

Precision Timing for CMS Phase 2 Upgrades, Monte Carlo Integration with Deep Learning

Josh Bendavid (CERN, LPC DR)

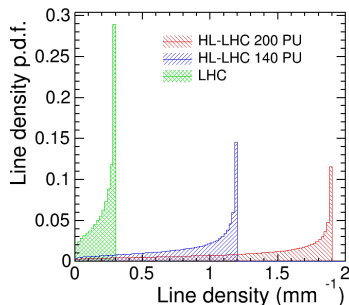
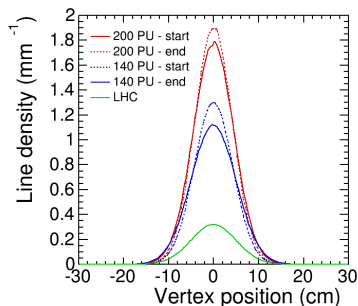


Jul. 27 2017

- **Precision Timing Simulation (Reconstruction) and Physics Performance Studies:**
 - Subgroup lead in Precision Timing WG for Simulation and Physics Performance (WG in transition to interim project)
 - Close collaboration with L. Gray (FNAL), physics studies for interim document in the fall including $H \rightarrow \tau\tau$ analysis impact with K. Kaadze (KSU, LPC DR)
- **Machine Learning Monte Carlo Integration**
 - Use machine learning to improve Monte Carlo integration efficiency in generators beyond what is achievable with VEGAS
 - Synergy with machine learning activities at LPC
 - [MC4BSM 2017 Talk](#) (SLAC)
 - J. Bendavid, “Efficient Monte Carlo Integration Using Boosted Decision Trees and Generative Deep Neural Networks”
<https://arxiv.org/abs/1707.00028>

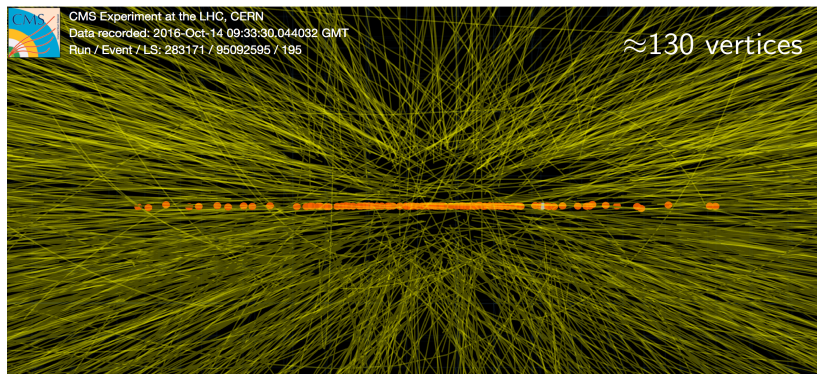
- Other Activities:
 - CMS DAS Convener (Statistics, Same-Sign Dilepton Search)
 - DS@HEP 2017 (Data Science in High Energy Physics): Local Organizing Committee, Speaker (“[Image Calorimetry: Status and Challenges](#)”)
 - Monte Carlo Generators (ME+PS matching), close contact with S. Prestel, S. Mrenna (FNAL)
 - SUSY/Anomalous Higgs Production Searches: Close collaboration with S.Xie (LPC-based, former DR)
 - W-Mass measurement (ongoing): Collaboration with J. Berryhill, A. Apyan (FNAL)
 - Incoming SMP convener
- Recently transitioned from Postdoc at Caltech to Research Staff LD at CERN (since July 1)

HL-LHC Luminous Region



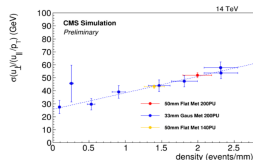
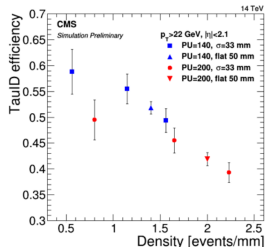
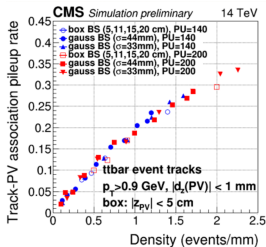
- Maximum pileup/density at HL-LHC determined by **Luminosity Levelling** working point
- Relaxing the constraint from 140→200 PU could extend integrated luminosity over HL-LHC lifetime from 3000→~ 4000 fb⁻¹
- Total amount of pileup and width of luminous region evolve over the course of a fill
- Linear **pileup density** further varies over the luminous region

