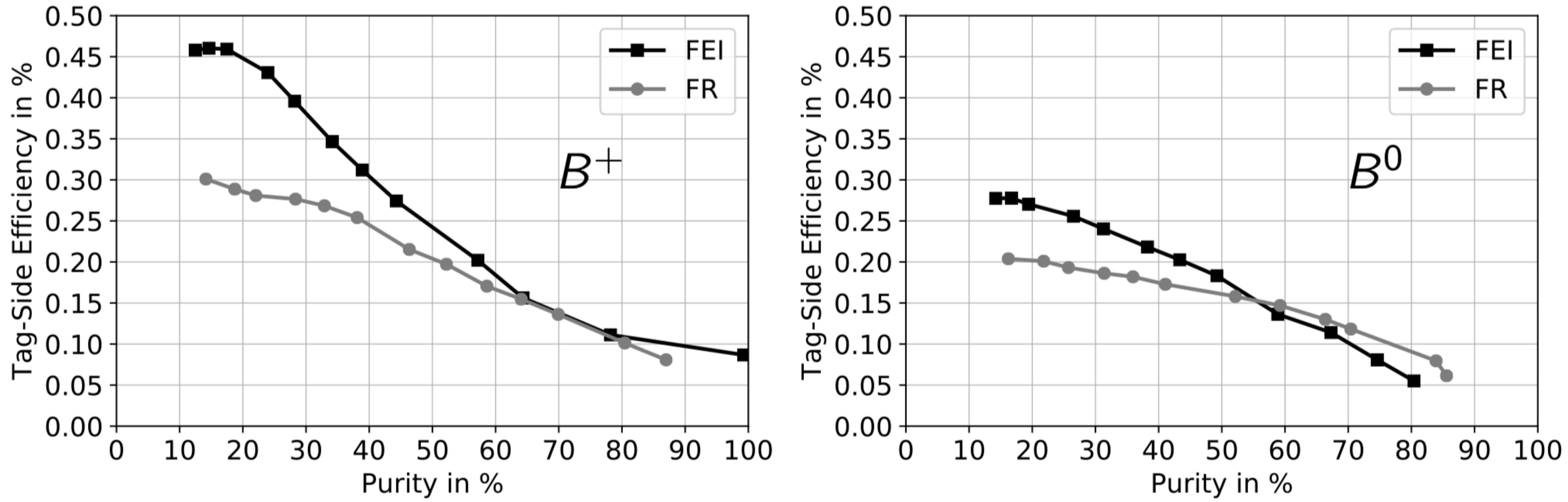
arxiv1807.08680, Keck, T. et al. Keck, T., PhD Thesis

William Sutcliffe | Full Event Interpretation at Belle II | 13 March 2019 | 6 / 20

6 stages of BDT

FEI at BELLE2

Tag-side efficiency again purity in Belle data



Maximum tag-side efficiency				
Tag	FR	SER	FEI Belle MC	FEI Belle II MC
Hadronic B^+	0.28%	0.4%	0.76%	0.66%
Hadronic B^0	0.18%	0.2%	0.46%	0.38%
SL B^+	0.31%	0.3%	1.80%	1.45%
SL B^0	0.34%	0.6%	2.04%	1.94%

FR = Full Reconstruction (Belle Algorithm), SER = Semi-Exclusive Reconstuction (BaBar Algorithm)

Search for $B^+ \to l^+ \nu\gamma$ with the FEI

- First physics analysis performed with the FEI on Belle data.
- Gained a factor of 3 in signal efficiency.

