

Underlying Event Fluctuations and Jet Background in $\sqrt{s_{NN}} = 200$ GeV Au+Au Collisions Using the sPHENIX Detector

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Tanner Mengel

May 2025

© by Tanner Mengel, 2025
All Rights Reserved.

To my love, Marisa.

Acknowledgments

My research at the University of Tennessee would not have been possible without the mentorship and support of many people. I was fortunate that my advisor, Christine Nattrass, provided me with the opportunity to pursue my interests, even when she didn't necessarily agree they would be fruitful endeavors. This thesis would not be possible without her guidance and mentorship — both in our field and within the community of experimental collaborations. Thank you for giving me the tools to become an independent researcher.

I would like to thank the many mentors who have guided me throughout my graduate career: Jo Schambach, Fredi Bock, Ken Read, and Soren Sorensen. Each of you has taught me an enormous amount about being a scientist, and you have sparked my curiosity in a wide range of topics, from silicon detectors to machine learning to chess. Your mentorship has been invaluable.

Over the past five years, I have had the privilege of working alongside several incredible post-docs and fellow graduate students at both the University of Tennessee and in the sPHENIX collaboration: Charles Hughes, Patrick Steffanic, Austin Schmier, Christal Martin, Antonio Da Silva, Niseem Magdy, Cameron Dean, Michael Peters, Hao-Ren Jheng, and Thomas Marshall. Thank you all for your support, for the great conversations (both physics-related and not), and for the vital collaboration that makes science possible.

I am grateful for my collaborators at sPHENIX: Rithya Kunnawalkam Elayavalli, Ming Liu, Rosi Reed, Virginia Bailey, and James Nagle. Your insights have continually

pushed me to deepen my own expertise and have motivated me to become a better physicist.

My journey as a physicist would not have begun without the unwavering love and support of my family. Mom, thank you for being my biggest fan and encouraging me to be myself. Dad, thank you for showing me what it means to be a leader and for instilling the values that guide my life. I am grateful to you both for your compassion and encouragement, especially during difficult times. Eli, I am so proud to be your big brother, and watching you grow into the incredible person you are has been one of my greatest joys. You all motivate me every day, and I cannot thank you enough for being in my life.

Finally, I want to thank my closest friend, partner, and wife, Marisa. You have had the greatest impact on my life, and I would not be who I am today without your love and support. Through both the good times and the difficult ones, you have never wavered in your belief in me, lifting me up when I have fallen. I am eternally grateful for you and for the life we share. This thesis would not exist without you.

Abstract

At the Relativistic Heavy Ion Collider at Brookhaven National Laboratory, heavy ions are collided at ultra-relativistic speeds. The motivation for these collisions is to study an extreme form of Quantum Chromodynamical matter called Quark-Gluon Plasma (QGP). The QGP can be studied using collimated sprays of particles that are products of initial hard-scatterings between colliding partons in the moments prior to the production of QGP. These collimated sprays of particles are called jets. Comparing how jets are modified in heavy ion collision systems to that of proton-proton collisions, where QGP production is not expected, provides a useful tool to probe the strongly interacting medium and test predictions of Quantum Chromodynamics.

Measurements of jets in heavy ion collisions must contend with the high entropy and the presence of a large, fluctuating background of soft particles unrelated to the hard-scattering event. This underlying event and its fluctuations must be well understood in order to correct for it in the most challenging jet kinematic regimes. Jet momentum reconstruction for low momenta jets in heavy ion collisions at the Relativistic Heavy Ion Collider is particularly difficult due to the low initial momentum transfer of the hard-scatterings, and the large fraction of gluon-initiated jets.

In this analysis, the underlying event in Au+Au collisions at a center of mass energy of 200 GeV is measured using the sPHENIX detector. The spatial and multiplicity correlations for underlying event fluctuations are presented and

the sources of underlying event fluctuations are determined and quantified. The fluctuations of the underlying event are found to be primarily driven by statistical fluctuations in particle multiplicity and non-stochastic contributions to the underlying event are found to be primarily due to initial geometric anisotropies in the colliding nuclei.

Direct comparisons between jet background subtraction methods are presented based on each jet background subtraction method's ability to suppress underlying event fluctuations. This is the first direct comparison between methods used by different experiments for jet background reconstruction. Machine learning applications to jet background subtraction are presented and their applicability to data is discussed. Interpretable machine learning techniques are applied to previously published jet background subtraction methods and a physics motivation is presented based on learned expressions. We apply a new jet background subtraction method, based on jet constituent multiplicity, to data for the first time. Our findings provide guidance for future jet measurements at sPHENIX.

Table of Contents

1	Introduction	1
1.1	The Standard Model	1
1.1.1	Quarks, Gluons, and The Strong Nuclear Force	4
1.2	QGP Production in Heavy Ion Collisions	11
1.3	Jets and Underlying Event	19
1.3.1	Jets in Heavy Ion Collisions	23
1.3.2	Jet Background Subtraction	27
2	Detector Equipment and Instrumentation	39
2.1	Relativistic Heavy Ion Collider	39
2.2	The sPHENIX Detector	41
2.2.1	MAPS-based Vertex Detector	43
2.2.2	Intermediate Silicon Tracking Detector	46
2.2.3	Time Projection Chamber	49
2.2.4	Time Projection Chamber Outer Tracker	52
2.2.5	Calorimeters in sPHENIX	55
2.2.6	sPHENIX Event Characterization	57
2.3	sPHENIX Acquisition and Signal Extraction	60
2.3.1	Calorimeter Towers Signal Extraction	60
2.3.2	Tracking in sPHENIX	61

3	Jet Background Studies	64
3.1	Multiplicity Based Background Subtraction	64
3.2	Machine Learning Studies	74
3.2.1	Machine Learning Interpretability	83
4	Analysis	92
4.1	Reconstructed Physics Quantities	93
4.2	Event Selection	102
4.3	Jet Background Subtraction Methods	103
4.3.1	Iterative Method	103
4.3.2	Area Method	104
4.3.3	Multiplicity Method	107
4.4	Underlying Event Characterization	109
4.4.1	Calorimeter Window Analysis	109
4.4.2	Random Cones Analysis	110
5	Results and Discussion	119
6	Conclusions	132
	Bibliography	137
	Appendix	150
A	Jet Background Subtraction	150
A.1	Energy Residuals	151
A.2	Multiplicity Method	154
A.3	Area Method	156
A.4	Iterative Method	158
A.5	Residual Fit Parameters	159
B	Supplemental Calorimeter Window Analysis Plots	163
B.1	Window Energy Distributions	164

B.2	Window Energy Fluctuations	186
B.3	Window Energy Fractions	208
B.4	Window Standard Deviations and Averages	219
B.5	Window Area Correlation Fits	223
Vita		224

List of Tables

2.1	Brief summary of the sPHENIX sub-systems.	45
2.2	MVTX layer parameters [110].	48
3.1	Parameters controlling artificial radiation for each of the three fragmentation cases. These parameters were chosen to reflect observations of jets and charged hadron yields in Pb+Pb collisions [18].	71
4.1	Area of calorimeter window size corresponding to a jet defined by the parameter R.	111
5.1	Width and mean of underlying event distributions δE_T and negative δE_T extrapolated via Gaussian fit in 0-5% central Au+Au events for both types of random cones and embedded high E_T probes subtracted with each background subtraction method.	126
1	Standard deviations and averages of δE_T distributions and Gaussian fits for the iterative method. All values are in units of GeV.	160
2	Standard deviations and averages of δE_T distributions and Gaussian fits for the area method. All values are in units of GeV.	161
3	Standard deviations and averages of δE_T distributions and Gaussian fits for the multiplicity method. All values are in units of GeV.	162

List of Figures

1.1	Elementary Particles of the Standard Model of Particle Physics. Figure taken from Ref. [57].	3
1.2	Possible QCD Feynman vertices. Figure taken from Ref [123].	6
1.3	The running of the strong coupling constant α_s as a function of momentum transfer Q . Figure taken from Ref. [8].	7
1.4	A quark-anti-quark pair being separated until the interaction strength becomes sufficient to produce a new quark-anti-quark pair. This figure is taken from Ref [56].	9
1.5	Quantum Chromodynamics matter phase diagram as a function of baryon chemical potential μ_b and temperature T . This figure is taken from ref [64].	12
1.6	Diagram showing the time Evolution of a Heavy Ion Collision. Figure taken from Ref. [91].	13
1.7	Cartoon illustrating the relationship between the average impact parameter, $\langle b \rangle$ and the charged particle multiplicity N_{ch} . Figure taken from ref. [96].	15
1.8	A light cone diagram showing the evolution of QGP from production at time τ_0 , transition to a hadron gas at temperature T_C , chemical freeze-out at T_{ch} , and kinetic freeze-out at temperature T_{fo} [58]. . . .	17
1.9	Diagram showing the translation of initial spatial anisotropies into pressure gradients of the expanding medium. Figure taken from Ref. [58].	18

1.10	Charged hadron differential v_2 in Au+Au $\sqrt{s_{NN}} = 200$ GeV collisions measured by PHENIX using particle cumulants (red), and event plane resolution methods (blue, black) [30].	20
1.11	Illustration of the outcome of the clustering process for the anti- k_t (left) and k_T (right) algorithms used in simulation [52].	22
1.12	Full jet spectra for $R=0.2$ (left), and $R=0.4$ (right) measured in p+p and 0-10% central Pb+Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV [17].	25
1.13	Jet R_{AA} from LHC run 1 data measured by ALICE for $R = 0.2$ jets at $ \eta < 0.5$, by ATLAS for $R = 0.4$ jets at $ \eta < 2.1$ and by CMS for $R = 0.2, 0.3, 0.4$ jets at $ \eta < 2.0$ in Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV [58].	26
1.14	Scale probed in the medium in $[1/\text{fm}]$ via high energy partons as a function of the local temperature in the medium. The red (black) curves are for different initial parton energies in the RHIC (LHC) medium.[27].	28
1.15	Comparison of the fraction of quark and gluon jets from leading order pQCD calculations for RHIC and LHC energies. [27].	29
1.16	Diagram showing the effect of a fluctuating background on a steeply falling spectra corrected with the average pedestal.	31
1.17	Nuclear modification factors of charged-particle jets for $R= 0.2, 0.4,$ and 0.6 jets in Pb+Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV using the area-based method (Equation 4.3) and the machine learning method [18].	32
1.18	δp_T for random cones in Pb-Pb events. A Gaussian fit to the left-hand-side and its extrapolation to positive δp_T . The solid line is a fit to the distribution given by Equation 1.9 [12].	34
1.19	Dependence of the standard deviation of the δp_T distributions on uncorrected charged particle multiplicity, compared to the limit derived for Poissonian (Equation 1.11) non-Poissonian (Equation 1.12) contributions to fluctuations [12].	36

1.20	(Top) Distributions of $E_T^{n \times m} - \langle E_T^{n \times m} \rangle$ for $n \times m = 3 \times 4, 7 \times 7$ and (Bottom) standard deviation $\sigma(E_T)^{3 \times 4}, \sigma(E_T)^{7 \times 7}$ compared between Pb+Pb $\sqrt{s_{NN}} = 2.76$ TeV collisions and HIJING [3].	38
2.1	RHIC collider complex at Brookhaven National Laboratory. Photo was taken from Ref. [125].	40
2.2	A schematic diagram of injection of ions to RHIC. The figure is taken from reference [124].	42
2.3	Overview of sPHENIX detector with labeled sub-systems [112].	44
2.4	MVTX stave diagram [110].	47
2.5	INTT cross section diagram [111].	50
2.6	Mechanical design of TPC [111].	51
2.7	TPC modules and FEE chips [111].	53
2.8	Schematic overview of a Micromegas chamber [43].	54
2.9	Cross section of sPHENIX calorimeters [85].	56
2.10	Electronics of an OHCAL sector [35].	58
2.11	(Left) The MBD array mounted on the MBD mechanical frame. (Right) The individual MBD counter module [111].	59
2.12	Schematic view of the sPHENIX tracking procedure using all available tracking detectors [43].	63
3.1	The background momentum in a group of N particles as predicted by Equation 3.3 compared to the background momentum in toy model jets with 30 constituents.	66
3.2	Comparisons of ratio of jet fragmentation functions measured in Pb+Pb to $p+p$ collisions $R_{D(p_T)}$ [1] to all artificial fragmentation modification cases for $R = 0.4$ jets at LHC energies.	72
3.3	Comparisons of ratio of jet fragmentation functions measured in Pb+Pb to $p+p$ collisions $R_{D(\xi)}$ [19] to all artificial fragmentation modification cases for $R = 0.4$ jets at RHIC energies.	73

3.4	Total feature scores for available jet input features relating to $p_{T,jet}^{Truth}$. Each score is the sum of the variance, mutual information, and Pearson Correlation Coefficient [69].	77
3.5	Each node and weight for every layer l in the deep neural network architecture. The nodes are colored by their probability of firing for the entire testing jet sample. The weights are normalized to the maximum weight from a given node to the next layer. The input features are listed by x_i where i ranges from 0 to 11 corresponding to the 12 input features for the deep neural network (uncorrected jet momentum, jet area, jet angularity, number of jet constituents, and eight leading constituent momenta). The truth jet momentum is labeled by y	79
3.6	Each node and weight for every layer l in for the first (a) and last (b) iteration of the shallow neural network produced through feature optimization and network pruning. The nodes are colored by their probability of firing for the entire testing jet sample. The weights are normalized to the maximum weight from a given node to the next layer. The input features are listed by x_i where i ranges from 0 to 4 corresponding to uncorrected jet momentum, leading constituent momenta, number of jet constituents, jet area, and jet angularity. The truth jet momentum is labeled by y	80
3.7	Comparisons of jet p_T residual width for each background subtraction method as a function of the reconstructed jet momentum for Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV and Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV for jet resolution parameters $R = 0.2, 0.4$, and 0.6 . Note the y -axes are zero suppressed for all panels.	82
3.8	Ratio of unfolded jet spectrum over truth spectrum for (a) Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV and (b) Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV for jet resolution parameter $R = 0.4$	84

3.9	Diagram showing the procedure of mapping a relationship between input jet features to the corrected jet momentum using a neural network and then extracting an analytical expression that describes that mapping with symbolic regression. The left panel shows fitting the space of possible functions that map input to output with the neural network. The middle panel shows sampling of this mapping across input jets. The right panel shows fitting this sample to an analytical expression using symbolic regression.	86
3.10	The eight highest scoring expressions from symbolic mapping of a deep neural network trained on $R = 0.4$ jets from $\sqrt{s_{NN}} = 200$ GeV Au+Au collisions. The score is defined as the negative derivative of the loss function with respect to equation complexity (equation 3.11).	88
3.11	PySR optimization constants compared to average value of multiplicity method parameters versus jet resolution parameter for Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV and Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV for jet resolution parameters $R = 0.2, 0.4$, and 0.6	90
4.1	Correlation between hadronic calorimeter energy deposits and MBD charge sum of both MBD arms the inner HCAL on the left and the outer HCAL on the right. Figures taken from Ref.[114].	94
4.2	(Top) MBD charge sum distribution is shown in the upper panel (open points). Also, shown are the fit results from the MC Glauber and negative binomial distribution best fit. (Bottom) The ratio of data to MC Glauber including the fit to the trigger efficiency turn-on region. Figure taken from Ref.[114].	96
4.3	Calorimeter tower event cross section display from a 0-10% central $\sqrt{s_{NN}} = 200$ GeV Au + Au event. The yellow, red, and blue towers correspond to the EMCAL, inner HCAL, and outer HCAL respectively. Event display from Ref. [113].	98

4.4	(Left) Outer hadronic calorimeter tower distribution of energy scale from cosmic muon calibration. (Right) Comparison of a single outer hadronic calorimeters tower's measured energy distribution and the corresponding GEANT4 simulation result. Figure taken from Ref.[114].	99
4.5	Calorimeter tower event cross section from a 50-60% central $Au + Au$ $\sqrt{s_{NN}} = 200$ GeV di-jet event display. The legend label EMCAL refers to the ReEMCAL towers. Event display from Ref. [113].	101
4.6	Average background calorimeter tower energy determined by the iterative background subtraction method in $Au + Au$ at $\sqrt{s_{NN}} = 200$ GeV.	105
4.7	ρ_A as function of total calorimeter energy in $\sqrt{s_{NN}} = 200$ GeV $Au + Au$ events.	106
4.8	ρ_{Mult} as function of total calorimeter energy in $\sqrt{s_{NN}} = 200$ GeV $Au + Au$ events.	108
4.9	Centrality dependence of event-by-event standard deviation of calorimeter windows of size $n \times m = 1 \times 1$ (Left), and the event average E_T of windows with size $n \times m = 13 \times 13$ (Right) for all events.	112
4.10	Window area dependence of event-averaged transverse energy $\langle \bar{E}_T \rangle^{n \times m}$ for windows of all window sizes (left), and the number of reconstructed windows of a given area (right) for 0-5% central events.	113
4.11	δE_T distributions for random cones reconstructed in 0-5% central events for area-based subtracted (black), multiplicity subtracted (red), and iteratively subtracted (blue) jets. (Left) Static random cones, (Right) randomized tower input random cones.	116
4.12	δE_T distributions for jets reconstructed containing a probe $E_{T,probe} = 30$ GeV in 0-5% central events for area-based subtracted (left), multiplicity subtracted (red), and iteratively subtracted (right) jets.	118

5.1	Example of the fit to the formula in equation 5.1 for 0-5% central (left) and 30-40% central (right) events. The x -axis is scaled by the area of a single tower and the points represent calorimeter window areas with dimensions presented in table 4.4.2.	120
5.2	The scale parameter k from equation 5.1 for all fits to the dependence of the mean standard deviation in calorimeter window transverse energy at electromagnetic scale to calorimeter window area, as a function of event centrality.	121
5.3	Underlying event quantification for the multiplicity method (left) and the area method (right) measured in Au+Au events for all event centrality classes.	123
5.4	δE_T for multiplicity subtracted (top left), iteratively subtracted (top right), and area subtracted (bottom) underlying event characterizations in 0-5% central Au+Au events at $\sqrt{s_{\text{NN}}} = 200$ GeV for both types of random cones and embedded high E_T probes.	124
5.5	Centrality dependence of $\sigma(\delta E_T)$ of both types of random cones for all background subtraction methods, compared to the Poissonian limit calculated with measured $E_{T, \text{Tower}}$ and to additional hydrodynamical flow contributions calculated with elliptical flow v_2 measured in [16] and triangular flow measured in [128].	128
5.6	Fractional contributions to underlying event fluctuations $\sigma(\delta E_T)$ for random cones subtracted with each background subtraction methods. The Poissonian (statistical) limit is used as a scale.	130
1	δE_T for standard rigid cones with $R = 0.4$ from each background subtraction method for all event centrality classes.	151
2	δE_T for rigid cones with randomized tower η, ϕ with $R = 0.4$ from each background subtraction method for all event centrality classes.	152

3	δE_T for embedded high energy probe (E_T^{probe})=30 GeV η, ϕ with $R = 0.4$ from each background subtraction method for all event centrality classes.	153
4	Probability distributions of ρ_M for each centrality bin.	154
5	δE_T for standard $R = 0.4$ random cones, $R = 0.4$ cones with randomized tower η, ϕ , and embedded high energy probe (E_T^{probe})=30 GeV, fit with a Gaussian on the left hand side and a Γ distribution globally. All underlying event characterizations are subtracted with the multiplicity method.	155
6	Probability distributions of ρ_A for each centrality bin.	156
7	δE_T for standard $R = 0.4$ random cones, $R = 0.4$ cones with randomized tower η, ϕ , and embedded high energy probe (E_T^{probe})=30 GeV, fit with a Gaussian on the left hand side and a Γ distribution globally. All underlying event characterizations are subtracted with the area method.	157
8	δE_T for standard $R = 0.4$ random cones, $R = 0.4$ cones with randomized tower η, ϕ , and embedded high energy probe (E_T^{probe})=30 GeV, fit with a Gaussian on the left hand side and a Γ distribution globally. All underlying event characterizations are subtracted with the iterative method.	158
9	Distribution of E_T^{1x1} for full calorimeter windows for all centralities. .	164
10	Distribution of E_T^{2x2} for full calorimeter windows for all centralities. .	165
11	Distribution of E_T^{3x4} for full calorimeter windows for all centralities. .	166
12	Distribution of E_T^{5x6} for full calorimeter windows for all centralities. .	167
13	Distribution of E_T^{7x8} for full calorimeter windows for all centralities. .	168
14	Distribution of E_T^{9x10} for full calorimeter windows for all centralities. .	169
15	Distribution of E_T^{11x12} for full calorimeter windows for all centralities. .	170
16	Distribution of E_T^{13x13} for full calorimeter windows for all centralities. .	171
17	Distribution of E_T^{15x15} for full calorimeter windows for all centralities. .	172

18	Distribution of E_T^{17x17} for full calorimeter windows for all centralities.	173
19	Distribution of E_T^{20x20} for full calorimeter windows for all centralities.	174
20	Distribution of E_T^{1x1} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	175
21	Distribution of E_T^{2x2} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	176
22	Distribution of E_T^{3x4} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	177
23	Distribution of E_T^{5x6} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	178
24	Distribution of E_T^{7x8} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	179
25	Distribution of E_T^{9x10} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	180
26	Distribution of E_T^{11x12} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	181
27	Distribution of E_T^{13x13} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	182
28	Distribution of E_T^{15x15} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	183
29	Distribution of E_T^{17x17} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	184
30	Distribution of E_T^{20x20} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	185
31	Distribution of $E_T^{1x1} - \langle E_T^{1x1} \rangle$ full calorimeter windows for all centralities.	186
32	Distribution of $E_T^{2x2} - \langle E_T^{2x2} \rangle$ full calorimeter windows for all centralities.	187

33	Distribution of $E_T^{3x4} - \langle E_T^{3x4} \rangle$ full calorimeter windows for all centralities.	188
34	Distribution of $E_T^{5x6} - \langle E_T^{5x6} \rangle$ full calorimeter windows for all centralities.	189
35	Distribution of $E_T^{7x8} - \langle E_T^{7x8} \rangle$ full calorimeter windows for all centralities.	190
36	Distribution of $E_T^{9x10} - \langle E_T^{9x10} \rangle$ full calorimeter windows for all centralities.	191
37	Distribution of $E_T^{11x12} - \langle E_T^{11x12} \rangle$ full calorimeter windows for all centralities.	192
38	Distribution of $E_T^{13x13} - \langle E_T^{13x13} \rangle$ full calorimeter windows for all centralities.	193
39	Distribution of $E_T^{15x15} - \langle E_T^{15x15} \rangle$ full calorimeter windows for all centralities.	194
40	Distribution of $E_T^{17x17} - \langle E_T^{17x17} \rangle$ full calorimeter windows for all centralities.	195
41	Distribution of $E_T^{20x20} - \langle E_T^{20x20} \rangle$ full calorimeter windows for all centralities.	196
42	Distribution of $E_T^{1x1} - \langle E_T^{1x1} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	197
43	Distribution of $E_T^{2x2} - \langle E_T^{2x2} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	198
44	Distribution of $E_T^{3x4} - \langle E_T^{3x4} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	199
45	Distribution of $E_T^{5x6} - \langle E_T^{5x6} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	200
46	Distribution of $E_T^{7x8} - \langle E_T^{7x8} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	201

47	Distribution of $E_T^{9x10} - \langle E_T^{9x10} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	202
48	Distribution of $E_T^{11x12} - \langle E_T^{11x12} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	203
49	Distribution of $E_T^{13x13} - \langle E_T^{13x13} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	204
50	Distribution of $E_T^{15x15} - \langle E_T^{15x15} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	205
51	Distribution of $E_T^{17x17} - \langle E_T^{17x17} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	206
52	Distribution of $E_T^{20x20} - \langle E_T^{20x20} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.	207
53	Distribution of fractional contribution $f(E_T^{1x1})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.	208
54	Distribution of fractional contribution $f(E_T^{2x2})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.	209
55	Distribution of fractional contribution $f(E_T^{3x4})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.	210
56	Distribution of fractional contribution $f(E_T^{5x6})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.	211
57	Distribution of fractional contribution $f(E_T^{7x8})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.	212
58	Distribution of fractional contribution $f(E_T^{9x10})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.	213
59	Distribution of fractional contribution $f(E_T^{11x12})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.	214
60	Distribution of fractional contribution $f(E_T^{13x13})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.	215

61	Distribution of fractional contribution $f(E_T^{15 \times 15})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.	216
62	Distribution of fractional contribution $f(E_T^{17 \times 17})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.	217
63	Distribution of fractional contribution $f(E_T^{20 \times 20})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.	218
64	$\langle \delta E_T \rangle$ for all calorimeter window sizes as a function event centrality. .	219
65	$\langle \delta E_T \rangle$ for all calorimeter window sizes as a function event centrality. .	220
66	$\langle \delta E_T \rangle$ for all calorimeter window sizes as a function event centrality. .	221
67	$\langle \delta E_T \rangle$ for all calorimeter window sizes as a function event centrality. .	222
68	Fits on average standard deviation dependence on calorimeter window area to Eq. 5.1.	223

Chapter 1

Introduction

The field of heavy ion physics can be summarized as the search for and study of the Quark-Gluon Plasma (QGP). Motivations for such an endeavor began in the 1970s with theoretical descriptions of a strongly interacting quark-gluon matter phase [118, 105]. These theoretical predictions evolved into fixed target experiments conducted in the 1990s at CERN’s Super Proton Synchrotron and the Alternating Gradient Synchrotron located at Brookhaven National Laboratory (BNL). Eventually, as accelerator technology advanced, collisions between ultra-relativistic beams of hadrons were studied with higher resolution and greater precision [50], culminating in the eventual discovery of signatures attributed to the phase of matter predicted decades earlier [103, 75]. In the early 2000s the Relativistic Heavy Ion Collider (RHIC) at BNL was constructed to further probe the properties of this novel nuclear phase of matter. Today, in 2025, RHIC and the Large Hadron Collider (LHC) at CERN serve as the world’s premier hadronic collision facilities.

1.1 The Standard Model

The Standard Model of particle physics is currently our best understanding of the smallest units of matter and how they interact. The model consists of 12 fundamental particles that make up all matter called fermions (and their anti-particle pairs) and

three forces which are mediated by six bosons. The fermions are classified into quarks and leptons, both of which have three generations. Figure 1.1 shows an overview of the fundamental particles, including their properties such as mass, spin, and charge. The generations are ordered from least massive and most common to most massive and least common with the first generation being fermions we associate with normal matter such as electrons, and the types of quarks that make up protons and neutrons (up quarks and down quarks). Each established generation has two flavors of quarks and two flavors of leptons. Leptons carry mass and charge so they interact via electromagnetic and weak interactions. The other fermions, quarks, also carry mass and charge in addition to color, which is the charge for the strong nuclear force. The interactions between fermions are mediated by gauge-bosons which include the photon γ , the Z and $W^\pm$ bosons, and the gluon g . The final boson and newest discovered member of the standard model [2] is the scalar Higgs boson H which gives mass to the fermions and both the Z and $W^\pm$ bosons.

The mathematical framework the standard model uses to describe the interactions between its fundamental particles is quantum field theory. Quantum field theory is an extension to classical field theory in which the degrees of freedom are quantized fields and the action, described by the Lagrangian, is required to be Lorentz invariant. Similar to classical field theory, quantum field theories use Noethers' theorem to establish conserved quantities from global symmetries of the action, such as spacetime symmetry or phase invariance. The quantum field theories in the standard model impose localized symmetries under a set of gauge transformations on the Lagrangian which introduces an interaction term and a gauge field. Requiring gauge invariance allows spin-1 boson fields to be coupled to the non-interacting spin-1/2 fermion fields and is the basis of the standard model. This framework has proven exceptionally useful at describing three out of the four of the known fundamental forces of nature, namely the electromagnetic force, the strong nuclear force, and the weak nuclear force. A notable exception to the standard model is the fourth fundamental force, gravity. Gravity is well described by Einstein's general theory of relativity. The framework of

Standard Model of Elementary Particles

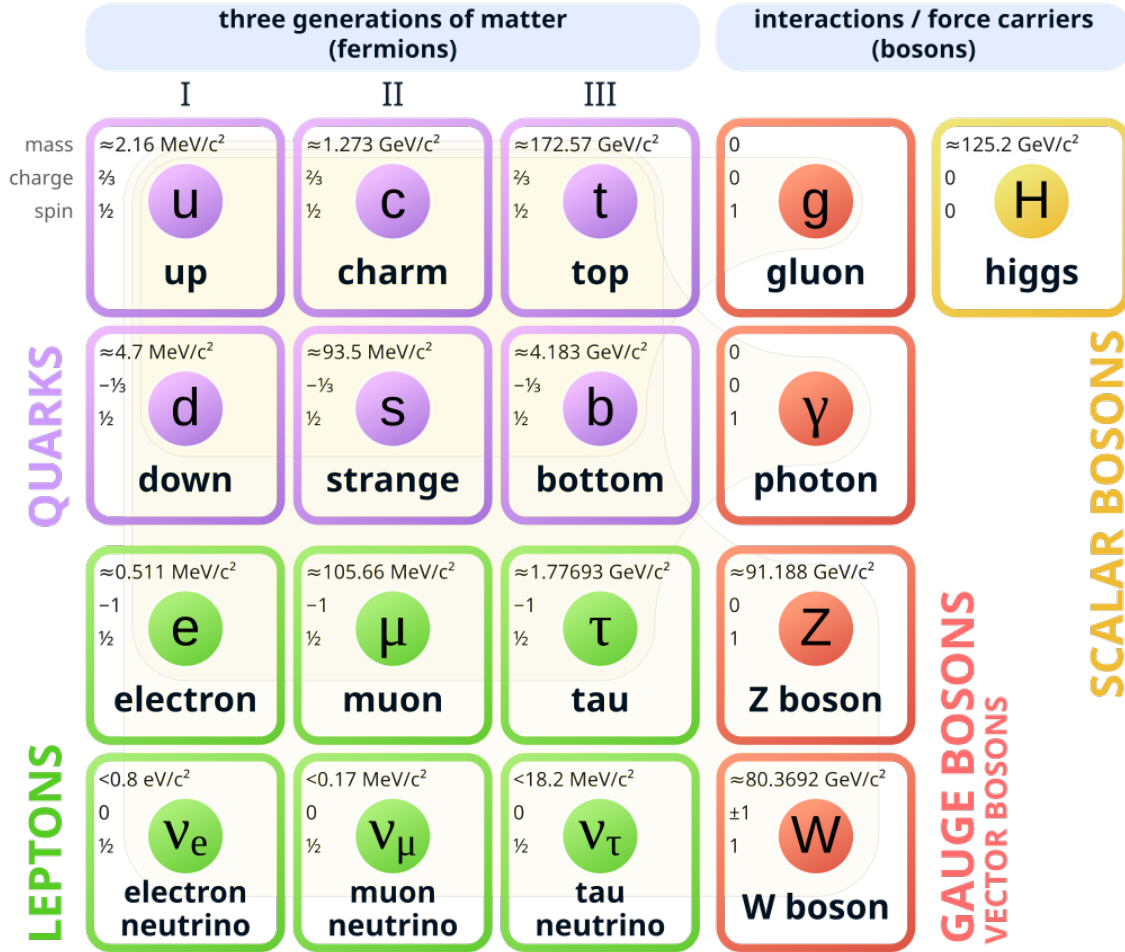


Figure 1.1: Elementary Particles of the Standard Model of Particle Physics. Figure taken from Ref. [57].

general relativity is differential geometry, which is not compatible with the quantum field theoretical framework of the standard model. Several efforts have been made to translate General Relativity into a quantum field theory framework but have been unsuccessful so far.

The standard model description of the electromagnetic interaction is quantum electrodynamics. In quantum electrodynamics, photons are the force mediating bosons and they interact with charged particles such as the charged leptons and all 6 flavors of quarks. The established charge carrying leptons, in order of least massive to most massive, are the electron e , muon μ , and tau τ . The weak nuclear force is mediated by the Z and $W^\pm$ bosons and is aptly named after its very short interaction range. In the standard model, the weak nuclear force is unified to the electromagnetic force and described with electro-weak theory [126] which provides a mechanism for the scalar Higgs boson to give mass to the Z and $W^\pm$ bosons via spontaneous symmetry breaking. The non-zero mass of the Z and $W^\pm$ gauge bosons is the main reason that the interaction range of the weak force is so small. The Higgs field also couples to fermionic fields giving the quarks and leptons mass.

1.1.1 Quarks, Gluons, and The Strong Nuclear Force

The standard model explains the strong nuclear force through Quantum Chromodynamics (QCD). QCD describes the interactions between fermionic quarks and force-carrying gluon fields, both of which possess the strong nuclear force's color charge. The color charge carried by both quarks and gluons comes in three charges, which are typically called red, green, and blue (and their anti-particle pairs). In addition to color charge, quarks also carry electric charge, weak hypercharge, and have mass. Quarks and gluons are referred to as partons because they comprise all hadronic matter. Experimental evidence of quarks was first observed in the 1960s in deep inelastic $e + p$ collisions [47, 48]. The gluon was discovered later in e^+e^- from gluon bremsstrahlung producing 3-jet events ($e^+e^- \rightarrow q\bar{q}g$) [45, 44]. Quarks come in

six flavors with two for each generation; up u , down d , charm c , strange s and top t , bottom b . The up and down quarks are those that make up everyday protons (uud) and neutrons (udd) and have masses of 2.16 MeV and 4.7 MeV respectively.

QCD is a non-Abelian quantum field theory with SU(3) gauge symmetry. The Lagrangian of QCD is given by

$$\mathcal{L}_{\text{QCD}} = \sum_q \left(\bar{\psi}_{q,i} i\gamma^\mu [\partial_\mu \delta_{ij} + ig_s (G_\mu^\alpha t_\alpha)_{ij}] \psi_{q,j} - m_q \bar{\psi}_{q,i} \psi_{q,i} \right) - \frac{1}{4} F_{\mu\nu}^a F_a^{\mu\nu}, \quad (1.1)$$

where ψ is the Dirac spinor for a quark field, m is the mass of the quark, G^μ is the four potential of the gluon field, t are 3×3 generators of the SU(3) group, $F^{\mu\nu}$ is the color-field tensor, and g_s (or $a_s = \frac{g_s^2}{4\pi}$) is the QCD coupling constant. In the equation, the q -index represents each of the six quark flavors ($q = 1, \dots, 6$), the i -index represents the three QCD color charges ($i = 1, 2, 3$), and the α -index is summed over for all eight gluon-fields ($\alpha = 1, \dots, 8$) [127]. The field gluon-field strength term in equation 1.1 is a consequence of gluons themselves carrying color charge. This term allows gluon self-interaction resulting in much more complex Feynman diagrams for QCD processes compared to quantum electrodynamics. Figure 1.2 shows these possible vertices with a simple quark-gluon interaction on the left and three and four-point gluon vertices in the center and right. These gluon-self interactions produce an anti-screening effect in which the effective field strength increases as the distance from the source increases [63, 127, 97].

The QCD coupling constant a_s is the only fundamental parameter in QCD, besides the quark masses [127]. In order to use QCD to make predictions the theory must be re-normalized by allowing the coupling constant that controls the strength of the interactions described by equation 1.1 to acquire a scale dependence $a_s(\mu_R^2)$ where μ_R is an arbitrary renormalization scale. Choosing a scale close to the momentum transfer Q for a QCD process makes the running coupling constant $a_s(Q)$ a measure of the strength of the strong interaction of that process [127, 63]. Figure 1.3 shows

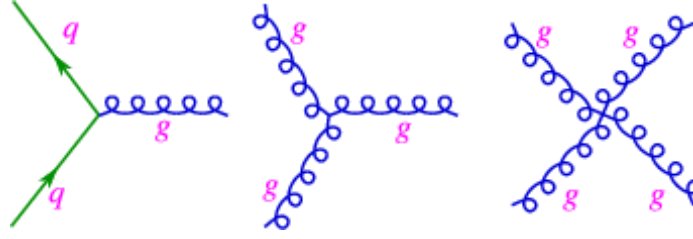


Figure 1.2: Possible QCD Feynman vertices. Figure taken from Ref [123].

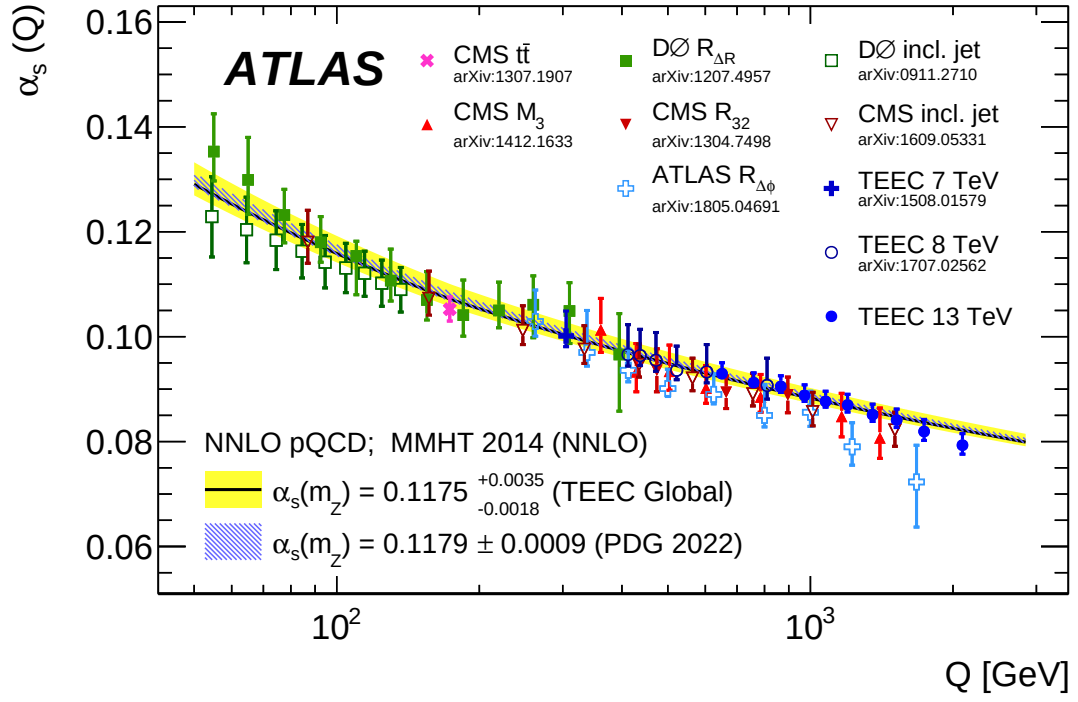


Figure 1.3: The running of the strong coupling constant α_s as a function of momentum transfer Q . Figure taken from Ref. [8].

the value of the running QCD coupling constant as a function of momentum transfer Q for different processes.

The behavior of $a_s(Q)$ highlights two pathologies of QCD at the low and high Q regimes. At short distance scales or high Q the strength of the interactions between quarks and gluons is weak. In the limit that Q^2 approaches infinity, $a_s(Q)$ will asymptotically approach zero. This is known as asymptotic freedom, where coupling becomes so weak that quarks and gluons are no longer confined to color neutral hadrons and the relevant degrees of freedom are the quarks and gluons themselves. In this regime of high energy and short distance scales, QCD is well described by perturbative expansions to arbitrary precision. This is the relevant energy scale for the heavy ion collision environment [63]

Perturbative QCD breaks down in the lower momentum transfer Q regime. This scale corresponds to the low energy or large length scale. Here, the coupling strength $a_s(Q)$ becomes large and no longer suppresses contributions from higher order diagrams in expansions. The increase in coupling strength for large distances is known as color confinement and is the reason that no free quarks can be observed outside of color neutral hadrons. Removing a quark from a hadron will increase the distance scale of the interaction between the other quarks and thus increase the coupling strength. The gluon field mediating this interaction will form a color flux tube between the quarks. As the coupling increases, the three-point and four-point gluon self-interaction terms will become more prevalent. As the energy increases, it becomes more energetically favorable to produce pairs of quark anti-quark pairs which will pair into two stable mesons. [63, 67]. Figure 1.4 shows a schematic of a quark anti-quark pair being separated until new pair production is achieved. In this strongly coupled regime, perturbative expansions become incalculable due to increasingly important next order effects. In this phase space, lattice QCD techniques are used to calculate the properties of QCD matter. Lattice QCD has been successful

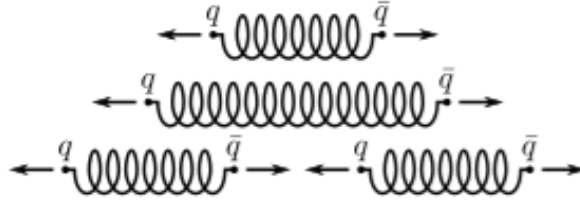


Figure 1.4: A quark-anti-quark pair being separated until the interaction strength becomes sufficient to produce a new quark-anti-quark pair. This figure is taken from Ref [56].

for calculations with small baryon chemical potential such as the calculation of the mass spectrum of hadrons [65].

As momentum transfers Q^2 between partons diminish, the strong coupling constant of the QCD Lagrangian (Equation 1.1) grows and eventually confines all color charge carrying particles into color neutral hadrons. When this hadronization process occurs, perturbative QCD is no longer an effective description of the interactions between hadrons [93]. Descriptions for non-perturbative interactions ($Q^2 \sim 2 \text{ GeV}^2$) are contained in distributions such as fragmentation functions, or parton distribution functions [92]. Parton distribution functions describe the momentum distribution of partons within a hadron at a given fraction of hadron momentum and Q^2 . Fragmentation functions describe the momentum fraction $z = p^h/p^p$ carried by a hadron from a parent parton [10]. Most processes in QCD fall into this category but can be separated out into sub-processes which can be evaluated at hard Q^2 scales. This interface between the soft and hard QCD processes is known as factorization which allows for the compartmental description of processes that involve both hard and soft interactions, such as particle formation, and partonic fragmentation. In the case of particle production, the cross section for particles produced from colliding hadrons is given by [102]

$$\frac{d^3\sigma^h}{dyd^2p_T} = \frac{1}{\pi} \int dx_a \int dx_b f_a^A(x_a) f_b^B(x_b) \frac{d\sigma_{ab \rightarrow cX}}{d\hat{t}} \frac{D_c^h(z)}{z}, \quad (1.2)$$

where $f_a^A(x_a)$ is the initial state parton distribution function for the incoming parton a which has momentum fraction x_a , $d\sigma_{ab \rightarrow cX}/d\hat{t}$ is the hard-scattering cross-section of partons $a + b \rightarrow c + X$, $\hat{t} = (p_c - p_a)^2$ is a Mandelstam scattering variable), and $D_c^h(z)$ is the fragmentation function for the outgoing parton c which describes the the distribution of outgoing momentum fraction $z = p_h/p$ carried by hadrons from parton c . The factorization theorem is based on the assumption that these sub-processes can be factorized and then re-scaled to hard scales. The non-perturbative components of equation 1.2 cannot be calculated but can be measured experimentally.

They are factorized from the hard scattering cross-section, while the hard scattering cross-section is re-scaled to the scale of momentum transfer for a given process.

The theoretical phases of QCD matter from the strongly coupled low energy regime to the weakly coupled high energy regime are shown in figure 1.5. This treats a system of quarks and gluons as a thermodynamic system with temperature T and baryon chemical potential μ_B as the parameters. Baryon chemical potential is a measure of the energy needed to add a baryon to the system. Ordinary nuclear matter exists with a baryon chemical potential of ~ 900 MeV and a temperature close to 0 MeV [68]. In this phase, nuclear matter forms a gas of bound color neutral hadrons. QCD predicts a phase transition at sufficient energy density when quarks and gluons become deconfined to form Quark-Gluon plasma (QGP). The universe is theorized to have been entirely QGP in the moments after the Big Bang [91].

1.2 QGP Production in Heavy Ion Collisions

A heavy ion collision can be partitioned into three distinct stages; initial state conditions, pre-equilibrium phase, and QGP production and evolution. A diagram showing these three stages is shown in Figure 1.6. The collision is initiated by the overlap of two Lorentz contracted nuclei. The initial state is defined by the positions of nucleons within the nuclei and cannot be precisely known due to fluctuations in the positions of incoming nucleons within the nucleus [58]. These conditions, such as the number of nucleons participating in the collision N_{part} , and the number of binary nucleon-nucleon collisions N_{coll} , will influence the initial energy density, final particle production, and final momenta distribution of a collision. The expectation values of N_{part} and N_{coll} are related to the collision impact parameter b which is defined as the transverse distance between the centers of the colliding nuclei.

The impact parameter cannot be directly measured for a given collision but can be inferred through the average multiplicity of produced final state particles. The number of final state particles produced during a collision is assumed to increase

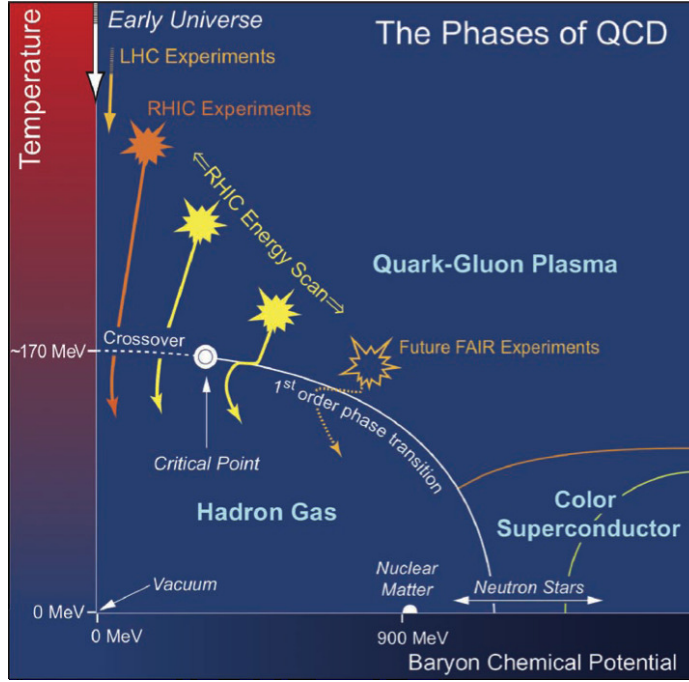


Figure 1.5: Quantum Chromodynamics matter phase diagram as a function of baryon chemical potential μ_b and temperature T . This figure is taken from ref [64].

