Massachusetts Institute of Technology

# Underlying Event Subtraction in Heavy Ion Collisions

Yue Shi Lai

MIT LNS

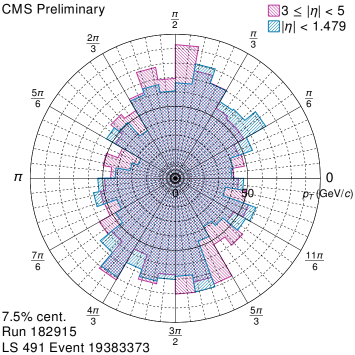
Mitigation of PU Effects at the LHC

# Outline

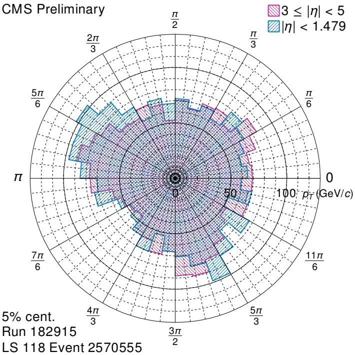
1. The “old” CMS iterative pile-up algorithm
2. The new CMS HF/Voronoi algorithm

- Emphasis is on pushing out physics results:
  - Little particle-level studies, write-up, public code
  - A lot of CMS GEANT results
  - Re-reconstruction campaign for 2011 PbPb data
  - ... plus first physics result to be shown next week (Quark Matter 2014 in Darmstadt, Germany)

# Heavy Ion Underlying Event



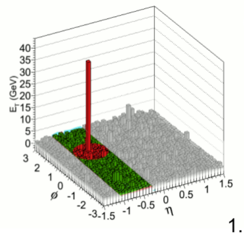

2nd harmonic

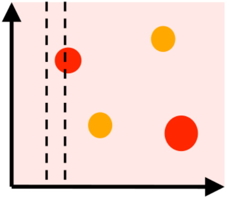



3rd harmonic

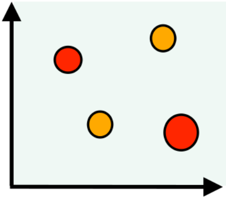
- $3 \leq |\eta| < 5$ corresponds to the CMS forward hadronic calorimeter (HF)
- Underlying event (UE) is not azimuthally symmetric, but has higher Fourier components correlated over $\eta$
- The constant UE is parametrized in heavy ion vs. the impact parameter $b$, quoted in percentage of total cross section (0% $\Rightarrow b \approx 0$, maximum UE; 100% $\Rightarrow$ pp-like, $b \to \infty$)

# CMS Iterative PU in Picture

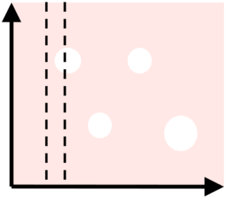


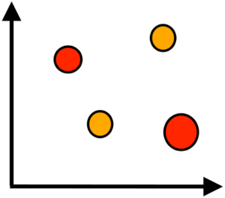
1. $\langle E_T \rangle$ calculated in strips of η. Subtract $\langle E_T \rangle + \sigma$



2. Run anti-$k_T$ algorithm on background-subtracted towers



3. Exclude reconstructed jets and re-estimate background



4. Re-run anti-$k_T$ algorithm to get final jets

For details see:

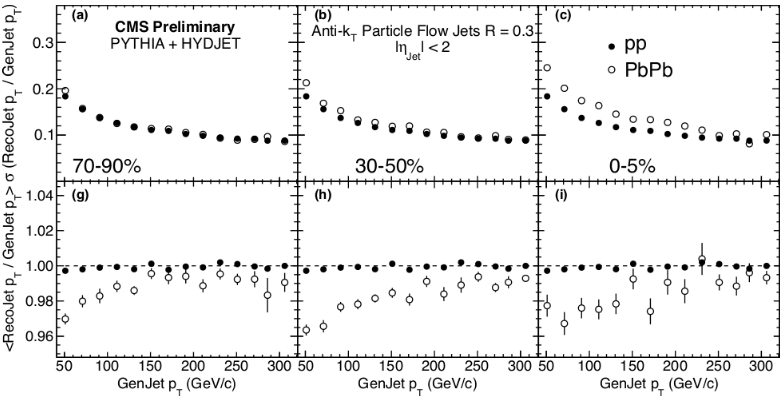
- CMS, arXiv:1102.1957
- Kodolova et al., EPJC 50 (2007) 117