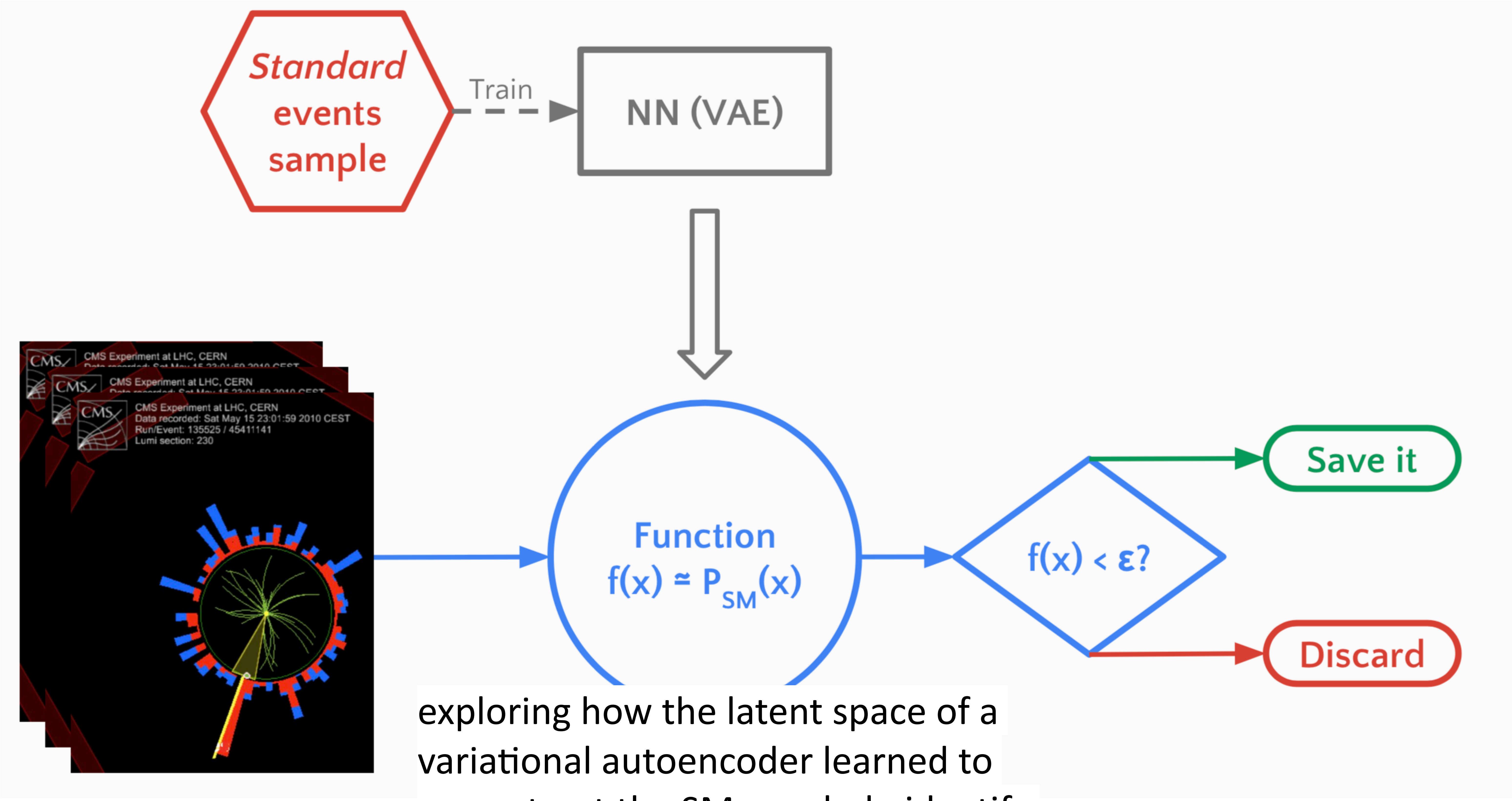
Phys. Rev. D 98, 112016

	$B^+ \to e^+\nu_e\gamma$	$B^+ \to \mu^+\nu_\mu\gamma$	Combined
N_{New}	24.8	25.7	50.5
$N_{\mathrm{Published}}$	8.0	8.7	16.5