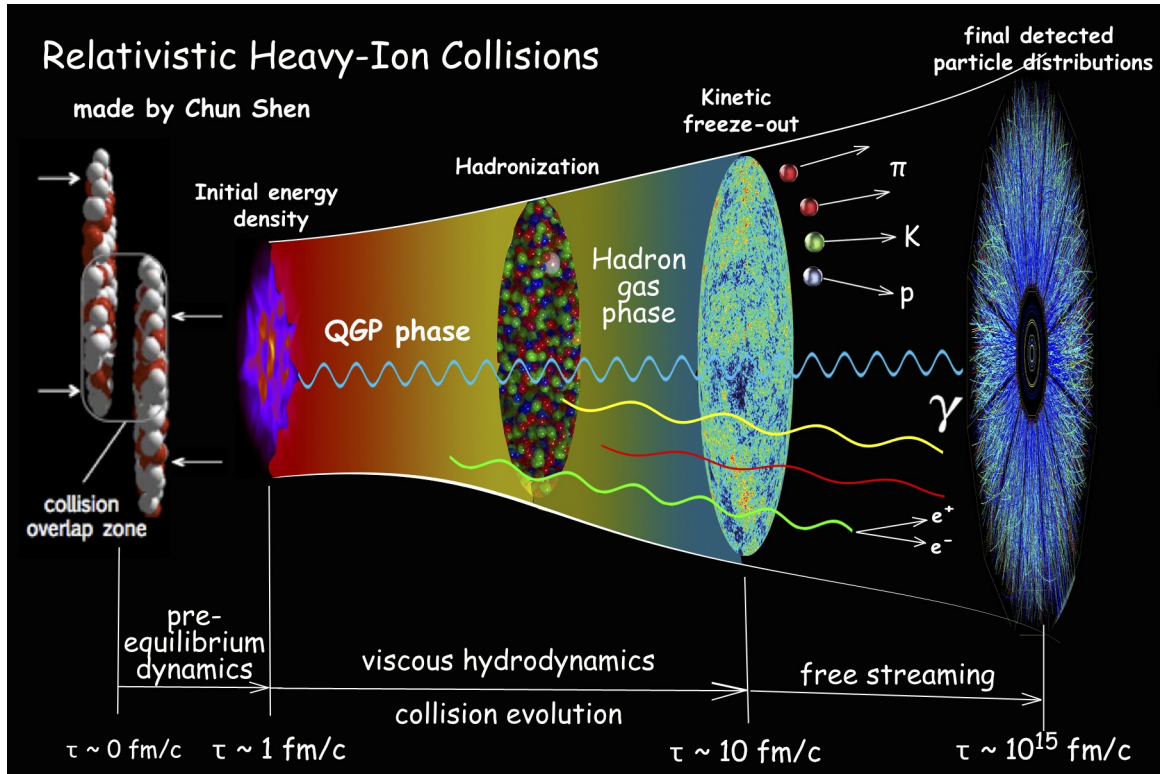


Figure 1.6: Diagram showing the time Evolution of a Heavy Ion Collision. Figure taken from Ref. [91].

monotonically as a function of the overlap volume between nuclei. This correlation is used to define the concept of event centrality as a fraction of total cross section. “Central events” have higher particle multiplicity and correspond to collisions with a smaller average impact parameter, whereas “peripheral events” have lower particle multiplicity and correspond to collisions with a larger average impact parameter. The relationship between impact parameter and particle multiplicity is determined by modeling the incoming nuclei as a collective of independent nucleons using a Glauber model [96]. Measurements of final state momentum distributions in collision systems with differing initial geometries indicate that initial conditions are well described by models based on nuclear geometry [32]. A Glauber model approximates the positions of nucleons within a nuclei and assumes that particle multiplicity is a function of N_{part} . Figure 1.7 demonstrates the practical method used for defining event centrality. The observed particle multiplicity is fit to determine a parametric description of the data, and the total energy measured in forward detectors is used as a calibration for N_{part} . After the fit, the data are binned by the fraction of the total cross section.

During the initial moments of the collision ($\sim 1 fm/c$) the energy density and temperature become sufficiently high to produce a medium of QGP. Current estimations for critical temperature and critical energy density are $T_c \approx 155 - 175$ MeV and $E_c \approx 1 \text{ GeV}/fm^3$ [116]. The initial energy density of a heavy ion collision can be estimated via the Bjorken formula [46, 106]

$$\epsilon_B = \frac{dE_T}{dy} \frac{1}{\tau_0 A_\perp}, \quad (1.3)$$

where $\frac{dE_T}{dy}$ is the transverse energy per unit rapidity y , τ_0 is the proper time shortly after the initial collision, and $A_\perp$ is the overlap transverse area of the collision or the initial size of the medium. In the most central Pb+Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ Equation 1.3 can yield energy densities up to $14 \text{ GeV}/fm^3$ at $\tau_0 = 1 \text{ fm}/c$, well above the expected critical energy density quark-gluon plasma production [54]. The QGP produced in heavy ion collisions has been found to be a strongly interacting liquid

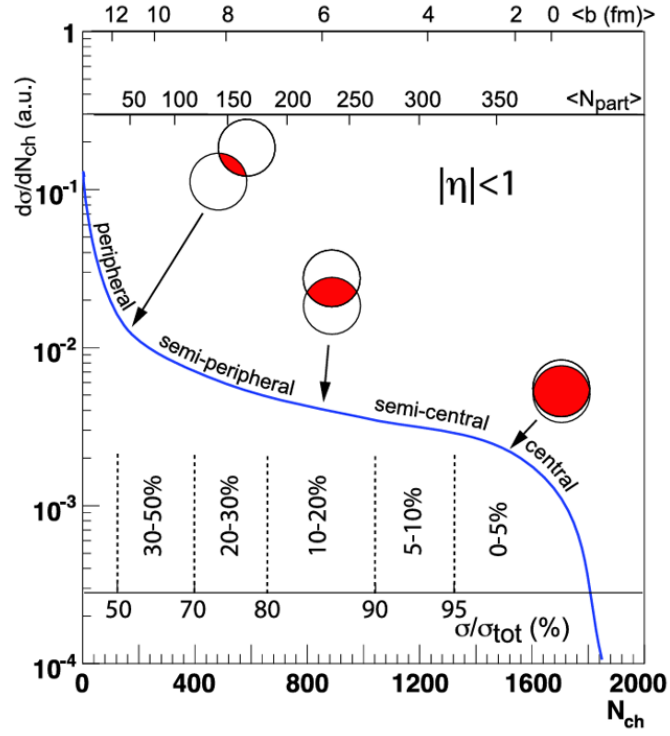


Figure 1.7: Cartoon illustrating the relationship between the average impact parameter, $\langle b \rangle$ and the charged particle multiplicity N_{ch} . Figure taken from ref. [96].

that is a nearly perfect fluid, with a shear-viscosity to entropy density ratio η/s to the theoretical limit perfected by QCD [9, 75, 62]. The QGP approaches near thermal equilibrium quickly after production $\sim 1\text{fm}/c$ which generates thermodynamic pressure within the medium, leading to rapid collective expansions of the reaction zone [74]. After $\tau \sim 5\text{ fm}/c$, the medium cools to the critical transition temperature T_c and begins to undergo a phase transition into a hadron gas. In this phase, the partons are confined back into color neutral hadrons and the degrees of freedom cease to be partonic and become hadronic. This transition continues until the medium reaches the chemical freeze out temperature T_{ch} when all partonic interactions end and the hadronic ratios are fixed [58]. Shortly after the formation of a hadron gas, the medium reaches the kinetic freeze-out temperature T_{fo} at which point elastic collisions within the medium cease and the hadrons reach their final energy and momenta. The kinetic freeze-out temperature is found to have a strong dependence on the impact parameter of the collision [76]. The phases of evolution of QGP from production to final state particles are shown in figure 1.8.

The collective expansion of the medium due to initial thermodynamic pressure is referred to as “flow”. This phenomenon results in an azimuthal anisotropy in the observed momenta of final state particles. This anisotropy in medium expansion, relative to the reaction plane, is caused by the initial geometry of the colliding nuclei. Measurements of final state azimuthal anisotropies in collision systems with different intrinsic initial geometries corroborate the geometric model for flow [32]. Azimuthal anisotropies observed in final state hadrons are a key signature for QGP production. This anisotropy can be effectively modeled by hydrodynamic flow, where the final momentum distribution of hadrons is a product of initial spatial anisotropies, which are translated into momentum space through pressure gradients [77, 81, 78]. The translation into momentum space through initial pressure gradients is shown in Figure 1.9.

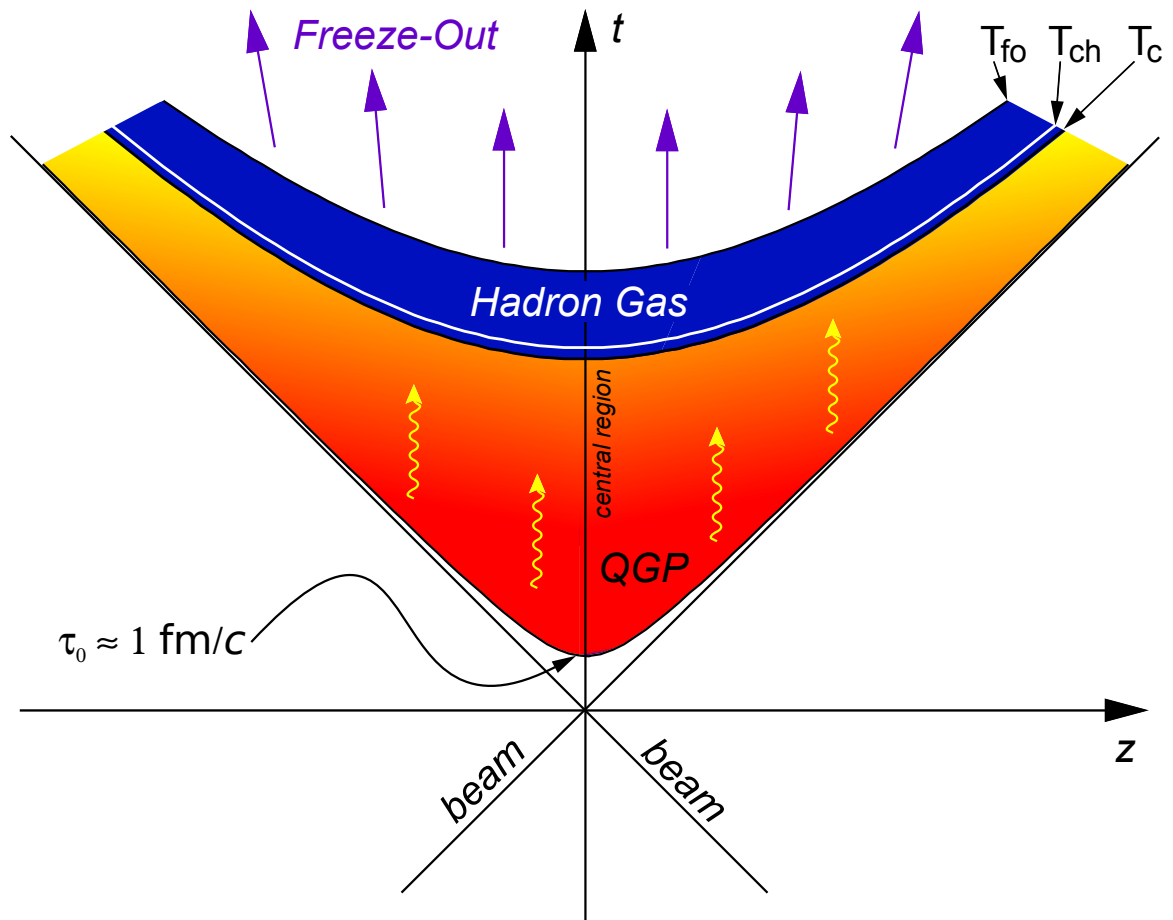


Figure 1.8: A light cone diagram showing the evolution of QGP from production at time τ_0 , transition to a hadron gas at temperature T_C , chemical freeze-out at T_{ch} , and kinetic freeze-out at temperature T_{fo} [58].

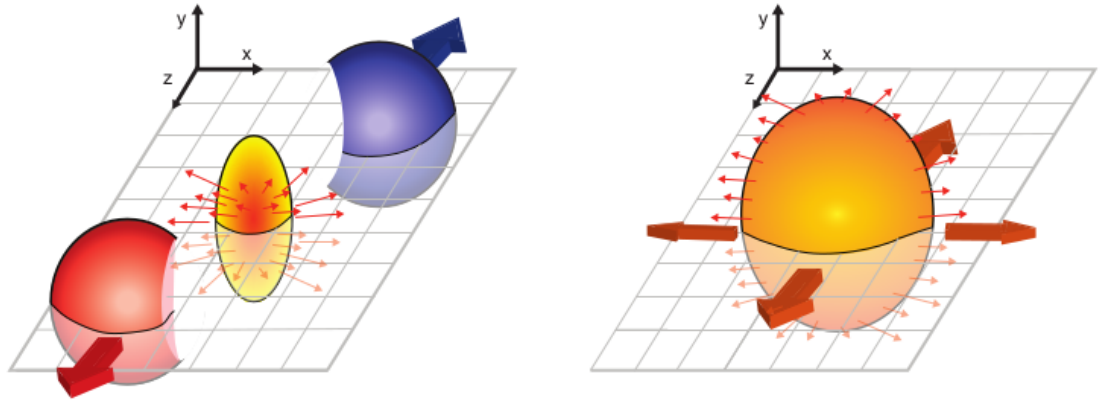


Figure 1.9: Diagram showing the translation of initial spatial anisotropies into pressure gradients of the expanding medium. Figure taken from Ref. [58].

The angular distribution of final state particles can be written as a Fourier-cosine expansion [121, 101]

$$\frac{dN}{d\phi} \propto \left[1 + 2 \sum_{n=1}^{\infty} v_n \cos(n(\phi - \psi_n)) \right], \quad (1.4)$$

where ϕ is the azimuthal angle of the particle, ψ_n is the n^{th} order symmetry plane, and v_n is the n_{th} order anisotropy coefficient. The magnitude of the Fourier coefficients decreases for increasing n [22]. The 2^{nd} -order harmonic v_2 , referred to as “Elliptical flow”, measured by the PHENIX collaboration at RHIC in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV for charged particles increases in less central events, shown in Figure 1.10. This is caused by an increased eccentricity ϵ_2 axis at a larger impact parameter for spherical nuclei. At low momenta ($p_T < 3$ GeV) the v_n mostly comes from particles unrelated to a hard process.

1.3 Jets and Underlying Event

Initial high momentum transfer collisions between partons in hadron collisions are called hard-scatterings. The high virtuality partons produced in hard-scattering events subsequently evolve into lower virtuality partons via the radiation of soft and collinear gluons. These partons and their children cannot exist as free particles due to color confinement so parton-showering continues until the partons form into color neutral hadrons. This process, referred to as hadronization, leads to a spray of particles originating from the initial parton, known as jets [104]. The hard-scattering process is in the asymptotically free regime of QCD, meaning it can be calculated perturbatively. This makes jet measurements an attractive way to compare QCD calculations to experimental data since measurements of jets produced in hadron collisions are a proxy for the scattered parton.

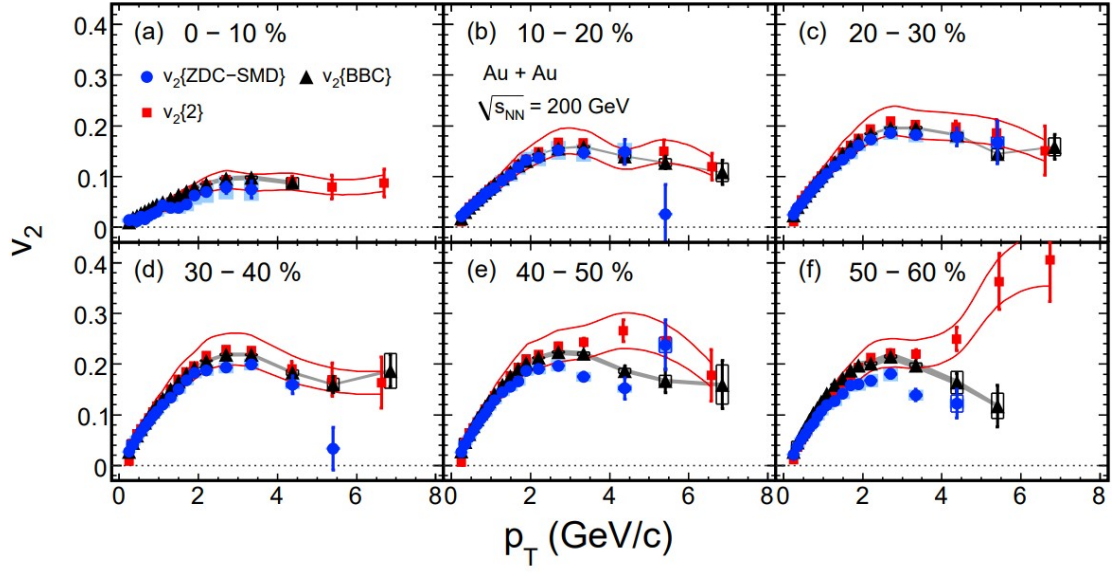


Figure 1.10: Charged hadron differential v_2 in Au+Au $\sqrt{s_{NN}} = 200$ GeV collisions measured by PHENIX using particle cumulants (red), and event plane resolution methods (blue, black) [30].

There are other physical processes besides hard-scattering events occurring in the simpler $p + p$ collision. These background effects are amplified when transitioning to higher entropy $A + A$ collisions. For soft gluon emission, the resulting hadrons can be soft enough that they could be mistakenly attributed to soft processes and excluded from the jet. This creates ambiguities in which final state particles are to be included as part of the final jet candidate. The Snowmass Accord [82] led to an agreement on the necessary criteria for well-defined jet measurements based on an algorithmic approach. This accord states that jet definitions must be equivalent to implement in both experimental analysis and theoretical calculations, well-defined to arbitrary order, and relatively insensitive to hadronization models. These criteria are met by requiring jet algorithms to be both infrared and collinear-safe. Infrared safety means the algorithm is unaffected by the presence of soft-gluon emission, and collinear safety refers to small angle splitting during the partonic shower.

There are four commonly used infrared and collinear safe jet-finding algorithms. These are the k_t , anti- k_t Cambridge/Aachen, and SISCone algorithms [40, 53]. The k_t , anti- k_t , Cambridge/Aachen algorithms are categorized as sequential recombination algorithms, while SISCone is a cone algorithm. This analysis uses both the k_t and anti- k_t algorithms. The anti- k_t [52] algorithm is less sensitive to the underlying event as it clusters the hardest final state particles first. This feature is useful when measuring jets in background-dominant heavy-ion collisions and for this reason, is the most commonly used algorithm. The k_t algorithm is useful for the exact opposite reason, as it starts with the softest splittings, which leads to large, asymmetric jets which are useful for characterizing properties of the underlying event in heavy ion collisions [52, 95]. Figure 1.11 shows an example of the differences in jets clustered with the anti- k_t and k_t algorithm in a simulated $p+p$ collision.

Sequential jet finding algorithms use the following process for clustering final state particles into jet candidates. First, the distance metric between each candidate particle pair (d_{ij}) and the distance between the particle and the beam-axis (d_{iB}) is

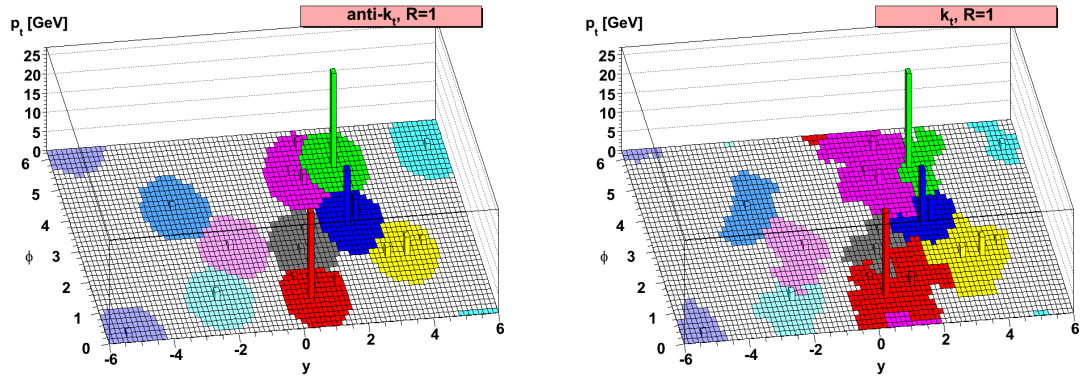


Figure 1.11: Illustration of the outcome of the clustering process for the anti- k_t (left) and k_T (right) algorithms used in simulation [52].

calculated for all particles. The distance metrics are defined as

$$\mathbf{d}_{iB} = p_{t,i}^{2n}, \quad (1.5)$$

for particle-beam axis distances

$$\mathbf{d}_{ij} = \min(p_{t,i}^{2n}, p_{t,j}^{2n}) \frac{\Delta_{ij}^2}{R}, \quad (1.6)$$

for distances between candidate particles where p_T is the transverse momentum of the particle, $\Delta_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$ is the spatial separation between final state hadrons i and j in detector coordinates, $\mathbf{d}_i$ is the distance between the beam-axis and particle i , and $\mathbf{d}_{ij}$ is the weighted momentum-space separation for all combinations of final state particles. The exponent n is different for each of the sequential clustering algorithms where $n = 1$ corresponds to k_t , $n = 0$ is the Cambridge-Aachen, and $n = -1$ is anti- k_t . Sequential algorithms are characterized by the jet resolution parameter, $R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, where η and ϕ are spatial coordinates of the center of mass frame. R is the effective termination parameter for cluster size, as when $\Delta_{ij} > R$ the distance metric between particles d_{ij} grows, making them less likely to be clustered. If the minimum distance $\mathbf{d}_{ij}$ between particles i and j is less than the distance of particle i to the beam-axis d_{iB} , then i and j are clustered into a jet-candidate. If the distance d_{iB} is less than d_{ij} then the object is a final jet candidate and is removed from the iteration list. This process continues until all objects have a distance d_{iB} less than the distance to the closest cluster.

1.3.1 Jets in Heavy Ion Collisions

Jet production in heavy ion collisions is a useful probe for properties of the medium, since the hard-scattering which produces them occurs prior to the production of QGP. While the hard-scattered partons fragment and hadronize, they will experience the full evolution of QGP and thus be modified when compared to those

produced in a vacuum. The production of jets in heavy ion collisions and those in vacuum are compared in Figure 1.12, which is a measurement of jet momentum production in Pb+Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV and $p+p$ from the ALICE collaboration [17]. The momentum spectra of jets in Pb+Pb are visibly lower than that measured in $p+p$ collisions which have been normalized by the ratio of binary nucleon-nucleon collisions to the inelastic nucleon-nucleon cross-section $\langle T_{AA} \rangle$. This suppression is due to in-medium energy loss for jets that experience QGP. During parton fragmentation within a strongly interacting medium, the partons lose energy from kinetic collisions with the medium and from gluon radiation induced by strong interactions [58]. This energy loss phenomenon is known as jet quenching.

One metric for jet energy loss in the medium is the jet R_{AA} which is a ratio of the yield of jets in A+A collisions to the scaled yield in $p + p$ collisions

$$R_{AA} = \frac{1}{T_{AA}} \frac{dN_{AA}}{dp_T} \bigg/ \frac{d\sigma_{pp}}{dp_T}, \quad (1.7)$$

where $\frac{dN_{AA}}{dp_T}$ is the heavy ion jet yield, $\frac{d\sigma_{pp}}{dp_T}$ is the pp cross-section, and T_{AA} is the nuclear thickness function. In the absence of jet medium modification R_{AA} would be close to unity. This is not the case as the jet R_{AA} has been measured to be less than one in Pb+Pb and Au+Au collisions at both RHIC and the LHC. Figure 1.13 shows the jet R_{AA} measured in Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV by the ATLAS [5], ALICE [21] and CMS [84] collaborations. Jet quenching is observed at all jet p_T regimes with the largest suppression occurring at low jet p_T corresponding to partons being more influenced by medium interactions. Observed jet momentum suppression is dependent on the path-length of the scattered parton through the QGP. Initial anisotropies in heavy ion collisions such as flow will influence the distance a parton traverses within the medium [7]. This suggests that reconstructed jets should experience higher suppression along the symmetry plane compared to jets oriented out of plane.

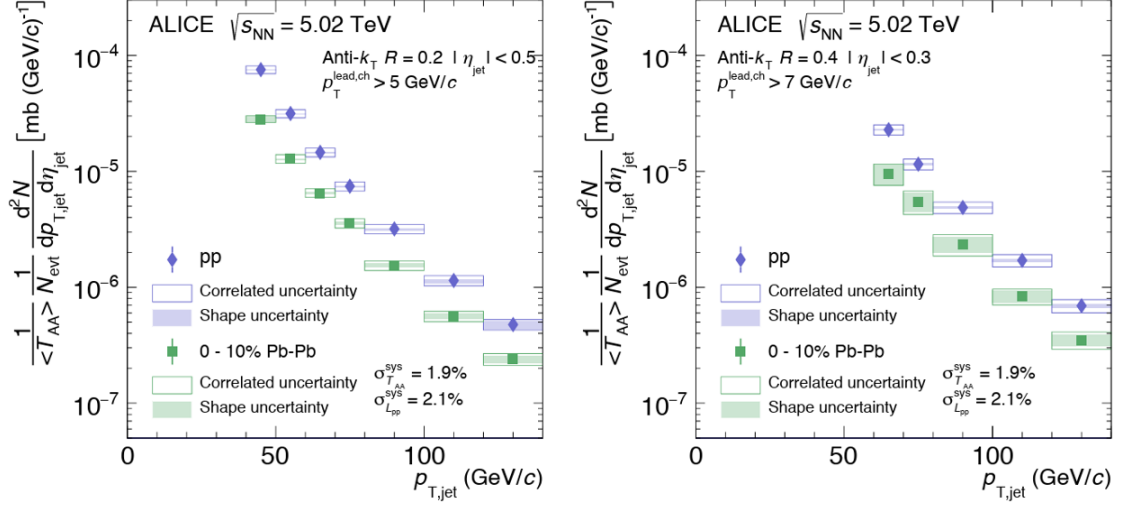


Figure 1.12: Full jet spectra for $R=0.2$ (left), and $R=0.4$ (right) measured in p+p and 0-10% central Pb+Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV [17].

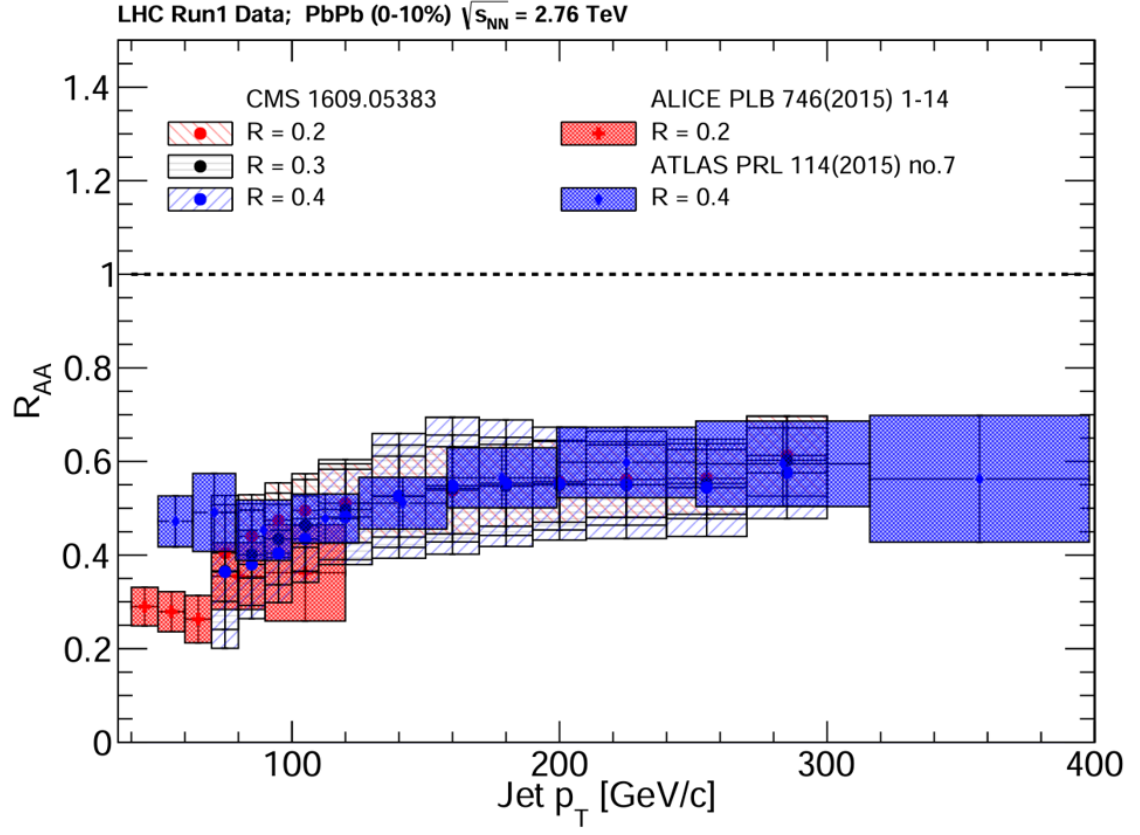


Figure 1.13: Jet R_{AA} from LHC run 1 data measured by ALICE for $R = 0.2$ jets at $|\eta| < 0.5$, by ATLAS for $R = 0.4$ jets at $|\eta| < 2.1$ and by CMS for $R = 0.2, 0.3, 0.4$ jets at $|\eta| < 2.0$ in Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV [58].

1.3.2 Jet Background Subtraction

Low momentum jets are most susceptible to the medium modifications observed in heavy ion collisions. Jets produced in lower heavy ion collisions with a lower center of mass collision energy, such as those at RHIC, are able to probe the medium properties of an entirely different phase space distribution than that produced in LHC collisions. The medium produced at RHIC is created with a lower initial temperature and thus has a smaller medium virtuality [27]. Calculations for the evolution of initial hard-scattered parton virtuality in units of $1/\text{fm}$ as a function of the local temperature of the QGP for RHIC and the LHC are shown in Figure 1.14. The red (black) curves correspond to different initial parton energies in the RHIC (LHC) medium. The thickened segments indicate regions where medium-induced virtuality significantly impacts parton splitting. The highest energy partons at the LHC remain dominated by vacuum virtuality evolution for over 10 fm/c. In contrast, lower-energy jets at RHIC are more influenced by the medium evolution and temperature [27].

The color-probes produced at RHIC are also a different population of jets than those produced from hard-scattering at the LHC. The fractions of initiating partons in jets with momentum greater than ~ 20 GeV are inverted to that of the jet population at the LHC. The fractions of quark and gluon-initiated jets at the LHC and RHIC are shown in Figure 1.15. Jet properties are influenced by the initiating parton species [87]. Gluon-initiated jets are typically wider, with a softer jet core and more long-range splittings due to gluon collinear radiation during parton fragmentation, whereas quark-initiated jets are typically more collimated with a well defined jet core. This makes measurements of gluon-initiated jets, particularly in the low jet momentum regime, more difficult in a heavy ion environment where underlying event fluctuations can be on the order of the hard-scattering parton momentum [18].

The precision of jet measurements and their kinematic range are restricted by corrections for this background and its fluctuations. Background fluctuations lead to a larger uncertainty in each individual jet's momentum. The fluctuations also

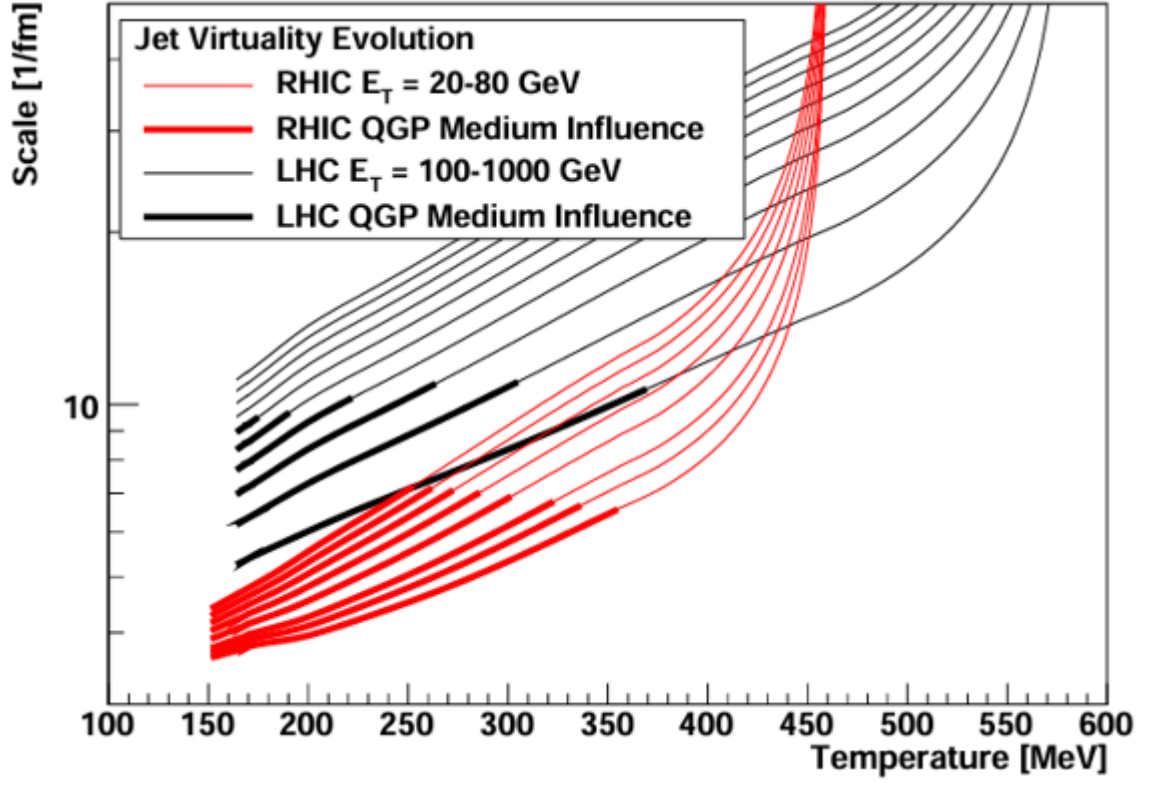


Figure 1.14: Scale probed in the medium in [1/fm] via high energy partons as a function of the local temperature in the medium. The red (black) curves are for different initial parton energies in the RHIC (LHC) medium.[\[27\]](#).

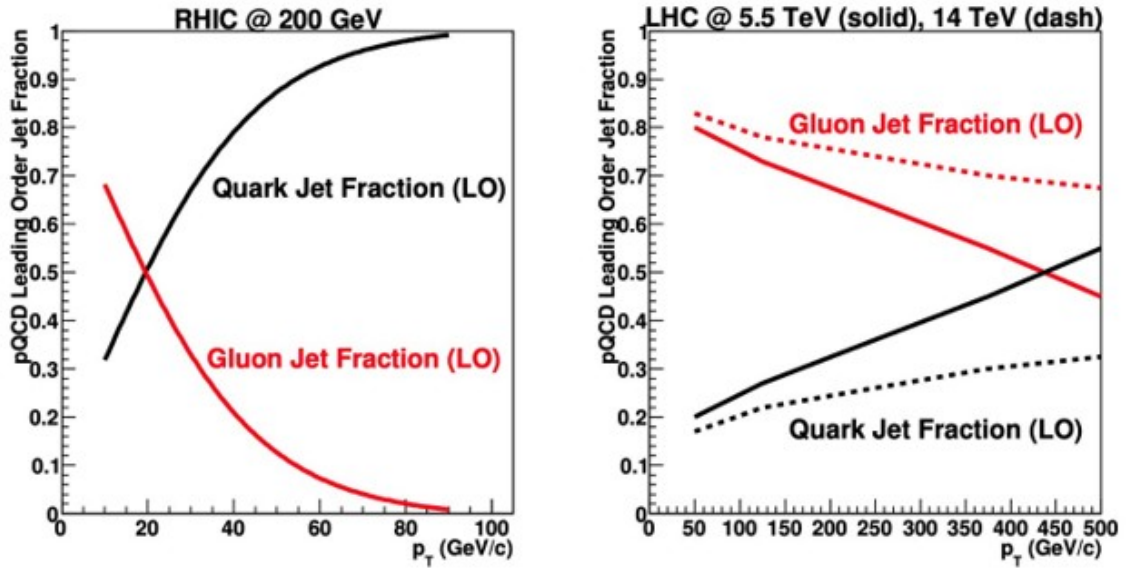


Figure 1.15: Comparison of the fraction of quark and gluon jets from leading order pQCD calculations for RHIC and LHC energies. [27].

make correcting for this background more challenging. The average pedestal can be corrected through a number of methods [52, 95, 71], but spatial and statistical fluctuations smear the corrected jet spectra. Figure 1.16 demonstrates the effect of correcting a fluctuating background with a pedestal subtraction technique.

Large fluctuations can be reconstructed by the jet-finding algorithm as purely statistical noise jets, known as combinatorial jets. The precise number of combinatorial jets is not well known, leading measurements to be restricted to a kinematic region where their contribution is negligible. Methods for characterizing and correcting for this underlying event are crucial to conducting jet measurements at the low momentum regime. Extending the kinematic range of jet measurements to lower momenta would improve constraints on partonic energy loss. Background fluctuations lead to smearing with a standard deviation independent of jet momentum. The combinatorial jet contribution is dominant at low momentum. The underlying event is the vast majority of particles present in heavy ion collisions.

This underlying event comprises soft particles unrelated to hard-scattering processes. The details of this background in jet measurements and its fluctuations are influenced by correlations arising from hydrodynamic flow and the shape of single particle spectra [122]. Naturally, machine learning has been applied to the problem of jet background subtraction, due to the complex nature of the underlying event in heavy ion collisions [70, 18]. A method using a neural network trained to predict truth jet momentum from reconstructed jet properties, such as area and angularity, has been developed and applied to Pb+Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV at ALICE. This method shows increased jet momentum resolution in for low momentum jets, allowing for an increased kinematic range. This technique is especially effective for suppressing the effects due to combinatorial jets. Figure 1.17 shows measurements of the jet R_{AA} in Pb+Pb collisions compared to the traditional area-based background subtraction method 4.3. The neural network is able to increase the kinematic reach of the R_{AA} measurement by 20 GeV for jets with resolution parameter $R=0.2$ and

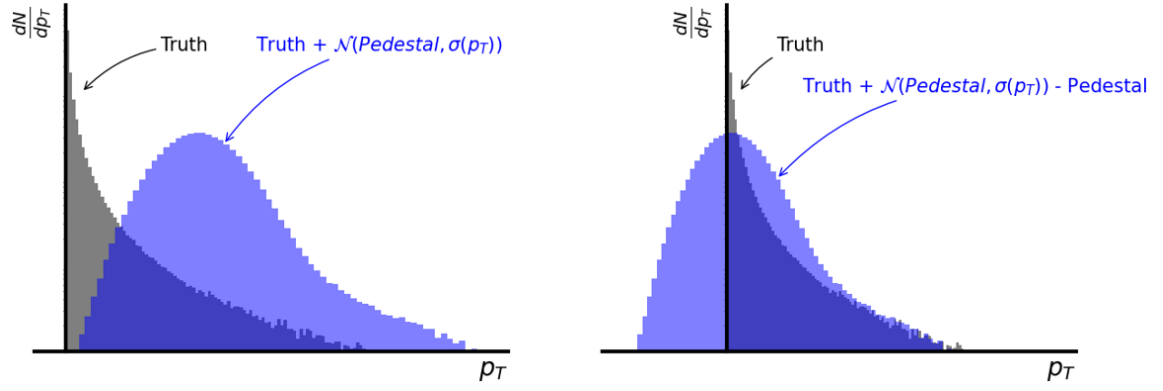


Figure 1.16: Diagram showing the effect of a fluctuating background on a steeply falling spectra corrected with the average pedestal.

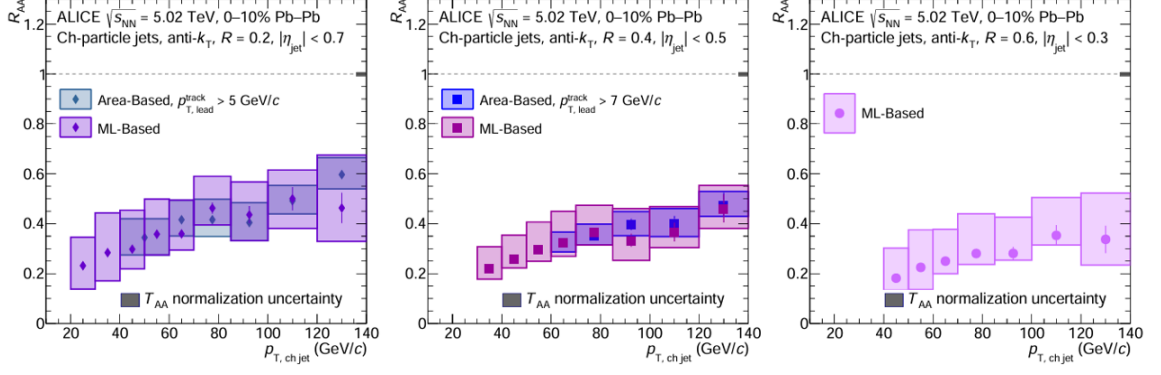


Figure 1.17: Nuclear modification factors of charged-particle jets for $R=0.2, 0.4$, and 0.6 jets in Pb+Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV using the area-based method (Equation 4.3) and the machine learning method [18].

allows for the first jet $R_a a$ measurements for larger $R = 0.6$ jets. This method will be discussed in further detail in Section 3.2.

Rigid random cones have been used to characterize underlying event fluctuations in Pb+Pb collisions by the ALICE collaboration. A random cone is placed within detector acceptance and then all charged particles within the cone radius $R = 0.4$ are summed into the constituent cone momenta. ALICE uses the area-based background subtraction method [52] to characterize the average pedestal of the underlying event. This technique was initially formulated to mitigate the underlying event in $p+p$ collisions under conditions of high pile-up [109], and has been subsequently adapted for application in heavy ion collisions as well [14, 21, 17, 23]. The underlying event fluctuations δp_T are calculated according to

$$\delta p_T = \sum p_{T,i} - \rho_A \cdot A, \quad (1.8)$$

where the cone area $A = \pi R^2$, $p_{T,i}$ is the individual momentum of charged particles within the cone, and ρ_A is the background momentum density per unit area. ρ_A is approximated as the median of $p_{T,\text{jet}}^{\text{tot}}/A$ for k_T jets, primarily due to the dominance of background within k_T jets [52, 53]. The underlying event fluctuations measured in 0-10% Pb+Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV for $R = 0.4$ cones are shown in Figure 1.18. This is consistent with the underlying event description of random particle emission from a single-particle momentum distribution described by

$$f_{\Gamma}(\delta p_T) = A \cdot \frac{b}{\Gamma(p)} \cdot (b \cdot \delta p_T + p)^{p-1} \cdot e^{-(b \cdot \delta p_T + p)}, \quad (1.9)$$

where Γ is the real-valued gamma function, b is the shape parameter, and p is the scale parameter [115, 80, 12].

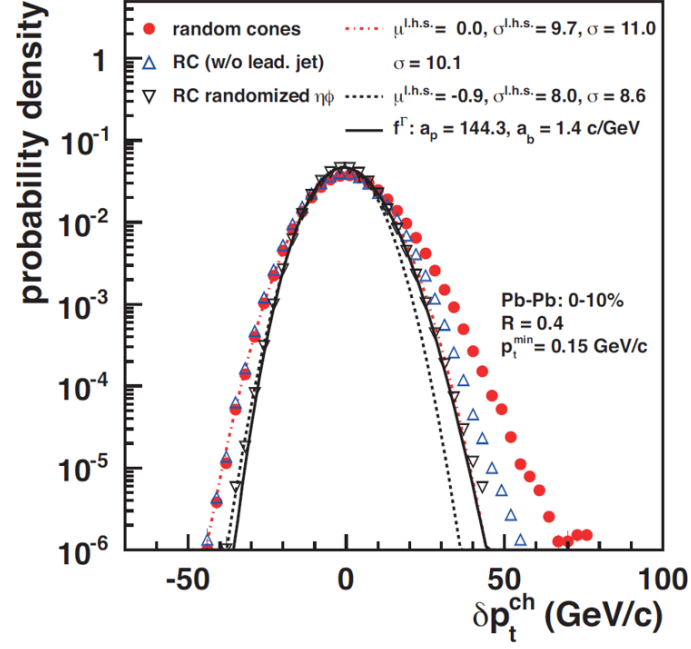


Figure 1.18: δp_T for random cones in Pb-Pb events. A Gaussian fit to the left-hand-side and its extrapolation to positive δp_T . The solid line is a fit to the distribution given by Equation 1.9 [12].

The standard deviation of δp_T in random cones is found to be a combination of the purely statistical limit and contributions from non-Poissonian sources

$$\sigma(\delta p_T) = \sqrt{\sigma^2(\delta p_T)_P + \sigma^2(\delta p_T)_{np}} \quad (1.10)$$

where $\sigma(\delta p_T)$ is the full standard deviation, $\sigma(\delta p_T)_P$ is the Poissonian statistical limit, and $\sigma(\delta p_T)_{np}$ is contributed from non-Poissonian sources, such as flow. The Poissonian limit is expressed as

$$\sigma(\delta p_T)_P = \sqrt{N\sigma_{p_T}^2 + N\langle p_T \rangle^2}, \quad (1.11)$$

where N is the expected number of particles within the cone radius, $\sigma_{p_T}^2$ is the variance of the single-particle momentum distribution, and $\langle p_T \rangle$ is the average momenta. The non-Poissonian contribution is approximated as

$$\sigma^2(\delta p_T)_{np} \approx N^2 \langle p_T \rangle^2 \sum v_n^2, \quad (1.12)$$

where v_n are the charged particle azimuthal anisotropy coefficients. This derivation assumes that each of the N particles is drawn from a single track momentum distribution given by equation 1.9 [115]. Figure 1.19 shows the dependence of the standard deviation of δp_T on uncorrected charged particle multiplicity. Slight variations in the width are observed due to deviations in the single track momentum distribution from a Gamma distribution and the momentum-dependent nature of v_n [80]. Several methods for correcting the underlying event in measurements attempt to incorporate azimuthal anisotropies into background estimations [71, 72, 84, 4].