Proof of Concept, Proof of Challenge



- Real-life event with HL-LHC-like pileup from special run in 2016 with individual high intensity bunches

Pileup Density Impact: Track-PV Association



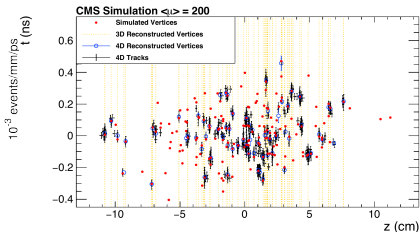
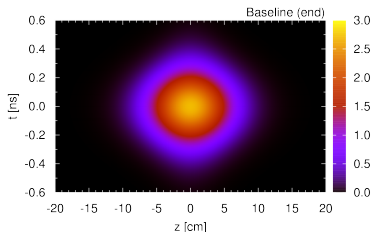
(a) Pileup Tracks in Hard PV

(b) Tau Isolation ϵ

(c) (Puppi) MET Resolution

- Primary vertex and tracks from hard interaction can still be reconstructed efficiently, but pileup contamination increases rapidly as a function of pileup density (pure geometric effect, independent of total pileup)
- **Quantities based on charged particles are currently nearly free of pileup, will not be so at HL-LHC**
- Corresponding effect visible in charged isolation efficiency for leptons, missing energy resolution, among others

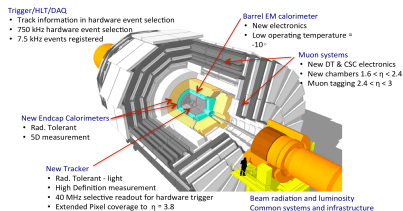
Mitigation with Precision Timing



- Interactions are also distributed in time with a spread of 150-200 ps
- A detector with 10's of ps timing resolution could meaningfully distinguish between interactions on the basis of timing
- With sufficient time resolution and coverage for charged particles, traditional three-dimensional vertex fit can be upgraded to a four-dimensional fit

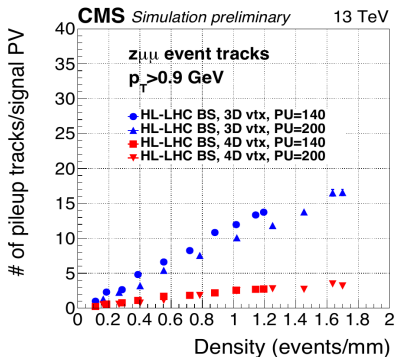
Additional Timing Capabilities

- Calorimeter upgrades can already provide precision timing for high energy photons in the central region, moderate energy photons, and higher energy hadrons in the forward region
- **Additional capabilities: MIP timing to cover large fraction of charged particles in the event**
- **Targeting $\sigma_t = 30$ ps**
- **Extension to Phase-II Upgrade: MIP timing layer**
- Concept for central region: Thin **LYSO + SiPM** layer built into tracker barrel support tube (in between tracker and ECal Barrel)
→ precision timing for charged particles and converted photons

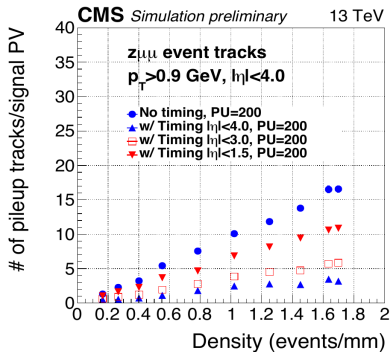


- Concept for forward region (more stringent radiation hardness requirements): **LGAD** (Silicon with Gain), with baseline location as additional final layer of strip tracker