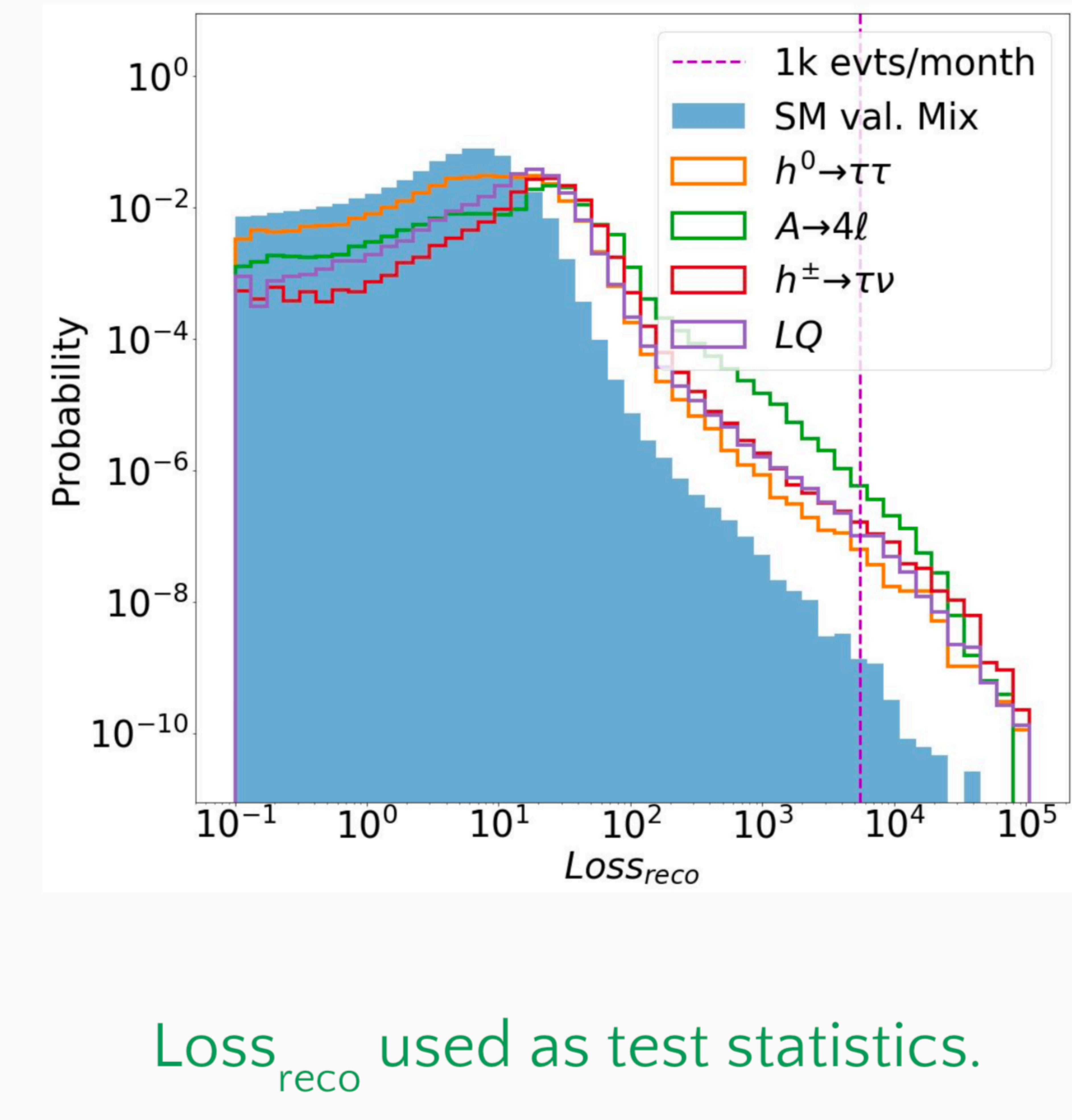
Plan to move to use from BDT Deep Learning