(M. Nguyen, Quark Matter 2012)

# CMS Iterative PU in Detail

- Subtraction on "pseudotower" (= HCAL granularity) level
- Iteration 1:
  - Determination of $\langle p_T \rangle$ and $\sigma(p_T)$ of the UE
  - Subtract $\langle p_T \rangle + \texttt{nSigmaPU} \times \sigma(p_T)$ ("noise suppression"), zero negative cells
- Iteration 2:
  - Exclude towers within `radiusPU` of a jet from iteration 1, with $p_T > \texttt{jetPtMin}$, redetermine $\langle p_T \rangle$ and $\sigma(p_T)$
  - Subtract $\langle p_T \rangle + \texttt{nSigmaPU} \times \sigma(p_T)$
- `nSigmaPU` serves dual purpose of actual noise suppression and a nonlinearity correction (caused by detector zero suppression).

# CMS Iterative PU Performance



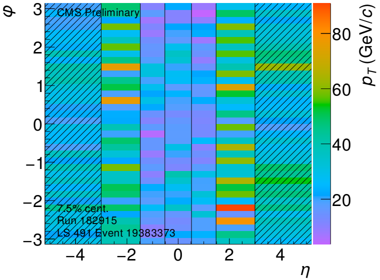

- Jets are always oversubtracted by $\approx$ 2 GeV$/c$
- Performance even poorer for calorimetric jets

CMS PAS-HIN-12-004

# CMS HF/Voronoi

- Really two separate problems:
  1. What amount of UE to subtract
  2. How to subtract
- Solution
  1. Singular vector decomposition (SVD) modeling of the UE
  2. Subtraction by Voronoi diagram
  3. Equalization of positive/negative fluctuation
- Advantages:
  - Arbitrary $v_n$ subtraction (not limited to $n = 2$ in data, but currently limited by MC)
  - Full particle flow hard event, not restricted to reconstructed jets
    - Jet ID, substructure
    - Event shape, missing $E_T$
    - Photon isolation
  - Best PbPb JER currently observed in CMS GEANT

# Step 1: SVD Modeling of the Underlying Event




- Predict midrapidity UE from measured UE in the forward detectors (little jet contamination)
- Forward calorimeter have vastly different radiation depth and segmentation
- Need to model the nonlinear difference

CMS DP-2013/018

# Step 1: SVD Modeling of the Underlying Event

- Goal (RHS) is the Fourier expansion for an arbitrary midrapidity range

$$\frac{dp_\mathrm{T}}{d\eta d\phi} = \mathrm{Re}\left(\sum_{k=0}^{N_\mathrm{Fourier}} a_k e^{ik\phi}\right) \quad (1)$$

- Approximate (predict) $a_k$ as function of the forward HCAL (HF) Fourier decomposition

$$a_k = f_k(a_0^\mathrm{HF}, a_1^\mathrm{HF}, \ldots, a_{N_\mathrm{Fourier}}^\mathrm{HF}). \quad (2)$$

- $f_k$ is expressed as

$$f_k = \sum_{k=0}^{N_\mathrm{Fourier}} \sum_{l=0}^{N_\mathrm{order}} c_l \frac{a_k^\mathrm{HF}}{s_k} \quad (3)$$

($s_k$ are numerical factors to scale 100 GeV$/c \lesssim a_k^\mathrm{HF} \lesssim$ 5000 GeV$/c$ to $\approx 1$)