Effect of Precision Timing on Track-PV Association



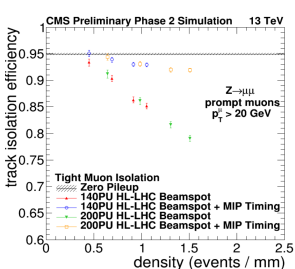
(a) Total # of tracks/hard PV



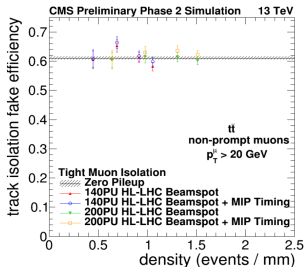
(b) η Coverage Variation

- **~ 5x reduction in effective pileup in terms of charge multiplicity**

Muon Charged Isolation



(a) Prompt μ Efficiency

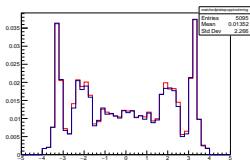


(b) Efficiency for fakes

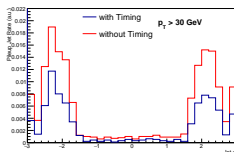
- Muon charged isolation efficiency in $Z \rightarrow \mu\mu$ (prompt) and $t\bar{t}$ (fake) events
- Efficiencies plotted at fixed working point (but applying timing does not increase the number of fakes within present uncertainties)

Pileup Jet Suppression

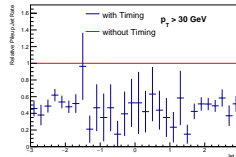
- Factor of two or better suppression of pileup jet rate across timing layer acceptance
- Efficiency for real signal jets (generator-matched) \sim unaffected
- **Timing information currently used only for charged particles, to be extended to photons in HGC, converted photons in BTL, relevant neutral hadrons in HGC in the future.**



(a) Signal Jets



(b) Pileup Jets

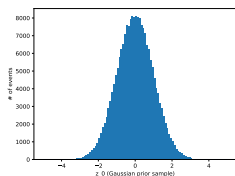


(c) Pileup Jets Ratio

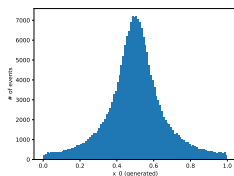
- **Monte Carlo integration:** Given an arbitrary/black box multidimensional function $f(\vec{x})$, find the integral $\int f(\vec{x})d\vec{x}$
- **Monte Carlo generation:** Given an arbitrary/black box multidimensional function $f(\vec{x})$, generate an unweighted set of vectors \vec{x} with a probability density $p(\vec{x}) = f(\vec{x}) / \int f(\vec{x})d\vec{x}$
- Typical HEP use case: Given a numerical implementation for a matrix element fully differential in incoming/outgoing four-vectors, compute the total cross section (integral), and generate a set of unweighted events

- Canonical approach: **Importance Sampling**: Construct an easily sampled from approximation to the target function
 - VEGAS: Product of adaptively-binned 1D histograms
 - FOAM: Sampling from a (single) binary decision tree → phase space divided into hyper-rectangles with optimized boundaries
 - BDT: Sampling from an additive series of decision trees → Gradient boosting used to improve performance wrt FOAM just like in classification and regression problems
 - DNN: Sampling from a generative deep-neural network

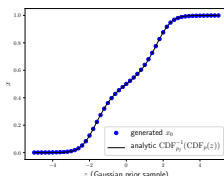
1D DNN Example with Analytic Solution



(a) Prior



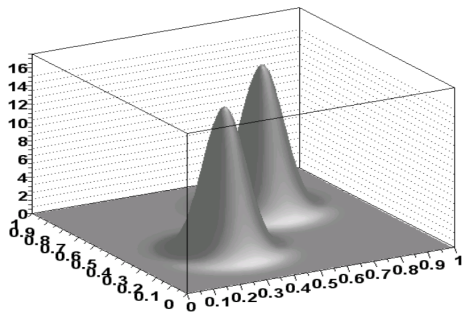
(b) Generated



(c) Generated vs Prior

- In 1D the generative network is essential just learning the inverse CDF of the target distribution (numerically)
- Technically the function is $x = CDF_{p_f}^{-1}(CDF_p(z))$
- For Cauchy distribution in this example, this can be computed analytically and compared to the trained DNN result