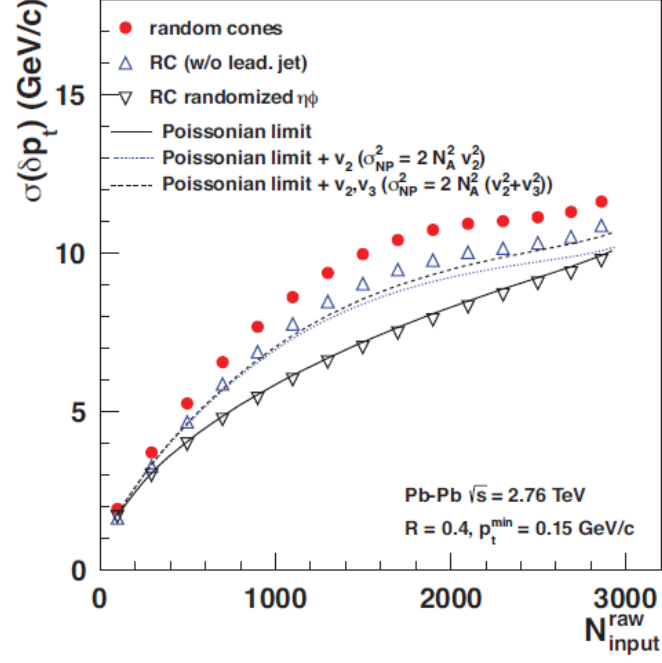


Figure 1.19: Dependence of the standard deviation of the δp_T distributions on uncorrected charged particle multiplicity, compared to the limit derived for Poissonian (Equation 1.11) non-Poissonian (Equation 1.12) contributions to fluctuations [12].

Measurements of the underlying event have been performed using calorimeter tower energy as input, rather than charged particles such as the ALICE random cone measurement. Characterizations of the underlying event and its spatial correlations, using calorimeter window areas, have been measured in Pb+Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV by the ATLAS collaboration [41, 3]. Figure 1.20 shows the comparisons between the fluctuations in calorimeter energy calibrated to the electromagnetic scale and compared to the HIJING event generator. There is good agreement between HIJING and measured window energy $E_T^{n \times m} - \langle E_T^{n \times m} \rangle$ up until very high energy windows. The standard deviation of the window energy is shown as a function of total energy in the forward calorimeter $\sum E_T^{FCal}$, which is a proxy for event centrality. There is a discrepancy for large windows $n \times m = 7 \times 7$ between HIJING and data at high $\sum E_T^{FCal}$, corresponding to central events [3].

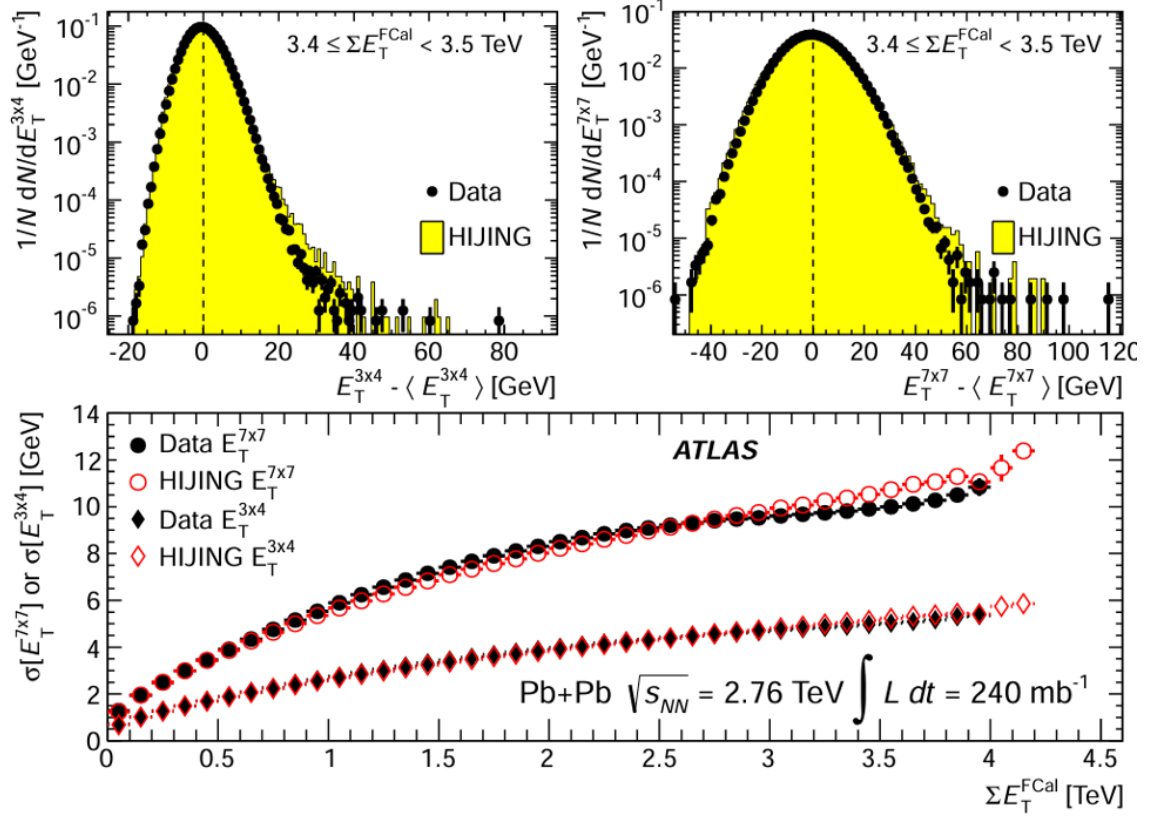


Figure 1.20: (Top) Distributions of $E_T^{n \times m} - \langle E_T^{n \times m} \rangle$ for $n \times m = 3 \times 4, 7 \times 7$ and (Bottom) standard deviation $\sigma(E_T)^{3 \times 4}, \sigma(E_T)^{7 \times 7}$ compared between Pb+Pb $\sqrt{s_{NN}} = 2.76$ TeV collisions and HIJING [3].

Chapter 2

Detector Equipment and Instrumentation

2.1 Relativistic Heavy Ion Collider

The Relativistic Heavy Ion Collider (RHIC) [73] is a synchrotron hadron accelerator located at Brookhaven National Laboratory (BNL) in Upton, New York. RHIC is a dual storage ring accelerator with two hexagonal rings, each 3.8 kilometers in circumference. The ring orbits beam species in opposite directions of each other with 6 interaction points where the collisions occur. The two rings are called “blue” for clockwise beams and “yellow” for counterclockwise beams. RHIC can collide a variety of ion species, with beam collision energies ranging from $7.7 \text{ GeV} < \sqrt{s_{NN}} < 200 \text{ GeV}$ for Au^{79+} ions and $200 \text{ GeV} < \sqrt{s} < 510 \text{ GeV}$ for polarized protons. RHIC can also collide mixed beam species such as $p + Au$, across a wide range of collision energies. Figure 2.1 shows an aerial overview of the RHIC complex.

The injection and storage of ions at RHIC involve several stages, starting from the ion source and progressing through a series of accelerators before reaching the collider/storage rings. The injection procedure for $Au + Au$ collisions is described in the following section. The injection chain for a Au atom at RHIC begins at the

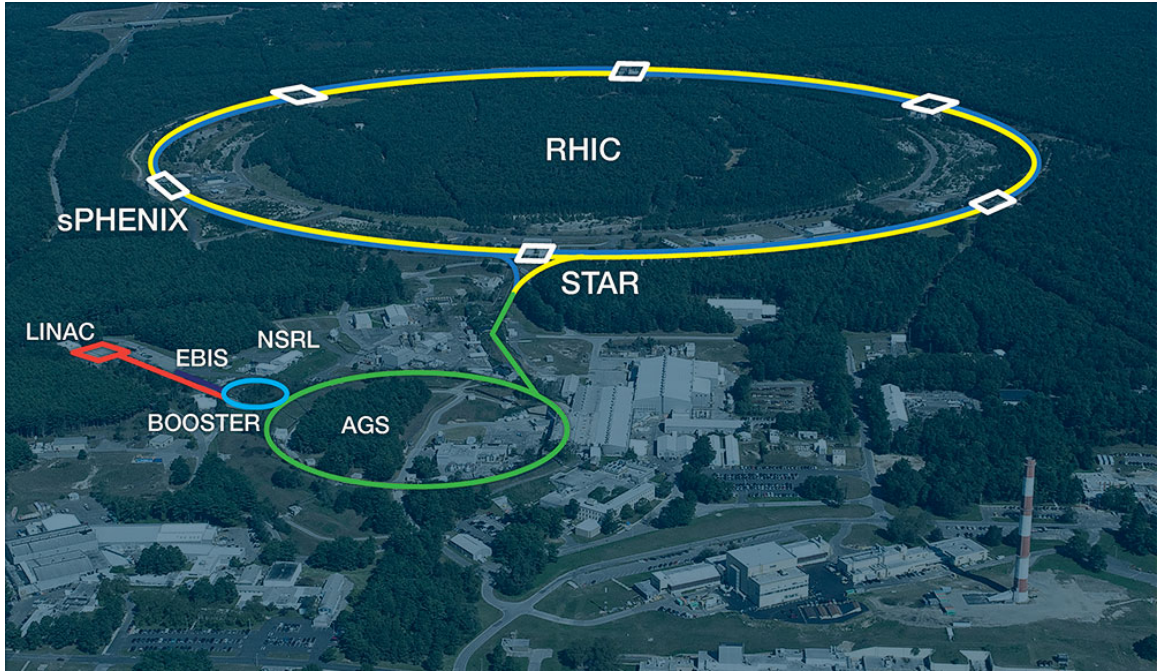


Figure 2.1: RHIC collider complex at Brookhaven National Laboratory. Photo was taken from Ref. [125].

Electron Beam Ion Source (EBIS). EBIS is an upgrade from the previous Tandem Van De Graff cesium sputter source and has been operational since 2012 [37]. The Au atoms start out as a cool gas in the EBIS, where they are bombarded by incident electrons. During the cycle in the EBIS, the atoms are stripped of 32 electrons and accelerated to a kinetic energy of 2 MeV/ u . After EBIS, the Au^{32+} atoms are transferred to the booster [38], which is a 200 m synchrotron accelerator ring. The gold atoms are injected into the booster in groups of 6 bunches, with the entire booster cycle lasting 200 ms [37]. In the booster, the Au^{32+} atoms are accelerated to 37% of the speed of light [120]. The atoms leave the booster with 100 MeV/ u and enter the Booster to AGS (BtA) line. In the BtA there is a silica stripping foil which strips another 45 electrons yielding Au^{77+} . The Au^{77+} ions arrive at the Alternating Gradient Synchrotron (AGS) in 24 bunches, which are split between the 12 AGS buckets. The AGS re-bunches these buckets into 4 bunches. The AGS gives the Au^{77+} ions the final kick up to 8.8 GeV/ u before splitting them into the AGS-to-RHIC line (AtR). The AtR is home to the final stripping foil which is a Al_2O_3 foil that strips off the final two electrons. Finally the Au^{79+} atoms are divided into the two counter-orbiting beams by a switching magnet. The switching magnet organizes the beam into 111 buckets for both storage rings. While in the rings, the ions are accelerated to the top collision energy, typically 100 GeV/ u by the 6 groups of 1740 niobium-titanium super conducting magnets located at the curve of the hexagonal rings. Beam-beam collisions occur at any of the 6 interaction regions which are labeled by position in the a clock. The 8 o'clock interaction region is home to the sPHENIX experiment. A diagram of the injection procedure of ions into RHIC is shown in figure 2.2.

2.2 The sPHENIX Detector

The super Pioneering High Energy Nuclear Interaction eXperiment (sPHENIX) is the newest high energy experiment in the world at the time of writing. Following its

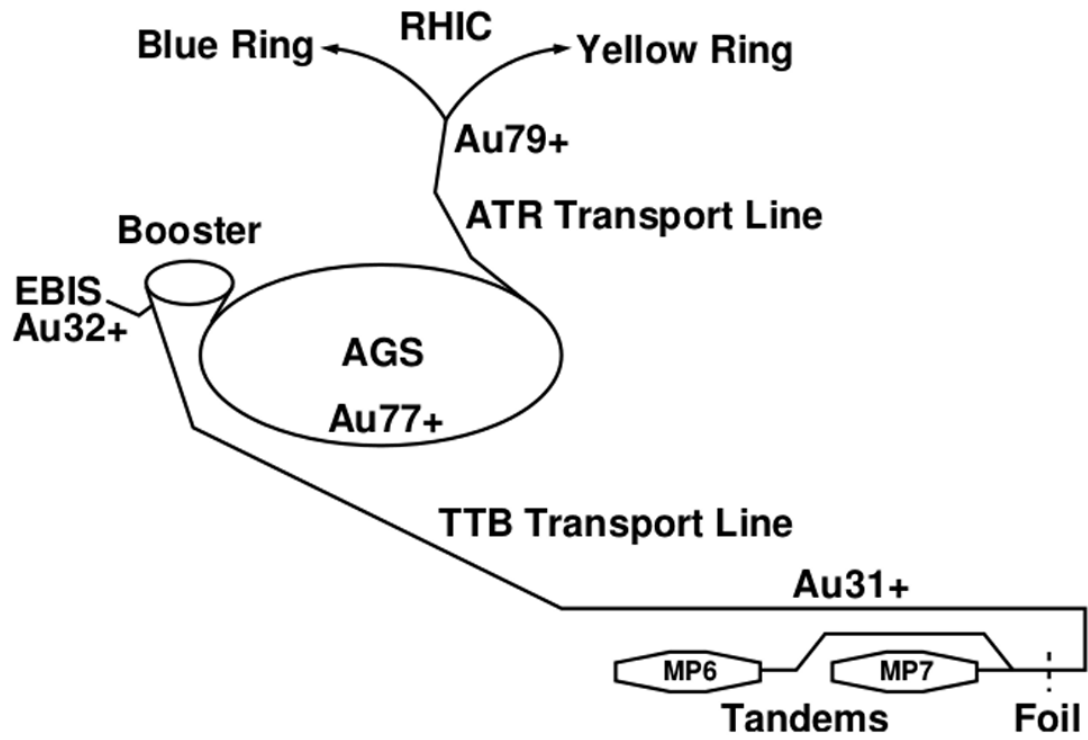


Figure 2.2: A schematic diagram of injection of ions to RHIC. The figure is taken from reference [124].

completion in late 2022, sPHENIX began commissioning in May 2023 with $Au + Au$ collisions at $\sqrt{s_{NN}} = 200$ GeV at RHIC. Originally envisioned as a major upgrade to its predecessor, PHENIX, the sPHENIX detector has evolved into an entirely new experiment with a focused mission: to deepen our understanding of the formation and properties of the QGP[27]. The design of sPHENIX, as depicted in figure 2.3, reflects this mission, integrating each sub-system to achieve its scientific goals.

sPHENIX is equipped with full azimuthal coverage and large pseudo-rapidity acceptance $|\eta| \leq 1.1$ electromagnetic and hadronic calorimeters. Notably, the mid-rapidity hadronic calorimeter system is the first of its kind at RHIC, enabling measurements of neutral hadronic contributions to jets. The central tracking volume, comprises the MAPS-based VerTeXing detector (MVTX), the Intermediate Silicon Tracking detector (INTT), the Time Projection Chamber (TPC), and the TPC Outer Tracker (TPOT) are designed with full streaming readout, meaning that the full luminosity provided by RHIC is sampled. Event characterization at sPHENIX is provided by three sub-systems, namely the Minimum Bias Detector (MBD), the Zero-Degree Calorimeter (ZDC), and the sPHENIX Event Plane Detector (sEPD). These event characterization detectors, along with the central calorimeters are read out at 15 kHz [27]. The detector’s subsystems are carefully integrated within a semiconducting solenoid magnet, repurposed from the BaBar experiment, which produces a central magnetic field of up to 1.4 T. Table 2.1 provides a basic definition and purpose of each sub-system in sPHENIX [111]. More detailed descriptions of the design and functionality of each of these sub-systems, starting from the inner-most MVTX, and continuing through the outer hadronic calorimeter and event characterization detectors, are provided in the following sections.

2.2.1 MAPS-based Vertex Detector

The inner-most tracking detector at sPHENIX is the MAPS based vertex detector (MVTX). The MVTX consists of 3 layers of silicon barrels, spanning the immediate

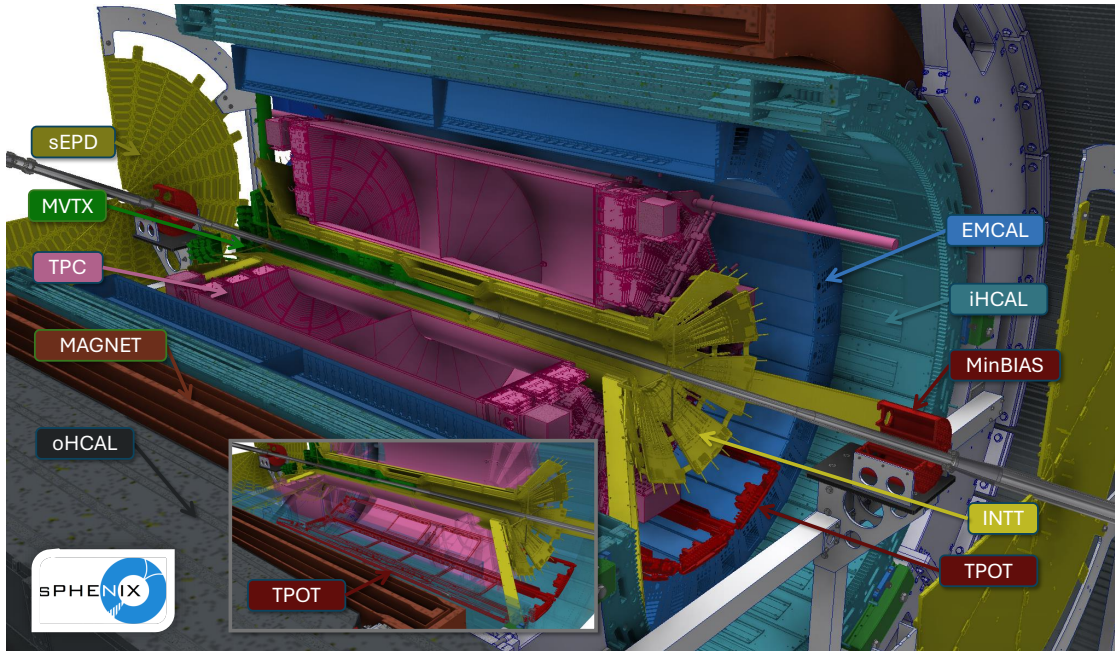


Figure 2.3: Overview of sPHENIX detector with labeled sub-systems [112].

Table 2.1: Brief summary of the sPHENIX sub-systems.

Name	Purpose
MVTX	Low momentum tracking, vertexing, and secondary vertex detection.
INTT	High momentum tracking, timing.
TPC	High momentum tracking, vertexing, and PID.
TPOT	Timing Calibration for TPC [42].
EMCal	Neutral particle contribution [34, 33].
HCal	Hadronic energy contribution [33].
sEPD	Event plane determination.
MBD	Spectator Neutron detector.
ZDC	Multiplicity measurements and pileup rejection

radial volume 2-5 cm from the beam pipe, $|\eta| < 1$ and full azimuth. The layers are made of multiple segments called staves constructed into a barrel. Each stave consists of a hybrid integrated circuit (HIC), made of 9 pixel chip sensors and a flexible polyimide circuit. The polyimide circuit supplies power and signal read out for the 9 pixel chips. The pixel chips are ALICE Pixel Detector (ALPIDE) sensors. The sensor strip is soldered to a cooling plate and a space frame for structural support [110]. Figure 2.4 illustrates the make up of each stave in a barrel layer.

The 3 layers in the MVTX contain 48 staves with a combined total of 432 ALPIDE sensors. ALPIDE sensors use monolithic active pixel sensor technology to produce high speed readout with spatial resolution in the order of $30\ \mu\text{m}$ [110]. The ALPIDE sensors are each $50\ \mu\text{m}$ thick with a surface area of $15\ \text{mm} \times 30\ \text{m}$.

That small surface area contains a matrix of 2^{19} pixels within a high-resistivity epitaxial layer [31]. When charged particles pass through the sensor, charge carriers are freed within the p-doped epitaxial layer. The charge carriers are confined to lateral diffusion towards the n-well signal diode (pixel). Therefore the charge carriers that reach the signal diode are only those deposited directly into the active sensor volume. This provides precise spatial resolution and limits readout to information about a particle crossing the sensor. The inner radius of the 3 layers of the MVTX is given in Table 2.2 as well as the number of staves in each barrel-layer.

2.2.2 Intermediate Silicon Tracking Detector

The Intermediate silicon Tracking (INTT) detector is a silicon strip detector installed outside of the MVTX. The INTT consists of 2 layers of semiconducting silicon barrels. The tracking volume of the INTT is a radial distance of 7-10 cm from the beam pipe with an acceptance of $|\eta| < 1.1$, full azimuth and $|z| < 10\ \text{cm}$ [111]. Each of the tracking layers is divided into two sub-layers which are staggered to achieve full azimuthal coverage.

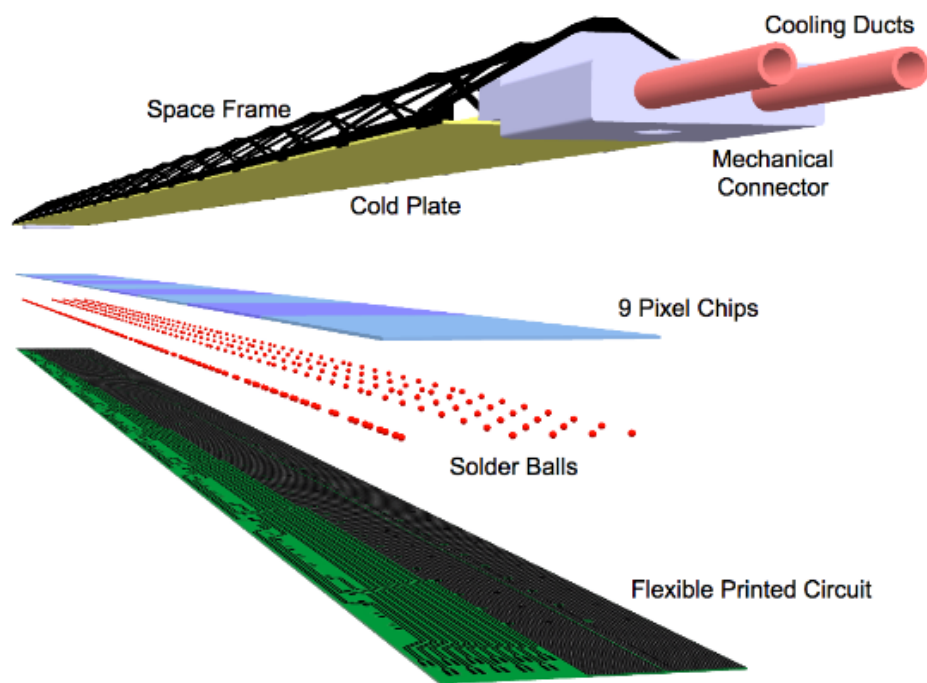


Figure 2.4: MVTX stave diagram [110].

Table 2.2: MVTX layer parameters [110].

Layer	Inner Radius [mm]	Number of staves	ALPIDE Sensors
1	23.7	12	108
2	31.4	16	144
3	39.1	20	180

Figure 2.5 shows a schematic of the ladders that make up each of the sub layers of the INTT. The ladders consist of two silicon modules attached to a mechanical support. The mechanical support supplies water cooling to the silicon tracking modules. The silicon modules are both read out from one side with a High Density Interconnect (HDI) board. Each of the silicon modules has two single-sided, AC coupled silicon sensors. The silicon sensors are $320\text{ }\mu\text{m}$ thick and depending on the type, Type-A or Type-B, which have $128\text{ mm} \times 20\text{ mm}$ or $100\text{ mm} \times 20\text{ mm}$ of active tracking area, respectively [111].

2.2.3 Time Projection Chamber

The main tracking sub-detector in sPHENIX is the Time Projection Chamber (TPC). The TPC is a continuous readout detector which consists of a cylinder spanning a radius of 20 cm to 80 cm from the beam pipe with full azimuth coverage and $|z| < 1.05\text{ m}$ along the beam pipe [85]. The TPC tracking volume is confined by the cylindrical volume between the inner and outer field cages and the readout planes on either end cap. The volume is separated into two symmetric “mirror-volumes” by the central membrane (CM) electrode, as shown in figure 2.6.

The central membrane electrode is set to high voltage, while the readout planes are set to ground. This creates an electric field within the TPC volume, which is filled with a gas mixture of Argon, Tetrafluoromethane, and Isobutane. As charged particles pass through the gas, they ionize the Ne producing free electrons and positive ions. The electric field within the TPC volume transports the electrons to the readout planes on either end cap of the TPC [111]. The positive ions remain within the TPC gas volume during the collection of charge that accumulates on readout channels from the drifting electrons. To achieve continuous readout Gas-Electron Multiplier (GEM) foil stacks limit the fraction of positive ions that remain in the active TPC tracking volume [111].

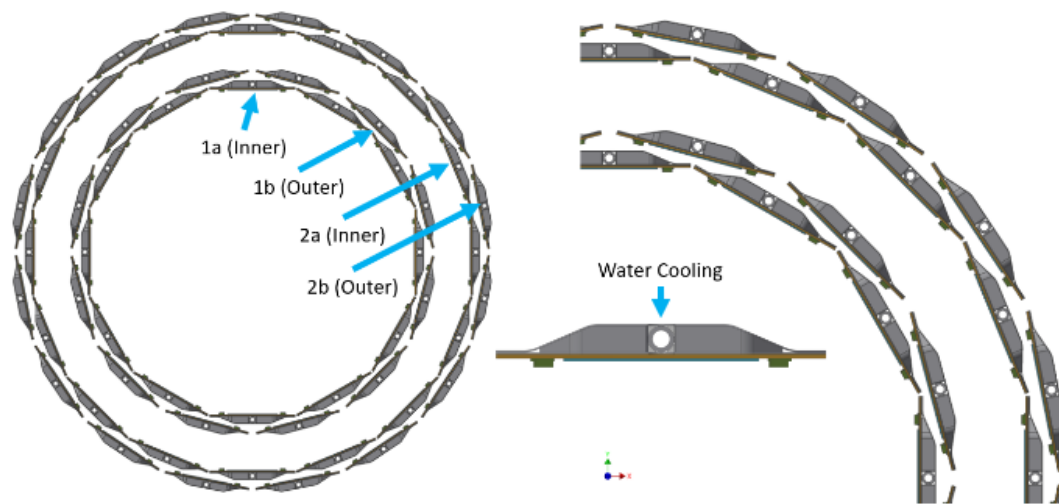


Figure 2.5: INTT cross section diagram [111].

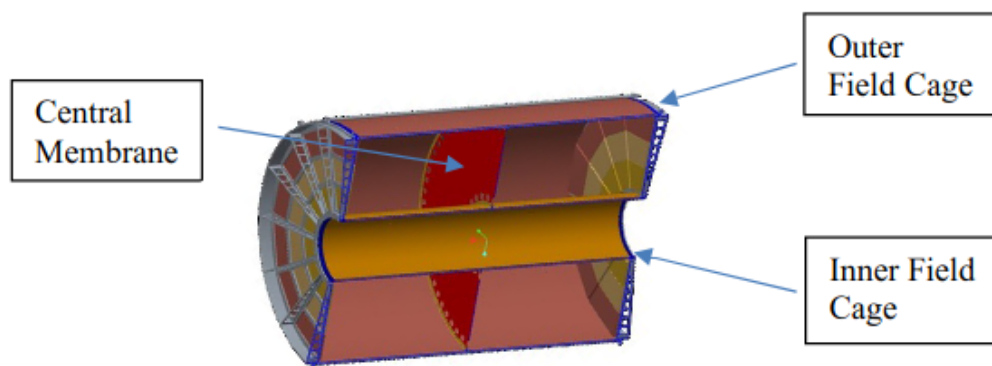


Figure 2.6: Mechanical design of TPC [111].

The end caps of the TPC are each segmented into 3 radial segments and 12 azimuthal segments called modules. Each module has a quad-GEMstack and 6, 8, or 12 FEE cards corresponding to the inner, middle, and outer radial segments, respectively. The FEE cards each have 8 SAMPA chips which collect drift charge accumulated on one of the 32 channels per SAMPA card. This corresponds to a total of 159744 channels for the TPC [111]. A hemisphere diagram of a segmented read out plane on either side of the TPC is shown in figure 2.7. The FEE chips can be seen in the corresponding radial segments.

2.2.4 Time Projection Chamber Outer Tracker

The Time Projection Outer Tracker (TPOT) is a Micromegas-based detector designed to aid in the calibration of the TPC by correcting distortions in the electron drift, which arise from time-averaged and beam-induced effects. The TPOT consists of eight modules, each containing two radially stacked Micromegas chambers. One chamber measures the longitudinal coordinate z along the beam axis, and the other measures the azimuthal coordinate ϕ . Each chamber's active area is 256 mm along ϕ and 512 mm along z . direction. A diagram of a single micromegas chamber is shown in figure 2.8. In total, the eight modules of the TPOT cover approximately 8% of the TPC's acceptance.

The following is an overview of how the Micromegas chambers in the TPOT work. A negative electric field is applied to guide electrons through the drift gap. The drift electrode, operated at a negative voltage, is 3 mm from the micromesh, which is grounded. The resistive layer, positioned 128 μm from the micromesh, is at a positive voltage and divided into four segments for independent High Voltage (HV) control. This segmentation allows the chamber to continue functioning even if part of it experiences a short circuit. The readout layer, beneath the resistive layer, has 256 strips that collect signals via capacitive coupling and connect to the Front-End Electronics (FEE) boards for data processing [43].

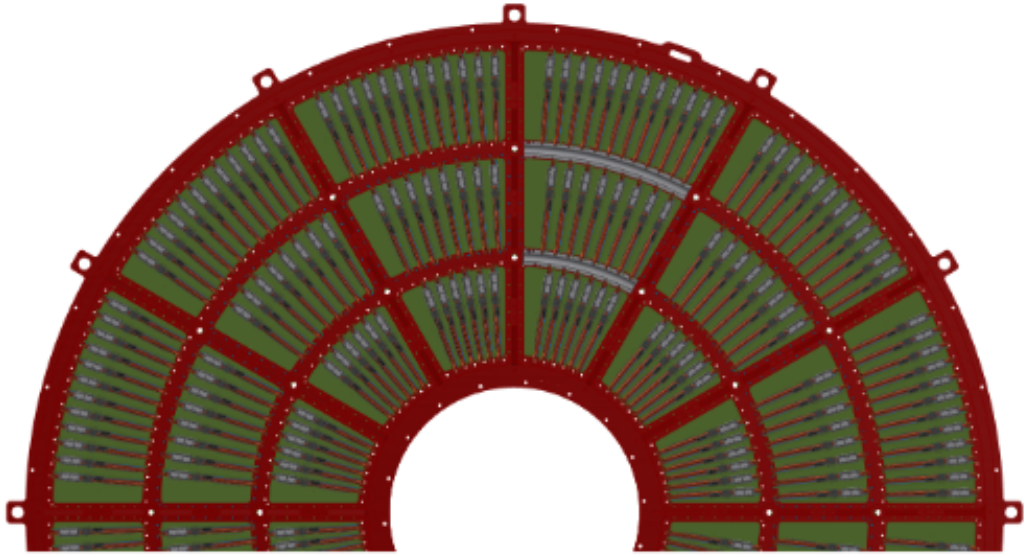


Figure 2.7: TPC modules and FEE chips [111].

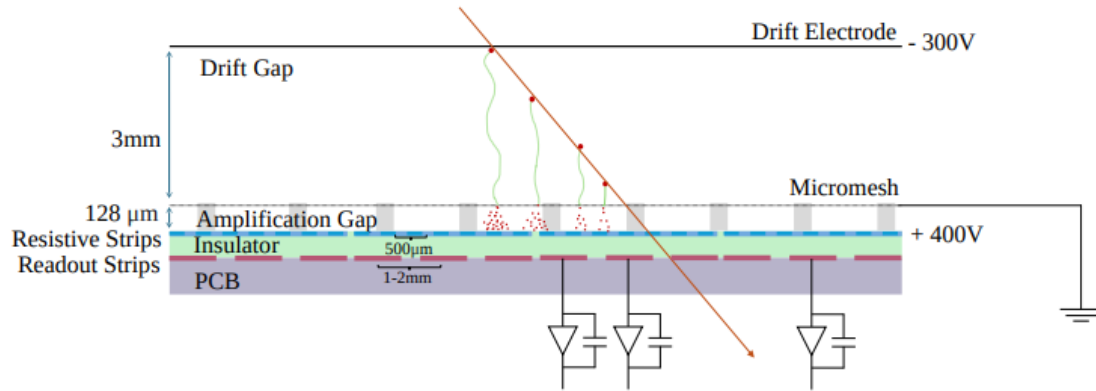


Figure 2.8: Schematic overview of a Micromegas chamber [43].

2.2.5 Calorimeters in sPHENIX

A cross section of the sPHENIX calorimeter system is shown in figure 2.9. The calorimeter system in sPHENIX consists of two types of scintillating sampling-calorimeters. The first is the electromagnetic calorimeter (EMCAL). The second is the hadronic calorimeter (HCAL) system. The HCAL is divided into two sub-sections. The inner most HCAL is installed within the cryostat cylinder of the BaBar magnetic. The larger outer HCAL is outside the solenoid at a radius of 182 cm from the beam pipe [111]. The EMCAL and both sub-sections of the HCAL cover $|\eta| < 1.1$ and full azimuth $0 < \phi < 2\pi$.

The EMCAL is used to measure energy from photons and electrons produced in heavy ion collisions. The EMCAL is installed within the BaBar magnet with an inner radius of 90 cm. The detector is made up of 64 sectors which are placed in two layers of 32 sectors each. There are 24,596 towers in the EMCAL with a segmentation of $\Delta\eta \times \Delta\phi = 0.025 \times 0.025$ [111]. The material design of the EMCAL consists of 3D-printed blocks of tungsten powder with scintillating fibers embedded within the matrix. The blocks each contain 4 towers arranged in a 2×2 matrix [35]. Tungsten powder has a small Molière radius so it is an efficient absorber in the limited space inside the solenoid. When high energy electrons and photons scatter off the tungsten powder matrix, the light is collected by the scintillating fibers and lead back to the light guide at the interface of the absorbing block. The light guide is read out by a silicon photomultiplier (SiPM) which converts the absorbed photons into an electric current. The four SiPM's per block are read out into a preamp that returns the signal via signal cables.

The inner HCAL is installed outside the magnet with an inner radius of 182 cm [111]. The hadronic calorimeters in sPHENIX are designed to measure the energy of hadrons produced in heavy ion collisions by measuring the shower from scatterings off the detector with scintillating tiles. Each section is a casing of absorber material with scintillating tiles inserted into the casing. The absorbing material is steel for the

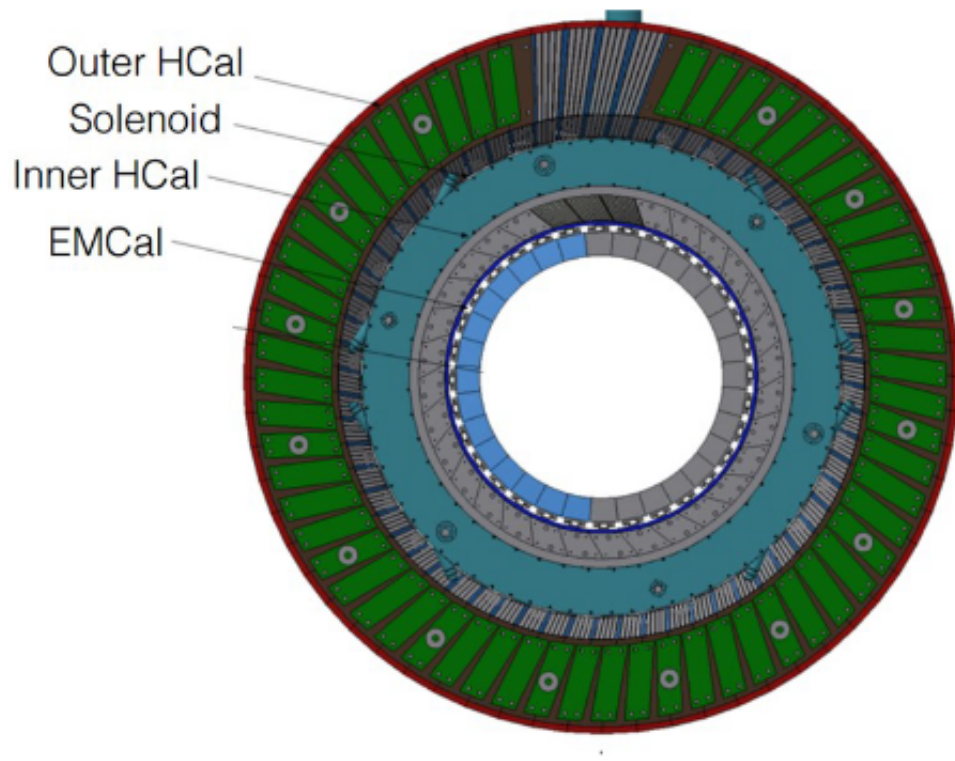


Figure 2.9: Cross section of sPHENIX calorimeters [85].

outer HCal and aluminum for the inner HCal. The reason for using aluminum for the inner HCal, despite the drop in performance, is the location of the inner HCal within the magnetic solenoid. The scintillating tiles in both the inner and outer HCal are p-doped polystyrene plastic, which is coated in an internally reflective paint. The tiles are embedded with a wavelength shifting fiber and then wrapped in an opaque foil to prevent light from escaping [111]. The HCal is divided into 64 total sections, 32 inner HCal sections, and 32 outer sections. The segmentation of the sections is approximately $\Delta\eta \times \Delta\phi = 0.1 \times 0.1$ [111]. Each section has 48 towers arranged in a 2×24 array. The towers are tilted 12° relative to a normal radial vector to be projective in pseudorapidity. Each tower comprises 5 (4) scintillating tiles in the outer (inner) HCal. When a hadron shower is incident on a tile, the tile will produce light which is measured by a silicon photo-multiplier. Each tile has a silicon photo-multiplier which couples all the tiles in a tower to a pre-amp board. The electronics of the HCal are shown in figure 2.10.

2.2.6 sPHENIX Event Characterization

The sPHENIX experiment has three global detectors to determine key collision properties, such as the reaction plane, collision vertex, and centrality. These detectors are the Minimum Bias Detector (MBD), the Zero Degree Calorimeter (ZDC), and the sPHENIX Event Plane Detector (sEPD). The MBD is the repurposed PHENIX Beam Beam Counter (BBC) and comprises two identical arrays of 64 counters positioned along the beam axis on either side of the collision point, with one set located on the North side and the other on the South side [83, 39]. Each counter features a one-inch diameter mesh-dynode photomultiplier tube coupled with a 3 cm thick quartz Cherenkov radiator mounted at the front of the photomultiplier tube. Figure 2.11 shows a picture of the sPHENIX MBD. Quartz was selected for its radiation-hard properties, essential for the BBC, which is situated near the beam-pipe in the forward regions where radiation levels are among the highest in PHENIX. Due to the inherent

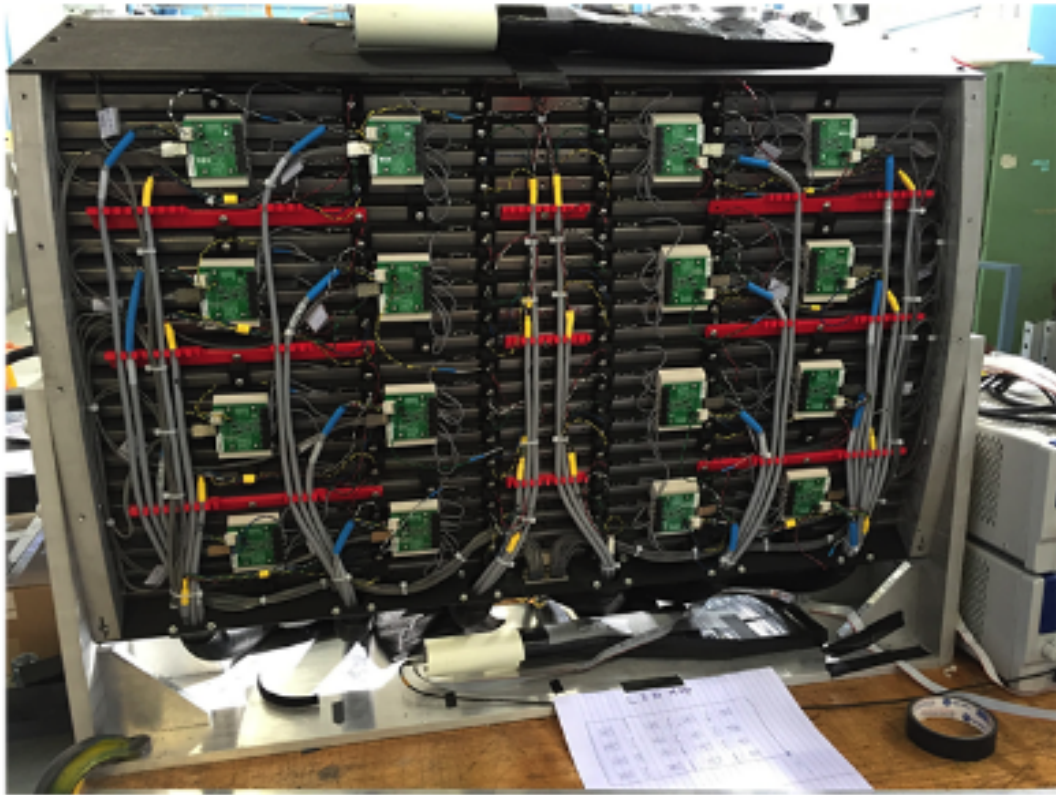


Figure 2.10: Electronics of an OHCAL sector [35].

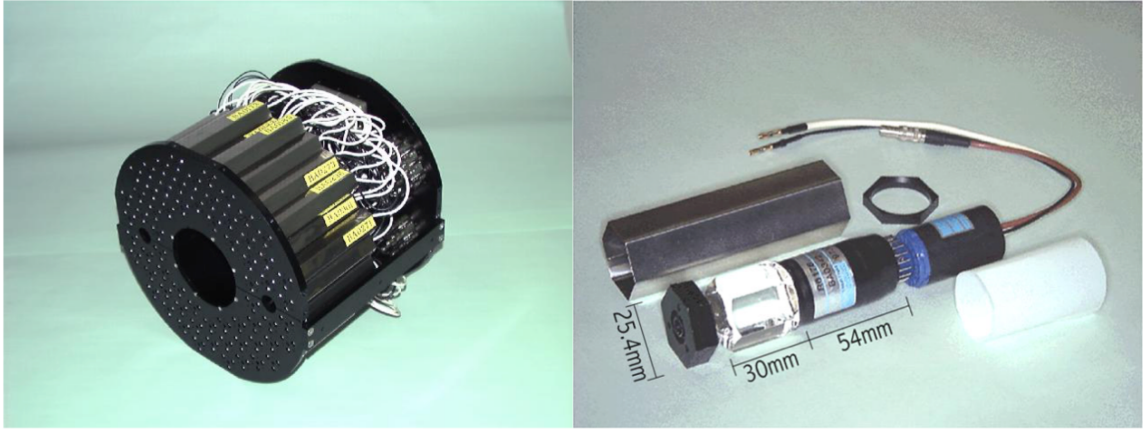


Figure 2.11: (Left) The MBD array mounted on the MBD mechanical frame. (Right) The individual MBD counter module [111].

radiation tolerance of the photomultiplier tube, the BBC system is highly resistant to radiation damage. Over 16 years of operation, no significant decline in BBC performance has been observed. The MBD provides full azimuthal coverage and measures pseudorapidity in the range of $3.0 < |\eta| < 3.9$. By measuring the timing of the first charged particles detected in each MBD, the vertex position along the beam-line can be determined. Additionally, the MBD is used to classify collision events by centrality based on the number of particles detected.

2.3 sPHENIX Acquisition and Signal Extraction

2.3.1 Calorimeter Towers Signal Extraction

During the run 2023 commissioning period, each calorimeter tower recorded 31 samples per event, with signal peaks timed to occur around the 7th sample. Raw tower signals were fit to a predefined waveform template. Towers that failed the template fit were rejected by imposing a χ^2 cut on the fit. All towers where the fraction of bad template fits for hits above threshold (> 0.5 GeV) exceeds $>1\%$ of events are masked for the entirety of a run. After signal processing noisy towers in the EMCal were systematically identified and flagged. The frequency of hits above a 15 analog digital current unit threshold was measured for each tower in a given run. Towers with hit frequencies deviating beyond a 2.5σ cutoff from the mean were flagged as noisy. These included both “hot” towers, which fired too frequently, and “cold” towers, which fired too infrequently. In some cases, anomalies were observed across entire interface boards. To mitigate this, interface boards exhibiting errant behavior were masked, ensuring that correlated issues affecting multiple towers were addressed. Typically, when an interface board was flagged, its constituent towers had already been identified as problematic. A fiducial cutoff of 50% hit frequency was also applied, automatically flagging towers and entire interface boards as bad.

2.3.2 Tracking in sPHENIX

The tracking system in sPHENIX allows for fast read out, precise momentum resolution, and secondary vertex detection. The three primary tracking detectors are the MVTX, INTT, and TPC [98]. The track reconstruction procedure at sPHENIX is implemented using the “A Common Tracking Software” library, combined with custom algorithms tailored for the sPHENIX detector. The process begins with the “A Common Tracking Software” seeding algorithm finding seeds in MVTX hits, called MVTX “triplets”. After forming the silicon track seeds, they are passed to the primary vertex-finding algorithm. The reconstruction then shifts to the TPC, where track seeds are found using a cellular automaton seeding algorithm, developed specifically for sPHENIX but inspired by the seeding algorithm used in the ALICE experiment [99].