Auto-encoders

- Variational Auto-encoders for un-expected events

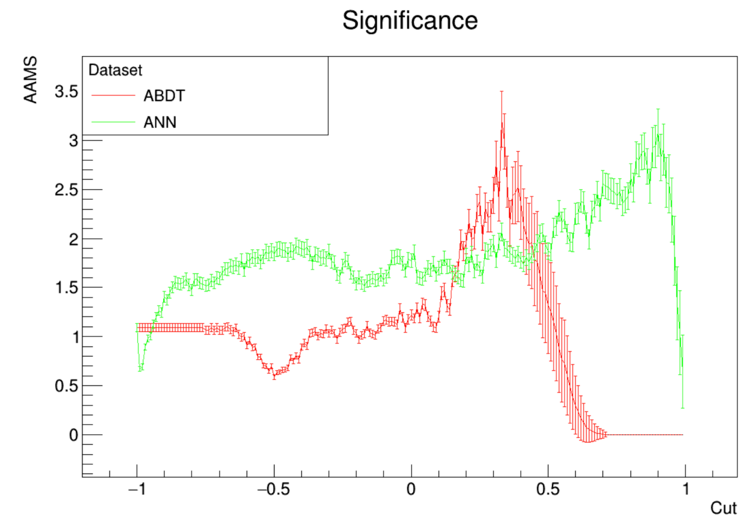


exploring how the latent space of a variational autoencoder learned to reconstruct the SM may help identify anomalous events, as a model-independent trigger



select 30 events / day

- Statistics and Uncertainty
  - several presentation focus more on method for dealing with uncertainties (e.g. systematic) and on statistical issues
    - Incorporation of systematic uncertainty for training of BDT
      - –comparison with adversarial neural network

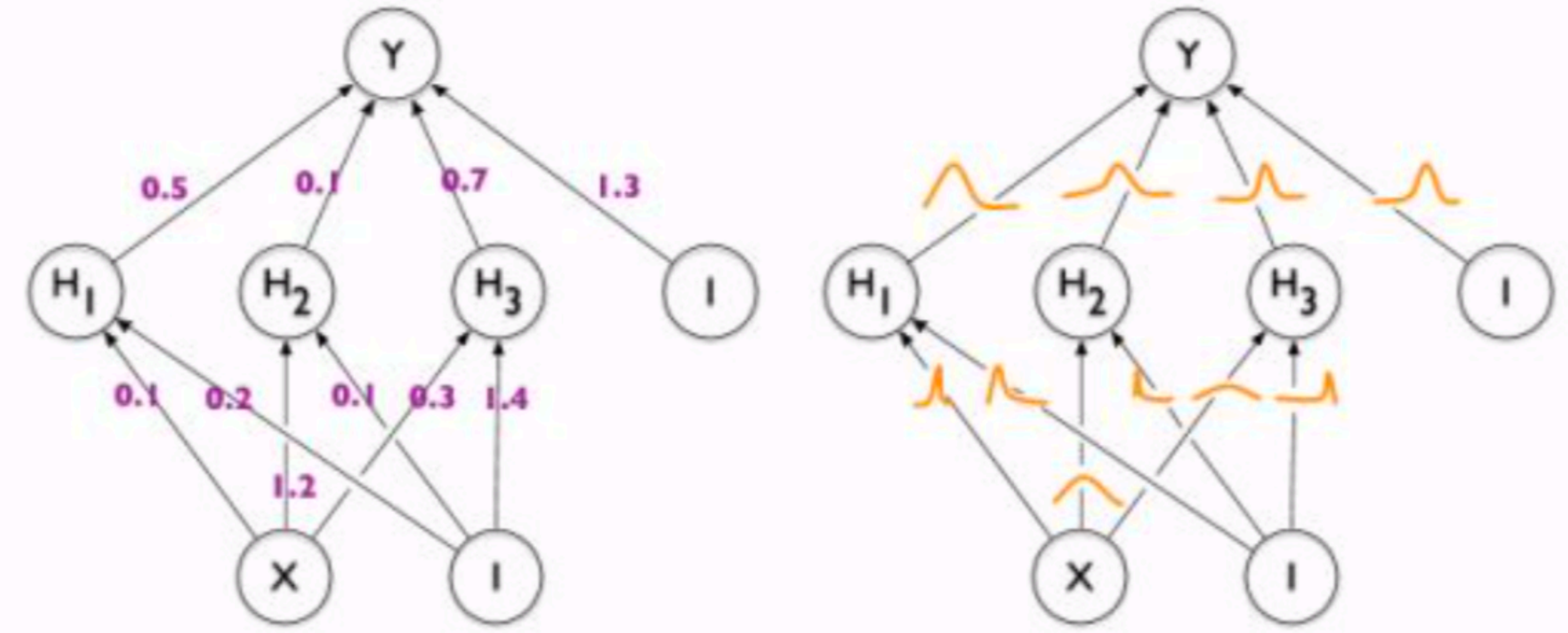


    - adversarial network for deriving data-data-simulation corrections for jet tagging in CMS
    - variational dropout for speed-up network
    - machine learning on s-weighted data set (negative weights)
    - Inference aware Neural optimisation (INFERNO).
      Learn summary statistics by directly minimising an approximation of the expected interval with accounting for the effect of nuisances

Variational Dropout Sparsification for Particle Identification speed-up

A.Ryzhikov suggests using Bayesian NN that gives hints which NN weights could be removed => inference speed increased.