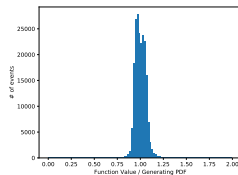
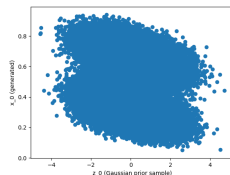
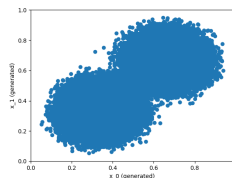
Monte Carlo Integration and Generation: Example Function



S. Jadach, physics/0203033

- This is the “camel” function from the original VEGAS paper, which can be generalized to N dimensions
- Factorized approach will not work well
- Significant low-density regions which cannot be easily excluded a-priori

DNN 4D Camel Function Example



(a) Generated (2D Slice) (b) Generated vs Prior (c) Integration Weight (1D pair)

- The multidimensional case can be considered a generalization of inverse CDF sampling
- This model has 17,220 free parameters

Some results - 4D Camel Function Integration

- Comparing Vegas, Foam, GBRIntegrator, Generative DNN for 4-dimensional camel function (since this appears in both VEGAS and Foam papers).
- Given relative weight variance $\sigma_w / \langle w \rangle$ after training/grid building, relative uncertainty on integral evaluated with N additional events is $\sigma_I / I = \frac{1}{\sqrt{N}} \sigma_w / \langle w \rangle$

Algorithm	# of Func. Evals	$\sigma_w / \langle w \rangle$	σ_I / I (2e6 add. evts)
VEGAS	300,000	2.820	$\pm 2.0 \times 10^{-3}$
Foam	3,855,289	0.319	$\pm 2.3 \times 10^{-4}$
GBRIntegrator	300,000	0.082	$\pm 5.8 \times 10^{-5}$
GBRIntegrator (staged)	300,000	0.077	$\pm 5.4 \times 10^{-5}$
Generative DNN	294,912	0.083	$\pm 5.9 \times 10^{-5}$
Generative DNN (staged)	294,912	0.030	$\pm 2.1 \times 10^{-5}$

- 3x smaller weight variance to foam with 10x less function evaluations
- For this particular function VEGAS performance saturates at relatively poor weight variance

Some results - 9D Camel Function Integration

- Comparing Vegas, GBRIntegrator, Generative DNN for 9-dimensional camel function

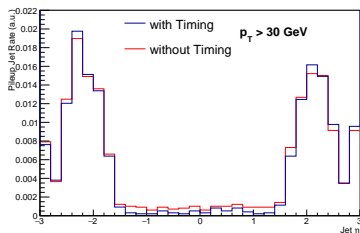
Algorithm	# of Func. Evals	$\sigma_w / \langle w \rangle$	σ_I / I (2e6 add. evts)
VEGAS	1,500,000	19	$\pm 1.3 \times 10^{-2}$
GBRIntegrator	3,200,000	0.63	$\pm 4.5 \times 10^{-4}$
GBRIntegrator (staged)	3,200,000	0.31	$\pm 2.2 \times 10^{-4}$
Generative DNN	294,912	0.15	$\pm 1.1 \times 10^{-4}$
Generative DNN (staged)	294,912	0.081	$\pm 5.7 \times 10^{-5}$

- 50x smaller weight variance to Vegas with 2x function evaluations (BDT)
- DNN approach scales much better with dimensionality (> 100x smaller weight variance than Vegas with 5x **fewer** function evaluations)

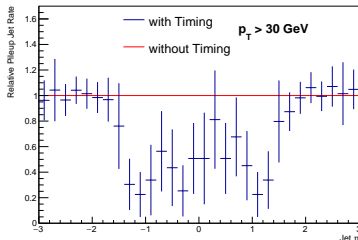
- **Precision Timing Simulation and Performance Studies**
 - Preliminary performance studies show large gains in pileup suppression for Jets, lepton isolation, etc
 - Migrate performance studies to full simulation geometry of timing layers
 - Integration with track reconstruction, track-calorimeter matching, etc
 - Further integration into high level reconstruction and physics objects
 - Impact of object-level improvements on high-level physics analyses for precision timing Interim Document in the fall
- **Machine Learning Monte Carlo Integration**
 - Large improvements with novel algorithms already demonstrated on test cases
 - Exploring alternative DNN architectures including auto-regressive models and convolutional elements
 - Integration into Madgraph_aMC@NLO and tests with QCD matrix elements in progress
- **Thank you for your attention and support!**