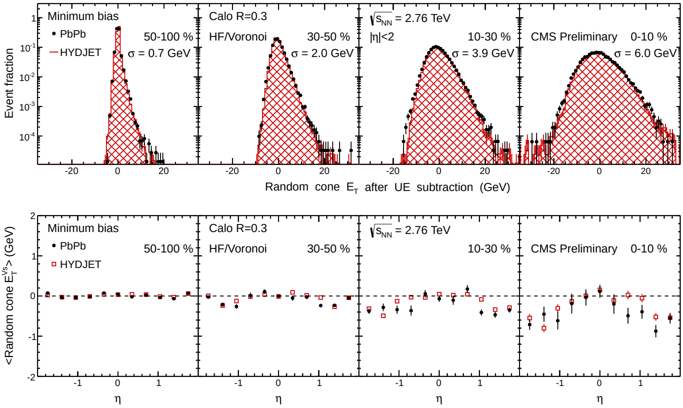
# Step 1: SVD Modeling of the Underlying Event

- Regression problem solved by SVD

$$\mathbf{f} = \mathbf{C}\hat{\mathbf{a}}^{\mathrm{HF}}, \quad \hat{\mathbf{a}}^{\mathrm{HF}} = \left(\frac{a_k^{\mathrm{HF}}}{s_k}\right)_k . \tag{4}$$

- This structure is intentionally chosen to be extended later to midrapidity "extrapolation beneath the jet" $\Rightarrow$ e.g. for pp PU.

# Data vs. MC Performance



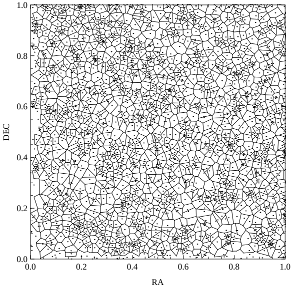

- For data, a separate training is used in step 1 (CMS GEANT is not accurate enough in forward region)
- The 0–10% there is a mean offset $\approx$ 50 GeV$/c$ being subtracted
- Some fluctuation difference mainly from the MC event generator tuning

CMS PAS-HIN-14-010

# Step 2: Subtraction by Voronoi Diagram

- Existing methods to avoid this problem:
  1. Cluster event including UE, remove soft scale at jet level (FastJet, ATLAS iterative mean)
  2. Convert final-state particles into a density, make the jet reconstruction commutative wrt. UE subtraction (RHIC-PHENIX Gaussian filter)
  3. Subtraction using fixed geometry (CMS iterative pile-up)
     Notice that this method also attempts produce a clustering that is restricted to the hard event.
- For particle-flow, need to extend (3) to:
  4. Associate an area to each final state particle (HF/Voronoi the first algorithm to attempt it)

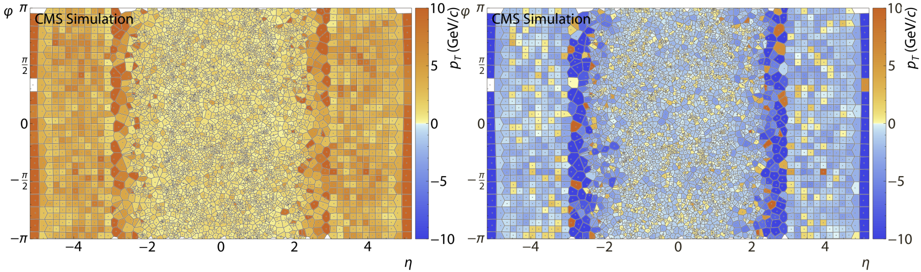
# Step 2: Subtraction by Voronoi Diagram




M. Soares-Santos et al., *Astrophys. J.* 727, 45 (2011)

- Density estimation by Voronoi diagram (point-like galaxies vs. mostly empty space) established for more than two decades in observational cosmology