In the TPC, a directed graph is created from clusters by assigning up to two edges to each cluster, connecting it to its spatially closest neighbors in adjacent layers to form the straightest possible triplet. This graph is then pruned to retain only mutual edges, which form chains of clusters, each representing a candidate track seed. A Kalman filter, also derived from ALICE [99], is then applied to estimate the initial track parameters, and the seeds are refined and extended by the sPHENIX-specific track propagation module. Only seeds with at least 20 clusters after propagation are retained for further reconstruction. Once TPC track seeds are generated, they are matched with the silicon seeds (MVTX + INTT) using azimuthal and η matching criteria. In cases where multiple silicon seeds match a single TPC seed, the TPC seed is duplicated, and combined full track seeds are created for each match. This step ensures that tracks are consistently linked across the different detector subsystems, forming complete track candidates. The final assembled track seeds are provided to the Kalman filter track fitting tool. This tool takes the full track seed, its estimated track parameters, and an initial vertex estimate to perform a fit on the tracks. The Kalman Filter provides refined track parameters and ensures accurate reconstruction

of the tracks. Figure [2.12](#) shows a diagram demonstrating the sPHENIX track reconstruction procedure.

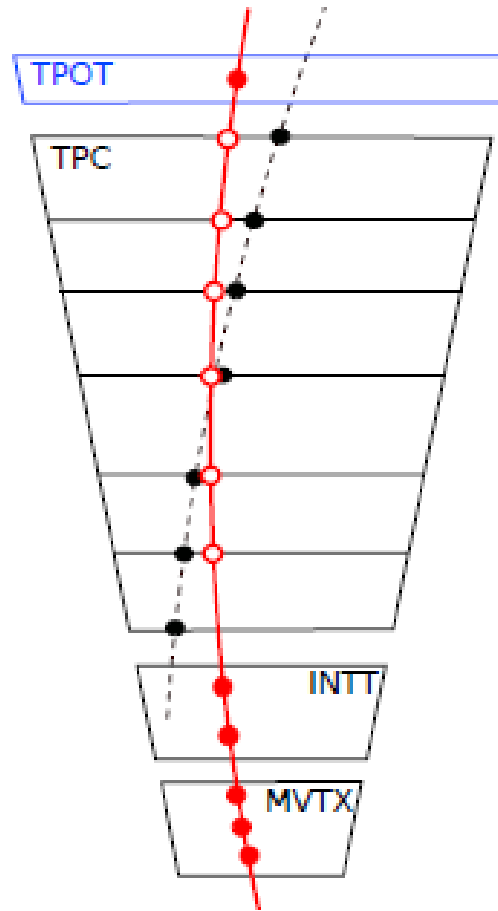


Figure 2.12: Schematic view of the sPHENIX tracking procedure using all available tracking detectors [43].

Chapter 3

Jet Background Studies

3.1 Multiplicity Based Background Subtraction

The multiplicity method was developed for use in this thesis [95]. The form of the multiplicity method can be obtained through a similar argument made to obtain the area-based background subtraction method [51] by considering a population of particles in a given event contains N_b particles originating from the underlying event and N_s particles originating from hard-scattering processes. The average number density of underlying event particles in the population is $\epsilon \equiv \langle \frac{N_b}{N} \rangle$ where N is the total number of particles in the event. The probability of the population containing N_b underlying event particles is

$$P(N_b) = \frac{(\epsilon N)^{N_b}}{N_b!} e^{-\epsilon N}, \quad (3.1)$$

where $\epsilon N_{\text{total}}$ is the expected number of underlying event particles in the event. The distribution for the momenta of the underlying event to the total momentum of the population will be a convolution of N_b single particle p_T gamma distributions

$$\frac{1}{P(N_b)} \frac{dP(N_b)}{dp_T} = \frac{1}{\langle p_T \rangle \Gamma(N_b)} \left(\frac{p_T}{\langle p_T \rangle} \right)^{N_b-1} e^{-p_T/\langle p_T \rangle}, \quad (3.2)$$

where $\langle p_T \rangle$ is the average momentum of a underlying event particle and $\Gamma(N_b) = (N_b - 1)!$ is the real valued gamma function. Integrating equation 3.2 over all multiplicities results in the probability distribution of the average underlying event momentum in the population

$$\begin{aligned} \frac{dP(N)}{dp_T} &= \sum_{N_b=1}^{\infty} e^{-\epsilon N - p_T / \langle p_T \rangle} \frac{(\epsilon N)^{N_b}}{N_b!} \frac{p_T^{N_b-1} \langle p_T \rangle^{-N_b}}{\Gamma(N_b)} \\ &= e^{-\epsilon N - p_T / \langle p_T \rangle} \sqrt{\frac{\epsilon N}{p_T \langle p_T \rangle}} \mathcal{I}_1 \left(\sqrt{\frac{4\epsilon N p_T}{\langle p_T \rangle}} \right), \end{aligned} \quad (3.3)$$

where $\mathcal{I}_1$ is the first order modified Bessel function of the first kind.

The probability distribution for the background momentum of a given population of particles is shown in figure 3.1. As ϵN becomes much greater than 1, the probability tends towards a Gaussian with $\mu = \epsilon N \langle p_T \rangle$ and standard deviation $\sigma = \sqrt{2\epsilon N} \langle p_T \rangle$. Since ϵ is the population density of background particles, it is equivalent to write $\epsilon = 1 - \epsilon_s$ where ϵ_s is the population density of signal particles. Using ϵ_s , the expected value of ϵN becomes $(1 - \epsilon_s)N \approx N - \langle N_{\text{signal}} \rangle$. This redefinition allows for easier application in experiments since the average number of signal particles $\langle N_{\text{signal}} \rangle$ can be estimated with accuracy, as models for jets in proton-proton collisions [15] can describe it within about 25%. From the mean and standard deviation of the distribution in eq. 3.3 we can write the average background momentum contribution $p_{T,\text{Bkdg.}}^N$ for a given population of particles as

$$p_{T,\text{Bkdg.}}^N = \langle p_T \rangle (N - \langle N_{\text{signal}} \rangle) \pm \sigma_{\delta N}, \quad (3.4)$$

where $\sigma_{\delta N} = \sqrt{2(N - \langle N_{\text{signal}} \rangle) \langle p_T \rangle}$ is the standard deviation of eq. 3.3. Using the characterization of the underlying event given by eq. 3.4, the multiplicity-based background subtraction method can be written as

$$p_{T,\text{Jet}}^{\text{Corr., N}} = p_{T,\text{Jet}}^{\text{Uncorr.}} - \rho_{\text{Mult}}(N_{\text{tot}} - \langle N_{\text{signal}} \rangle), \quad (3.5)$$

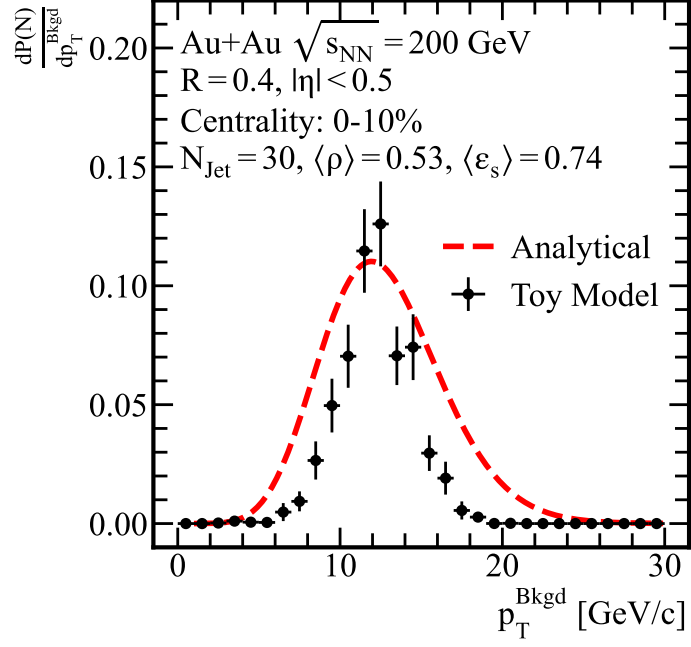


Figure 3.1: The background momentum in a group of N particles as predicted by Equation 3.3 compared to the background momentum in toy model jets with 30 constituents.

where N_{tot} is the observed number of particles within the jet, $\langle N_{\text{signal}} \rangle$ is the average number of particles in the signal, and N_b is estimated as $N_b = N_{\text{tot}} - \langle N_{\text{signal}} \rangle$. Additionally, ρ_{Mult} within an event represents the mean transverse momentum per background particle $\langle p_T \rangle$. Similar to the area method ρ_{Mult} is approximated as the median of $p_{T,\text{Jet}}^{\text{Total}}/N_{\text{tot}}$ for k_T jets. This approach capitalizes on the fact that the natural variable in the standard deviation is the number of background particles, eliminating the second and third terms in Equation 1.10. The standard deviations of the momentum residual for the multiplicity method are therefore approximately

$$\sigma_{\delta p_T} = \sqrt{N} \sigma_{p_T}. \quad (3.6)$$

There will still be some influence from deviations in the shape of the single particle spectrum, a more accurate accounting of flow, and the added uncertainty due to estimates of N_{signal} , but contributions from number fluctuations are eliminated.

While this method introduces an additional systematic uncertainty because $\langle N_{\text{signal}} \rangle$ must be estimated with a model, this may still decrease the overall uncertainties because of the reduction in background fluctuations. Since the dependence of $\langle N_{\text{signal}} \rangle$ on $p_{T,\text{Jet}}^{\text{Truth}}$ and $p_{T,\text{Jet}}^{\text{Uncorr.}}$ is similar and PYTHIA agrees well with the data, a reasonable uncertainty on $\langle N_{\text{signal}} \rangle$ for unmodified jets is projected to be less than one for $R \leq 0.4$ and around two for $R = 0.6$.

The additional uncertainty in jet energy introduced by the multiplicity method is proportional to $\sigma_{N_{\text{signal}}} \cdot \rho_{\text{Mult}}$, where $\sigma_{N_{\text{signal}}}$ is the uncertainty in the count of signal particles in the jet. Given that $\rho_{\text{Mult}} \approx 0.5 \text{ GeV}/c$, this uncertainty would approximately amount to 1.5 GeV for medium modifications and around 0.5–1.0 GeV for uncertainties in the count of signal particles in unmodified jets. These values are small compared to typical experimental jet momentum resolutions of 10-20%.

The procedure we use to estimate the systematic uncertainties from the multiplicity method was introduced in previous studies into fragmentation biases of the machine learning based jet background subtraction method [18]. We use a toy-model consisting of jets generated with PYTHIA 8.307 [107] combined with

a heavy ion background approximated by the TennGen [79, 94] model. TennGen simulates a heavy ion environment by randomly generating $\pi^\pm$, $K^\pm$, p and $\bar{p}$ hadrons with momenta and azimuthal positions sampled from fits to measured momentum spectra [25, 13] and azimuthal anisotropies [28, 22]. By construction, TennGen does not include correlations beyond flow effects. This choice is motivated by the observation that jet background fluctuations are largely consistent with expectations from randomly sampled single-particle momentum distributions [12, 80, 24]. Proton-proton collisions at $\sqrt{s} = 200$ GeV and 2.76 TeV were simulated using the Monash 2013 tune [108], with 25 p_T^{hard} bins starting at 10 GeV, each containing 1 million $p+p$ events. Only final-state charged particles from PYTHIA were merged with a TennGen background event. A minimum particle momentum cut $p_T > 150$ MeV/ c and a pseudorapidity acceptance $|\eta| < 0.9$. cut are applied to both PYTHIA and TennGen particles.

Jet reconstruction is performed on the combined event using the anti- k_T algorithm with the standard recombination scheme, as implemented in FastJet 3.4.0 [53]. The jet resolution parameters, R , are set to 0.2, 0.4, and 0.6. Before merging, PYTHIA events are clustered separately for all jet resolution parameters, providing the truth jet momentum. After merging, the combined event is re-clustered, and jets from the PYTHIA +TennGen event are geometrically matched to PYTHIA jets if $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2} < 0.75R$, where $\Delta\eta$ and $\Delta\phi$ are the differences in pseudorapidity and azimuthal angle, respectively. The matched PYTHIA jet’s momentum is then assigned as the true momentum of the combined jet, denoted as $p_{T,\text{Jet}}^{\text{Truth}} \equiv p_{T,\text{Jet}}^{\text{PYTHIA}}$. Kinematic cuts of $p_{T,\text{Jet}}^{\text{Uncorr.}} > 10$ GeV and $|\eta_{\text{jet}}| < 0.9 - R$ are applied to all jet candidates within the combined event.

We explore the fragmentation dependence of the multiplicity method by artificially modifying the fragmentation of jets generated with PYTHIA 8.3. The modifications to PYTHIA jet fragmentation can be separated into two classes; parent parton biasing, and artificial radiating of constituents. For clarity, the unmodified jet sample is referred to as “nominal”. We estimate $\langle N_{\text{Signal}} \rangle$ from the nominal jet sample and use

the modified jets to quantify the variation of $\langle N_{\text{Signal}} \rangle$ for jets that undergo medium modification.

Quark- and gluon-initiated jets exhibit different fragmentation characteristics, leading to variations in the multiplicity distribution of jet constituents [36, 6]. Quark jets tend to be more collimated, meaning they have fewer constituents, with a few carrying a significant fraction of the total jet momentum, defined as $z \equiv \frac{p_{T,\text{Track}}}{p_{T,\text{Jet}}} \cdot \cos(\Delta R)$. In contrast, gluon jets are broader and contain a higher number of constituents, resulting in a more evenly distributed fragmentation function. Since real-world jet samples consist of a mixture of both quark- and gluon-initiated jets, studying pure parton-initiated jet samples provides an upper bound on parton-dependent fragmentation biases in the multiplicity method. This approach offers a conservative estimate of the potential impact of fragmentation differences.

Artificial radiating of jet constituents is used to mimic the effects of medium modification on jet fragmentation. For artificially modified toy-model samples, anti- k_T jets with resolution parameters $R = 0.2, 0.4$, and 0.6 were found from PYTHIA final state charged particles prior to merging with the TennGen background. Jet candidates that satisfied $p_{T,\text{Jet}}^{\text{PYTHIA}} > 5 \text{ GeV}$ and $|\eta_{\text{Jet}}| < 0.9 - R$ were selected for the modification procedure. Constituents in these jets were randomly selected with probability $P_{\text{Rad.}}$. If a constituent was selected, a splitting angle from the jet axis was generated using a uniform distribution from 0 to ΔR_{Max} . The new “radiated” particle carries momentum $p_{T,\text{Track}}^{\text{Rad}} = f_{\text{Rad.}} \cdot p_{T,\text{Track}}^{\text{Original}}$. The radiated momentum is subtracted from the original constituent’s momentum (except in the Medium Response scenario) and both particles are added back to the population of final state charged particles of the PYTHIA event. The PYTHIA event is then re-clustered and merged with the TennGen background. We have implemented three variations of artificial fragmentation modification to describe different aspects of medium modification [18]. Details and motivations for the values of $P_{\text{Rad.}}$, $f_{\text{Rad.}}$, and ΔR_{max} are provided in [18]. The three types of artificial modifications we used are Fractional Collinear, Fractional Large Angle, and Medium Response. The

parameters governing the radiation procedure for each of these cases are presented in Table 3.1. In the Fractional Collinear case, the radiated particle falls primarily within the jet cone and in the Fractional Large Angle case, the radiated particle falls primarily outside the jet cone. In both Fractional Collinear and Fractional Large Angle the radiated constituent momentum is subtracted from the original constituent's momentum $(p_{\text{Original}}^\mu)' = p_{\text{Original}}^\mu - p_{\text{Rad.}}^\mu$. The Medium Response case has the same ΔR_{Max} as the Fractional Collinear case, however, the radiated momentum is not subtracted from the original constituent's momentum. This simulates the addition of low momentum particles from the medium into the jet cone.

The fragmentation modifications in this study are consistent with observations of jets in $A+A$ collisions. Measurements of jet fragmentation functions with $p_{T,\text{Jet}} > 120$ GeV/ c in Pb+Pb collisions [55, 1] suggest the presence of at most one additional particle. The ratio of jet fragmentation functions measured in Pb+Pb to $p+p$ collisions $R_{D(p_T)}$ reported as a function of constituent track p_T shows that additional particles largely appear in the low momentum regime. Figure 3.2 shows comparisons of the modified jet fragmentation observed in our toy-model to $R_{D(p_T)}$ in Pb+Pb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV measured by ATLAS [1]. The modification observed in data is largely consistent with that seen in our Pb+Pb $\sqrt{s_{\text{NN}}} = 2.76$ TeV toy-model.

Low momentum measurements of gamma-hadron correlations at RHIC [19], covering a much lower kinematic range, likewise indicate that on average there are fewer than one additional particle due to medium modifications. The ratio of jet fragmentation functions measured in Au+Au to $p+p$ collisions $R_{D(\xi)}$ reported as a function of $\xi = -\ln(z)$ shows that additional particles largely appear in the soft fragmentation regime. Figure 3.3 shows comparisons of the modified jet fragmentation observed in our toy-model to $R_{D(\xi)}$ in Au+Au collisions at $\sqrt{s_{\text{NN}}} = 200$ GeV measured by PHENIX [19]. Again we see the modification observed in data is largely consistent with that seen in our model.

Table 3.1: Parameters controlling artificial radiation for each of the three fragmentation cases. These parameters were chosen to reflect observations of jets and charged hadron yields in Pb+Pb collisions [18].

Modification Class	$\mathbf{P}_{\text{Rad.}}$	$\mathbf{f}_{\text{Rad.}}$	$\Delta\mathbf{R}_{\text{Max}}(R = 0.2, 0.4, 0.6)$
Fractional Collinear	1.0	0.25	0.1, 0.2, 0.4
Fractional Large Angle	1.0	0.25	0.4, 0.6, 0.8
Medium Response	0.5	0.1	0.1, 0.2, 0.4

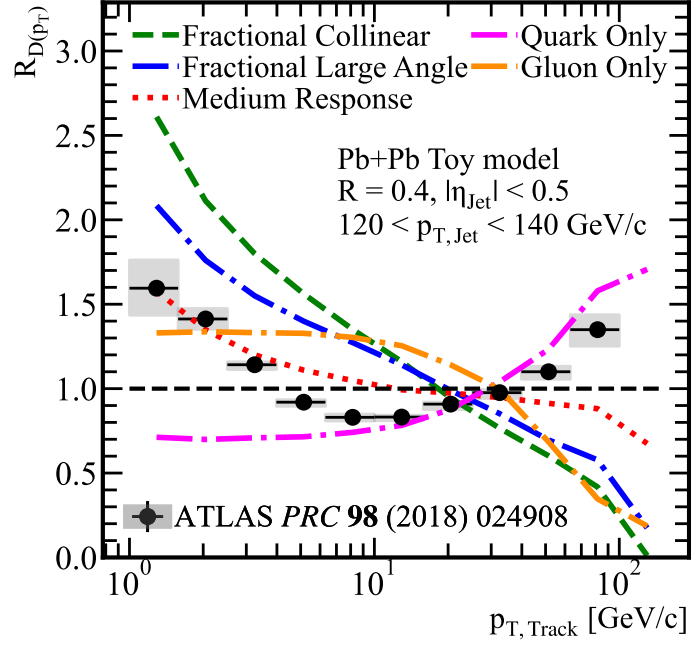


Figure 3.2: Comparisons of ratio of jet fragmentation functions measured in Pb+Pb to $p+p$ collisions $R_{D(p_T)}$ [1] to all artificial fragmentation modification cases for $R = 0.4$ jets at LHC energies.

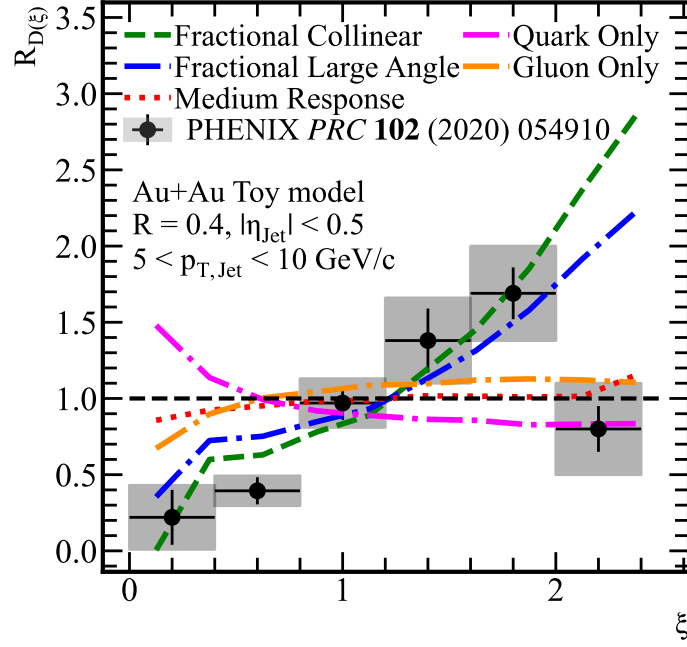


Figure 3.3: Comparisons of ratio of jet fragmentation functions measured in Pb+Pb to $p+p$ collisions $R_{D(\xi)}$ [19] to all artificial fragmentation modification cases for $R = 0.4$ jets at RHIC energies.

In both comparisons between data and our fragmentation modification, we find good agreement in intermediate ranges while the most aggressive fragmentation modification procedures often over-estimate the modification towards low track p_T and soft fragmentation high ξ , indicating that our estimate for the uncertainty due to medium modification to the jet constituent multiplicity is conservative.

3.2 Machine Learning Studies

In this section, we present an overview of machine learning studies performed during the development of the multiplicity method used in this thesis. These methods include a deep neural network inspired by [70], a pruned shallow neural network that highlights the dimensionality of the jet background subtraction problem, and a deep symbolic regression approach that enables direct interpretation of the mappings learned by each of the neural networks [66, 11]. In all cases, the TennGen [80, 79, 94] model was used to approximate the background seen in heavy ion collisions, while PYTHIA jets simulated in p+p collisions served as the signal.

Neural networks are a type of machine learning model that have demonstrated remarkable performance in diverse applications, including image recognition, natural language understanding, and predictive regression tasks. The deep neural network employed in this study consists of multiple hidden layers, where each neuron is linked to those in the subsequent layer through adjustable weights. The transformation of information from one layer to the next follows the general form

$$\psi_i = \mathcal{A}(\mathbf{W}_i \cdot \psi_{i-1} + b_i), \quad (3.7)$$

where ψ_i represents an n -dimensional transformation of the preceding layer's output ψ_{i-1} , $\mathbf{W}_i$ is an $n \times m$ matrix of trainable weights applied to the m -dimensional input, b_i is the bias term, and $\mathcal{A}$ is the activation function. The activation function controls

whether a neuron transmits its signal forward and is used to introduce non-linearity which allows the neural network to describe intricate data structures.

Following the approach presented in Ref. [70], we train a neural network to predict the corrected jet momentum using a set of input jet features: the uncorrected jet momentum, jet area, jet angularity, the number of jet constituents, and the transverse momenta of the eight leading constituents. We refer to this as the deep neural network model which consists of three hidden layers with 100, 100, and 50 nodes, respectively. The loss function is defined as the mean squared error between the predicted jet momentum and the PYTHIA truth jet momentum, $p_{T,\text{jet}}^{\text{Truth}}$. We investigate the validity of the neural network architecture and the input jet features presented in Ref. [70] through network pruning and feature engineering.

To determine the most effective jet input features for the truth jet regression task, we apply input feature engineering. This process helps identify the most relevant and informative input features, allowing for the development of more precise models and potentially enhancing the overall performance of the regression task. A composite feature score is derived from multiple metrics, including the mutual information between each input feature and the truth momentum, the variance threshold of the feature, and its Pearson correlation with other inputs. A description of each of these feature scoring techniques is presented in the following paragraphs.

Shared mutual information identifies input features that are statistically dependent on the truth jet p_T , helping to select those with high predictive significance. Mutual information between two random variables X and Y is defined as the distance between the joint entropy distribution $H(x, y)$ and the sum of the individual entropy $H(x)$ and $H(y)$,

$$I(x; y) \equiv H(x) + H(y) - H(x, y), \quad (3.8)$$

where the $H(n)$ is the Shannon entropy, $H(x) \equiv -\sum_i p(x_i) \log(p(x_i))$, for variable n . Mutual information is applied to feature selection as a way to quantify how relevant an input feature is for predicting the target variable [119].

The variance threshold method evaluates the variability of features within the dataset by setting a minimum variance criterion. Features with variance, $\sigma^2 = \sum_i \frac{(x_i - \mu)^2}{N}$, below this threshold are removed from consideration. This approach is particularly useful in high-dimensional datasets, as it filters out low-variance features that contribute little meaningful information or exhibit minimal variation [69]. By reducing the number of input features, the model’s complexity decreases, which can enhance both training efficiency and inference performance.

The Pearson correlation coefficient measures the linear relationship between two variables. The Pearson correlation coefficient between two variables is defined as

$$\rho_{x,y} \equiv \frac{\mathbb{E}[(x - \mu_x)(y - \mu_y)]}{\sigma_x \sigma_y}, \quad (3.9)$$

where $\mathbb{E}$ is the expectation value, μ_n is the mean of variable n and σ_n is the standard deviation of variable n . By calculating the correlation between each feature and the $p_{T,jet}^{Truth}$, we can identify features that exhibit strong linear associations with the target variable [100, 69]. High absolute correlation values indicate features that are more likely to contribute to the regression model’s predictive power. Pearson correlations provide insights into the strength and directionality of the relationships, helping to select features that have significant predictive capabilities.

All three feature selection techniques are applied to a total of 29 available input jet features, with a regression target of the truth jet momentum $p_{T,jet}^{Truth}$. By integrating these different selection methods the resulting composite score serves as a reliable metric to identify the input jet features that consistently rank highly across multiple criteria, signaling their strong relevance to the regression task [129]. This score is used to apply constraints on the training phase space limiting training to the most informative input features [86]. The results of this feature optimization are shown in figure 3.4.

To evaluate the necessity of the connections formed by the deep neural network, we apply network pruning to reduce complexity and remove redundant connections [88].

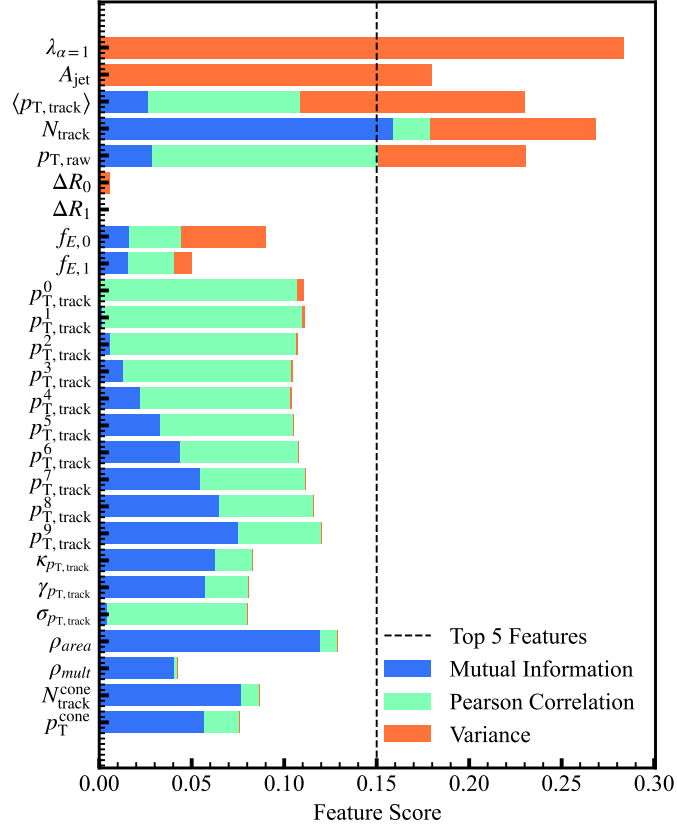


Figure 3.4: Total feature scores for available jet input features relating to $p_{T,\text{jet}}^{\text{Truth}}$. Each score is the sum of the variance, mutual information, and Pearson Correlation Coefficient [69].

Our pruning approach begins by training the network and then analyzing the probability of each node firing across the entire test dataset. Nodes that do not activate beyond a specified threshold are eliminated. In this study, we set the threshold to zero, meaning only completely inactive nodes were removed. After pruning, the network is retrained, and the resulting loss is compared to the original, un-pruned network. If a layer retains fewer than three active nodes, those nodes are merged into the preceding layer, and the now-empty layer is removed. This iterative process continues until the loss exceeds 5% of the original value. A schematic of the fully trained deep neural network is shown in figure 3.5. Prior to pruning the network fails to use all connections available to it. While this network offers improved jet momentum resolution, its over-complex architecture lacks interpretability, hindering our ability to extract meaningful insights from the learned representations [66].

To further streamline the network pruning procedure, we incorporate kernel regularization, which discourages unnecessary weight parameters. This technique introduces a penalty term into the loss function,

$$\mathcal{L}' = \mathcal{L}_o + \lambda \sum_{l=1}^L \|\mathbf{W}_l\|^2, \quad (3.10)$$

where $\mathcal{L}_o$ is the original loss function, $\mathbf{W}_l$ represents the weight matrix of layer l , and λ is the regularization parameter. The sum extends over all layers, applying an $L2$ norm penalty to the weights. This method encourages the network to emphasize only the most meaningful connections, leading to a more compact and interpretable architecture [117]. We refer to the resulting neural network architecture after the application of both network pruning and kernel regularization as the shallow neural network. Figure 3.6 illustrates the network structure at both the initial and final iterations of the pruning process with $\lambda = 0.01$. After pruning, the resulting shallow neural network retains only 29 learned parameters, a drastic reduction from the 16,501 parameters available in the original deep neural network.

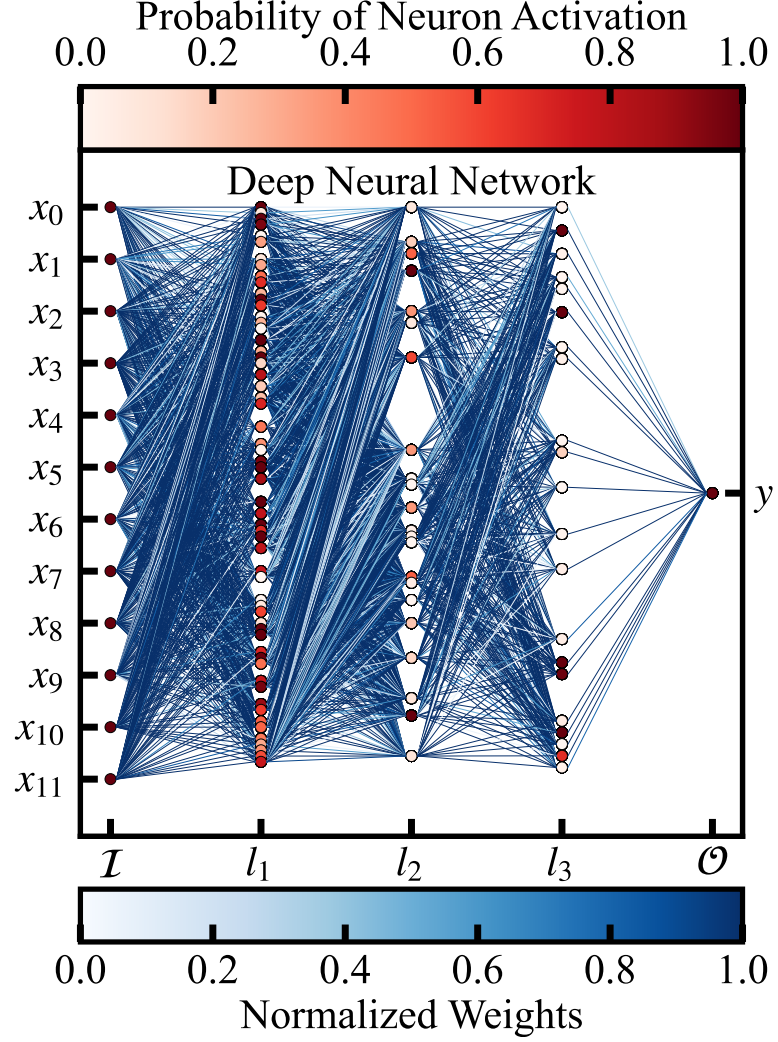


Figure 3.5: Each node and weight for every layer l in the deep neural network architecture. The nodes are colored by their probability of firing for the entire testing jet sample. The weights are normalized to the maximum weight from a given node to the next layer. The input features are listed by x_i where i ranges from 0 to 11 corresponding to the 12 input features for the deep neural network (uncorrected jet momentum, jet area, jet angularity, number of jet constituents, and eight leading constituent momenta). The truth jet momentum is labeled by y .

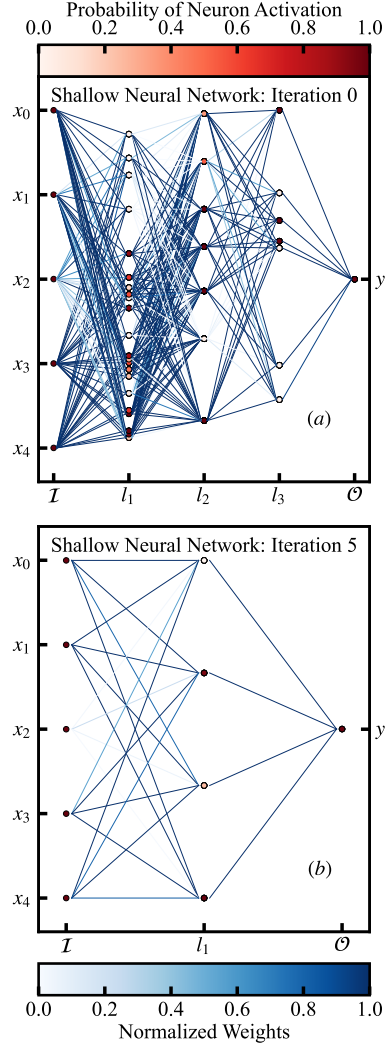


Figure 3.6: Each node and weight for every layer l in for the first (a) and last (b) iteration of the shallow neural network produced through feature optimization and network pruning. The nodes are colored by their probability of firing for the entire testing jet sample. The weights are normalized to the maximum weight from a given node to the next layer. The input features are listed by x_i where i ranges from 0 to 4 corresponding to uncorrected jet momentum, leading constituent momenta, number of jet constituents, jet area, and jet angularity. The truth jet momentum is labeled by y .

Figure 3.7 presents the dependence of the jet momentum residual distribution width on jet momentum for background subtraction techniques including the deep neural network, the shallow neural network, the multiplicity method presented in section 3.1, and a standard area method in both Au+Au collisions at $\sqrt{s_{\text{NN}}} = 200$ GeV and Pb+Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV. As the jet resolution parameter increases, the width of underlying event fluctuations also grows, indicating a larger background contribution for larger jets. Additionally, $\sigma_{\delta p_T}$ increases with $\sqrt{s_{\text{NN}}}$, reflecting the higher particle multiplicity at elevated collision energies.

These findings align with previous studies [70, 95], demonstrating that the deep neural network significantly improves momentum reconstruction resolution compared to the area-based method. The shallow neural network captures some of this improvement but does not fully match the deep network’s performance. In Au+Au collisions with small jet resolution parameters, the multiplicity method achieves a resolution comparable to that of the deep neural network, while in Pb+Pb collisions, it surpasses the deep network’s performance at high jet momentum. As the jet resolution parameter increases, the jet momentum resolution of the multiplicity method and the shallow neural network become increasingly similar.

Enhancing jet momentum resolution is expected to expand the kinematic reach of the measurement to lower jet momenta. This is evaluated by unfolding the uncorrected jet momentum spectra using five iterations of the Bayesian unfolding method [61] implemented in RooUnfold 2.0.0 [49]. Unfolding is used to correct for smearing effects caused by finite resolution in the uncorrected jet momentum spectra. We construct a response matrix by matching PYTHIA jets (truth jets) to PYTHIA +TennGen jets (uncorrected jets). The uncorrected spectra have no matching criteria between PYTHIA +TennGen jets and PYTHIA jets. After five iterations, we observe that the change in χ^2 between the unfolded and truth spectra falls below the uncertainties of the unfolded spectra, indicating that additional iterations would not meaningfully alter the result within uncertainties. The lower unfolding threshold is typically set at five times the width of the jet momentum residual, $\sigma_{\delta p_T}$ [14, 18]. The

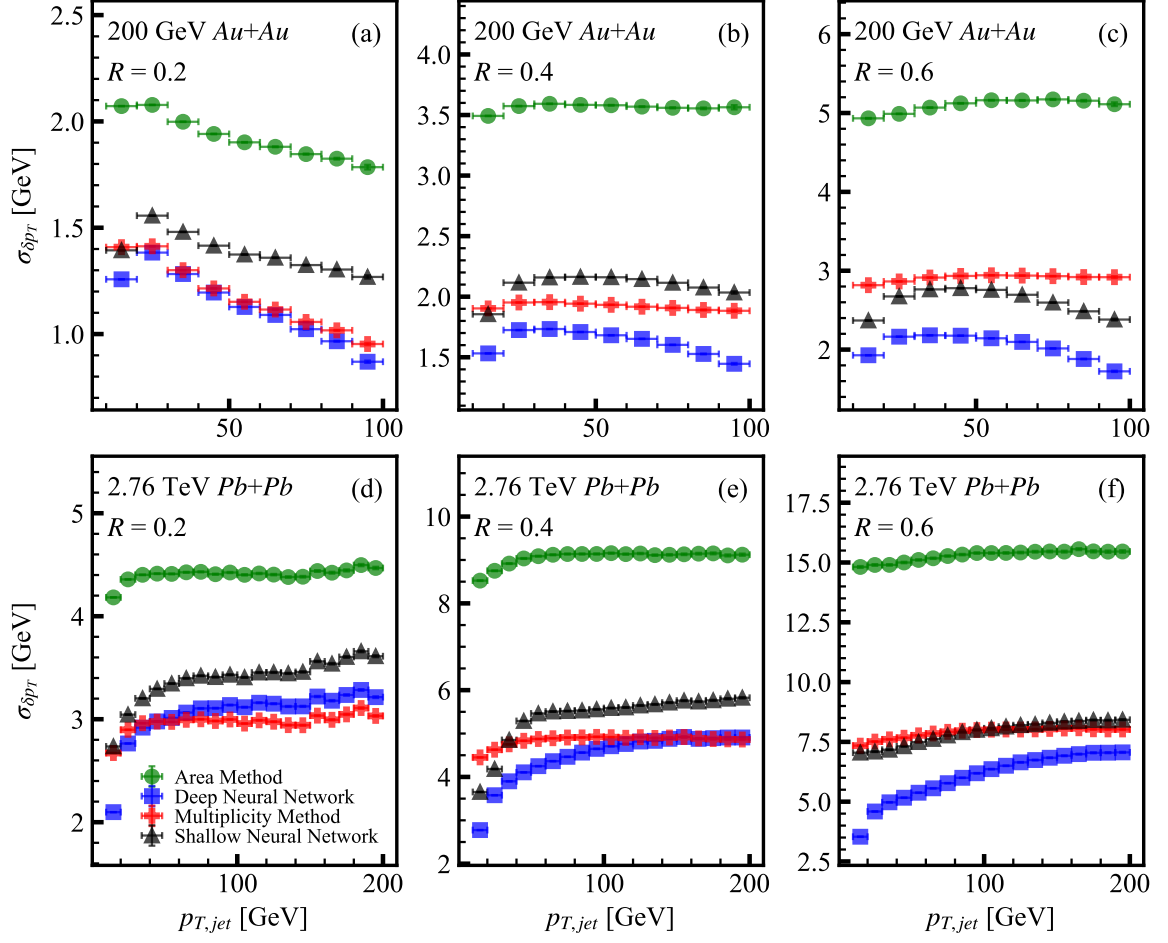


Figure 3.7: Comparisons of jet p_T residual width for each background subtraction method as a function of the reconstructed jet momentum for Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV and Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV for jet resolution parameters $R = 0.2, 0.4$, and 0.6 . Note the y -axes are zero suppressed for all panels.

ability of each method to suppress combinatorial jets at low momentum determines the kinematic limit of the unfolded jet spectra.

The ratios of unfolded to true jet spectra are presented in Figure 3.8. We find that the multiplicity method extends the kinematic range equivalently to the deep neural network method. Deviations from unity at lower jet momenta indicate the region where the method becomes unstable due to significant contributions from the combinatorial background. Since reconstructed jet spectra do not include kinematic cuts to suppress combinatorial jet contributions, any extension in the lower kinematic range is solely due to the momentum resolution of the background subtraction method. For all jet resolution parameters and collision systems, both the multiplicity and deep neural network methods remain stable at least 10 GeV/ c lower in momentum than the area method.

3.2.1 Machine Learning Interpretability

Applying machine learning techniques to background subtraction requires careful consideration, as existing background models struggle to accurately capture fluctuations in heavy-ion collisions [80]. The usefulness of non-interpretable machine learning methods is particularly limited in cases where training models may be imprecise, where understanding the method itself is essential for interpreting results, or when predictions are needed beyond the scope of the training dataset. Additionally, while non-interpretable machine learning approaches can outperform traditional background subtraction techniques, they often provide little insight into the physical relationships that drive these improvements.

Neural networks with sufficient complexity, such as those discussed in section 3.2, can approximate any function. These neural networks are also prone to bias, as their predictions become unreliable outside the training domain, and their performance may degrade when extrapolated beyond this range. Due to their opacity, neural networks offer limited transparency regarding where and why their predictions may

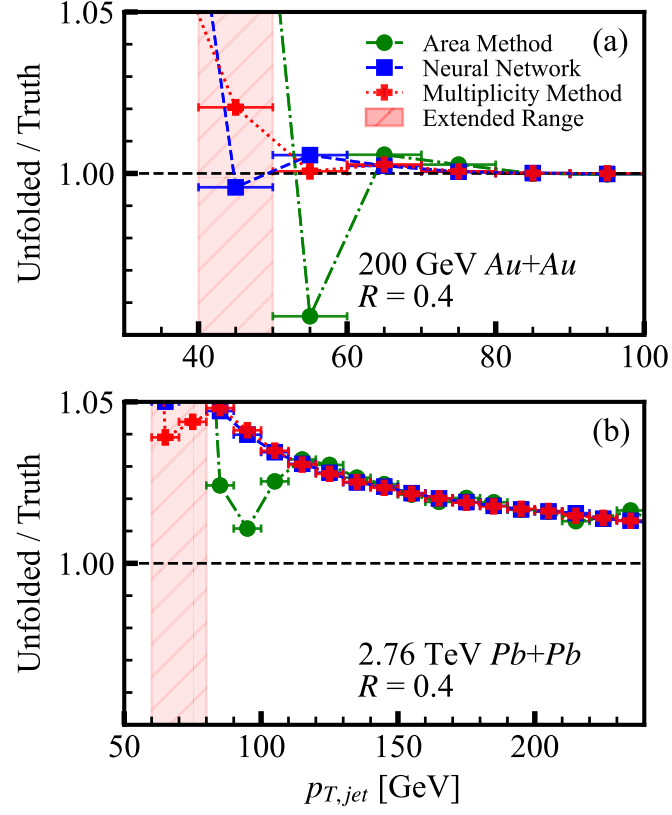


Figure 3.8: Ratio of unfolded jet spectrum over truth spectrum for (a) Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV and (b) Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV for jet resolution parameter $R = 0.4$.

fail. Simpler network architectures, such as the shallow neural network can provide a clearer understanding of training biases and applicability limits.

This lack of transparency in current machine learning techniques presents a challenge in physics applications, where understanding model predictions and potential biases is essential. To address this issue, we employ symbolic regression to extract mathematical expressions from the neural networks described in section 3.2 using the PySR 0.11.11 [59] package. These equations serve as effective representations of the network’s learned mapping between inputs and outputs. By restricting the available mathematical operations, we can enforce constraints on complexity and smoothness. A schematic representation of the process used to extract functional expressions from trained deep neural networks is shown in Figure 3.9.

The PySR model samples the space of analytic expressions by applying genetic programming to minimize an objective function, using a predefined set of operators, input features, and constants. The input features remain the same as those used in the deep neural network, while the available operations include arithmetic, exponential, trigonometric, and exponentiation functions. The model evolves over 50 generations, with each generation consisting of 20 separate population samples, each containing 33 individuals. The loss function for the PySR model is the mean squared error between the prediction from PySR momentum prediction $p_{T,jet}^{PySR}$ and the corrected jet momentum predicted by the deep neural network $p_{T,jet}^{DNN}$. PySR evaluates expressions based on a score S defined as

$$S = -\frac{d \ln \mathcal{L}}{dC}, \quad (3.11)$$

where the equation complexity C is defined as the total number of operations, variables, and constants used in an equation [60]. This score rewards minimizing the loss function $\mathcal{L}$ while penalizing equation complexity C . The highest scoring PySR expression is a functional representation of the mapping from input jet features to corrected jet momentum learned by the deep neural network.

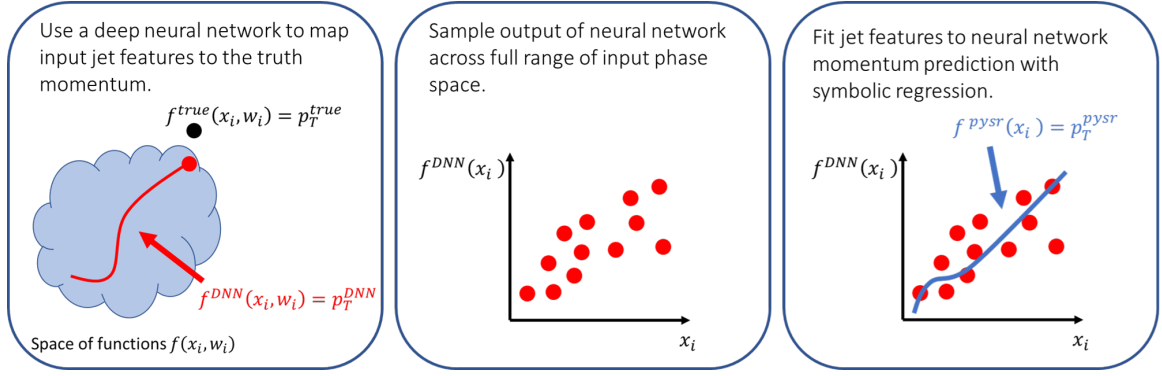


Figure 3.9: Diagram showing the procedure of mapping a relationship between input jet features to the corrected jet momentum using a neural network and then extracting an analytical expression that describes that mapping with symbolic regression. The left panel shows fitting the space of possible functions that map input to output with the neural network. The middle panel shows sampling of this mapping across input jets. The right panel shows fitting this sample to an analytical expression using symbolic regression.