Pileup Jet Suppression - Reduced $|\eta|$ Coverage

- Reducing coverage to $|\eta| < 1.5$, “local” behaviour of timing for jets as expected



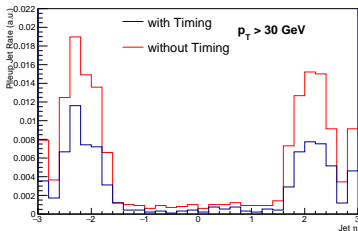
(a) Pileup Jets



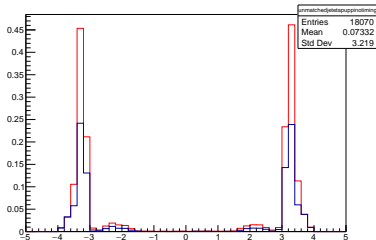
(b) Pileup Jets Ratio

Pileup Jet Suppression - Extended $|\eta|$ Coverage

- Extending coverage to full tracking acceptance, “local” behaviour of timing for jets as expected, large potential gains (but tracking, particle flow, and puppi likely still not optimal in forward region)



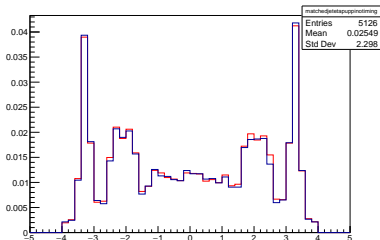
(a) Pileup Jets



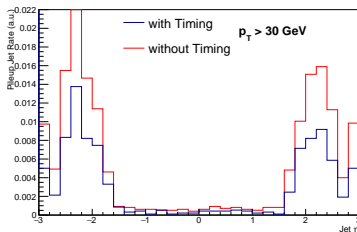
(b) Pileup Jets

Pileup Jet Suppression - Alternate Track-Vtx Assignment Logic

- Repeating study with PV Assignment Algo for track-vtx assignment instead of $|\Delta z|, |\Delta t|$ (qualitatively similar behaviour, more detailed comparisons in progress)

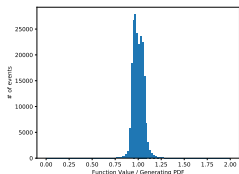


(a) Signal Jets

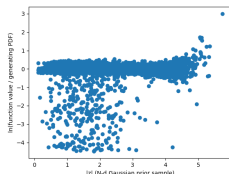


(b) Pileup Jets

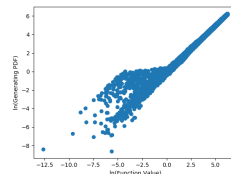
Some Diagnostic Plots - 4D Generative DNN



(a) Integration Weight



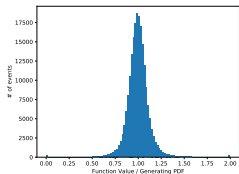
(b) $\ln W$ vs $|z|$



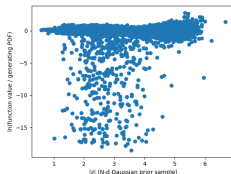
(c) Gen PDF $g(\bar{x})$ vs $f(\bar{x})$

- Reasonably good behaviour
- Some tails in weight distribution in the tails of the prior distribution (biasing the sampling of the prior is straightforward if needed)
- Generating pdf $g(\bar{x})$ tracks the target function $f(\bar{x})$ down to low values, and then sometimes overshoots (not a big issue for either integration or un-weighting)

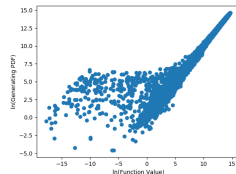
Some Diagnostic Plots - 9D Generative DNN



(a) Integration Weight



(b) $\ln W$ vs $|z|$



(c) Gen PDF $g(\bar{x})$ vs $f(\bar{x})$

- Qualitatively similar behaviour to $4D$ case
- This generative model has 17,865 free parameters