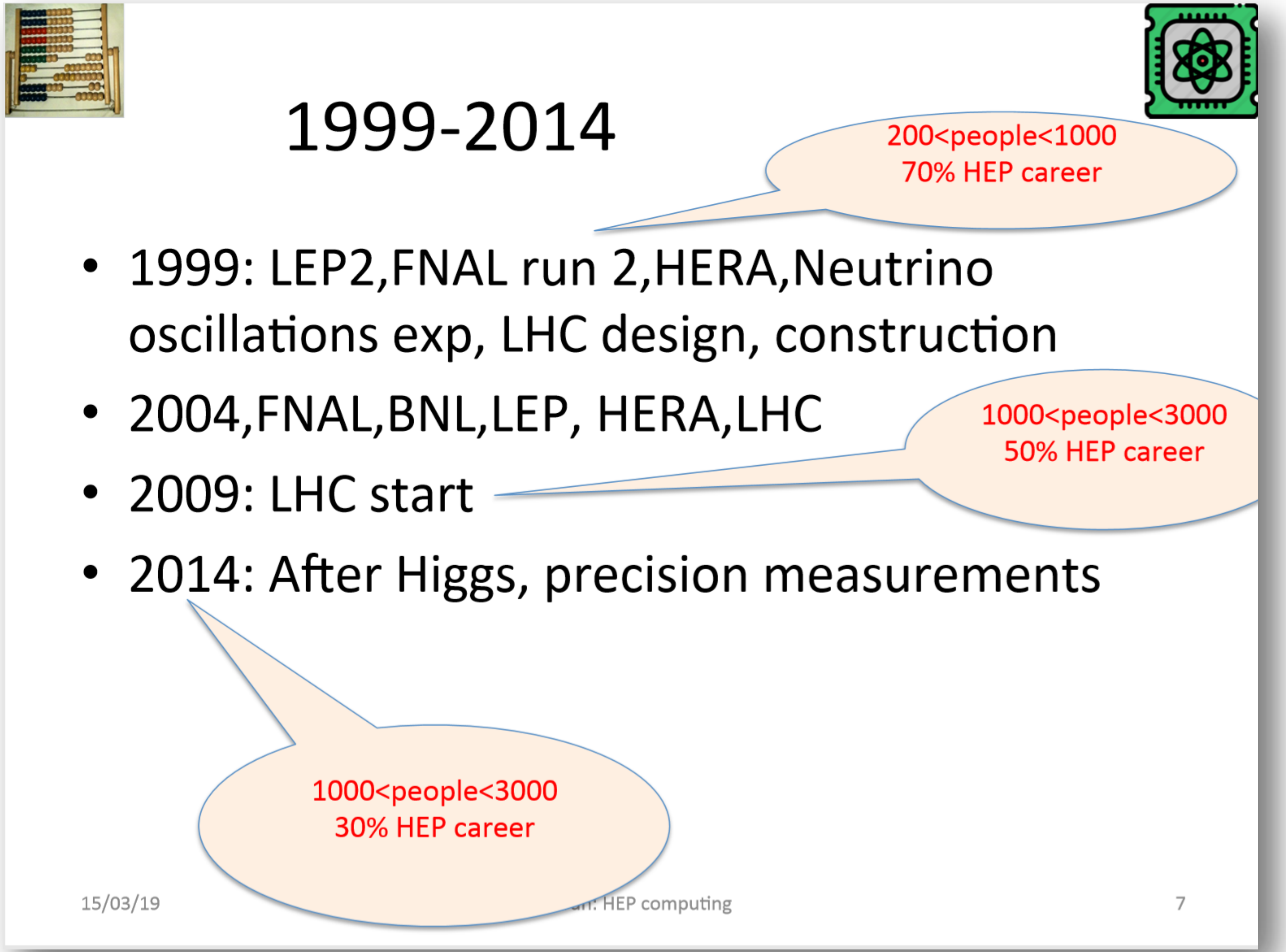
Significant improvement for LHCb PID!