To illustrate the output of PySR, figure 3.10 displays the eight highest-scoring expressions derived from symbolic mapping of a deep neural network trained on $R = 0.4$ jets in $\sqrt{s_{\text{NN}}} = 200$ GeV Au+Au collisions. This mapping process was also performed for deep and shallow neural networks trained on jet samples from both $\sqrt{s_{\text{NN}}} = 200$ GeV Au+Au collisions and $\sqrt{s_{\text{NN}}} = 2.76$ TeV Pb+Pb collisions.

We find that the functional representation of the relationship between input jet features and truth jet momentum, as learned by both the deep and shallow neural networks, closely aligns with the multiplicity method. For all jet resolution parameters and both collision energies, symbolic regression identified the optimal functional form describing the deep neural network as

$$p_{T,jet}^{Corr.PySR} = p_{T,jet}^{Uncorr.} - C_1 \cdot (N_{tracks} - C_2), \quad (3.12)$$

where the two parameters, C_1 and C_2 , are optimization constants defined by PySR. The result of the symbolic regression implies that each of the neural networks relies heavily on input features such as the total number of particles in the jet N_{tot} and the uncorrected jet momentum $p_{T,Jet}^{tot}$, both variables which are native to the multiplicity method.

The similarity between the predicted PySR expression and the multiplicity method is reinforced by investigating these learned parameters. Figure 3.11 presents these parameters alongside the average values used in the multiplicity method for comparison. We find that the symbolic regression parameters, C_1 and C_2 , extracted from both the shallow and deep neural networks, closely correspond to the multiplicity method's $\langle \rho_{Mult} \rangle$ and $\langle N_{signal} \rangle$, respectively. However, larger deviations are observed at LHC energies and for increased jet resolution parameters, R . The parameters obtained from the shallow neural network show even greater discrepancies from those in the multiplicity method at smaller jet resolution parameters and in Pb+Pb collision systems. Notably, neither the shallow nor deep neural networks were explicitly provided with the average background particle momentum, ρ_{Mult} , as an input. Despite

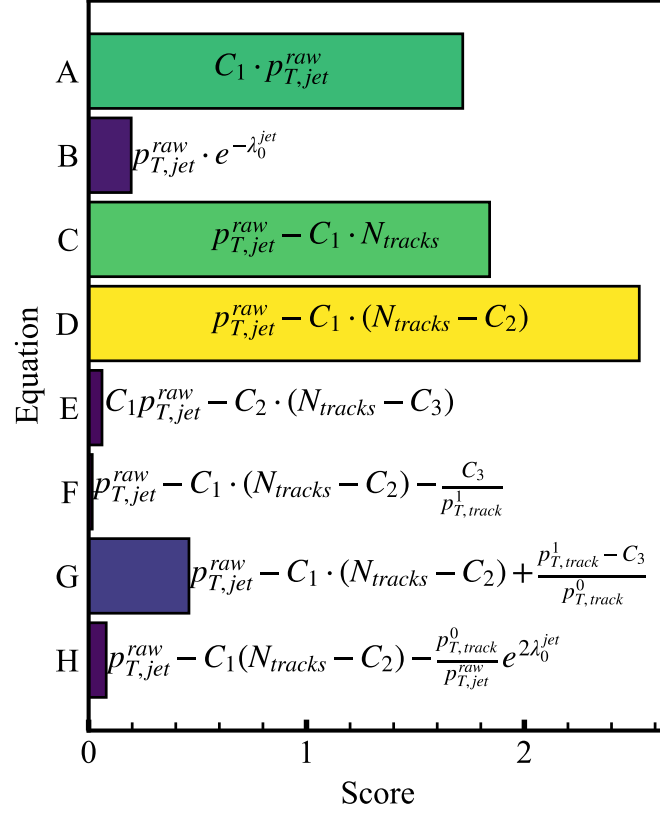


Figure 3.10: The eight highest scoring expressions from symbolic mapping of a deep neural network trained on $R = 0.4$ jets from $\sqrt{s_{NN}} = 200$ GeV Au+Au collisions. The score is defined as the negative derivative of the loss function with respect to equation complexity (equation 3.11).

this, the networks were able to infer the median per-particle momentum in the event, as well as the average number of signal particles in the jet during training. This suggests that both neural networks independently learned a relationship that closely mirrors the multiplicity method for predicting jet momenta.

In nuclear physics, there has been a strong focus on developing machine learning analysis techniques that prioritize transparency, robustness, and precise uncertainty quantification, all while maintaining explainability [20]. We argue that for machine learning to be effectively applied to scientific problems, it must be interpretable. While the definition of interpretable machine learning remains an area of active research, several perspectives on interpretability are discussed in [89]. Previously, we introduced a preliminary set of criteria that a machine learning method must satisfy to be considered interpretable and, therefore, suitable for applications in nuclear physics [95]. Here, we present these criteria in a more explicit and refined manner.

Transparent: A machine learning method is considered transparent if its operations and decision-making processes are easily understandable and explainable to humans.

- The method allows for a measurable uncertainty for its predictions, allowing users to assess the confidence and reliability of the prediction. The model’s training uncertainty should not be mistaken as a substitute for measurement uncertainty.
- The internal mechanics of the method—including decision-making steps, feature relevance, and model parameters—should be accessible and interpretable.

Transferable: A machine learning method is transferable if the insights, learned representations, and results can be effectively applied to related problems or domains.

- Predictions should remain comprehensible beyond the domain of the training set, meaning the model should generalize well and its behavior should be predictable when applied to new data.

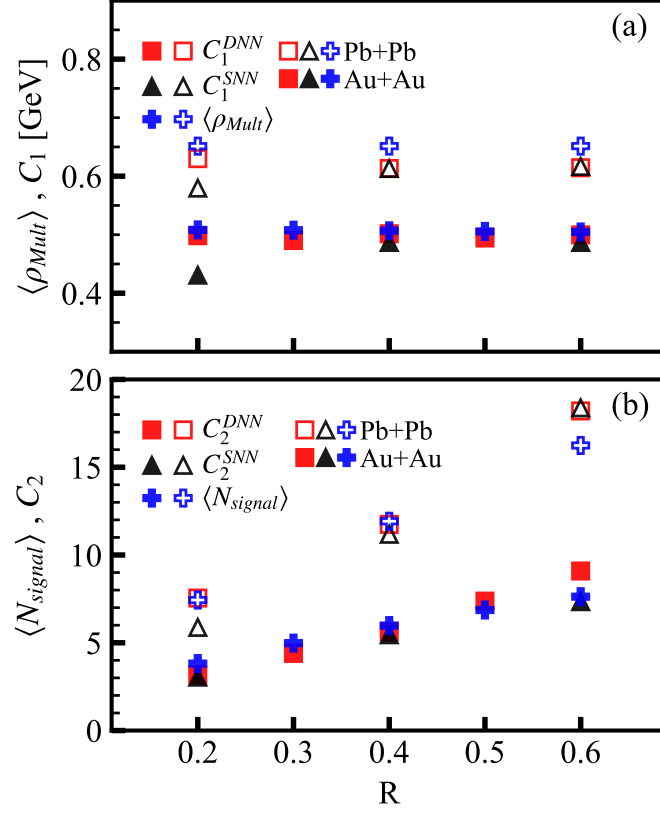


Figure 3.11: PySR optimization constants compared to average value of multiplicity method parameters versus jet resolution parameter for Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV and Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV for jet resolution parameters $R = 0.2, 0.4$, and 0.6 .

- The learned relationships and principles should extend to similar tasks, demonstrating robustness and versatility.

Trustworthy: A machine learning method is trustworthy if its predictions are reliable and the model’s behavior is consistent and justifiable.

- The method should be applicable equivalently to both experimental data and simulations, ensuring consistency in interpretation.
- The method should demonstrate robustness, and reproducibility, ensuring that it can be relied upon in applications to data.

Using interpretable forms of machine learning, such as symbolic mapping of a neural network, allows us to make use of expert domain knowledge. The optimization parameters from PySR would otherwise not have a clear physical interpretation. Since these parameters are understood in the multiplicity method, it is possible to assign a physically motivated uncertainty to them. Assumptions inherent in the method can then be understood.

Chapter 4

Analysis

Four methods are used to quantify the fluctuations of the underlying event observed in Au+Au at $\sqrt{s_{\text{NN}}} = 200$ GeV collisions measured with the sPHENIX calorimeter systems. Each of these characterization methods increases in complexity and allows us to quantify several aspects of the underlying event in heavy ion collisions, its contributions, and how it can be corrected in future jet measurements. A base-line characterization of the underlying event, both the pedestal and fluctuations, is found with calorimeter area windows, following the method described in [41]. In addition to calorimeter window areas, fixed area rigid cones placed randomly in central calorimeter acceptance $|\eta| < 1.1$ are used to quantify the underlying event in terms of jet-like objects without the bias of a given jet-clustering algorithms [12]. We explore the interplay of jet-clustering algorithms and the underlying event by placing high E_T probes into single calorimeter towers and embedding full PYTHIA8 di-jet events into minimum bias Au+Au collisions at $\sqrt{s_{\text{NN}}} = 200$ GeV. The underlying event and its fluctuations can then be studied in the presence of jets and jet-like objects.

The average pedestal of the underlying event can be measured and subtracted from jets on an event-by-event basis. We compare the impact of jet background fluctuations on jet reconstruction between various jet background subtraction methods. The precision of jet measurements and their kinematic range are restricted by corrections

for this background and its fluctuations. Extending the kinematic range of jet measurements to lower momenta would improve constraints on partonic energy loss. In particular, since a greater fraction of gluon-like jets are at low momenta, this would increase sensitivity to modifications of gluon-like jets. Jet background fluctuations are compared for three methods: the area-based method [52], the multiplicity-based method [95], and an iterative approach proposed in Ref. [71].

4.1 Reconstructed Physics Quantities

This analysis uses measured energy deposits from all three central calorimeters, as well as event-level information from the MBD and ZDC. First, the raw signal from all subsystems is organized into physics events using timing information from a granulated timing module. Every level-1 trigger issues a timestamp, which is synchronized with the RHIC bunch crossing orbit time of 107 ns, to the front-end electronics of each sPHENIX sub-detector. Figure 4.1 shows the correlation between energy deposits in the inner and outer hadronic calorimeters to the MBD charge sum after the offline event combining procedure.

Minimum bias events are defined as those triggering the MBD hardware level-1 minimum bias trigger, which requires at least two hits above the threshold in both the north and south MBD arms ($n^{MBD}N/S > 2$). Minimum bias events must also have at least two hits with a charge signal above 0.4 in both MBD arms, and the charge sum in one MBD arm must exceed 10 while the other arms charge sum remains below 150 ($\sum Q^{MBD}(N/S) > 10$ and $\sum Q^{MBD}(N/S) < 150$). These criteria exclude events where the charge signal in one MBD arm significantly exceeds the other by an order of magnitude. To ensure the presence of spectator neutrons on both sides, the sum of energy in each ZDC arm must exceed 40 GeV ($\sum E^{ZDC}N > 40$ GeV and $\sum E^{ZDC}S > 40$) [26, 29]. Finally, the z -vertex is required to be within 10 cm of the center of the sPHENIX detector volume ($|z_{vtx,MBD}| < 10$ cm). The tight z -vertex requirement suppresses events which triggered the MBD trigger but occur outside of

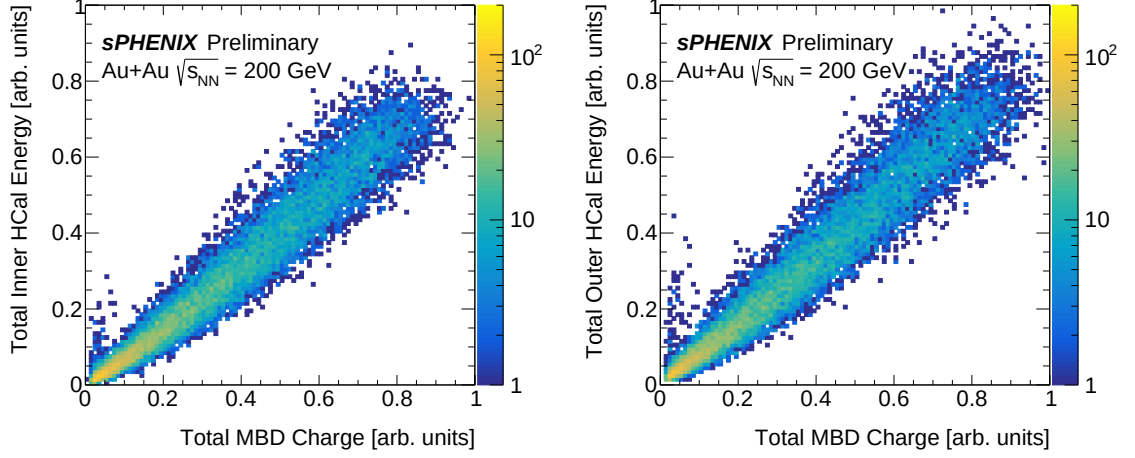


Figure 4.1: Correlation between hadronic calorimeter energy deposits and MBD charge sum of both MBD arms the inner HCAL on the left and the outer HCAL on the right. Figures taken from Ref.[114].

the mid-rapidity acceptance of the sPHENIX calorimeters ($|\eta| < 1.1$). The MBD z -vertex is based on the analog digital current charge channel information. Ideally, the calculation would use the MBD timing channels, but the alignment of these was not yet stable enough to perform calibration in Run 2023. The “hit-time” for each MBD arm is calculated using a digital constant fraction discriminator to convert the sampled waveforms from MBD photomultiplier tubes into time in nanoseconds. The channel to channel timing offset corrections are obtained by fitting the bunch time structure for each photomultiplier tube. After corrections, the time of arrival to each MBD arm is the mean of a Gaussian fit of all photomultiplier tubes of a given arm with a charge above the threshold. The $|z_{vtx,MBD}|$ can then be derived via

$$|z_{vtx,MBD}| = \frac{c}{2}(t_S - t_N)$$

where $t_{S/N}$ is the time of arrival in the south/north arm of the ZDC and c is the speed of light. The systematic uncertainty of measuring the time of arrival for each MBD arm is due to slight differences in the transit time of electrons in the photomultiplier tubes and signal cable transmission. This gives an uncertainty of the order of 2-3 cm on the $|z_{vtx,MBD}|$ determination.

The background fluctuations and pedestal are dependent on the charged particle multiplicity. Therefore this analysis is performed differentially as a function of event centrality classifications. Centrality percentiles were derived by fitting the MBD charge distribution to a convolution of particle production and event sampling based on a negative binomial distribution from a Glauber simulation [90, 26]. Figure 4.2 shows how centrality intervals are defined from MBD charge sub-distributions. Event selection achieved complete efficiency within 0-80% central events.

Calorimeter towers used in this analysis are the reconstructed signal corresponding to either five HCAL scintillating tiles or two EMCAL tungsten blocks. The raw signal extraction from calorimeters is described in detail in section 2.3.1. Each tile of the HCAL is read out at the outer radius, and the analog signals from five tiles are summed to a single preamplifier channel to form a single calorimeter tower. In the

08/30/2023

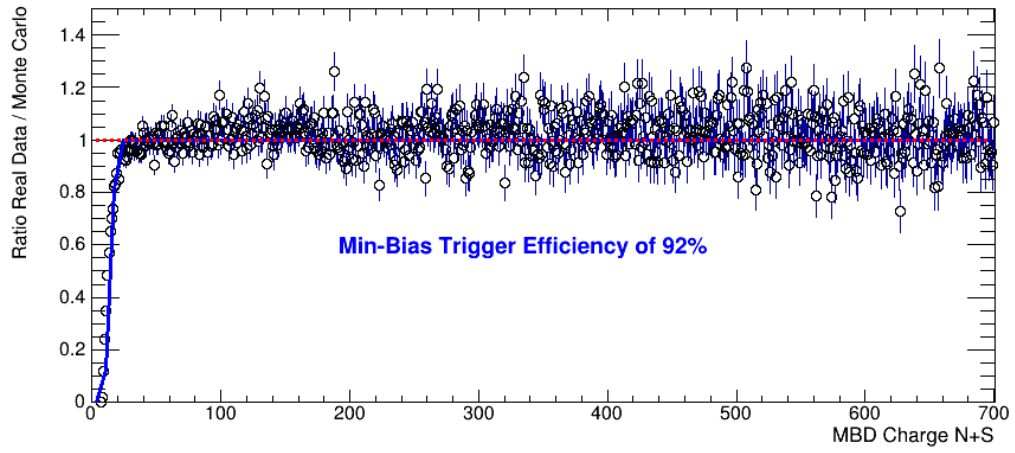
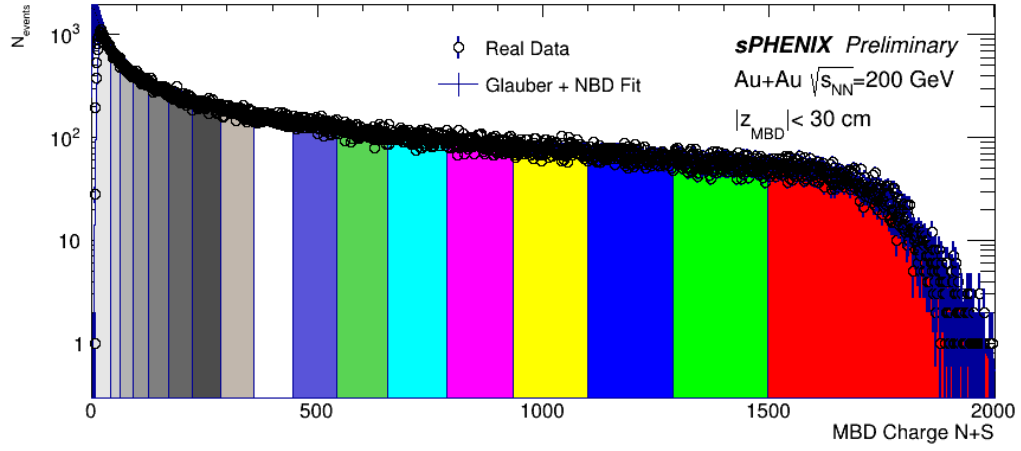


Figure 4.2: (Top) MBD charge sum distribution is shown in the upper panel (open points). Also, shown are the fit results from the MC Glauber and negative binomial distribution best fit. (Bottom) The ratio of data to MC Glauber including the fit to the trigger efficiency turn-on region. Figure taken from Ref.[114].

EMCAL, the scintillating fibers embedded in the tungsten absorber material are read out as one tower. The corresponding subsystem calorimeter towers will be referred to as EMCAL towers, iHCAL towers, and OHcal towers. Figure 4.3 shows an example of reconstructed calorimeter towers in a 0-10% Au+Au event. Calorimeter towers have an energy and a position in center-of-mass coordinates η , ϕ . The azimuthal position is found from the detector geometry, and the pseudorapidity η is found in reference to the reconstructed $|z_{vtx,MBD}|$.

Both the inner and outer HCALs were calibrated using cosmic ray muons to establish a tower-by-tower absolute energy scale at the electromagnetic scale. The calibration began with relative corrections derived from test stand data collected during assembly, which normalized tower performance and accounted for temperature variations. The energy distributions were compared to GEANT4 simulations of cosmic muon flux, and the most probable value of the energy deposition for each tower was extracted using a gamma function fit. The resulting energy scale was determined to be 283 analog digital current units per MeV. The calibration validation showed strong agreement between the in situ data and the simulation, with discrepancies limited to approximately 2% in most of the detectors. Larger deviations, up to 5%, were observed in specific regions due to geometry and temperature effects. These were mitigated using additional temperature corrections derived from LED calibration data, which accounted for variations in silicon photomultiplier gain. Figure 4.4 shows the single tower energy scale distributions and comparisons of the cosmic ray muons in simulation and data.

The EMCAL was calibrated to the electromagnetic scale with a combination of tower energy spectra, reconstructed π^0 mass, and Minimally Ionizing Particle distributions. Tower responses in ϕ are flattened via the Tower Slope Calibration, which compares the slope of the single tower distributions at low energy in one ring in ϕ to obtain calibration constants which calibrate each tower's slope to the ring's average. The tower response in η is calibrated using reconstructed neutral pions combined with single particle simulations which provide the absolute energy

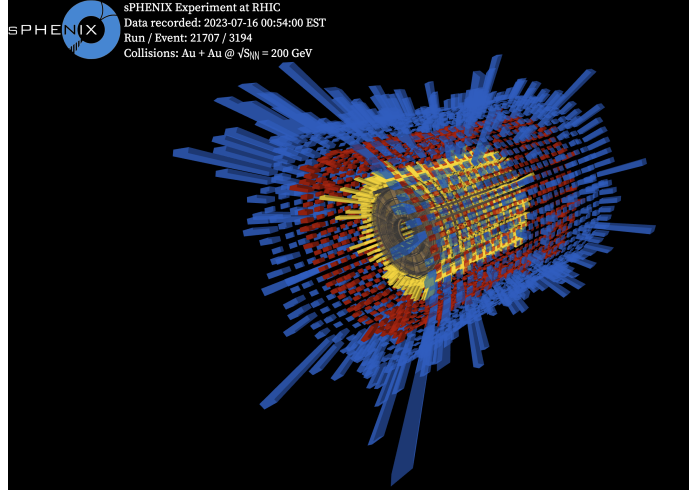


Figure 4.3: Calorimeter tower event cross section display from a 0-10% central $\sqrt{s_{NN}} = 200$ GeV $Au + Au$ event. The yellow, red, and blue towers correspond to the EMCAL, inner HCAL, and outer HCAL respectively. Event display from Ref. [113].

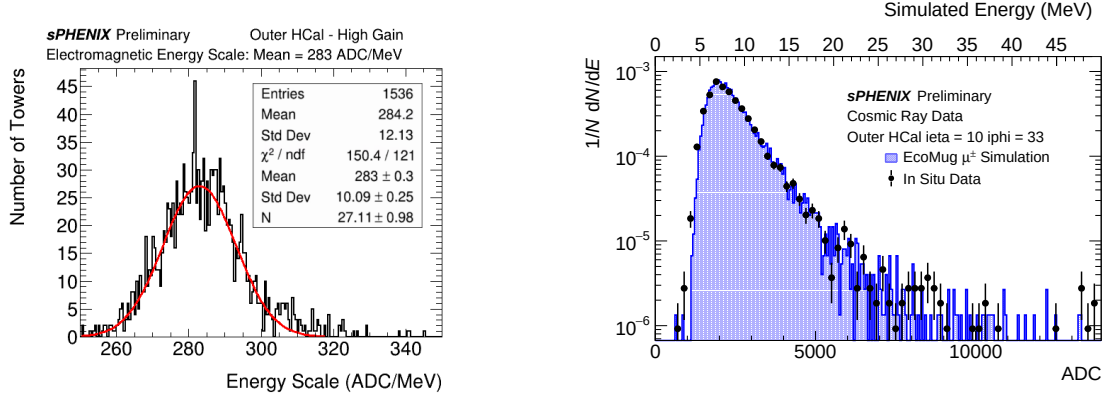


Figure 4.4: (Left) Outer hadronic calorimeter tower distribution of energy scale from cosmic muon calibration. (Right) Comparison of a single outer hadronic calorimeters tower's measured energy distribution and the corresponding GEANT4 simulation result. Figure taken from Ref.[114].

scale taking into account resolution effects. The calibration was conducted in two stages: relative scale calibration (tower slope) and absolute energy scale calibration. The relative scale calibration aligned the responses of towers within the same η slice, assuming ϕ symmetry. Parameters for gain and scale were set by fitting each tower's energy to a reference spectrum. The calibration was performed iteratively. In the first iteration, a reference spectrum was defined using all towers in the slice. Towers identified as outliers were then masked, and the calibration was repeated. The absolute energy scale calibration aligned energy scales between different η and was calibrated to match a target π^0 mass in simulation. The calibration and noise reduction methods were validated by comparing the results to simulation.

Besides the primary calorimeter towers corresponding to the reconstructed subsystems (EMCAL, iHCAL, oHCAL), a “Retowered” EMCAL tower is also used in several parts of this analysis. For clarity, these will be referred to as ReEMCAL towers. The reason for this second level of reconstruction over the EMCAL towers is the extreme segmentation of the EMCAL compared to the hadronic calorimeters. The iterative method of background reconstruction (described further in section 4.3.1) requires a smooth distribution of energy as a function of pseudorapidity to estimate local background energy densities. This is difficult with highly granular ($0.025 \Delta\eta \times 0.025 \Delta\phi$) EMCAL towers when only 50% of towers are occupied in central $Au + Au$ collisions. Therefore, the concept of ReEMCAL towers is to combine EMCAL towers into “towers” that match the geometry of the inner HCAL. There are roughly 16 EMCAL towers ($0.025 \Delta\eta \times 0.025 \Delta\phi$) per 1 ReEMCAL tower ($0.1 \Delta\eta \times 0.1 \Delta\phi$). The match in geometry between the ReEMCAL, iHCAL, and oHCAL towers is shown in figure 4.5. The ReEMCAL tower is a combination of multiple EMCAL towers. EMCAL towers are masked based on the previous cuts prior to the reclustering. We require that at least 50% of the constituent EMCAL towers in a ReEMCAL tower be unmasked. If more than 50% of constituent towers are masked, the ReEMCAL tower is masked.

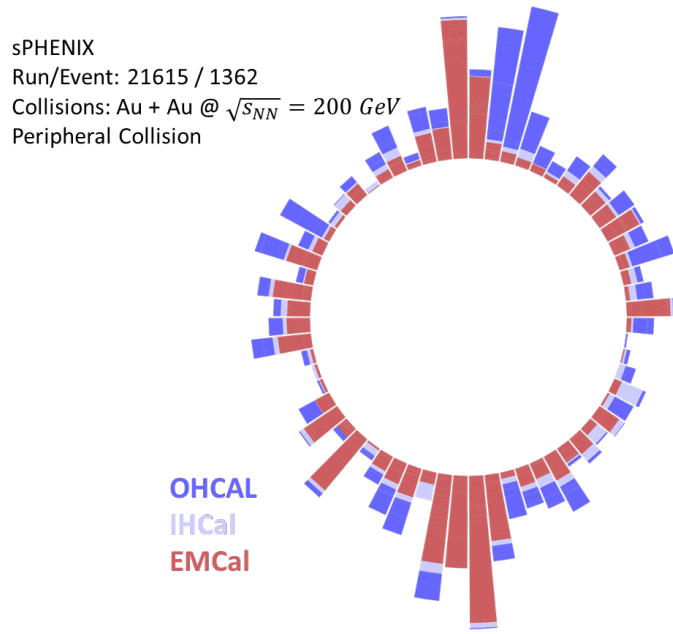


Figure 4.5: Calorimeter tower event cross section from a 50-60% central $Au + Au$ $\sqrt{s_{NN}} = 200 \text{ GeV}$ di-jet event display. The legend label EMCAL refers to the ReEMCAL towers. Event display from Ref. [113].

The final calibrated and zero suppressed calorimeter towers are converted into vectors of input jet objects for the random cone, high E_T probes, and embedded PYTHIA8 di-jet studies. The Cartesian coordinates for a tower are corrected for vertex location on an event-by-event basis. This correction is used to determine the η of the calorimeter tower,

$$\eta_{\text{corr.}} = \sinh^{-1} \left(\frac{z_{\text{tower}} - z_{\text{MBD}}}{r_{\text{calo}}} \right), \quad (4.1)$$

where $z_{\text{tower}} - z_{\text{MBD}}$ is the corrected z-position of the tower in a given event. Only the $z_{\text{vrtx, MBD}}$ was used in this analysis. The ϕ position of the towers is based on detector geometry,

$$\phi_{\text{tower}} = \arctan \left(\frac{y_{\text{tower}}}{x_{\text{tower}}} \right), \quad (4.2)$$

where y_{tower} and x_{tower} are defined to be the center of the tower. The four momentum of the jet input object is calculated with the corrected coordinates,

$$p_x = E_T \cos \phi_{\text{tower}},$$

$$p_y = E_T \sin \phi_{\text{tower}},$$

$$p_z = E_T \sinh \eta_{\text{corr}},$$

$$E_T = E_{\text{tower}} / \cosh(\eta_{\text{corr}}),$$

where E_{tower} is the calibrated tower energy.

4.2 Event Selection

Events selected in this analysis are $Au + Au \sqrt{s_{NN}} = 200$ GeV minimum bias events with a restricted z -vertex acceptance of $|z_{\text{vrtx, MBD}}| < 20$ cm, and a centrality classification between 0–80%. In addition, all calorimeter towers in the event must pass the waveform template fit and the event must contain the full possible EMCAL. A total of 10^6 events were selected from stable running periods in July 2023 and

September 2024. This measurement is not statistically demanding so the relatively small number of events is sufficient. Removing the maximum EMCAL acceptance criteria would lead to more events but at the cost of introducing event-by-event fluctuations in the EMCAL active area and related systematic uncertainties. The input calorimeter towers are restricted to those within $|\eta| < 1.1$. All towers must have good status, and a software level zero suppression of 60 (30) analog digital current counts is implemented for EMCAL (HCAL) towers. Towers from the ReEMCAL must have $> 50\%$ of constituent towers with good status.

4.3 Jet Background Subtraction Methods

4.3.1 Iterative Method

The iterative subtraction method is based on reference [71]. This method is motivated by the method used by ATLAS and is similar to the iterative approach used by CMS. The underlying event is determined individually for each layer of the calorimeter in strips of $\Delta\eta = 0.1$. First, the EMCAL towers are “retowered” into the ReEMCAL towers. After retowering, the anti- k_T algorithm is run over the ReEMCAL, iHCAL, and oHCAL towers with $R = 0.2$ to determine candidate jets in the event, referred to as “seeds”. For the first iteration, seed jets are defined as $R = 0.2$ jets with a maximum constituent energy divided by the mean constituent energy to be greater than 3. The average energy per tower is determined for each η strip, excluding towers within $\Delta R < 0.4$ of a seed from the calculation. The determination is done individually in each layer of the calorimeter. This average energy is then subtracted from each tower in the event, and the collection of $R = 0.2$ jets is re-analyzed after the first iteration of subtraction to determine the seeds for the second iteration. For the second iteration, the seeds are required to have a $p_T^{\text{sub}} > 7$ GeV. The average energy per tower is again determined, excluding the new set of seeds, and the towers are subtracted. These background subtracted towers are then used as the input to

the anti- k_T algorithm used to reconstruct jets, or in the case of this analysis are used as inputs to either random cones or calorimeter windows. Figure 4.6 shows a linear dependence of the event averaged tower background on the total central calorimeter energy.

4.3.2 Area Method

The area method is based on reference [52] and is primarily used in STAR and ALICE jet measurements. The method corrects jet transverse momentum, or in our case energy, by estimating the average background energy density per unit area. In the area method used in this analysis, pseudorapidity dependencies or local fluctuations are not taken into account. The method can be expressed as

$$E_{T,\text{jet}}^{\text{Corr, A}} = E_{T,\text{jet}}^{\text{Uncorr, A}} - \rho_{\text{Area}} A_{\text{jet}}, \quad (4.3)$$

where $E_{T,\text{jet}}^{\text{Corr, A}}$ is the correct transverse energy of the jet, $E_{T,\text{jet}}^{\text{Uncorr, A}}$ is the uncorrected transverse energy, A_{jet} is the area of the jet, and ρ_{Area} is the background energy density. The area of the jet is computed by adding uniformly distributed, extremely low energy “ghost” ($E_{T,\text{ghost}} = 1$ MeV) particles into the FastJet [53] clustering algorithm. The resulting jets are not affected by these “ghosts” due to infrared safety of jet-finding algorithms. The area of the jet is proportional to the number of “ghosts” that are clustered within the jet cone. The background density is an event-level quantity estimated using jets clustered with the k_T algorithm. First, k_T jets with resolution parameter $R = 0.4$ are clustered from towers in the electromagnetic calorimeter, and both hadronic calorimeters. The two hardest of these k_T jets are omitted to suppress biasing ρ_{Area} with objects originating from a hard-scattering. The remaining k_T jets are used to create a distribution of $E_{T,\text{jet}}^{k_T}/A_{\text{jet}}^{k_T}$ where the median value is assumed to be ρ_{Area} . Figure 4.7 shows the distribution of ρ_{Area} for $\sqrt{s_{NN}} = 200$ GeV $Au + Au$ events observed in sPHENIX.

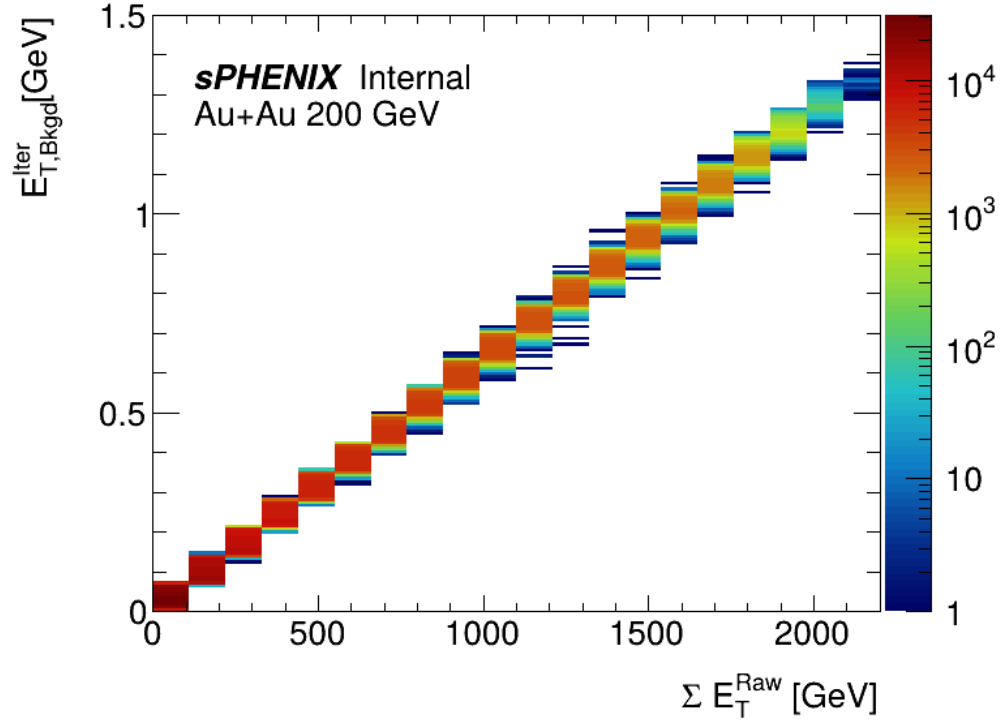


Figure 4.6: Average background calorimeter tower energy determined by the iterative background subtraction method in $Au + Au$ at $\sqrt{s_{NN}} = 200$ GeV.

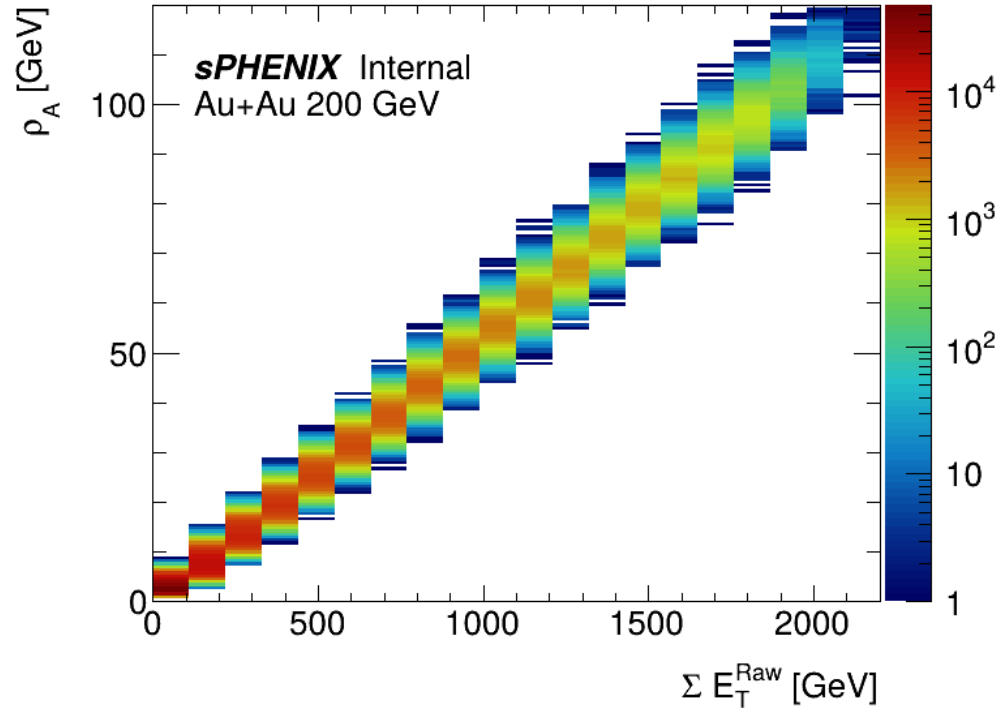


Figure 4.7: ρ_A as function of total calorimeter energy in $\sqrt{s_{NN}} = 200$ GeV $Au + Au$ events.

4.3.3 Multiplicity Method

The multiplicity method is based on reference [95]. This method was proposed in 2023 and has not been applied to data until this measurement. It is a ρ -based background subtraction method similar to the area method. The multiplicity method uses the average momentum of background particles $\langle p_T \rangle$ and the average excess multiplicity originating from the underlying event within the jet cone to subtract the average background energy. This approach uses the natural variable in the standard deviation, N , eliminating the second and third terms in equation 1.10.

In this analysis, the average transverse energy of underlying event calorimeter towers $\langle E_T \rangle$ and the average excess tower multiplicity originating from background within the jet to subtract the background energy

$$E_{T,\text{jet}}^{\text{Corr, N}} = E_{T,\text{jet}}^{\text{Uncorr}} - \rho_{\text{Mult}}(N_{\text{towers}} - \langle N_{\text{signal}} \rangle), \quad (4.4)$$

where N_{towers} is the observed number of towers within the jet, $\langle N_{\text{signal}} \rangle$ is the average number of towers in a signal jet of a given $E_{T,\text{jet}}^{\text{raw}}$, and ρ_{Mult} represents the mean transverse energy per background tower.

The average background tower energy ρ_{Mult} in an event is determined using a calculation similar to ρ_{Area} . Jets are found using the k_T algorithm with resolution parameter $R = 0.4$ from towers in the electromagnetic calorimeter, and both hadronic calorimeters. The two hardest of these k_T jets are omitted from the ρ_{Mult} calculation. The average energy per tower is calculated from the remaining k_T jets and the median value of this distribution is taken to be the average transverse energy per background tower $\langle E_T^{\text{background}} \rangle$. Figure 4.8 shows the distribution of ρ_{Mult} observed in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV.

The parameter average number of signal towers in a jet $\langle N_{\text{signal}} \rangle$ can be estimated with accuracy, as models for jets in proton-proton collisions [15] can adequately describe it. Our estimate of $\langle N_{\text{signal}} \rangle$ is done by matching signal PYTHIA8 jets to the same jets which have been embedded in Au+Au minimum bias events. The matching

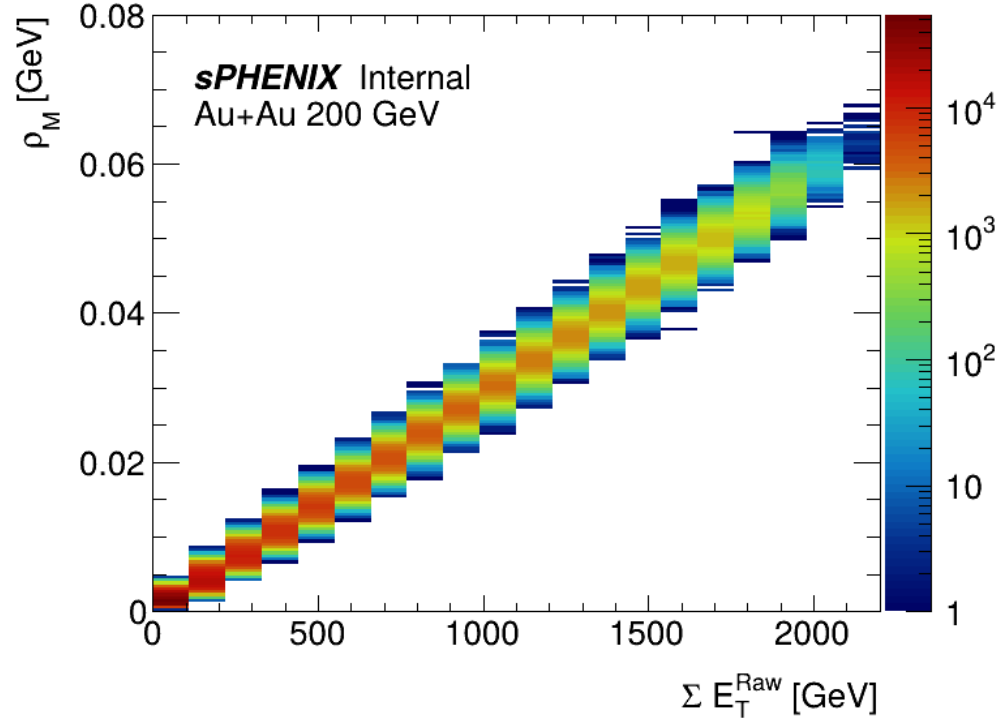


Figure 4.8: ρ_{Mult} as function of total calorimeter energy in $\sqrt{s_{NN}} = 200$ GeV $Au+Au$ events.

criteria is geometric, requiring that $\Delta R = \sqrt{(\eta_t - \eta_r)^2 + (\phi_t - \phi_r)^2} < 0.75R$ and that the reconstructed jet energy $E_{T, reco} > 5$ GeV. The number of towers in the matched PYTHIA jets are then binned according to the raw uncorrected E_T of anti- k_T jets reconstructed from the mixed event sample.

4.4 Underlying Event Characterization

To quantify the fluctuations of the underlying event seen in $Au + Au \sqrt{s_{NN}} = 200$ GeV data, we use a multi-faceted approach. Each of these characterizations increases in complexity and allows us to quantify several aspects of the underlying event and how it can be corrected in jet measurements. These analyses start with a baseline characterization of the underlying event with calorimeter windows. In order to quantify the underlying event in terms of jet-like objects without the bias of a given jet clustering algorithm a random cone analysis is performed [12]. Finally, high energy probes are embedded into minimum bias data to investigate the interplay between the jet-finding algorithm and underlying event fluctuations. The background fluctuations seen are described by Poissonian number fluctuations with contributions from hydrodynamical flow [115].

4.4.1 Calorimeter Window Analysis

The calorimeter window analysis is based on previous studies done by ATLAS [41]. In each event, the calorimeter is divided into $n \times m$ tower sized windows, and the mean and standard deviation of the transverse energy in each of the windows is determined. This is done separately for the ReEMCAL (as described in section 4.1), iHCAL, and oHCAL towers. The final window energy $E_T^{m \times n}$ is the sum of overlapping windows in each of the calorimeter layers. A corresponding calorimeter window must exist in all calorimeter layers to be considered as a full window. Windows that contain masked towers in either the ReEMCAL, iHCAL, or oHCAL within the window area

are excluded. The acceptance for calorimeter towers to be included in a calorimeter window is $|\eta| < 1.1$. Table 4.4.2 shows the correspondence between a given jet radius and a window of approximately the same area.

We use a sliding window to maximize the number of $n \times m$ window areas in a given event and to avoid edge effects from run-by-run masked calorimeter areas. In the case that windows are not symmetric in $\eta \times \phi$, (i.e. $n \times m = 11 \times 12$), the reciprocal window ($m \times n$) area is calculated with a sliding window acceptance and included in the window $E_T^{m \times n}$ distribution. The standard deviation of the $E_T^{m \times n}$ distribution is then calculated within each event for all calorimeter window area sizes according to

$$\sigma^{n \times m} = \sqrt{\langle E_{T,n \times m}^2 \rangle - \langle E_{T,n \times m} \rangle^2}. \quad (4.5)$$

The values for $\langle E_{T,n \times m}^2 \rangle$ and $\langle E_{T,n \times m} \rangle^2$ are calculated using the sum of all E_T from windows of size $n \times m$ and the number of accepted windows. Figure 4.9 shows example distributions for $\sigma^{n \times m}$ and $\langle E_T^{n \times m} \rangle$ of different sized window area.

The event-averaged standard deviation $\bar{\sigma}^{n \times m}$, average energy $\langle \bar{E}_T \rangle^{n \times m}$ are used to quantify average pedestal and fluctuations of the underlying event. Due to finite η acceptance, each increase in window area size results in fewer calorimeter windows being accepted for a given event. The event-averaged $\langle \bar{E}_T \rangle^{n \times m}$ maintains a linear dependence with calorimeter window area for all event centralities. The decrease in window population can be seen in figure 4.10, in comparison to the linear dependence of $\langle \bar{E}_T \rangle^{n \times m}$.

4.4.2 Random Cones Analysis

Random cones enable the allow us to identify contributions to the fluctuations in terms of jet-like objects without the bias imposed by jet clustering algorithms [12]. This random cone analysis is modeled after a similar analysis performed in ALICE to measure the underlying event fluctuations at LHC energies using charged particles [12]. The background fluctuations are sampled by drawing a single rigid cone

Table 4.1: Area of calorimeter window size corresponding to a jet defined by the parameter R.