# Step 2: Subtraction by Voronoi Diagram




Before UE subtraction

After UE subtraction

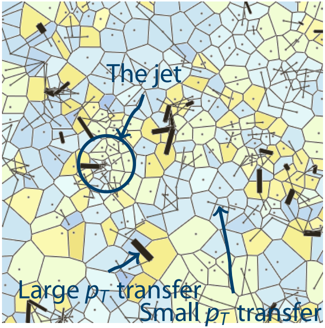
- Subtraction algorithm matches the particle position with the area where the nearest neighbor of a given point is that particle-flow candidate
- Note the 2nd Fourier component with $\Psi^{EP} \approx 0$ that is subtracted away

CMS DP-2013/018

# Step 3: Equalization

- An idea that is inspired from ATLAS early "noise tool" (now abandoned)
- Try to get a right "noise tool" without the issues how ATLAS implemented it:
  - Make it non-iterative
  - Define the goal as same energy at given (jet algorithm) distance scale, and not just remove negativity

# Step 3: Equalization in Pictures




- Ingredient #3: “Equalization” to produce positive event
- Purely a band-aid for iterative, positive-$p_T$ jet reconstruction algorithms
- Balance out negative cells to represent the same event using positive cells only
- Implement a delocalized underlying event on particle level

- Similar in idea to a mosaic:
  - Manipulation of color tiles on a small scale
  - “Reconstructs” to a correct visual impression on a coarse scale

# Step 3: Equalization in Equations

- Set of variables describing the equalization problem:
  1. Geometry
     - $R_{kj}$ is the distance between cells $k, j$
  2. Energy/energy transfer
     - $p_j$ is the $p_T$ within a positive cell $j$
     - $\hat{p}_j$ is the equalized $p_T$
     - $n_k$ is the $p_T$ within a negative cell $k$
     - $t_{kj}$ is the energy transfer cell $k \leftarrow j$
  3. "Service variables"
     - $c$ is minimax cost variable
     - $u$ is the globally subtracted residual energy

# Step 3: Equalization in Equations

- Rewrite this as linear optimization

$$
\begin{aligned}
\underset{c,t_{kj}}{\text{minimize}} \quad & c + \varepsilon \sum_{k,j,\Delta R_{kj}<R_{\text{jet}}} t_{kj} \Delta R_{kj} \\
\text{subject to} \quad & \sum_{R_{\text{jet}}} \hat{p}_i - \sum_{R_{\text{jet}}} p_i + \sum_{R_{\text{jet}}} n_i \leq c \quad \text{around all positive regions} \\
& \sum_{R_{\text{jet}}} \hat{p}_i - \sum_{R_{\text{jet}}} p_i + \sum_{R_{\text{jet}}} n_i \geq -c \quad \text{around all positive regions} \\
& p_j - \sum_k t_{kj} - u = \hat{p}_i \quad \text{for positive cells } p_j \\
& n_k + \sum_j t_{kj} - u \leq 0 \quad \text{for negative cells } n_k \\
& c \geq 0 \\
& \hat{p}_i \geq 0 \\
& t_{kj} \geq 0 \\
& u \geq 0
\end{aligned}
$$

# Step 3: Equalization in Equations

- Target function

$$\underbrace{c}_{p_T \text{ nonclosure}} + \underbrace{\varepsilon \sum_{k,j,\Delta R_{kj} < R_{\text{jet}}} t_{kj} \Delta R_{kj}}_{\text{Energy transfer cost over distance}} \tag{6}$$

- The $\varepsilon$ term is a "tie-breaker" to handle the degeneracy, i.e. recombination over large distances are (slightly) more costly. Value e.g. $\varepsilon = 10$ MeV (hide it in a $p_T$ range we don't really care about).
- The minimax condition

$$-c \leq \underbrace{\sum_{R_{\text{jet}}} \hat{p}_i - \sum_{R_{\text{jet}}} p_i + \sum_{R_{\text{jet}}} n_i}_{p_T \text{ nonclosure}} \leq c \tag{7}$$

automatically finds the worst nonclosure and attempts to remove it

- With the CMS PbPb nonlinearity the minimax condition is modified to:

$$-c \leq \ldots \leq \frac{c}{4} \tag{8}$$

# Step 3: Equalization in Equations

- Energy transfer and residual subtraction

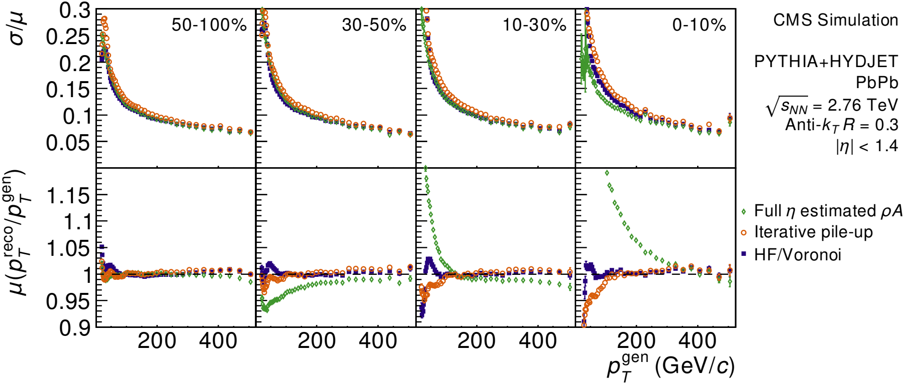
$$
\begin{aligned}
\underbrace{p_j - \sum_k t_{kj}}_{p_T \text{ after transferring out}} &- \underbrace{u}_{\text{residual subtraction}} \underbrace{= \hat{p}_i}_{\text{positive-only event } p_T} && \text{for positive cells } p_j \\
\underbrace{n_k + \sum_j t_{kj}}_{p_T \text{ after transferring in}} &- \underbrace{u}_{\text{residual subtraction}} \underbrace{\leq 0}_{\text{any } p_T \text{ that will be truncated to 0}} && \text{for negative cells } n_k
\end{aligned}
\tag{9}
$$

- $\hat{p}_i$ is the solution
- $c, \hat{p}_i, t_{kj}, u$ all variables that are positively constrained
- In practice we segment the event into staggered tiles to allow different $\eta$ region having different nonclosure (staggered tiling is the “cost trick” applied twice)

# Free Parameters

- Step 1: Many ($\approx$ 50k) parameters in the UE model
- Step 2–3: Exactly two parameters
  - $R_{\text{jet}}$, same as in the jet reconstruction algorithm
  - The $\frac{1}{4}$ that is for the CMS nonlinearity

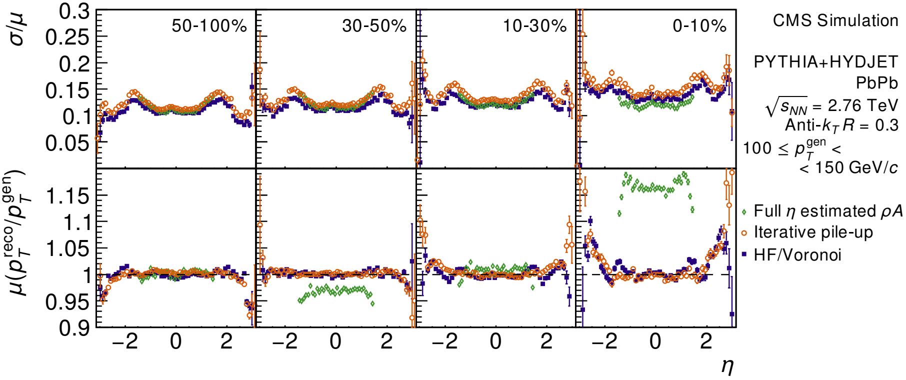
# MC Generator + GEANT Performance: Vs. $p_T^{truth}$




- $\approx$ 10% improved mid-rapidity jet energy resolution for the new HF/Voronoi algorithm vs. old CMS' iterative pileup (PU)
- Median-$\rho A$ is the "global rescaled" variant (there is a still-CMS-internal version with FastJet-"doughnut")