Method	# Neurons	Electron	Ghost	Kaon	Muon	Pion	Proton	Speed-Up
6xDNN	45-48	0.9855	0.9485	0.9148	0.9844	0.9346	0.9178	x1
1xDNN	150	0.9863	0.9570	0.9145	0.9889	0.9463	0.9167	x1
Ternary	Auto	0.9843	0.9435	0.9154	0.9834	0.9352	0.9110	x5
1xDNN	30	0.9871	0.9557	0.9158	0.9893	0.9427	0.9125	x5
BDNN	Auto	**0.9881**	0.9548	**0.9244**	**0.9896**	**0.9509**	**0.9228**	**x16**

Plenaries

- Rene: Future of HEP computing



1999-2014

- 1999: LEP2,FNAL run 2,HERA,Neutrino oscillations exp, LHC design, construction
- 2004,FNAL,BNL,LEP, HERA,LHC
- 2009: LHC start
- 2014: After Higgs, precision measurements

15/03/19 …n: HEP computing 7

Conditions for success

- Top priority: Instantaneous user support
- Understand & prioritize users requirements
- Project members must follow all branches
- Stability & continuity even with revolutions
- Creativity with short term validation
- Code quality: dash boards, coverity, etc
- Do not duplicate user interfaces
- Simplify installation
- Tools for beginners
- Last but not least

- D. Rezende (DeepMind) : Deep Generative Models

Popular Generative Models

Algorithm	Model	Learning Principle
Pixel(C/R)NN	Autoregressive	Maximum-likelihood = ApproximationOf(Bayesian posterior)
Normalizing Flows	InvertibleMap(simple distribution)	Maximum-likelihood
VAEs / Bayesian (deep) networks	Gaussian Latent-variable model	ELBO = ApproximationOf(Maximum-likelihood)
GANs	Gaussian Latent-variable model	ApproximationOf(JS, Wasserstein, Energy)
GPs	GP	Maximum-likelihood(meta params) + Bayesian posterior
Sparse / Deep GPs	GP(GP(GP(...)))	ELBO
Boltzmann Machines	Unnormalized energy function	

This observation is the key insight of GANs, allowing us to train models from which we can sample from but not evaluate its likelihood

Plenaries

- S. Chintala: *Automatic Differentiation and Deep Learning with examples from PyTorch*

Forward-mode autodiff

- Computes Jacobian-vector products
- Each evaluation gives one row of the Jacobian
- Typically used when: dimensionality of y >> dimensionality of x

- Popular software implementations:
  - HIPS/autograd
  - JAX by Google
  - Flux.jl