R	A_{cone}	Window Size	A_{window}
0.2	0.13	3×4	0.11
0.3	0.28	5×6	0.27
0.4	0.5	7×8	0.5
0.5	0.79	9×10	0.81
0.6	1.13	11×12	1.19
0.7	1.54	13×13	1.52
0.8	2.01	15×15	2.02

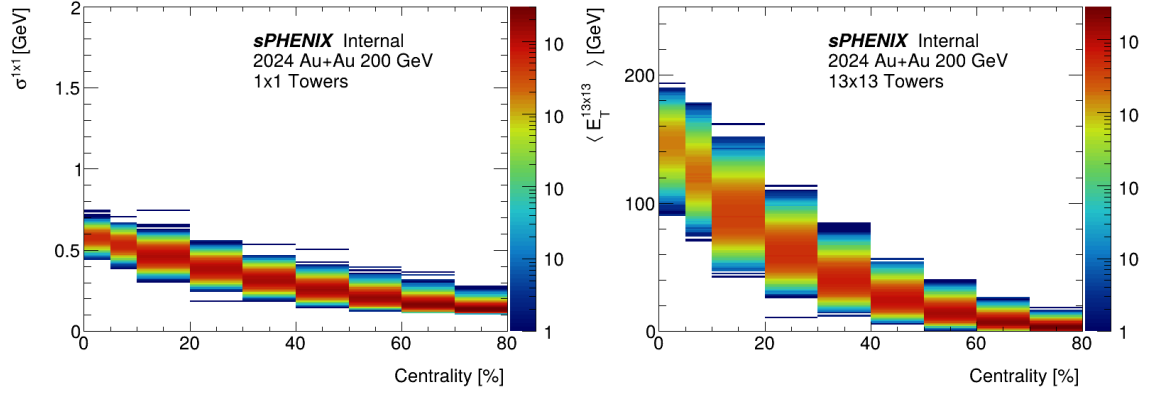


Figure 4.9: Centrality dependence of event-by-event standard deviation of calorimeter windows of size $n \times m = 1 \times 1$ (Left), and the event average E_T of windows with size $n \times m = 13 \times 13$ (Right) for all events.

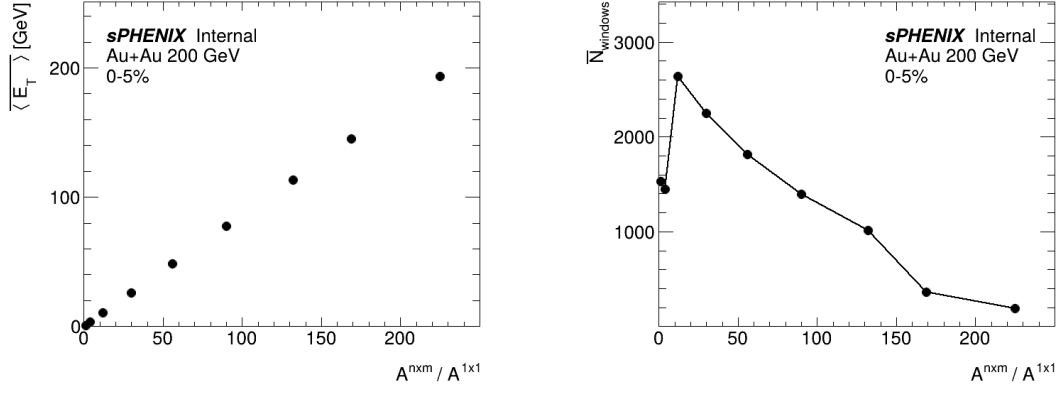


Figure 4.10: Window area dependence of event-averaged transverse energy $\langle E_T \rangle^{n \times m}$ for windows of all window sizes (left), and the number of reconstructed windows of a given area (right) for 0-5% central events.

with radius $R=0.4$ in a random direction such that the full cone is within calorimeter acceptance $|\eta_{\text{cone}}| < 1.1 - R$, $0 \leq \phi_{\text{cone}} < 2\pi$. Calorimeter towers from the EMCAL, iHCAL, and oHCAL within $\Delta R < 0.4$ of the cone axis are included in the cone energy

$$E_T^{\text{Cone}} = \sum_{i=0}^N E_{T,i},$$

where $E_{T,i}$ is the transverse energy of the i^{th} calorimeter tower. The background fluctuations are characterized by the difference in E_T^{Cone} to the average underlying event pedestal. The expected pedestal is determined using three jet background subtraction methods described in section 4.3. The underlying event fluctuations δE_T for each subtraction method are defined as:

$$\delta E_T^{\text{Area}} = \sum_{i=0}^N E_{T,i} - \rho_A \cdot A_{\text{cone}},$$

$$\delta E_T^{\text{Mult}} = \sum_{i=0}^N E_{T,i} - \rho_M \cdot N,$$

$$\delta E_T^{\text{iter}} = \sum_{i=0}^N E_{T,i}^{\text{Sub}},$$

where ρ_A and ρ_M are the average background densities defined in section 4.3, $A_{\text{cone}} = \pi R^2$, N is the number of towers within the cone radius, and $E_{T,i}^{\text{Sub}}$ is the subtracted calorimeter towers after iterative subtraction. The multiplicity method is altered in the case of random cone subtraction because $\langle N_{\text{signal}} \rangle$ is assumed to be zero, rather than estimated from the reconstructed cone E_T , due to the assumption that there are no jets present in the minimum bias data.

In addition to static random cones, the analysis is repeated for events in which the tower (η, ϕ) has been randomized. This destroys any event level correlation between tower input objects and allows for a sampling of the underlying event fluctuations near the expected statical limit derived from equation 1.9. The randomization is done while maintaining the locations of the masked calorimeter towers within the event. Masked towers remain constant under randomization so the calorimeter acceptance

remains unchanged. The two samples of random cones will be denoted as “Basic” for standard static random cones, and “Randomized” or “Randomized $\eta\phi$ ” for random cones constructed with randomized tower input objects.

Each δE_T distribution is characterized by the standard deviation $\sigma(\delta E_T)$ for different event centrality bins. The δE_T distributions from random cones subtracted with the area, multiplicity, and iterative subtraction methods in 0-5% central Au+Au events are shown in figure 4.11. The distributions are centered near 0 for all methods, indicating each method describes the average underlying event pedestal.

Probe and Jet embedding Analysis

The jet probe analysis quantifies the underlying event in jets clustered around a point-like massless probe. This analysis takes into account the effects of jet reconstruction on the underlying event, without the effect of a realistic jet. The jet embedding analysis quantifies the underlying event by looking at simulated jets embedded into minimum bias $Au + Au$ data, and then reconstructing the embedded jets. The jets are found via the anti- k_T jet finding algorithm with resolution parameter $R=0.4$. Jet candidates are required to be within $|\eta_{jet}| < 1.1 - R$ and have an uncorrected energy greater than $E_{T,jet}^{Uncorr.} > 5$ GeV. The accepted candidates are geometrically matched to the truth level definition (either $\eta_{probe}, \phi_{probe}$ or the particle level PYTHIA8 jet $\eta_{truth}, \phi_{truth}$) requiring that $\Delta R = \sqrt{(\eta_t - \eta_r)^2 + (\phi_t - \phi_r)^2} < 0.75R$. Each background subtraction method is applied to both embedded and high E_T probe matched jet. The δE_T residual is calculated for each of the background subtraction methods according to

$$\delta E_{T,Area} = E_{T,jet}^{Uncorr.} - \rho_A \cdot A_{jet} - E_{T,truth},$$

$$\delta E_{T,Mult} = E_{T,jet}^{Uncorr.} - \rho_M \cdot (N_{const} - \langle N_{signal} \rangle) - E_{T,truth},$$

$$\delta E_{T,Iter} = E_{T,jet}^{sub.} - E_{T,truth},$$

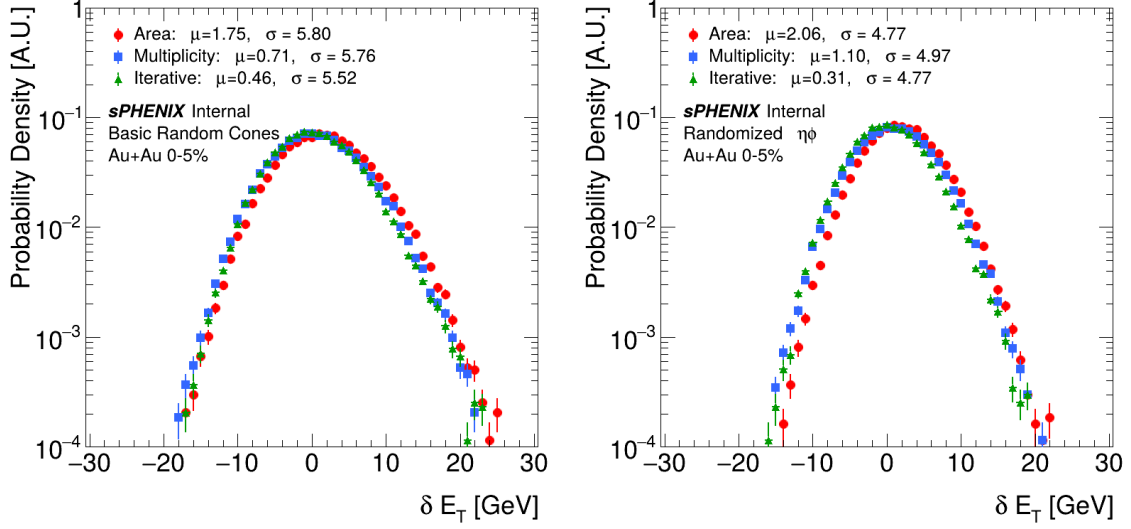


Figure 4.11: δE_T distributions for random cones reconstructed in 0-5% central events for area-based subtracted (black), multiplicity subtracted (red), and iteratively subtracted (blue) jets. (Left) Static random cones, (Right) randomized tower input random cones.

where the A_{jet} is defined by the Fastjet via introducing ghost particles, $\langle N_{signal} \rangle$ is the expectation value of the number of signal constituents in the jet (1 for E_T probes and $\langle N_{PYTHIA} \rangle(E_T)$, $E_{T,jet}^{sub.}$ is the reconstructed jet from the iteratively subtracted calorimeter towers, and $E_{T,truth}$ is the energy of either the high E_T probe or the truth energy of the PYTHIA8 jet.

We use $E_{T,probe} = 30$ GeV for our high E_T jet input probe. This is done by adding a particle randomly within the calorimeter acceptance $|\eta_{probe}| < 0.6$, $0 \leq \phi_{cone} < 2\pi$ after tower calibration and iterative tower subtraction in the case of $E_{T,Iter.}$. The resulting jets are highly circular with radius $\approx R$. We assume that $\langle N_{signal} \rangle = 1$ for jets that contain a probe particle. Figure 4.12 shows the δE_T distributions for jets reconstructed containing a probe $E_{T,probe} = 30$ GeV in 0-5% central events.

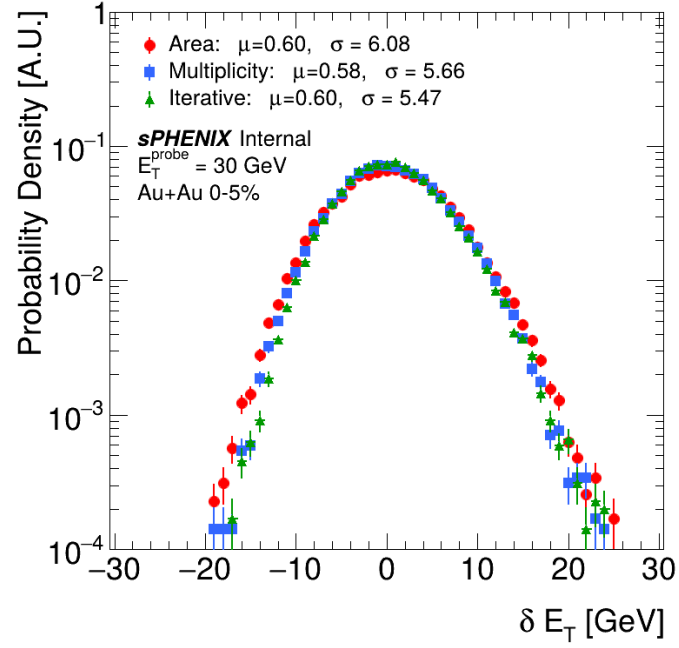


Figure 4.12: δE_T distributions for jets reconstructed containing a probe $E_{T,\text{probe}} = 30$ GeV in 0-5% central events for area-based subtracted (left), multiplicity subtracted (red), and iteratively subtracted (right) jets.

Chapter 5

Results and Discussion

The correlation of underlying event fluctuations between calorimeter regions is characterized using the dependence of the event-averaged standard deviation $\bar{\sigma}^{n \times m}$ for calorimeter windows with area $A_{window} = dA_{tower} \cdot (n \times m)$, where dA_{tower} is the area of a single tower in HCAL geometry ($\Delta\eta \times \Delta\phi = 0.1 \times 0.1$). If the fluctuations are uncorrelated between regions, then $\bar{\sigma}^{n \times m}$ will increase with calorimeter window area proportional to $\sqrt{n \times m}$. Correlated underlying event fluctuations will cause $\bar{\sigma}^{n \times m}$ to increase faster than $\sqrt{n \times m}$. The dependence of the fluctuations on window area is determined by fitting

$$\bar{\sigma}^{n \times m} / \bar{\sigma}^{1 \times 1} = (n \times m)^k, \quad (5.1)$$

where $\bar{\sigma}^{n \times m} / \bar{\sigma}^{1 \times 1}$ is event-averaged standard deviation scaled by the event-averaged standard deviation for single calorimeter tower windows in a given centrality bin, and k is scaling parameter. Figure 5.1 shows examples of these fits for 0-5% and 30-40% central events. Fits for all event centrality bins are available in App. B.5.

The correlation parameter k is shown for all calorimeter window sizes are shown as a function of event centrality in Figure 5.2. The scaling parameter in peripheral collisions is found to be close to the expected $\sqrt{n \times m}$ scaling, but for central to semi-central collisions the exponent increases to about 0.57 indicating a possible role

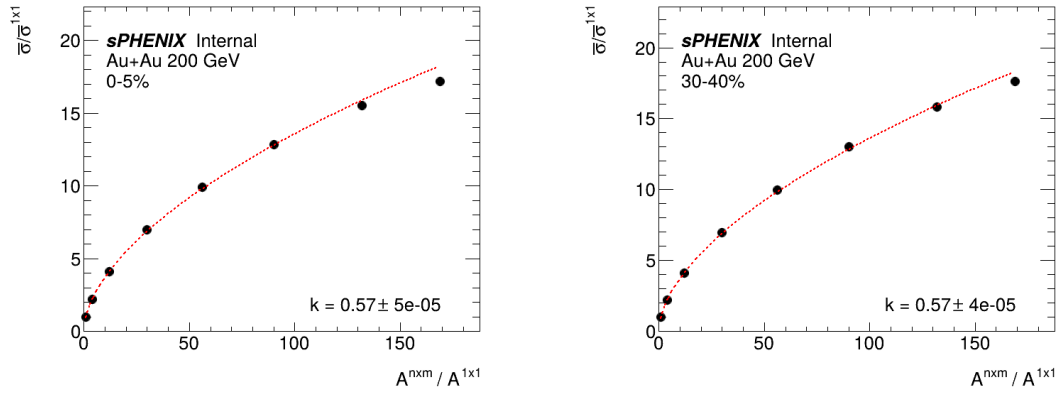


Figure 5.1: Example of the fit to the formula in equation 5.1 for 0-5% central (left) and 30-40% central (right) events. The x -axis is scaled by the area of a single tower and the points represent calorimeter window areas with dimensions presented in table 4.4.2.

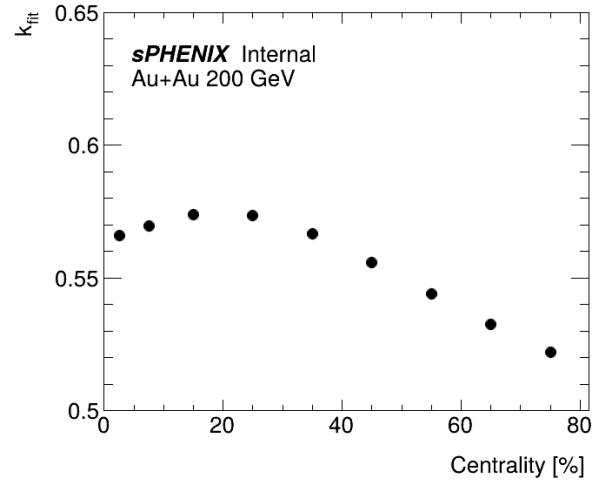


Figure 5.2: The scale parameter k from equation 5.1 for all fits to the dependence of the mean standard deviation in calorimeter window transverse energy at electromagnetic scale to calorimeter window area, as a function of event centrality.

of correlated fluctuations. The observed scaling with centrality could be indicative of short-range correlations and collective flow. Further study is needed to understand this quantitatively, both in terms of detector response and physics.

The dominant contribution to the jet energy resolution in heavy ion collisions comes from the fluctuations of the underlying event. The pedestal and fluctuations of the underlying event, characterized by the ρ -based background subtraction methods, is shown in Figure 5.3. These methods take no regional or long-range correlations into their background estimations. In the case of random cones where N is the equation to the number of towers within the cone and $A = \pi R^2$, both methods are expected to produce similar estimations of the underlying event. The similarity of $\rho_A \cdot A$ and $\rho_M \cdot N$ is seen for all centrality classes.

The multiplicity method introduced in Sec 3.1 was developed for charged particle jets using tracks as the input object to jet-finding and ρ_N calculations. This method capitalizes on the fact that the natural variable in the standard deviation is the number of background particles, eliminating the second and third terms in equation 1.10. In the case of sPHENIX calorimeter towers as the input to jet-finding, the dynamic range of jet/cone constituent multiplicity is significantly decreased due to the high occupancy of calorimeters in Au+Au events.

Distributions of underlying event fluctuations δE_T are characterized by the standard deviation $\sigma(\delta E_T)$ for different event centrality bins. The δE_T distributions for random cones subtracted with the area, multiplicity, and iterative subtraction methods in 0-5% central Au+Au events are shown in Figure 5.4. The distributions are centered near zero for all methods, indicating each method describes the average underlying event pedestal. The largest discrepancy from a mean value of zero is the area method.

The negative tail of the δE_T distributions are iteratively fit with a Gaussian function to determine the mean $\mu^{l.h.s}$ and standard deviation $\sigma^{l.h.s}$ of the left-hand side. This Gaussian fit is extrapolated to positive δE_T for both types of random cones. The positive tail of δE_T is found to be wider than that of a Gaussian distribution for

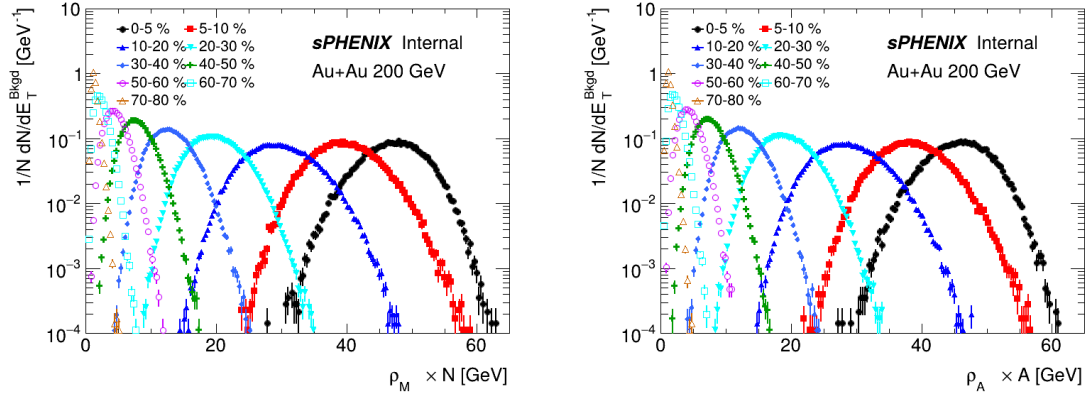


Figure 5.3: Underlying event quantification for the multiplicity method (left) and the area method (right) measured in Au+Au events for all event centrality classes.

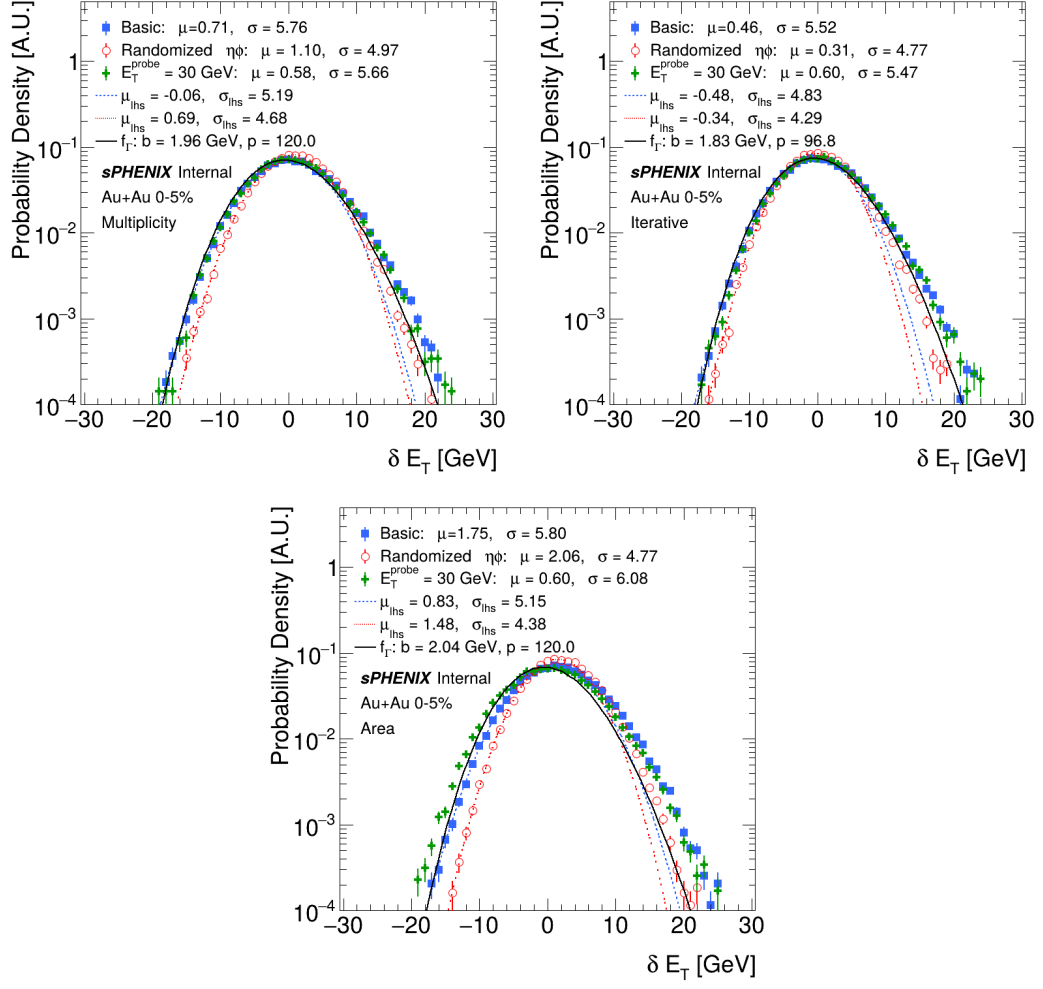


Figure 5.4: δE_T for multiplicity subtracted (top left), iteratively subtracted (top right), and area subtracted (bottom) underlying event characterizations in 0-5% central Au+Au events at $\sqrt{s_{\text{NN}}} = 200$ GeV for both types of random cones and embedded high E_T probes.

both types of cones. The values of the mean and standard deviation for both types of cones and embedded high energy probes are detailed in Table 5.1 for 0-5% central Au+Au events. The values for the left-hand side negative δE_T tail are not used in comparisons of underlying event fluctuations but rather presented to characterize the differences between the positive and negative δE_T tails. The standard deviation for central events using unbiased sampling is found to be $\sigma = 5.5 \pm 0.5$ GeV.

The shape of δE_T is not expected to be Gaussian, even in the limit of purely statistical fluctuations without contributions from hard-scattering or hydrodynamical flow. Approximating this purely statistical limit can be done by randomizing the charged particle coordinates η and ϕ within an event and calculating the underlying event fluctuations with the same procedure. This randomization destroys any long-range spatial correlations from flow, as well as short-ranged correlations from hard-scattering within the event [80, 95]. Fitting the left-hand side of the randomized particle random cones δE_T and extrapolating to the positive δE_T side clearly shows that the statistical limit has a non-Gaussian positive tail. This distribution is better described by a Γ distribution, which matches the descriptions of the underlying event of random particles sampled from a single particle emission spectra [115, 12, 80]. The single particle emission Γ -distribution in the case of calorimeter towers can be expressed as

$$f_{\Gamma}(\delta E_T) = A \cdot \frac{b}{\Gamma(p)} \cdot (b \cdot \delta E_T + p)^{p-1} \cdot e^{-(b \cdot \delta E_T + p)}, \quad (5.2)$$

where Γ is the real-valued gamma function, b is the shape parameter, and p is the scale parameter. This corresponds to a Γ -distribution with a mean value of $\mu = p/b$ and a standard deviation of $\sigma = \sqrt{p}/b$. The solid line in Figure 5.4 is a fit to the $f_{\Gamma}(\delta E_T)$ function (equation 5.2) for the randomized tower $\eta\phi$ events.

The width of the δE_T distribution due to purely statistical fluctuations without event correlations is given by [12]

$$\sigma(\delta E_T) = \sqrt{\langle N \rangle \cdot \langle E_T^2 \rangle}, \quad (5.3)$$

Table 5.1: Width and mean of underlying event distributions δE_T and negative δE_T extrapolated via Gaussian fit in 0-5% central Au+Au events for both types of random cones and embedded high E_T probes subtracted with each background subtraction method.

	μ [GeV]	σ [GeV]	$\sigma^{l.h.s.}$ [GeV]	$\mu^{l.h.s.}$ [GeV]
Area Method				
Basic Cone	1.75	5.8	5.15	0.83
Randomized $\eta\phi$	2.1	4.8	4.4	1.5
High Energy Probe	0.6	6.1		
Multiplicity Method				
Basic Cone	0.71	5.8	5.2	-0.057
Randomized $\eta\phi$	1.1	5	4.7	0.69
High Energy Probe	0.58	5.7		
Iterative Method				
Basic Cone	0.46	5.5	4.8	-0.48
Randomized $\eta\phi$	0.31	4.8	4.3	-0.34
High Energy Probe	0.6	5.5		

where $\langle N \rangle$ is the average constituent tower multiplicity for all random cones in events with a given centrality, computed with the basic random cones, and $\langle E_T^2 \rangle = \sigma^2(E_T) + \langle E_T \rangle$ is the simplified sum of the variance of single tower energy spectra and the average single tower energy. This expression is derived from the single tower E_T distribution, similar to Equation 1.12. We assume that the single-tower transverse energy spectra are similar in shape to that of the single particle p_T spectra. Equation 5.3 describes the lower limit for underlying event fluctuations in the absence of contributions due to long-range correlations from flow or short-range correlations from hard-scatterings within the collision. Similar to Equation 1.12, we estimate the contributions to the underlying event fluctuations by including a term that accounts for hydrodynamical flow. The non-Poissonian prediction for the width of δE_T is given by

$$\sigma(\delta E_T) = \sqrt{\langle N \rangle \cdot \langle E_T^2 \rangle + \langle N \rangle^2 \cdot \langle E_T \rangle^2 \cdot \sum 2v_n^2}, \quad (5.4)$$

where we use previously measured azimuthal anisotropy coefficients in our estimations for additional multiplicity fluctuations. The elliptical flow v_2 measured in Au+Au collisions at 200 GeV for 0-80% central events [16] and triangular flow v_3 was measured in Au+Au collisions at 200 GeV for 0-70% central events [128].

Figure 5.5 compares the centrality dependence of $\sigma(\delta E_T)$ for both types of random cones. The distribution of purely statistical fluctuations, as described by Equation 5.3, accurately represents the events with randomized tower $\eta\phi$ for all background subtraction methods.

Two non-Poissonian curves are predicted to quantify the contributions from higher order azimuthal anisotropies, according to Equation 5.4. The elliptical flow contribution is estimated by

$$\sigma_{\text{NP}}^2 \approx 2v_2^2 \langle N \rangle^2. \quad (5.5)$$

This approximate inclusion of v_2 effects qualitatively accounts for the increased fluctuations observed in mid-central collisions compared to randomized tower $\eta\phi$

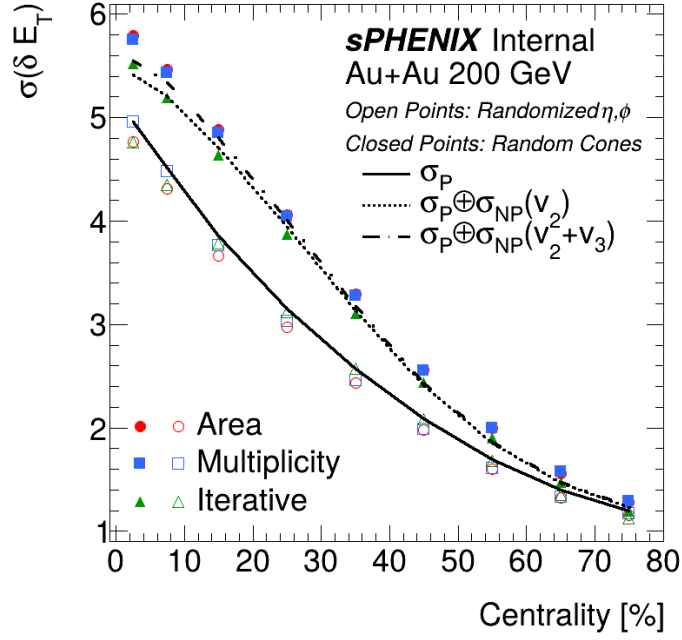


Figure 5.5: Centrality dependence of $\sigma(\delta E_T)$ of both types of random cones for all background subtraction methods, compared to the Poissonian limit calculated with measured $E_{T,Tower}$ and to additional hydrodynamical flow contributions calculated with elliptical flow v_2 measured in [16] and triangular flow measured in [128].

events. The triangular flow contribution is estimated by

$$\sigma_{\text{NP}}^2 \approx 2\langle N \rangle^2(v_2^2 + v_3^2). \quad (5.6)$$

The inclusion of v_3 effects has minimal effects on the predicted fluctuations for the increased fluctuations observed in central collisions compared to randomized tower $\eta\phi$ events. The contribution from hard-scatterings in events to the positive δE_T tail cannot be cleanly separated from multiplicity fluctuations caused by flow.

The discrepancies between the three background subtraction methods calculated in events with randomized tower $\eta\phi$ and the predicted limit due to purely statistical fluctuations seen in Figure 5.5 in central events is due to the fact that Equation 5.3 does not take into account any correlated changes to $\langle E_T \rangle$ due to long-range correlations.

The three background subtraction methods appear to perform similarly in both the basic random cones and for random cones reconstructed in events with randomized tower $\eta\phi$. All methods are within 0.5 GeV of each other for all event centrality classes for both the basic and randomized random cones. The iterative method is able to suppress long-range correlations due to flow better than the event averaged ρ based methods. To investigate this further, the ratio underlying event fluctuations for each method in both the basic random cones and random cones reconstructed in events with randomized tower $\eta\phi$ to the statistical limit calculated in Equation 5.3 was taken. Figure 5.6 shows these ratios for both types of random cones.

The random cones reconstructed in events with randomized tower $\eta\phi$ all are within 5% of unity to the distribution of purely statistical fluctuations, as described by Equation 5.3, showing that this description accurately represents the events with randomized tower $\eta\phi$ for all background subtraction methods. The largest discrepancies between the three background subtraction methods calculated in events with randomized tower $\eta\phi$ and the predicted limit due to purely statistical fluctuations occur at different event centrality for each background subtraction method. The

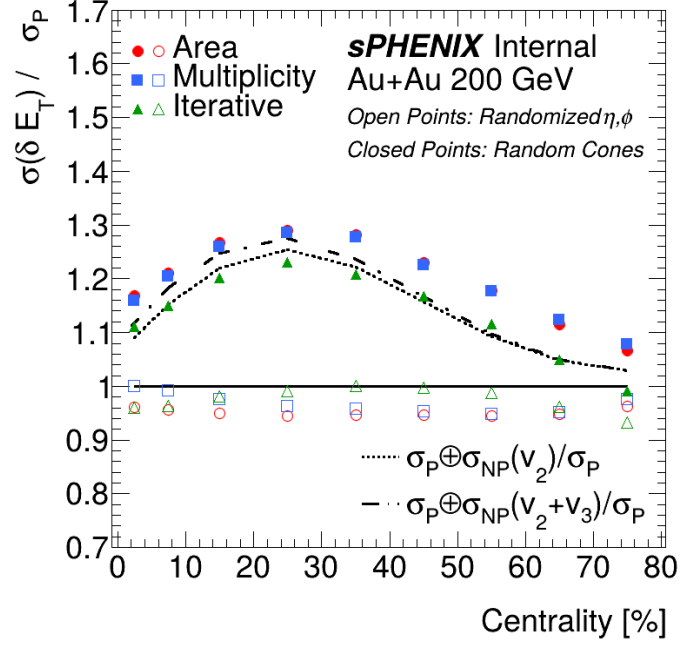


Figure 5.6: Fractional contributions to underlying event fluctuations $\sigma(\delta E_T)$ for random cones subtracted with each background subtraction methods. The Poissonian (statistical) limit is used as a scale.

multiplicity method for random cones in randomized tower $\eta\phi$ events is furthest from unity in semi-peripheral events. The area method follows a similar pattern but shifted down from the multiplicity method. The iterative method is furthest from the statistical description in both central and peripheral events, with the best agreement occurring in mid-central to mid-peripheral events.

The iterative method appears to have the best performance in terms of suppressing non-statistical fluctuations due to upward multiplicity fluctuations from flow. There is a constant 5% decrease in the width of underlying event fluctuations characterized by the iterative method to those characterized by the area and multiplicity methods, which is seen in all event centrality classes. This is due to the nature of the area and multiplicity methods, which estimate the average underlying event density within an event in a pseudorapidity window of $\Delta\eta = 2.2$. The iterative method estimates the average underlying event density in slices of pseudorapidity $\Delta\eta = 0.1$. This gives the iterative method a better description of the regional fluctuations in multiplicity caused by flow.

Chapter 6

Conclusions

This analysis presents the first measurements of jet background and the underlying event from the sPHENIX detector, as well as the first direct comparison between different jet background subtraction techniques within the same experimental setup and on the same data. The sources of underlying event fluctuations are systematically determined. Their contributions are characterized in Au+Au collisions at $\sqrt{s_{\text{NN}}} = 200$ GeV, allowing for initial comparisons to fluctuations measured in Pb+Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV. The fluctuations in the underlying event are found to be largely driven by multiplicity fluctuations. This correlation is used to develop a new jet background subtraction method based on jet constituent multiplicity [95]. The relationship between jet constituent multiplicity and jet background is found to be the primary cause for the increased performance achieved by neural network background subtraction methods.

Due to the intended impact of this analysis: to characterize and identify sources of underlying event fluctuations, motivate the future of jet reconstruction, and understand limits to unfolded jet measurements, we have chosen to report the values with only statistical uncertainties. The reconstructed energy of calorimeter towers will be affected by uncertainties in the electromagnetic scale calibration for both the hadronic and electromagnetic calorimeters. The tower zero-suppression and energy

resolution affect the total tower multiplicity and shape of the observed tower E_T spectrum at low E_T . The total effect on the width of underlying event fluctuations from calorimeter calibration uncertainties can be estimated to be 1.2-1.3% depending on event centrality, however, a complete correction for the finite energy resolution would typically be handled through the construction of a response matrix to account for detector effects including finite calorimeter resolution and sampling fraction. This study is intended to inform future measurements of jets, where subtraction of the underlying event would be one of many corrections. Since this is a correction applied to measurements and does not have a physical meaning independent of another measurement, we do not determine systematic uncertainties on it.

The fluctuations in the underlying event fluctuations are found to be primarily driven by multiplicity fluctuations due to stochastic fluctuations and hydrodynamic correlations in particle production. The contributions to the underlying event fluctuations from short-range correlations due to hard-processes were found to be significantly less than those seen in minimum bias Pb+Pb collisions at $\sqrt{s_{\text{NN}}} = 2.76$ TeV [12]. We have quantified the contribution of long-range correlations from flow, which constitute up to 30% of underlying event fluctuations in semi-peripheral collisions, a proportion that steadily decreases with increasing event centrality. These findings align with previous measurements from ALICE and ATLAS in Pb+Pb collisions $\sqrt{s_{\text{NN}}} = 2.76$ TeV, establishing a consistent picture of underlying event dynamics across collision systems and energies [12, 41, 3]. The dependence of underlying event fluctuations on event centrality is compared to predictions derived from the description of the underlying event presented in Ref. [115]. Our results are consistent with an underlying event described by the emission of particles from a single-particle momentum spectra (Eq. 1.9) with additional contributions due to hydrodynamic flow which are dependent on initial state geometry. The magnitude of the underlying event pedestal in Pb+Pb collisions is two times that measured in Au+Au but the relative dependence of underlying event fluctuations on the calorimeter window area is the same. Our results show that the spatial correlations of

the underlying event fluctuations are dependent on the regional fluctuations caused by odd v_n coefficients. Both long-range correlations due to flow, and short-range correlations due to hard processes were found to be correlated to calorimeter area size. The spatial correlation extracted from underlying event fluctuations was found to have a strong dependence on event centrality, indicating that this dependence is proportional to the odd-ordered flow harmonics, which contribute to regional fluctuations within an event.

We have demonstrated that the neural network techniques used by ALICE achieve their performance advantage primarily by exploiting the correlation between jet momenta and constituent multiplicity. This relationship is understood through descriptions of the underlying event and was used in the derivation of the multiplicity method [115, 95, 80]. Through the neural mapping procedure presented in section 3.2, we demonstrated that representations learned by a neural network trained to perform jet background subtraction were effective approximations of our physically motivated multiplicity method. Comparisons in performance of the multiplicity method against machine learning-based approaches, we have shown that this physically motivated technique achieves 92% of the improvement in underlying event fluctuation suppression achieved by a neural network while utilizing 8,250 times fewer parameters. This result underscores the necessity for a paradigm shift in the applications of machine learning in both heavy ion physics and general scientific data analysis: rather than prioritizing black-box performance, we must emphasize interpretable models that enhance our understanding of the underlying physics. Our method of neural network mapping provides the first example of using machine learning to better the understanding of the underlying physics relationships in the context of heavy ion collisions. The implications of these findings extend beyond sPHENIX. The development of the multiplicity method provides an optimal strategy for suppressing both stochastic and non-Poissonian multiplicity fluctuations in the underlying event, making it widely applicable to future jet measurements in heavy ion collisions.

The performance of three jet background subtraction methods was assessed by their abilities to suppress underlying event fluctuations [52, 95, 71]. The background subtraction methods investigated represent techniques used by several major experiments within the field, as well as the first application of the multiplicity method, discussed in section 3.1, to data. In our comparisons between these three jet background subtraction methods, we found that all were able to reliably estimate and subtract the underlying event pedestal on an event-by-event basis. The multiplicity fluctuations in the underlying event are best suppressed with background subtraction methods that take local fluctuations into account in their estimates of underlying event density. This accounts for the increase of 5% to random cone E_T resolution using the iterative background subtraction method. The conclusions of these comparisons are valid within the context of the sPHENIX detector using only calorimeter towers as input. In this context, the effectiveness of the multiplicity method is limited by the nearly complete occupancy in the sPHENIX central calorimeters, regardless of event centrality. The effect of this is that the area and multiplicity methods obtain roughly equal performance since there is little dynamic range available to the multiplicity method with calorimeter tower multiplicity. The effect of this is that the area and multiplicity methods obtain roughly equal performance. Our results suggest that incorporating tracking information would enhance the performance of the multiplicity method by at least 50% in central to semi-central collisions effectively extending the lower limit of jet measurements at sPHENIX by 5–10 GeV. The projected improvement corresponds to the suppression of nearly all contributions to the underlying event from flow, and a conservative estimate for the ability of the method to suppress stochastic fluctuations based on our findings in chapter 3.

A thorough understanding of the underlying event and its fluctuations is crucial for accurate jet measurements, particularly at low momentum, where background contributions are most significant. This study finds that the contributions of stochastic fluctuations and long-range correlations to the underlying event fluctuations are consistent with those measured by ALICE in Pb+Pb collisions at $\sqrt{s_{\text{NN}}}$

= 2.76 TeV. Future work should build on our findings regarding the correlation between jet momenta and constituent multiplicity to further refine jet background subtraction methods. Incorporating tracking information, such as particle flow, into the multiplicity method is expected to significantly improve jet energy resolution and extend jet measurements to lower transverse momenta. This advancement would provide more stringent constraints on partonic energy loss, particularly in the low-momentum regime, where medium-induced modifications to jets remain poorly understood.

Our comparisons of the performance of jet background subtraction methods using calorimeter tower inputs have established constraints on jet reconstruction at sPHENIX based on measured underlying event fluctuations, which is the dominant contribution to the jet energy resolution in heavy ion collisions. These results will shape the future of jet reconstruction and inform the unfolding of jet measurements at sPHENIX. More broadly, this work presents an effective approach for suppressing multiplicity fluctuations in the underlying event in the multiplicity method, which can be used to improve jet background subtraction in heavy ion collisions. Additionally, our findings underscore the importance of interpretability in machine learning applications for physics data analysis. Moving forward, the development of machine learning techniques should prioritize not only performance but also their ability to enhance our fundamental understanding of the underlying physics.