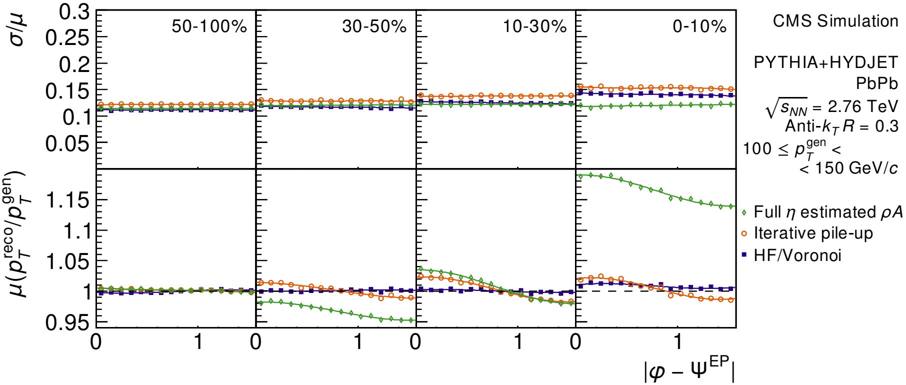
CMS DP-2013/018

# MC Generator + GEANT Performance: Vs. $\eta$




- Nonclosure in 0–10%/endcap due to factorization between $\eta$ and centrality dependent JEC

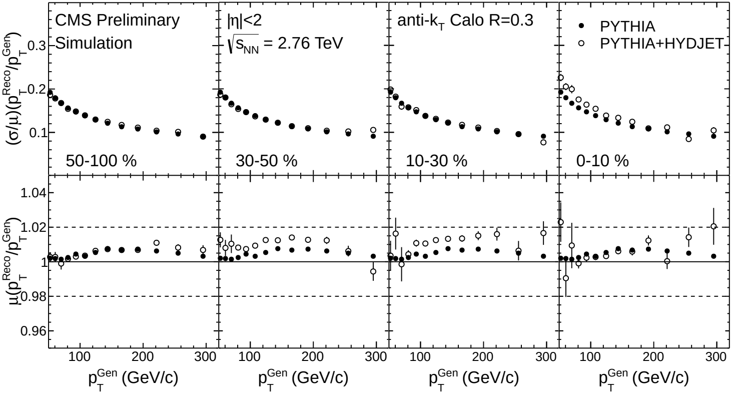
CMS DP-2013/018




- $\Psi^{EP}$ is the arctan of the 2nd Fourier coefficients of $dE_T/d\varphi$ in the HF (direction of "event plane" with maximum energy)
- Improves the previous 5–6% jet energy scale (JES) variation vs. orientation in iterative PU down to $\approx$ 1%
- Fit to $A + B|\varphi - \Psi^{EP}| + C\cos(2|\varphi - \Psi^{EP}|)$ for visual aid

CMS DP-2013/018

# MC Generator + GEANT Performance: Calorimetric Jets




- The first analysis for 2014 that uses the HF/Voronoi algorithm happens to be jet–track correlation that needs it calorimetric jets (track bias in particle-flow)
- Performance better than past iterative PU with particle-flow

CMS PAS-HIN-14-010

# Summary

- First algorithm for a "non-jet" UE subtraction in particle-flow environment
- New ideas for UE determination:
  - Data-driven prediction of midrapidity UE from forward detectors
  - Applicable for calorimeter with nonlinearity
- New ideas for UE subtraction:
  - Arbitrary geometry (e.g. particle-flow)
  - Full conversion to a hard event without negative particles, that clusters into the correct jet energy
- Studied and commissioned in CMS, first physics results already public
- Standalone/particle-level code and write-up in preparation (apologies!)

# Part I

# Backup

# The "Full $\eta$ estimated $\rho A$" Being Compared to

- Stock "global rescaled" $\rho A$-subtraction in FastJet 3.0.3
- $\rho$ estimation using $k_T$ $R = 0.6$, within $|\eta| < 5.191$
- $\rho$ rescaling for $|\eta| < 1.479$ (largest acceptance still like an order-4 polynomial)
- Same calibration procedure (note: L1–L2 are factorized)