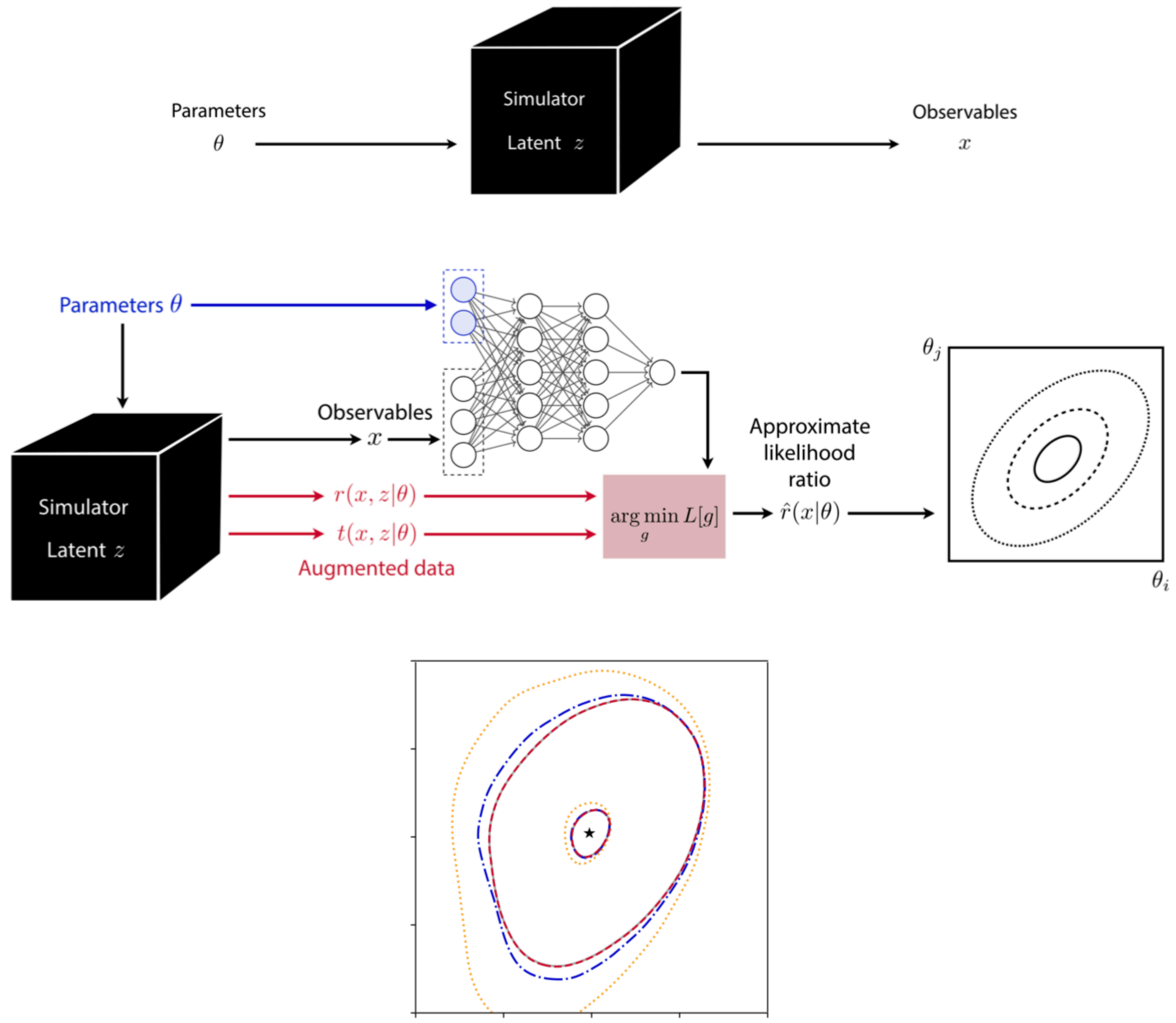
Reverse-mode autodiff

- Computes Vector-Jacobian products
- Each evaluation gives one column of the Jacobian
- Typically used when: dimensionality of x >> dimensionality of y
  - Like in deep learning

- Popular software implementations
  - All deep learning frameworks (PyTorch, TensorFlow, MXNet, Caffe, etc.)
  - HIPS/autograd
  - Jax by Google
  - Flux.jl

Plenaries

- J. Brehemer (NYU): Constraining effective field theories with machine learning
- **A new approach to LHC measurements**



- LHC measurements with high-dimensional observations have "likelihood-free" structure
- New multivariate inference techniques: Leverage information in matrix elements + power of machine learning to…
  - estimate the full likelihood function
  - learn optimal summary statistics
- First tests show potential to substantially increase sensitivity to new physics

39/

ACAT 2019

Conclusions

- Many very interesting presentations
  - both in computing and algorithms sessions
  - seeing more and more different applications of modern machine learning techniques (e.g graph networks, GAN, CNN, VAE, etc…)
  - Many excellent plenary speakers
  - Very good atmosphere: many discussions, questions
    - continue ACAT tradition of being an excellent conference
  - Very well organized conference
    - great location in Swiss mountains
    - nice conference center (nice walking every day to reach conference venue)
    - not the best luck with the weather (windy and lots of snow on Thursday night)
  - Next ACAT will be in Korea (Daejon) in summer 2020