Bibliography

- [1] Aaboud, M. et al. (2018). Measurement of jet fragmentation in Pb+Pb and pp collisions at $\sqrt{s_{NN}} = 5.02$ TeV with the ATLAS detector. *Phys. Rev. C*, 98(2):024908. [xiv](#), [70](#), [72](#)
- [2] Aad, G. et al. (2012). Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC. *Phys. Lett. B*, 716:1–29. [2](#)
- [3] Aad, G. et al. (2013a). Measurement of the jet radius and transverse momentum dependence of inclusive jet suppression in lead-lead collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS detector. *Phys. Lett. B*, 719:220–241. [xiv](#), [37](#), [38](#), [133](#)
- [4] Aad, G. et al. (2013b). Measurement of the jet radius and transverse momentum dependence of inclusive jet suppression in lead-lead collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS detector. *Phys. Lett. B*, 719:220–241. [35](#)
- [5] Aad, G. et al. (2015). Measurements of the Nuclear Modification Factor for Jets in Pb+Pb Collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS Detector. *Phys. Rev. Lett.*, 114(7):072302. [24](#)
- [6] Aad, G. et al. (2016). Measurement of the charged-particle multiplicity inside jets from $\sqrt{s} = 8$ TeV pp collisions with the ATLAS detector. *Eur. Phys. J. C*, 76(6):322. [69](#)

- [7] Aad, G. et al. (2021). Measurements of azimuthal anisotropies of jet production in Pb+Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV with the ATLAS detector. ””. [24](#)
- [8] Aad, G. et al. (2023). Determination of the strong coupling constant from transverse energy–energy correlations in multijet events at $\sqrt{s} = 13$ TeV with the ATLAS detector. *JHEP*, 07:085. [xii](#), [7](#)
- [9] Aamodt, K. et al. (2011). Higher harmonic anisotropic flow measurements of charged particles in Pb-Pb collisions at $\sqrt{s_{NN}}=2.76$ TeV. *Phys. Rev. Lett.*, 107:032301. [16](#)
- [10] Aaron, F. D. et al. (2010). Combined Measurement and QCD Analysis of the Inclusive e+- p Scattering Cross Sections at HERA. *JHEP*, 01:109. [10](#)
- [11] Abadi, M. and et. al. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org. [74](#)
- [12] Abelev, B. et al. (2012). Measurement of Event Background Fluctuations for Charged Particle Jet Reconstruction in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *JHEP*, 03:053. [xiii](#), [33](#), [34](#), [36](#), [68](#), [92](#), [109](#), [110](#), [125](#), [133](#)
- [13] Abelev, B. et al. (2013). Centrality dependence of π , K, p production in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev.*, C88:044910. [68](#)
- [14] Abelev, B. et al. (2014). Measurement of charged jet suppression in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *JHEP*, 03:013. [33](#), [81](#)
- [15] Abelev, B. B. et al. (2015). Charged jet cross sections and properties in proton-proton collisions at $\sqrt{s} = 7$ TeV. *Phys. Rev. D*, 91(11):112012. [65](#), [107](#)
- [16] Abelev, B. I. et al. (2008). Centrality dependence of charged hadron and strange hadron elliptic flow from s(NN)**(1/2) = 200-GeV Au + Au collisions. *Phys. Rev. C*, 77:054901. [xviii](#), [127](#), [128](#)

- [17] Acharya, S. et al. (2020a). Measurements of inclusive jet spectra in pp and central Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV. *Phys. Rev. C*, 101(3):034911. [xiii](#), [24](#), [25](#), [33](#)
- [18] Acharya, S. et al. (2024). Measurement of the radius dependence of charged-particle jet suppression in Pb-Pb collisions at $\sqrt{s_{NN}}=5.02$ TeV. *Phys. Lett. B*, 849:138412. [xi](#), [xiii](#), [27](#), [30](#), [32](#), [67](#), [69](#), [71](#), [81](#)
- [19] Acharya, U. et al. (2020b). Measurement of jet-medium interactions via direct photon-hadron correlations in Au+Au and $d + Au$ collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. C*, 102(5):054910. [xiv](#), [70](#), [73](#)
- [20] Achenbach, P. et al. (2023). The Present and Future of QCD. In ””. [89](#)
- [21] Adam, J. et al. (2015). Measurement of jet suppression in central Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Lett. B*, 746:1–14. [24](#), [33](#)
- [22] Adam, J. et al. (2016). Higher harmonic flow coefficients of identified hadrons in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *JHEP*, 09:164. [19](#), [68](#)
- [23] Adam, J. et al. (2020). Measurement of inclusive charged-particle jet production in Au + Au collisions at $\sqrt{s_{NN}}=200$ GeV. *Phys. Rev. C*, 102(5):054913. [33](#)
- [24] Adamczyk, L. et al. (2014). Jet-Hadron Correlations in $\sqrt{s_{NN}} = 200$ GeV $p + p$ and Central Au + Au Collisions. *Phys. Rev. Lett.*, 112(12):122301. [68](#)
- [25] Adare, A. et al. (2013). Spectra and ratios of identified particles in Au+Au and $d+Au$ collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. C*, 88(2):024906. [68](#)
- [26] Adare, A. and et. al. (2014). Centrality categorization for $R_{p(d)+A}$ in high-energy collisions. *Phys. Rev. C*, 90:034902. [93](#), [95](#)
- [27] Adare, A. et al. (2015). An Upgrade Proposal from the PHENIX Collaboration. . [xiii](#), [27](#), [28](#), [29](#), [43](#)

- [28] Adare, A. et al. (2016). Measurement of the higher-order anisotropic flow coefficients for identified hadrons in Au+Au collisions at $\sqrt{s_{NN}} = 200$ GeV. *Phys. Rev. C*, 93(5):051902. [68](#)
- [29] Adler, S. S. et al. (2005). Systematic studies of the centrality and $s(NN)^{1/2}$ dependence of the $d E(T) / d \eta$ and $d (N(ch) / d \eta$ in heavy ion collisions at mid-rapidity. *Phys. Rev. C*, 71:034908. [Erratum: *Phys.Rev.C* 71, 049901 (2005)]. [93](#)
- [30] Afanasiev, S. et al. (2009). Systematic Studies of Elliptic Flow Measurements in Au+Au Collisions at $s^{1/2} = 200$ -GeV. *Phys. Rev. C*, 80:024909. [xiii](#), [20](#)
- [31] Aglieri Rinella, G. (2017). The ALPIDE pixel sensor chip for the upgrade of the ALICE Inner Tracking System. *Nucl. Instrum. Meth. A*, 845:583–587. [46](#)
- [32] Aidala, C. et al. (2019). Creation of quark–gluon plasma droplets with three distinct geometries. *Nature Phys.*, 15(3):214–220. [14](#), [16](#)
- [33] Aidala, C. A. and Et., a. (2018). Design and beam test results for the sphenix electromagnetic and hadronic calorimeter prototypes. *IEEE Transactions on Nuclear Science*, 65(12):2901–2919. [45](#)
- [34] Aidala, C. A. and Et., a. (2021). Design and beam test results for the 2-d projective sphenix electromagnetic calorimeter prototype. *IEEE Transactions on Nuclear Science*, 68(2):173–181. [45](#)
- [35] Aidala, C. A. et al. (2018). Design and Beam Test Results for the sPHENIX Electromagnetic and Hadronic Calorimeter Prototypes. *IEEE Trans. Nucl. Sci.*, 65(12):2901–2919. [xiv](#), [55](#), [58](#)
- [36] Akers, R. et al. (1995). A Model independent measurement of quark and gluon jet properties and differences. *Z. Phys. C*, 68:179–202. [69](#)

- [37] Alessi, J. and Et., a. (2007). High performance EBS for RHIC. In *2007 IEEE Particle Accelerator Conference (PAC)*, pages 3782–3785. [41](#)
- [38] Alessi, J. e. (2005). Booster Design Manual. Technical report, Brookhaven National Laboratory, Upton, NY. [41](#)
- [39] Allen, M. et al. (2003). PHENIX inner detectors. *Nucl. Instrum. Meth.*, A499:549–559. [57](#)
- [40] Atkin, R. (2015). Review of jet reconstruction algorithms. *Journal of Physics: Conference Series*, 645(1):012008. [21](#)
- [41] ATLAS Collaboration (2012). Transverse energy fluctuations in Pb+Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS detector at the LHC. ATLAS-CONF-2012-045. [37](#), [92](#), [109](#), [133](#)
- [42] Aune, S. and et. al. (2024). The sphenix micromegas outer tracker. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1066:169615. [45](#)
- [43] Aune, S. et al. (2024). The sPHENIX Micromegas Outer Tracker. *Nucl. Instrum. Meth. A*, 1066:169615. [xiv](#), [52](#), [54](#), [63](#)
- [44] Bartel, W. et al. (1980). Observation of Planar Three Jet Events in $e^+ e^-$ Annihilation and Evidence for Gluon Bremsstrahlung. *Phys. Lett. B*, 91:142–147. [4](#)
- [45] Berger, C. et al. (1979). Evidence for Gluon Bremsstrahlung in $e^+ e^-$ Annihilations at High-Energies. *Phys. Lett. B*, 86:418–425. [4](#)
- [46] Bjorken, J. D. (1983). Highly Relativistic Nucleus-Nucleus Collisions: The Central Rapidity Region. *Phys. Rev. D*, 27:140–151. [14](#)
- [47] Bloom, E. D. et al. (1969). High-Energy Inelastic $e p$ Scattering at 6-Degrees and 10-Degrees. *Phys. Rev. Lett.*, 23:930–934. [4](#)

- [48] Breidenbach, M., Friedman, J. I., Kendall, H. W., Bloom, E. D., Coward, D. H., DeStaebler, H. C., Drees, J., Mo, L. W., and Taylor, R. E. (1969). Observed behavior of highly inelastic electron-proton scattering. *Phys. Rev. Lett.*, 23:935–939. [4](#)
- [49] Brenner, L., Balasubramanian, R., Burgard, C., Verkerke, W., Cowan, G., Verschuuren, P., and Croft, V. (2020). Comparison of unfolding methods using RooFitUnfold. *Int. J. Mod. Phys. A*, 35(24):2050145. [81](#)
- [50] Busza, W., Rajagopal, K., and van der Schee, W. (2018). Heavy Ion Collisions: The Big Picture, and the Big Questions. *Ann. Rev. Nucl. Part. Sci.*, 68:339–376. [1](#)
- [51] Cacciari, M., Salam, G. P., and Sapeta, S. (2010). On the characterisation of the underlying event. *JHEP*, 04:065. [64](#)
- [52] Cacciari, M., Salam, G. P., and Soyez, G. (2008). The Catchment Area of Jets. *JHEP*, 04:005. [xiii](#), [21](#), [22](#), [30](#), [33](#), [93](#), [104](#), [135](#)
- [53] Cacciari, M., Salam, G. P., and Soyez, G. (2012). FastJet User Manual. *Eur. Phys. J. C*, 72:1896. [21](#), [33](#), [68](#), [104](#)
- [54] Chatrchyan, S. et al. (2012). Measurement of the Pseudorapidity and Centrality Dependence of the Transverse Energy Density in PbPb Collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. Lett.*, 109:152303. [14](#)
- [55] Chatrchyan, S. et al. (2014). Measurement of Jet Fragmentation in PbPb and pp Collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. C*, 90(2):024908. [70](#)
- [56] Commons, W. (2020). File:quark confinement.svg — wikimedia commons, the free media repository. [Online; accessed 6-Jan-2025]. [xii](#), [9](#)

- [57] Commons, W. (2024). File:standard model of elementary particles.svg — wikimedia commons, the free media repository. [Online; accessed 5-December-2024]. [xii](#), [3](#)
- [58] Connors, M., Nattrass, C., Reed, R., and Salur, S. (2017). Review of Jet Measurements in Heavy Ion Collisions. *arXiv*. [xii](#), [xiii](#), [11](#), [16](#), [17](#), [18](#), [24](#), [26](#)
- [59] Cranmer, M. (2020). Pysr: Fast & parallelized symbolic regression in python/julia. [85](#)
- [60] Cranmer, M., Sanchez-Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., and Ho, S. (2020). Discovering symbolic models from deep learning with inductive biases. *NeurIPS 2020*. [85](#)
- [61] D’Agostini, G. (1995). A multidimensional unfolding method based on bayes’ theorem. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 362(2):487–498. [81](#)
- [62] Demir, N. and Bass, S. A. (2009). Shear-Viscosity to Entropy-Density Ratio of a Relativistic Hadron Gas. *Phys. Rev. Lett.*, 102:172302. [16](#)
- [63] Deur, A., Brodsky, S. J., and de Teramond, G. F. (2016). The QCD Running Coupling. *Nucl. Phys.*, 90:1. [5](#), [8](#)
- [64] Du, L., Sorensen, A., and Stephanov, M. (2024). The qcd phase diagram and beam energy scan physics: A theory overview. *International Journal of Modern Physics E*. [xii](#), [12](#)
- [65] Fodor, Z. and Hoelbling, C. (2012). Light Hadron Masses from Lattice QCD. *Rev. Mod. Phys.*, 84:449. [10](#)
- [66] Goodfellow, I. J., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, MA, USA. <http://www.deeplearningbook.org>. [74](#), [78](#)

- [67] Gross, D. J. and Wilczek, F. (1973). Ultraviolet Behavior of Nonabelian Gauge Theories. *Phys. Rev. Lett.*, 30:1343–1346. [8](#)
- [68] Guenther, J. N. (2021). Overview of the QCD phase diagram: Recent progress from the lattice. *Eur. Phys. J. A*, 57(4):136. [11](#)
- [69] Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *J. Mach. Learn. Res.*, 3(null):1157–1182. [xv](#), [76](#), [77](#)
- [70] Haake, R. and Loizides, C. (2019). Machine Learning based jet momentum reconstruction in heavy-ion collisions. *Phys. Rev. C*, 99(6):064904. [30](#), [74](#), [75](#), [81](#)
- [71] Hanks, J. A. and Et., a. (2012). Method for separating jets and the underlying event in heavy ion collisions at the bnl relativistic heavy ion collider. *Phys. Rev. C*, 86:024908. [30](#), [35](#), [93](#), [103](#), [135](#)
- [72] Hanks, J. A. et al. (2012). Method for separating jets and the underlying event in heavy ion collisions at the BNL Relativistic Heavy Ion Collider. *Phys. Rev. C*, 86:024908. [35](#)
- [73] Harrison, M., Ludlam, T., and Ozaki, S. (2003). RHIC project overview. *Nucl. Instrum. Meth. A*, 499:235–244. [39](#)
- [74] Heinz, U. (2004). Thermalization at rhic. *AIP Conference Proceedings*, 739(1):163–180. [16](#)
- [75] Heinz, U. W. and Jacob, M. (2000). Evidence for a new state of matter: An Assessment of the results from the CERN lead beam program. . [1](#), [16](#)
- [76] Heinz, U. W. and Kestin, G. (2008). Jozso’s legacy: Chemical and kinetic freeze-out in heavy-ion collisions. *Eur. Phys. J. ST*, 155:75–87. [16](#)
- [77] Hirano, T., Heinz, U. W., Kharzeev, D., Lacey, R., and Nara, Y. (2006). Hadronic dissipative effects on elliptic flow in ultrarelativistic heavy-ion collisions. *Phys. Lett. B*, 636:299–304. [16](#)

- [78] Hirano, T. and Tsuda, K. (2002). Collective flow and two pion correlations from a relativistic hydrodynamic model with early chemical freezeout. *Phys. Rev. C*, 66:054905. [16](#)
- [79] Hughes, C. and Mengel, T. (2019). Github: TennGen. [68](#), [74](#)
- [80] Hughes, C., Oliveira Da Silva, A. C., and Nattrass, C. (2022). Model studies of fluctuations in the background for jets in heavy ion collisions. *Phys. Rev. C*, 106(4):044915. [33](#), [35](#), [68](#), [74](#), [83](#), [125](#), [134](#)
- [81] Huovinen, P., Kolb, P. F., Heinz, U. W., Ruuskanen, P. V., and Voloshin, S. A. (2001). Radial and elliptic flow at RHIC: Further predictions. *Phys. Lett. B*, 503:58–64. [16](#)
- [82] Huth, J. E. et al. (1990). Toward a standardization of jet definitions. In *1990 DPF Summer Study on High-energy Physics: Research Directions for the Decade (Snowmass 90)*, pages 0134–136. [21](#)
- [83] Ikematsu, K. et al. (1998). A Start - timing detector for the collider experiment PHENIX at RHIC-BNL. *Nucl. Instrum. Meth.*, A411:238–248. [57](#)
- [84] Khachatryan, V. et al. (2017). Measurement of inclusive jet cross sections in pp and PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV. *Phys. Rev. C*, 96(1):015202. [24](#), [35](#)
- [85] Klest, H. (2020). Overview and design of the sPHENIX TPC. *J. Phys. Conf. Ser.*, 1498:012025. [xiv](#), [49](#), [56](#)
- [86] Kraskov, A., Stögbauer, H., and Grassberger, P. (2004). Estimating mutual information. *Phys. Rev. E*, 69:066138. [76](#)
- [87] Larkoski, A. J. and Metodiev, E. M. (2019). A Theory of Quark vs. Gluon Discrimination. *JHEP*, 10:014. [27](#)
- [88] Li, H., Kadav, A., Durdanovic, I., Samet, H., and Graf, H. P. (2017). Pruning filters for efficient convnets. [76](#)

- [89] Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57. [89](#)
- [90] Loizides, C., Nagle, J., and Steinberg, P. (2015). Improved version of the PHOBOS Glauber Monte Carlo. *SoftwareX*, 1-2:13–18. [95](#)
- [91] Lopes, A. (2020). Recreating big bang matter on earth. [Online; accessed 5-January-2025]. [xii](#), [11](#), [13](#)
- [92] Majumder, A. (2007). A Comparative study of Jet-quenching Schemes. *J. Phys. G*, 34:S377–388. [10](#)
- [93] Majumder, A. and van Leeuwen, M. (2011). The theory and phenomenology of perturbative qcd based jet quenching. *Progress in Particle and Nuclear Physics*, 66(1):41–92. [10](#)
- [94] Mengel, T. and Hughes, C. (2022). Github: tenngen200. [68](#), [74](#)
- [95] Mengel, T., Steffanic, P., Hughes, C., da Silva, A. C. O., and Nattrass, C. (2023). Interpretable machine learning methods applied to jet background subtraction in heavy-ion collisions. *Phys. Rev. C*, 108(2):L021901. [21](#), [30](#), [64](#), [81](#), [89](#), [93](#), [107](#), [125](#), [132](#), [134](#), [135](#)
- [96] Miller, M. L., Reygers, K., Sanders, S. J., and Steinberg, P. (2007). Glauber modeling in high energy nuclear collisions. *Ann. Rev. Nucl. Part. Sci.*, 57:205–243. [xii](#), [14](#), [15](#)
- [97] Narison, S. (2005). *QCD as a Theory of Hadrons : From Partons to Confinement*, volume 17. Oxford University Press. [5](#)
- [98] Osborn, J. D. (2020). Requirements, Status, and Plans for Track Reconstruction at the sPHENIX Experiment. In . [61](#)

- [99] Osborn, J. D. and Et., a. (2021). Implementation of ACTS into sPHENIX Track Reconstruction. *Comput. Softw. Big Sci.*, 5(1):23. [61](#)
- [100] Pearson, K. (1895). VII. Note on regression and inheritance in the case of two parents. *Proceedings of the Royal Society of London*, 58(347-352):240–242. Publisher: Royal Society. [76](#)
- [101] Poskanzer, A. M. and Voloshin, S. A. (1998). Methods for analyzing anisotropic flow in relativistic nuclear collisions. *Phys. Rev. C*, 58:1671–1678. [19](#)
- [102] Putschke, J. et al. (2019). The JETSCAPE framework. *JETSCAPE*. [10](#)
- [103] Rafelski, J. (2015). Melting Hadrons, Boiling Quarks. *Eur. Phys. J. A*, 51(9):114. [1](#)
- [104] Schwartz, M. D. (2014). *Quantum Field Theory and the Standard Model*. Cambridge University Press. [19](#)
- [105] Shuryak, E. (2024). *Quark-Gluon Plasma, Heavy Ion Collisions and Hadrons*. World Scientific. [1](#)
- [106] Sirunyan, A. M. et al. (2019). Centrality and pseudorapidity dependence of the transverse energy density in pPb collisions at $\sqrt{s_{\text{NN}}} = 5.02$ TeV. *Phys. Rev. C*, 100(2):024902. [14](#)
- [107] Sjostrand, T., Mrenna, S., and Skands, P. Z. (2008). A Brief Introduction to PYTHIA 8.1. *Comput. Phys. Commun.*, 178:852–867. [67](#)
- [108] Skands, P., Carrazza, S., and Rojo, J. (2014). Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J. C*, 74(8):3024. [68](#)
- [109] Soyez, G. (2009). Jet areas as a tool for background subtraction. In *25th Winter Workshop on Nuclear Dynamics*. [33](#)

- [110] sPHENIX collaboration (2018). A Monolithic Active Pixel Sensor Detector for the sPHENIX Experiment. ””. [xi](#), [xiv](#), [46](#), [47](#), [48](#)
- [111] sPHENIX collaboration (2019). sPHENIX Technical Design Report. . [xiv](#), [43](#), [46](#), [49](#), [50](#), [51](#), [52](#), [53](#), [55](#), [57](#), [59](#)
- [112] sPHENIX Collaboration (2023a). Sphenix renderings. [xiv](#), [44](#)
- [113] sPHENIX Collaboration (2023b). ssphenix event displays. [Online; accessed 5-December-2024]. [xvi](#), [xvii](#), [98](#), [101](#)
- [114] sPHENIX Collaboration (2023c). ssphenix performance plots. [Online; accessed 5-December-2024]. [xvi](#), [xvii](#), [94](#), [96](#), [99](#)
- [115] Tannenbaum, M. (2001). The distribution function of the event-by-event average pt for statistically independent emission. *Physics Letters B*, 498(1):29 – 34. [33](#), [35](#), [109](#), [125](#), [133](#), [134](#)
- [116] Tawfik, A. (2009). Cosmological Consequences of QCD Phase Transition(s) in Early Universe. *AIP Conf. Proc.*, 1115(1):239–247. [14](#)
- [117] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288. [78](#)
- [118] Van Hove, L. (1986). THEORETICAL PREDICTION OF A NEW STATE OF MATTER, THE ‘QUARK - GLUON PLASMA’ (ALSO CALLED ‘QUARK MATTER’). In *17th International Symposium on Multiparticle Dynamics*, pages 801–818. [1](#)
- [119] Vergara, J. R. and Estévez, P. A. (2013). A review of feature selection methods based on mutual information. *Neural Computing and Applications*, 24(1):175–186. [75](#)

- [120] Vogt, R. (2007). Chapter 1 - kinematics and invariants. In Vogt, R., editor, *Ultrarelativistic Heavy-Ion Collisions*, pages 3–24. Elsevier Science B.V., Amsterdam. [41](#)
- [121] Voloshin, S. and Zhang, Y. (1996). Flow study in relativistic nuclear collisions by Fourier expansion of Azimuthal particle distributions. *Z. Phys. C*, 70:665–672. [19](#)
- [122] Voloshin, S. A., Poskanzer, A. M., and Snellings, R. (2010). Collective phenomena in non-central nuclear collisions. *Landolt-Bornstein*, 23:293–333. [30](#)
- [123] Walet, N. (2010). Nuclear and particle physics the feynman diagrams of qcd. Accessed on January 11, 2025. [xii](#), [6](#)
- [124] Walsh, K. M. (2020). Rhic gets ready to smash gold ions for run 23. <https://science.osti.gov/np/Facilities/User-Facilities/RHIC>. Accessed: 2024-08-26. [xiv](#), [42](#)
- [125] Walsh, K. M. (2023). Rhic gets ready to smash gold ions for run 23. <https://www.bnl.gov/newsroom/news.php?a=221190>. Accessed: 2024-08-26. [xiv](#), [40](#)
- [126] Weinberg, S. (1967). A model of leptons. *Phys. Rev. Lett.*, 19:1264–1266. [4](#)
- [127] Workman, R. L. et al. (2022). Review of Particle Physics. *PTEP*, 2022:083C01. [5](#)
- [128] Yi, L. (2014). Azimuthal Anisotropy Measurements by STAR. *Nucl. Phys. A*, 926:198–204. [xviii](#), [127](#), [128](#)
- [129] Yu, L. and Liu, H. (2003). Feature selection for high-dimensional data: A fast correlation-based filter solution. In ””, volume 2, pages 856–863. [76](#)

Appendix

A Jet Background Subtraction

This appendix provides relevant figures for each of the three jet background subtraction methods used in the main analysis, but not included in the thesis body. The full δE_T distributions are included for each method as well as fits. Figures [1,2](#), and [3](#) show the δE_T distributions for basic random cones, random cones reconstructed with randomized tower η, ϕ , and high energy embed probes, each subtracted with the area (red), multiplicity (blue), and iterative methods (green). Sections [A.2](#), [A.3](#), and [A.4](#) include distributions with a Gaussian fit to the left-hand-side and its extrapolation to positive δE_t . The solid black line is a Γ – distribution from Equation. [5.2](#). Section [A.5](#) shows the standard deviate and mean for δE_T distributions and parameters for the Gaussian fits.

A.1 Energy Residuals

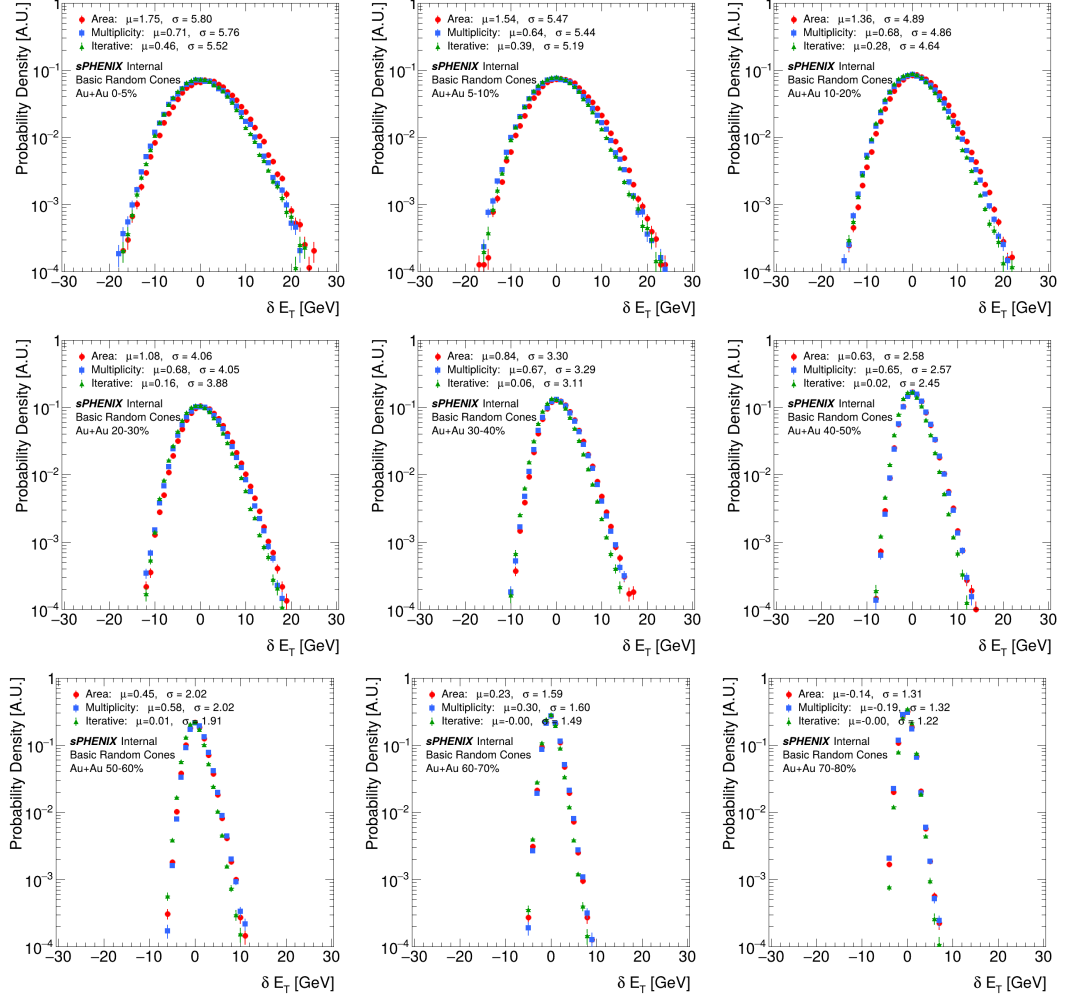


Figure 1: δE_T for standard rigid cones with $R = 0.4$ from each background subtraction method for all event centrality classes.

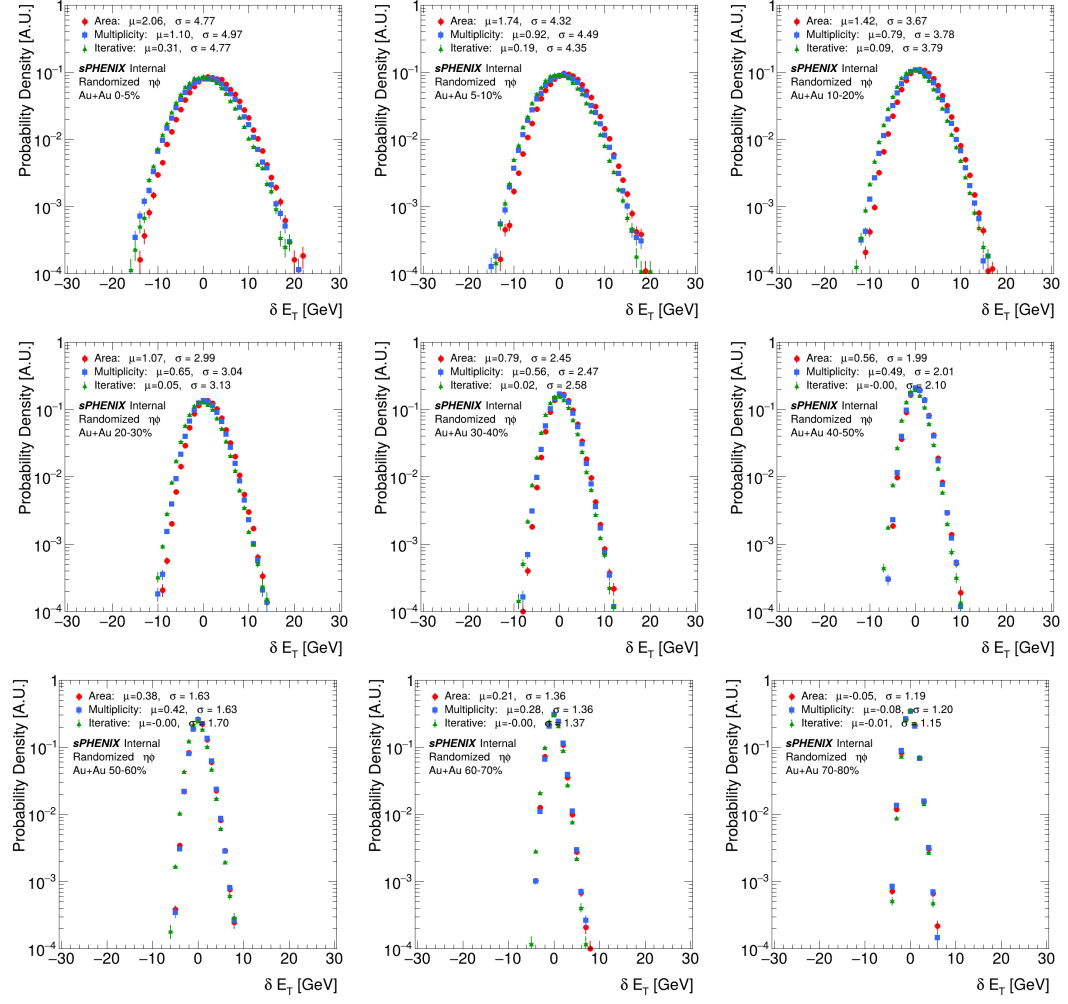


Figure 2: δE_T for rigid cones with randomized tower η, ϕ with $R = 0.4$ from each background subtraction method for all event centrality classes.

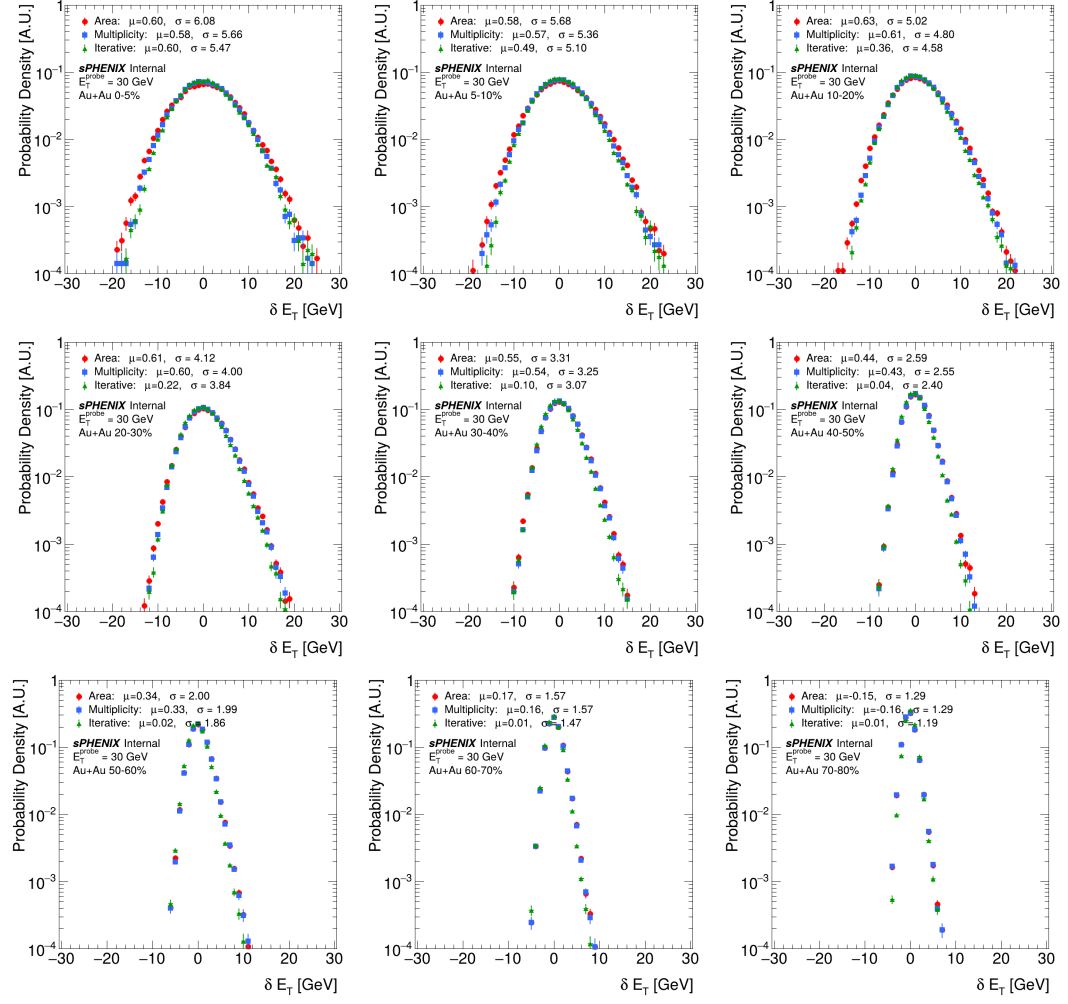


Figure 3: δE_T for embedded high energy probe ($E_T^{probe} = 30$ GeV η, ϕ with $R = 0.4$ from each background subtraction method for all event centrality classes.

A.2 Multiplicity Method

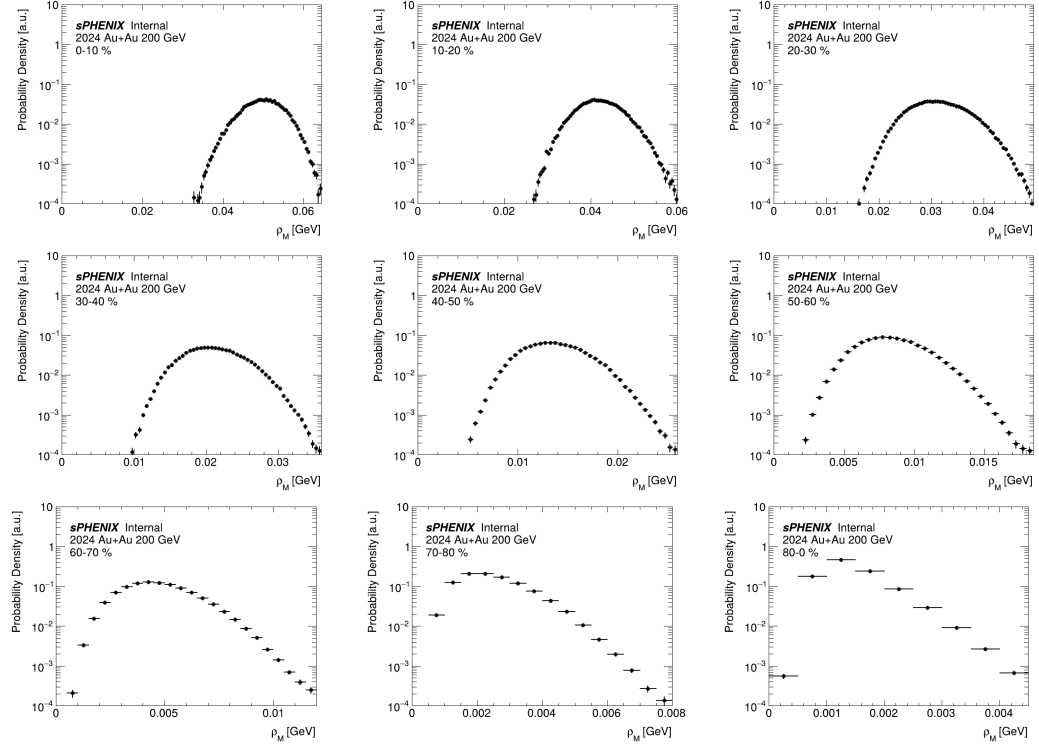


Figure 4: Probability distributions of ρ_M for each centrality bin.

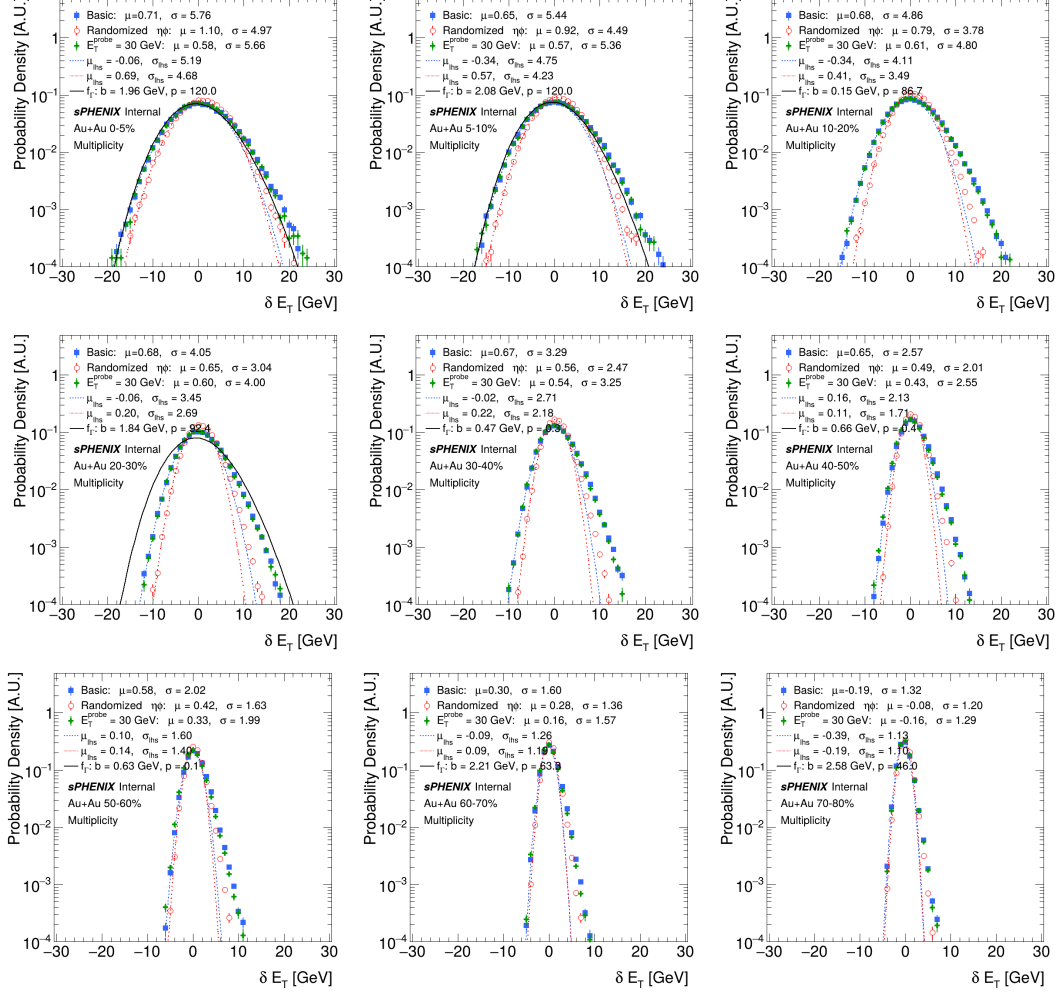


Figure 5: δE_T for standard $R = 0.4$ random cones, $R = 0.4$ cones with randomized tower η, ϕ , and embedded high energy probe ($E_T^{probe} = 30$ GeV), fit with a Gaussian on the left hand side and a Γ distribution globally. All underlying event characterizations are subtracted with the multiplicity method.

A.3 Area Method

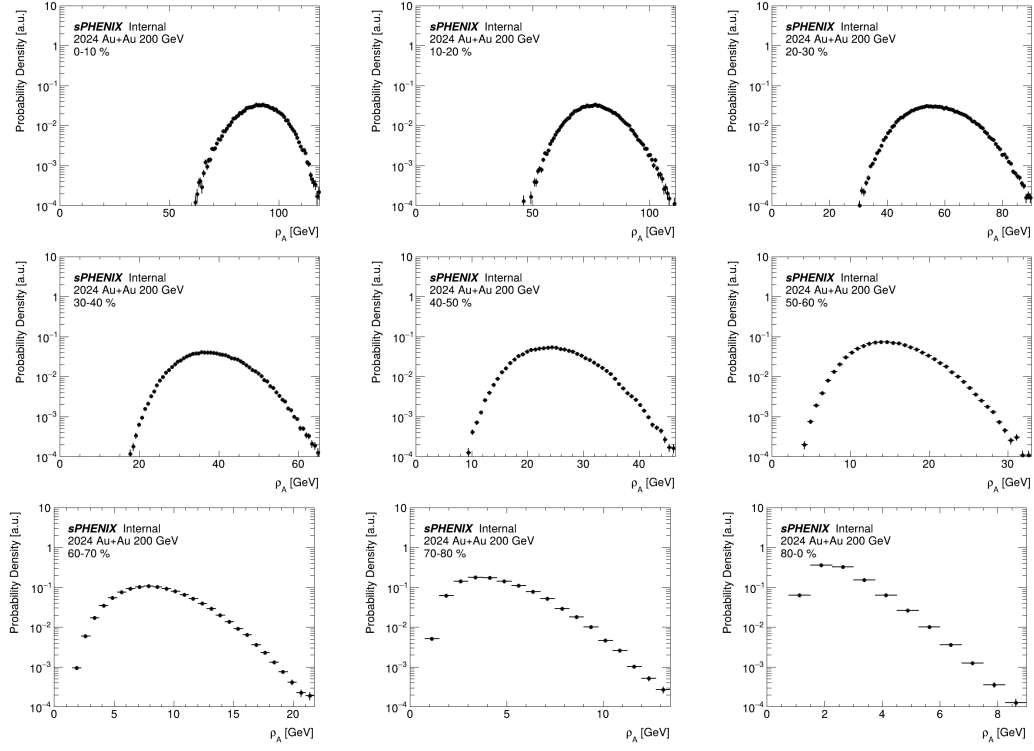


Figure 6: Probability distributions of ρ_A for each centrality bin.

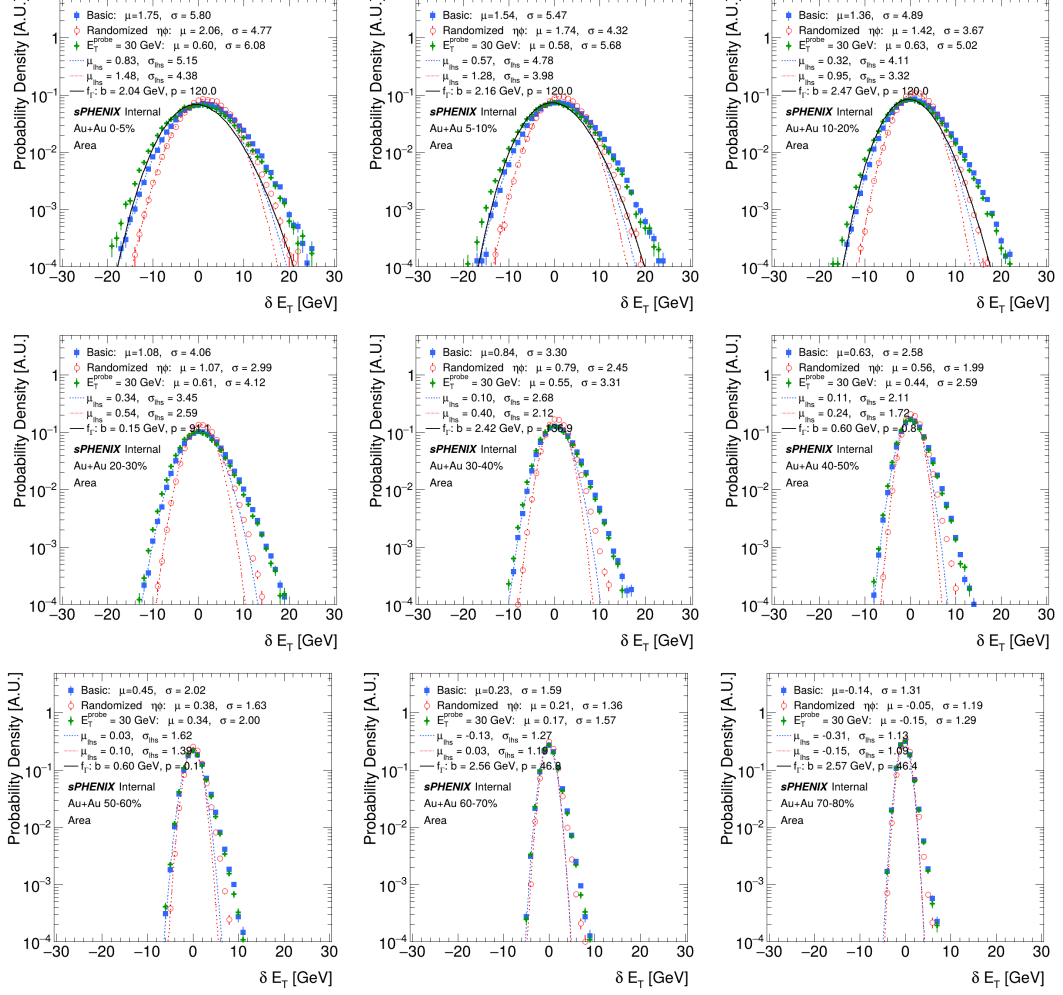


Figure 7: δE_T for standard $R = 0.4$ random cones, $R = 0.4$ cones with randomized tower η, ϕ , and embedded high energy probe ($E_T^{probe} = 30$ GeV), fit with a Gaussian on the left hand side and a Γ distribution globally. All underlying event characterizations are subtracted with the area method.

A.4 Iterative Method

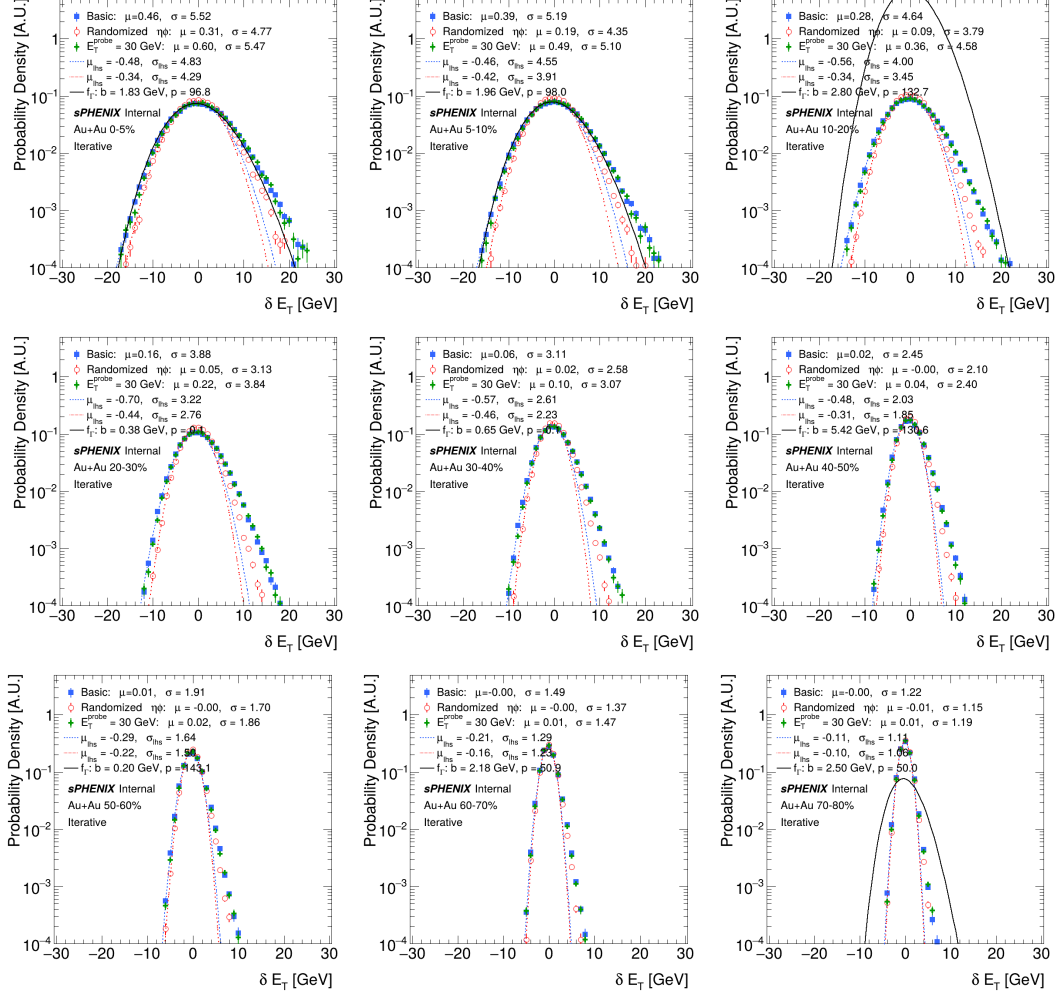


Figure 8: δE_T for standard $R = 0.4$ random cones, $R = 0.4$ cones with randomized tower η, ϕ , and embedded high energy probe ($E_T^{probe} = 30$ GeV), fit with a Gaussian on the left hand side and a Γ distribution globally. All underlying event characterizations are subtracted with the iterative method.

A.5 Residual Fit Parameters

Table 1: Standard deviations and averages of δE_T distributions and Gaussian fits for the iterative method. All values are in units of GeV.

	μ	σ	$\sigma^{l.h.s.}$	$\mu^{l.h.s.}$
Centrality: 0-5%				
Basic Cone	0.46	5.5	4.8	-0.48
Randomized $\eta\phi$	0.31	4.8	4.3	-0.34
High Energy Probe	0.6	5.5		
Centrality: 5-10%				
Basic Cone	0.39	5.2	4.5	-0.46
Randomized $\eta\phi$	0.19	4.4	3.9	-0.42
High Energy Probe	0.49	5.1		
Centrality: 10-20%				
Basic Cone	0.28	4.6	4	-0.56
Randomized $\eta\phi$	0.09	3.8	3.5	-0.34
High Energy Probe	0.36	4.6		
Centrality: 20-30%				
Basic Cone	0.16	3.9	3.2	-0.7
Randomized $\eta\phi$	0.05	3.1	2.8	-0.44
High Energy Probe	0.22	3.8		
Centrality: 30-40%				
Basic Cone	0.058	3.1	2.6	-0.57
Randomized $\eta\phi$	0.016	2.6	2.2	-0.46
High Energy Probe	0.1	3.1		
Centrality: 40-50%				
Basic Cone	0.016	2.4	2	-0.48
Randomized $\eta\phi$	-0.0022	2.1	1.8	-0.31
High Energy Probe	0.043	2.4		
Centrality: 50-60%				
Basic Cone	0.0067	1.9	1.6	-0.29
Randomized $\eta\phi$	-5.5e-05	1.7	1.5	-0.22
High Energy Probe	0.016	1.9		
Centrality: 60-70%				
Basic Cone	-0.00055	1.5	1.3	-0.21
Randomized $\eta\phi$	-0.0019	1.4	1.2	-0.16
High Energy Probe	0.013	1.5		
Centrality: 70-80%				
Basic Cone	-0.0033	1.2	1.1	-0.11
Randomized $\eta\phi$	-0.0067	1.2	1.1	-0.1
High Energy Probe	0.0058	1.2		

Table 2: Standard deviations and averages of δE_T distributions and Gaussian fits for the area method. All values are in units of GeV.

	μ	σ	$\sigma^{l.h.s.}$	$\mu^{l.h.s.}$
Centrality: 0-5%				
Basic Cone	1.8	5.8	5.2	0.83
Randomized $\eta\phi$	2.1	4.8	4.4	1.5
High Energy Probe	0.6	6.1		
Centrality: 5-10%				
Basic Cone	1.5	5.5	4.8	0.57
Randomized $\eta\phi$	1.7	4.3	4	1.3
High Energy Probe	0.58	5.7		
Centrality: 10-20%				
Basic Cone	1.4	4.9	4.1	0.32
Randomized $\eta\phi$	1.4	3.7	3.3	0.95
High Energy Probe	0.63	5		
Centrality: 20-30%				
Basic Cone	1.1	4.1	3.4	0.34
Randomized $\eta\phi$	1.1	3	2.6	0.54
High Energy Probe	0.61	4.1		
Centrality: 30-40%				
Basic Cone	0.84	3.3	2.7	0.1
Randomized $\eta\phi$	0.79	2.4	2.1	0.4
High Energy Probe	0.55	3.3		
Centrality: 40-50%				
Basic Cone	0.63	2.6	2.1	0.11
Randomized $\eta\phi$	0.56	2	1.7	0.24
High Energy Probe	0.44	2.6		
Centrality: 50-60%				
Basic Cone	0.45	2	1.6	0.032
Randomized $\eta\phi$	0.38	1.6	1.4	0.097
High Energy Probe	0.34	2		
Centrality: 60-70%				
Basic Cone	0.23	1.6	1.3	-0.13
Randomized $\eta\phi$	0.21	1.4	1.2	0.035
High Energy Probe	0.17	1.6		
Centrality: 70-80%				
Basic Cone	-0.14	1.3	1.1	-0.31
Randomized $\eta\phi$	-0.054	1.2	1.1	-0.15
High Energy Probe	-0.15	1.3		

Table 3: Standard deviations and averages of δE_T distributions and Gaussian fits for the multiplicity method. All values are in units of GeV.

	μ	σ	$\sigma^{l.h.s.}$	$\mu^{l.h.s.}$
Centrality: 0-5%				
Basic Cone	0.71	5.8	5.2	-0.057
Randomized $\eta\phi$	1.1	5	4.7	0.69
High Energy Probe	0.58	5.7		
Centrality: 5-10%				
Basic Cone	0.65	5.4	4.8	-0.34
Randomized $\eta\phi$	0.92	4.5	4.2	0.57
High Energy Probe	0.57	5.4		
Centrality: 10-20%				
Basic Cone	0.68	4.9	4.1	-0.34
Randomized $\eta\phi$	0.79	3.8	3.5	0.41
High Energy Probe	0.61	4.8		
Centrality: 20-30%				
Basic Cone	0.68	4.1	3.4	-0.059
Randomized $\eta\phi$	0.65	3	2.7	0.2
High Energy Probe	0.6	4		
Centrality: 30-40%				
Basic Cone	0.67	3.3	2.7	-0.016
Randomized $\eta\phi$	0.56	2.5	2.2	0.22
High Energy Probe	0.54	3.2		
Centrality: 40-50%				
Basic Cone	0.65	2.6	2.1	0.16
Randomized $\eta\phi$	0.49	2	1.7	0.11
High Energy Probe	0.43	2.6		
Centrality: 50-60%				
Basic Cone	0.58	2	1.6	0.095
Randomized $\eta\phi$	0.42	1.6	1.4	0.14
High Energy Probe	0.33	2		
Centrality: 60-70%				
Basic Cone	0.3	1.6	1.3	-0.09
Randomized $\eta\phi$	0.28	1.4	1.2	0.09
High Energy Probe	0.16	1.6		
Centrality: 70-80%				
Basic Cone	-0.19	1.3	1.1	-0.39
Randomized $\eta\phi$	-0.083	1.2	1.1	-0.19
High Energy Probe	-0.16	1.3		

B Supplemental Calorimeter Window Analysis Plots

This appendix includes the full set figures from the calorimeter window analysis which were not used in the body of the thesis. This includes the distributions of the E_T distributions (Section B.1), the fluctuations about average E_T (Section B.2), calorimeter fractional energy (Section B.3), the two dimensional distributions for average energy and standard deviation (Section B.4), and fits to average standard deviation dependence on calorimeter window area (Sec.B.5) for all windows sizes and each calorimeter when available. The example distributions for $\sigma^{n \times m}$ and $\langle E_T^{n \times m} \rangle$ of different sized window area are included below. Finally, the fits of the average standard deviation as a function of calorimeter window area are included. All window sizes listed in 4.4.2 are included for all event centralities.

B.1 Window Energy Distributions

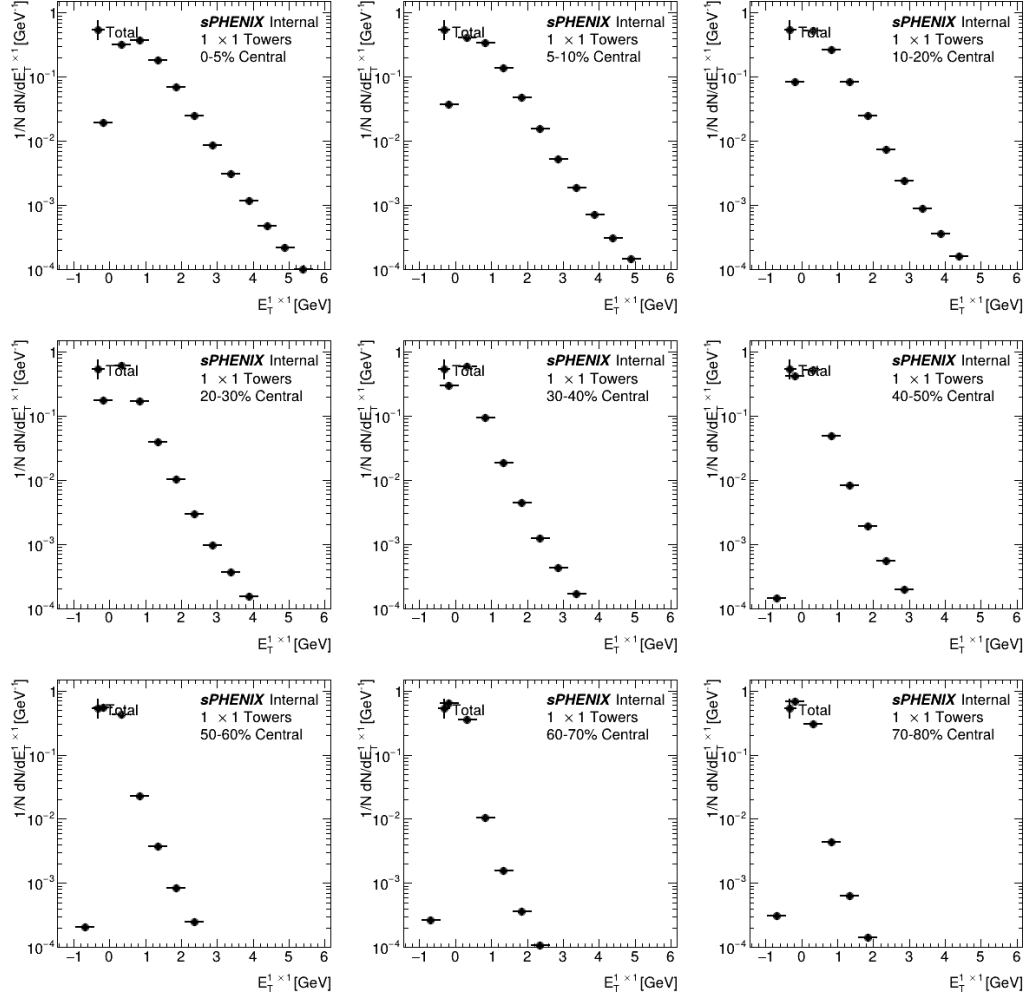


Figure 9: Distribution of E_T^{1x1} for full calorimeter windows for all centralities.

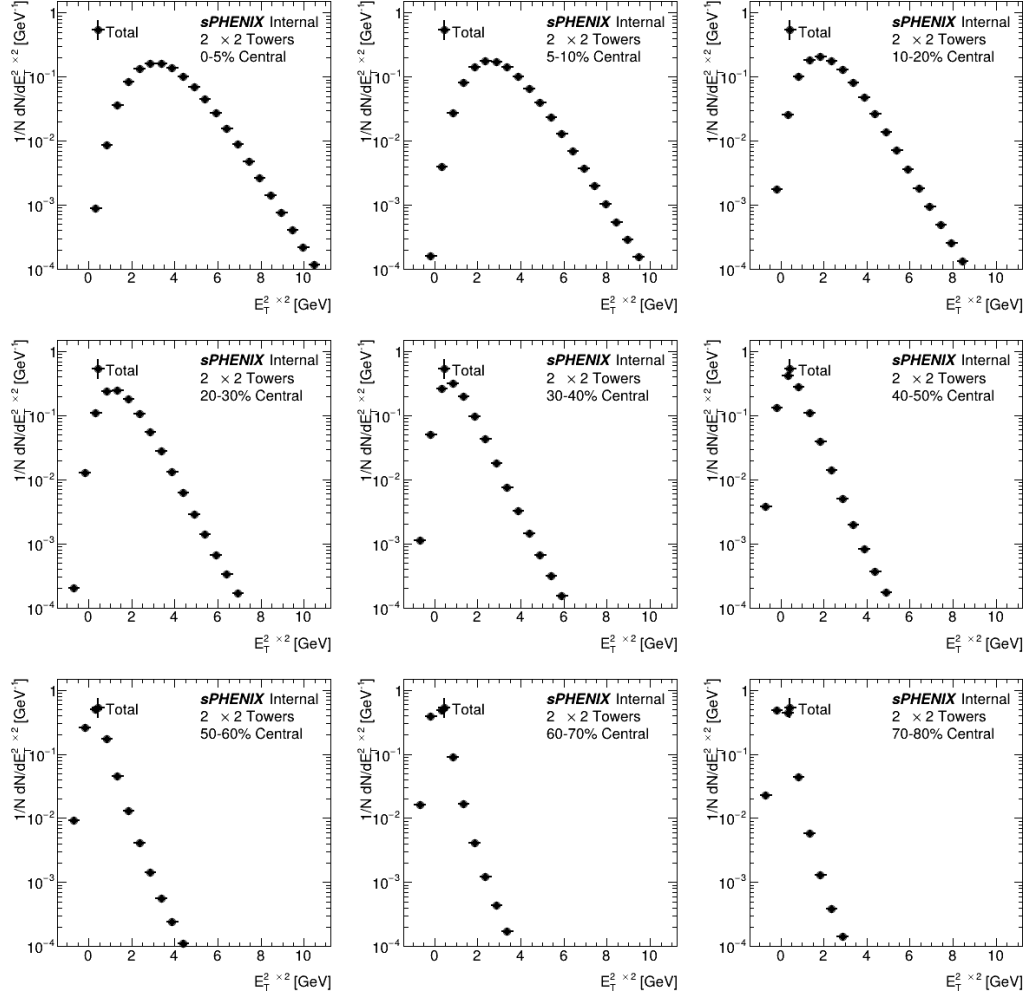


Figure 10: Distribution of E_T^2 for full calorimeter windows for all centralities.

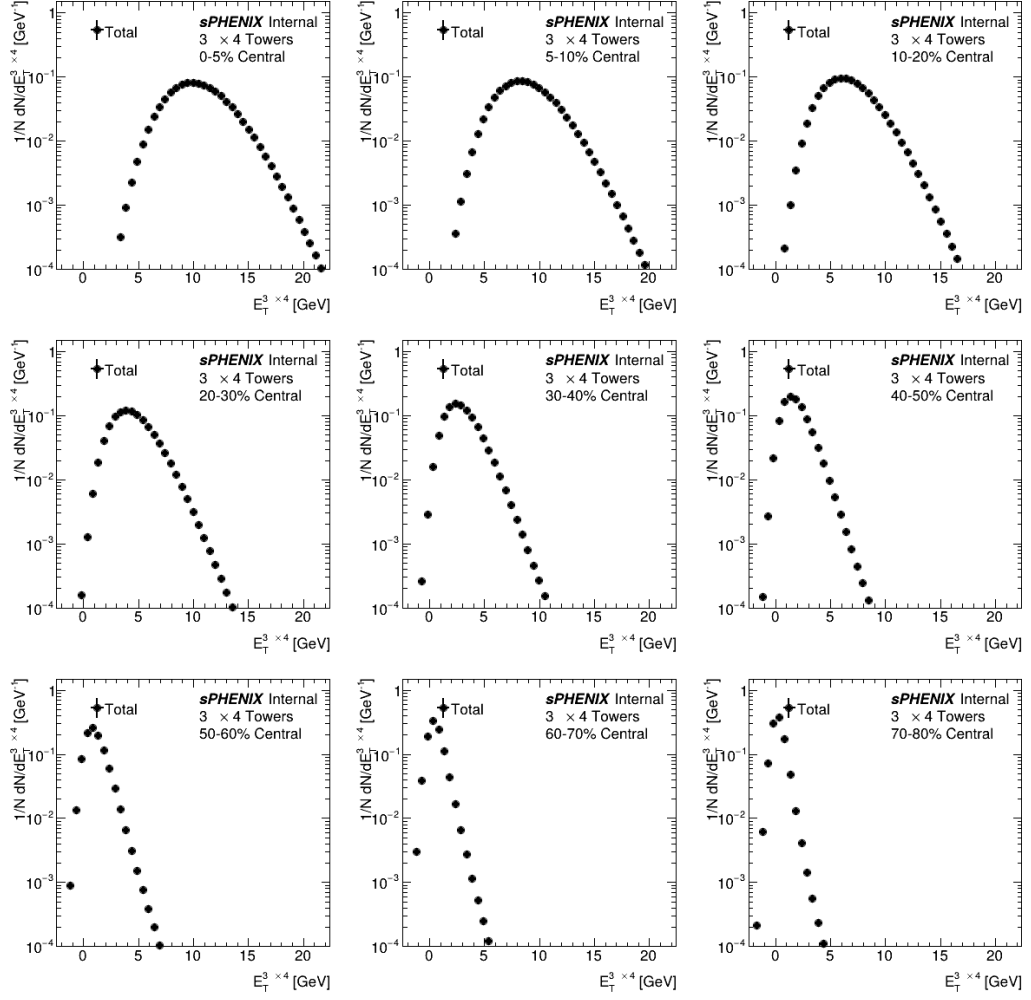


Figure 11: Distribution of E_T^3 for full calorimeter windows for all centralities.

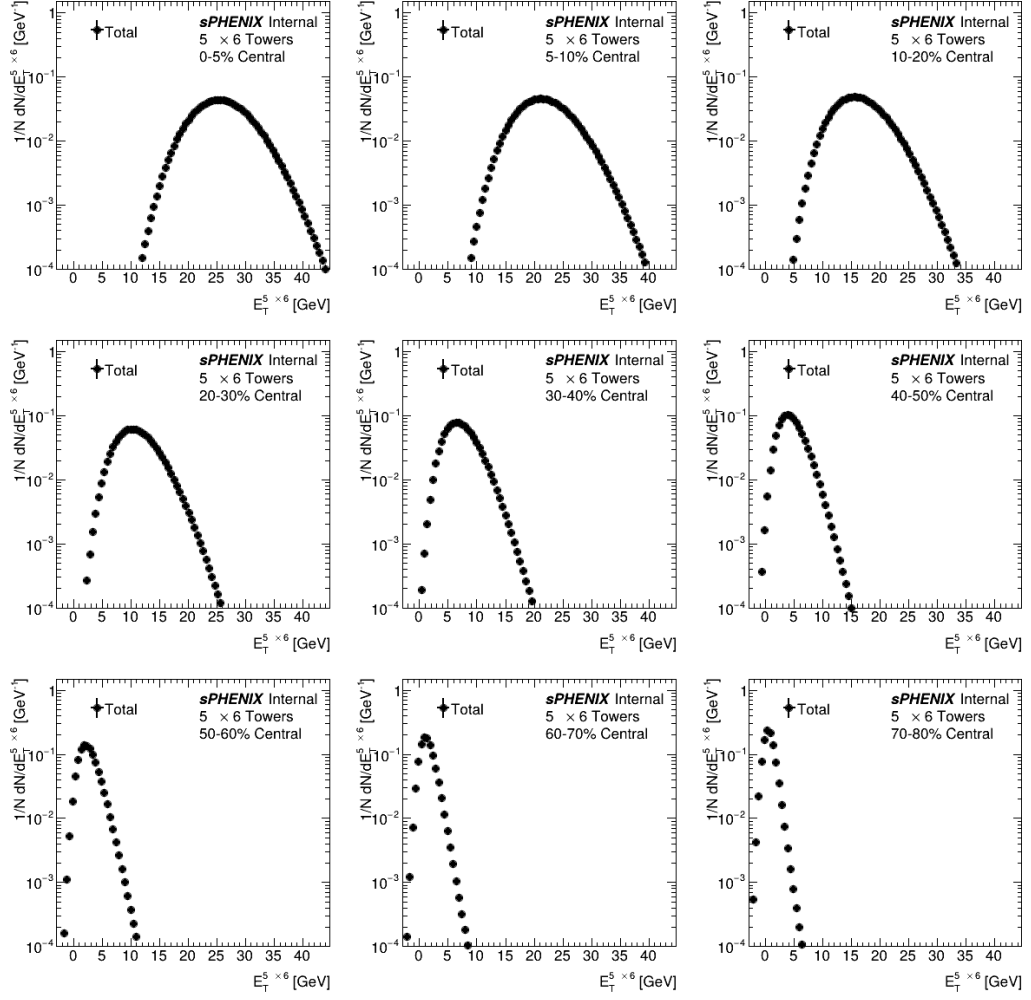


Figure 12: Distribution of E_T^{5x6} for full calorimeter windows for all centralities.

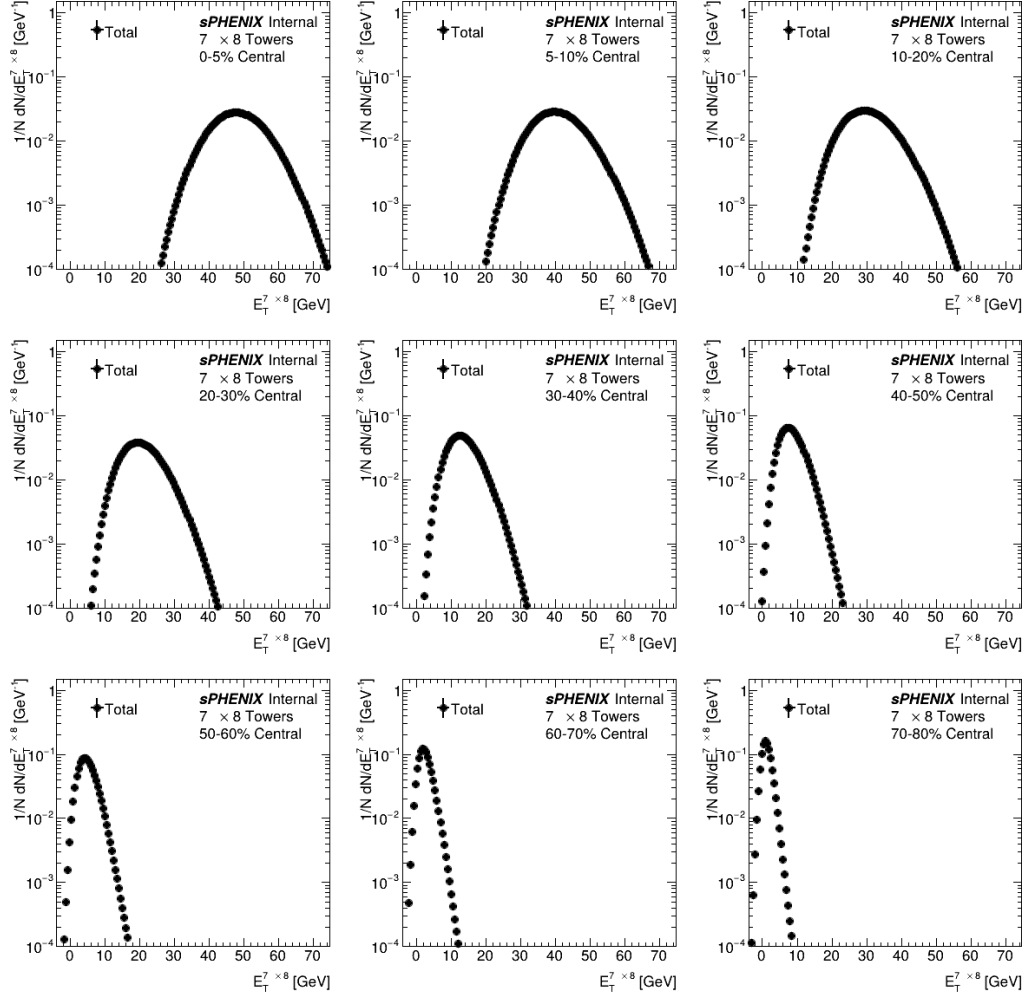


Figure 13: Distribution of E_T^{7x8} for full calorimeter windows for all centralities.

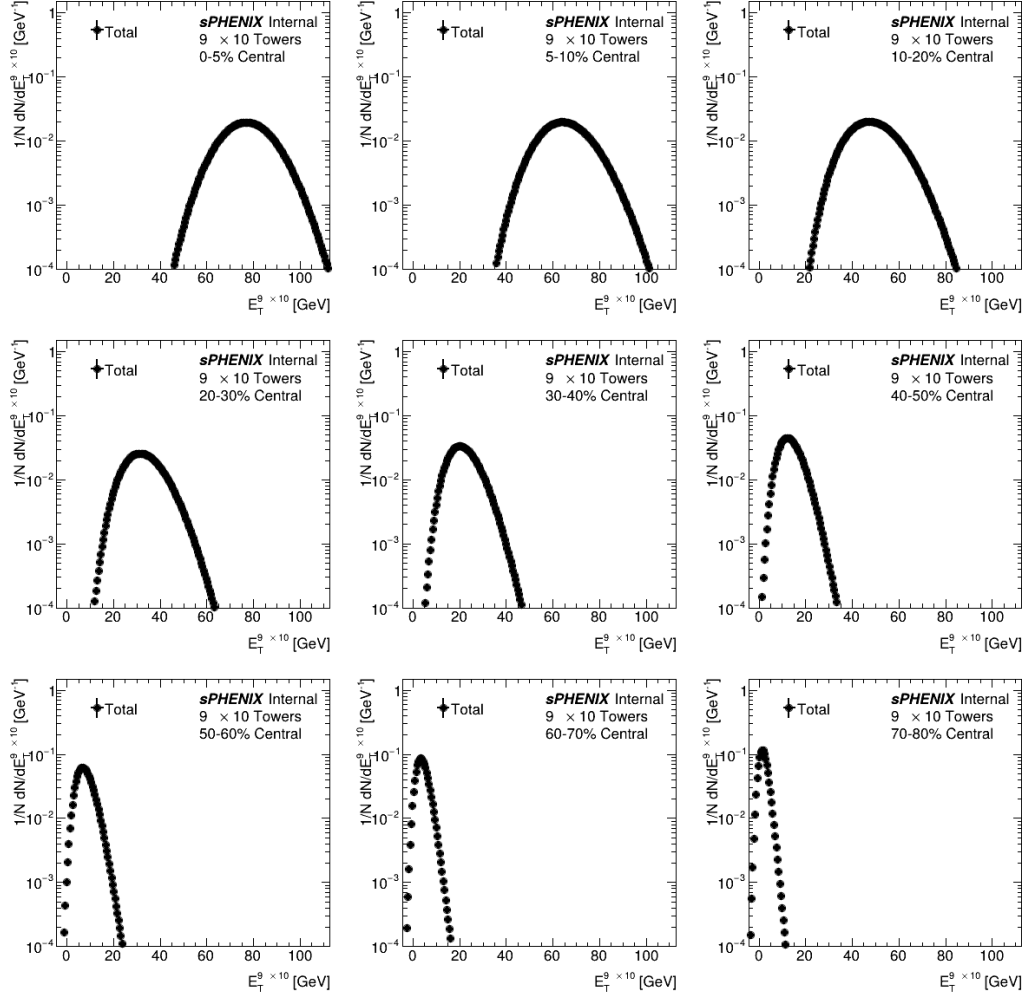


Figure 14: Distribution of $E_T^{9 \times 10}$ for full calorimeter windows for all centralities.

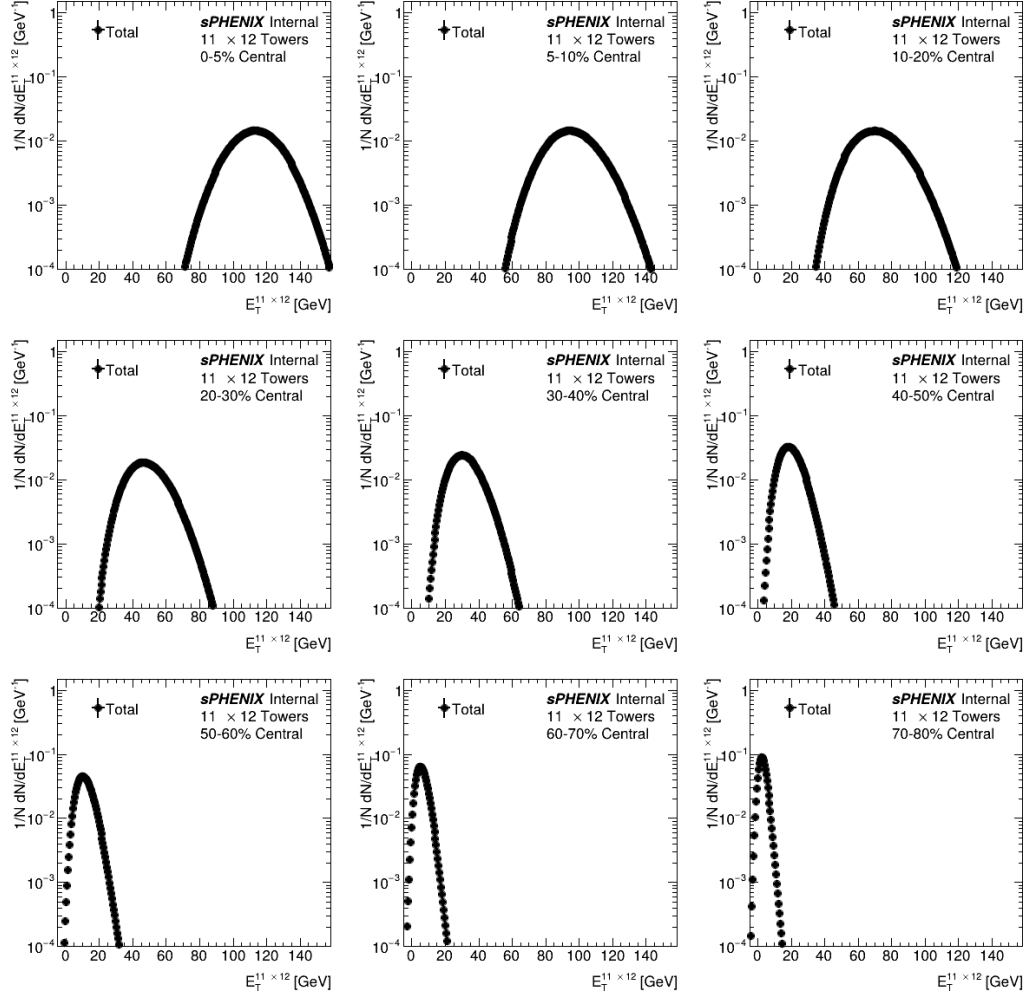


Figure 15: Distribution of $E_T^{11 \times 12}$ for full calorimeter windows for all centralities.

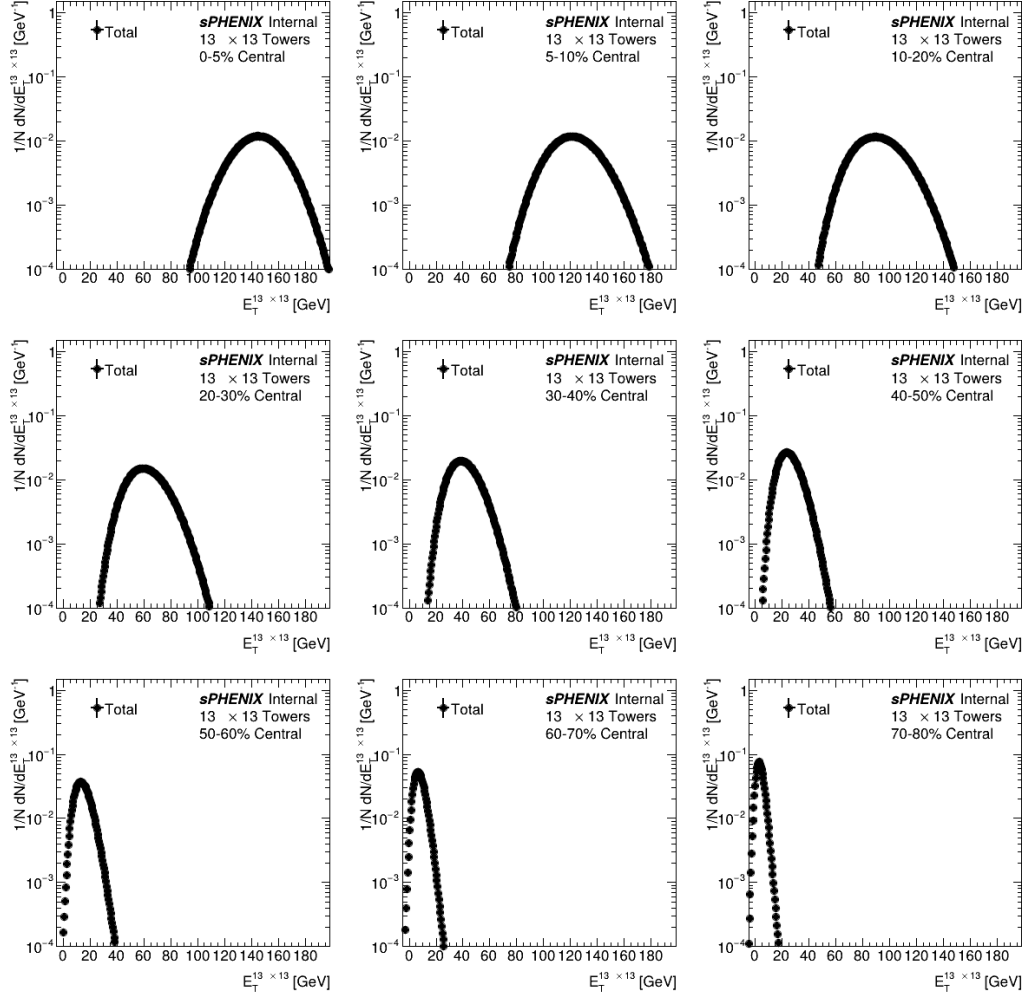


Figure 16: Distribution of $E_T^{13 \times 13}$ for full calorimeter windows for all centralities.

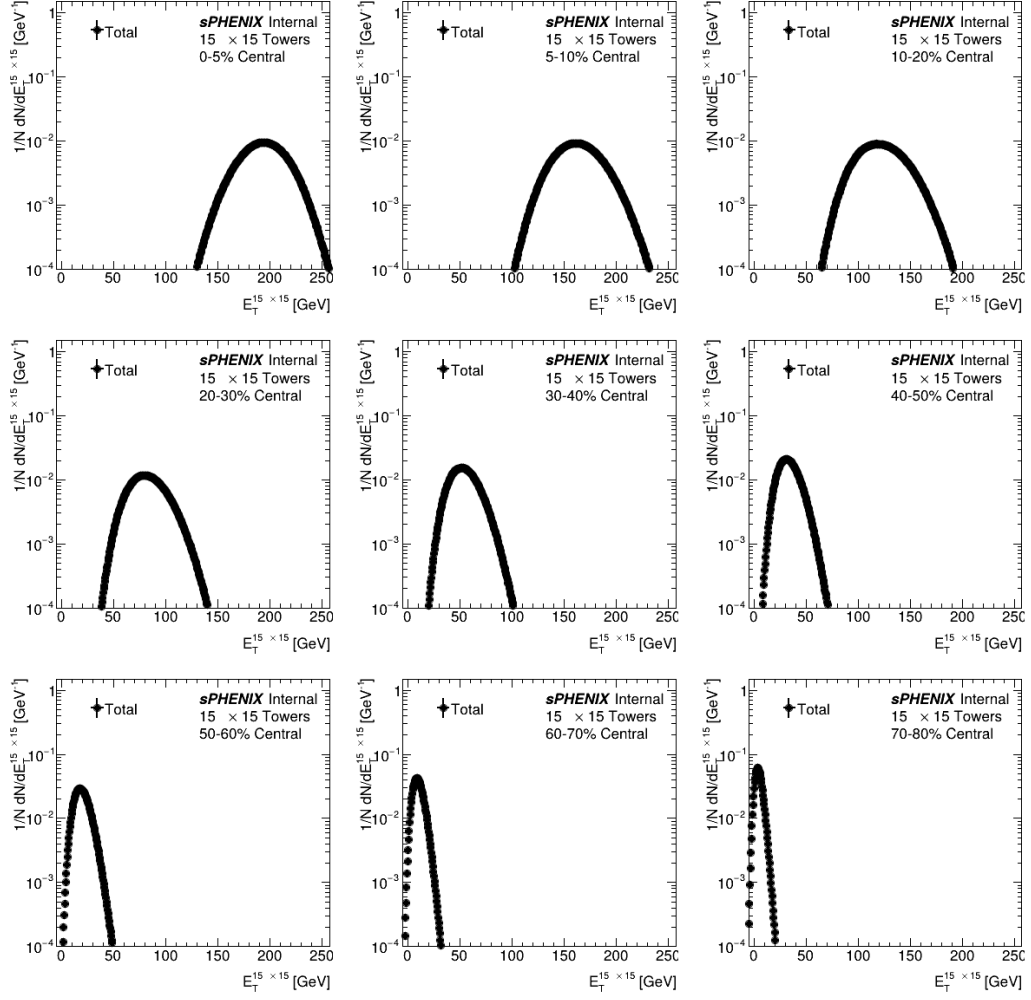


Figure 17: Distribution of $E_T^{15 \times 15}$ for full calorimeter windows for all centralities.

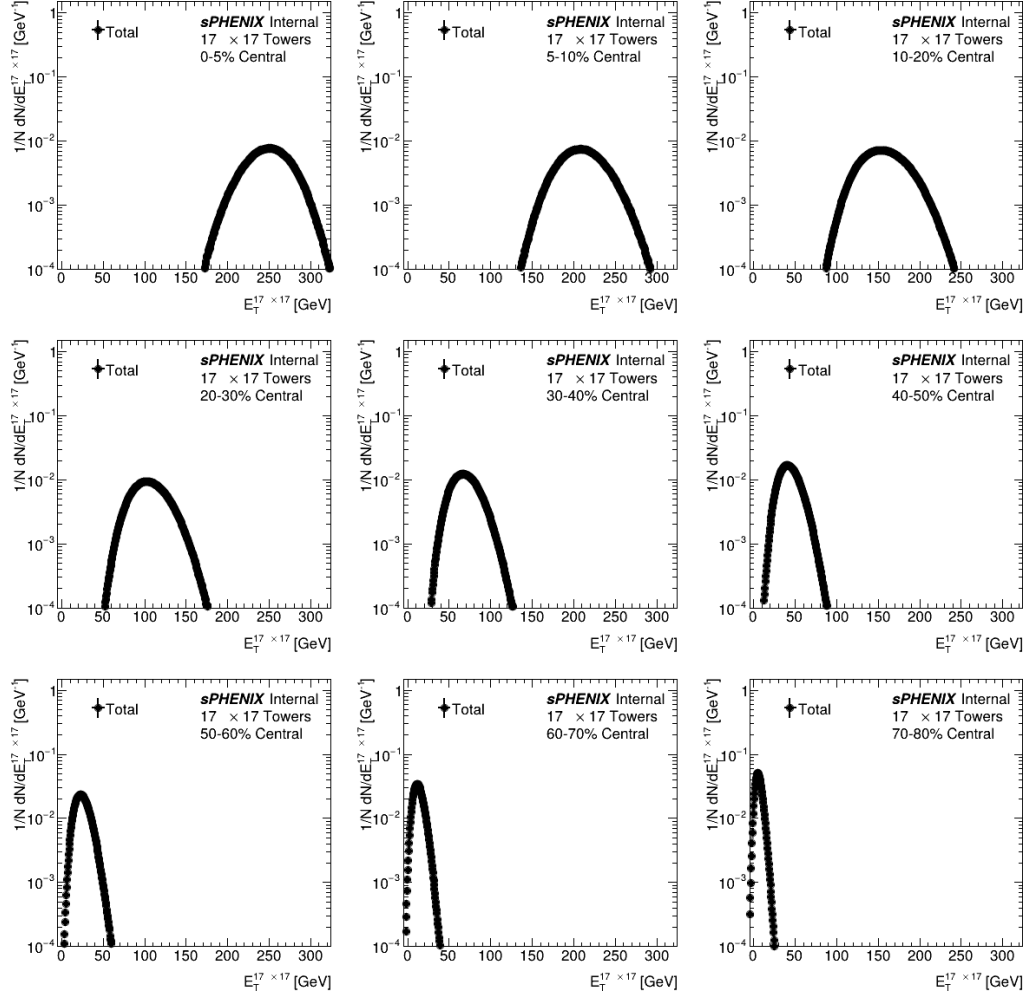


Figure 18: Distribution of $E_T^{17 \times 17}$ for full calorimeter windows for all centralities.

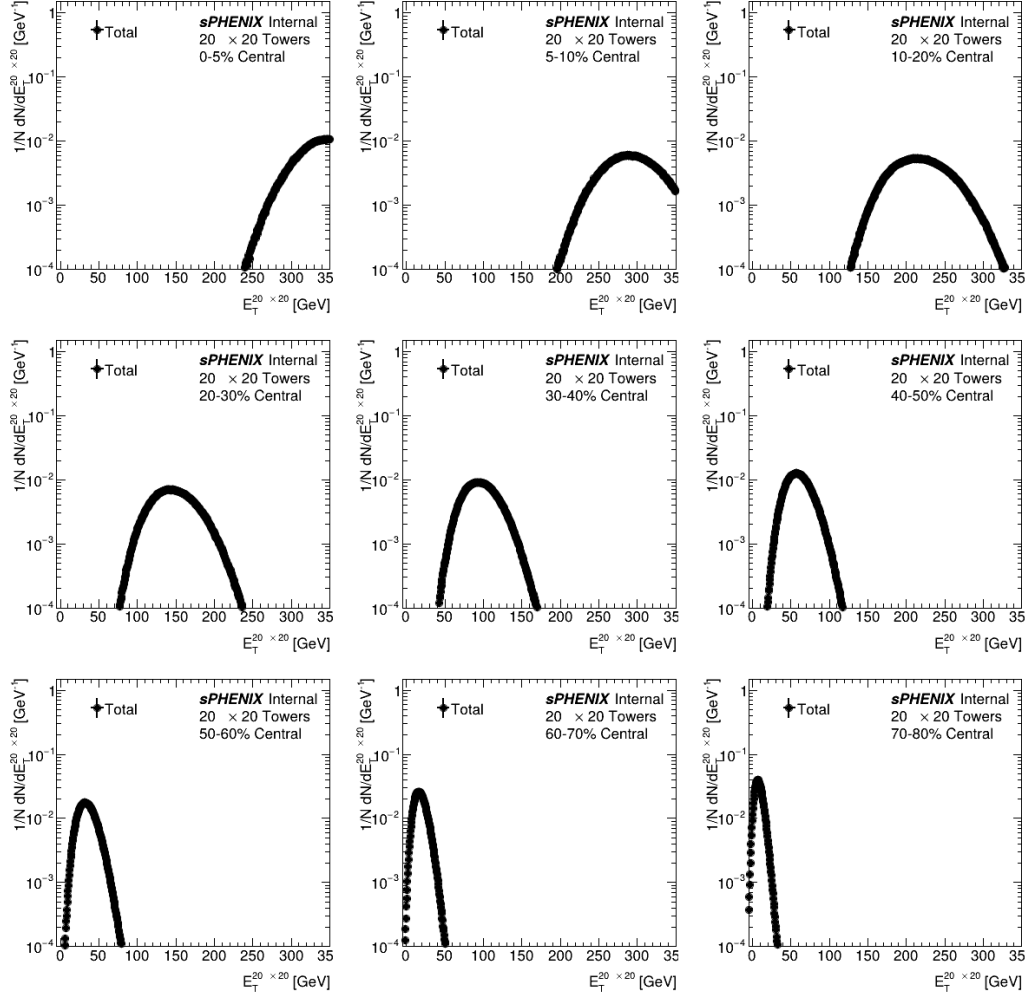


Figure 19: Distribution of $E_T^{20 \times 20}$ for full calorimeter windows for all centralities.

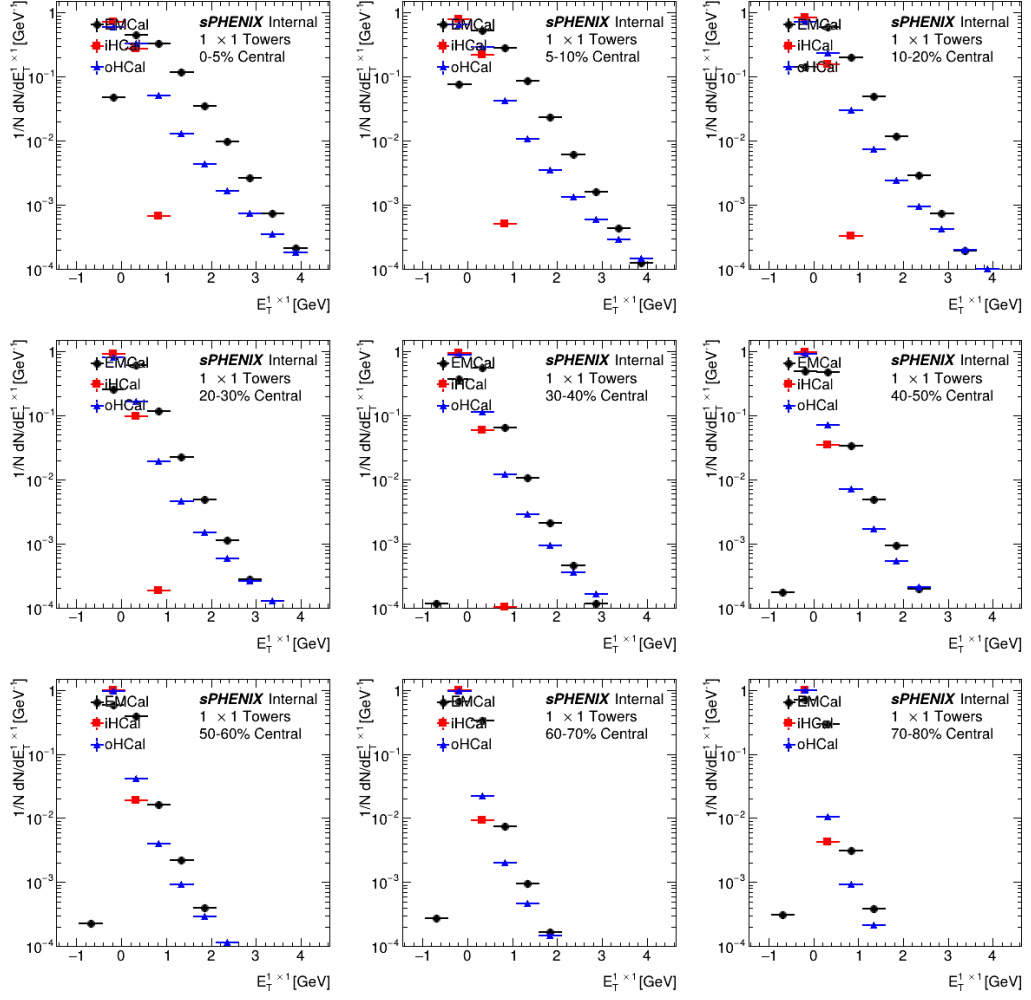


Figure 20: Distribution of E_T^{1x1} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

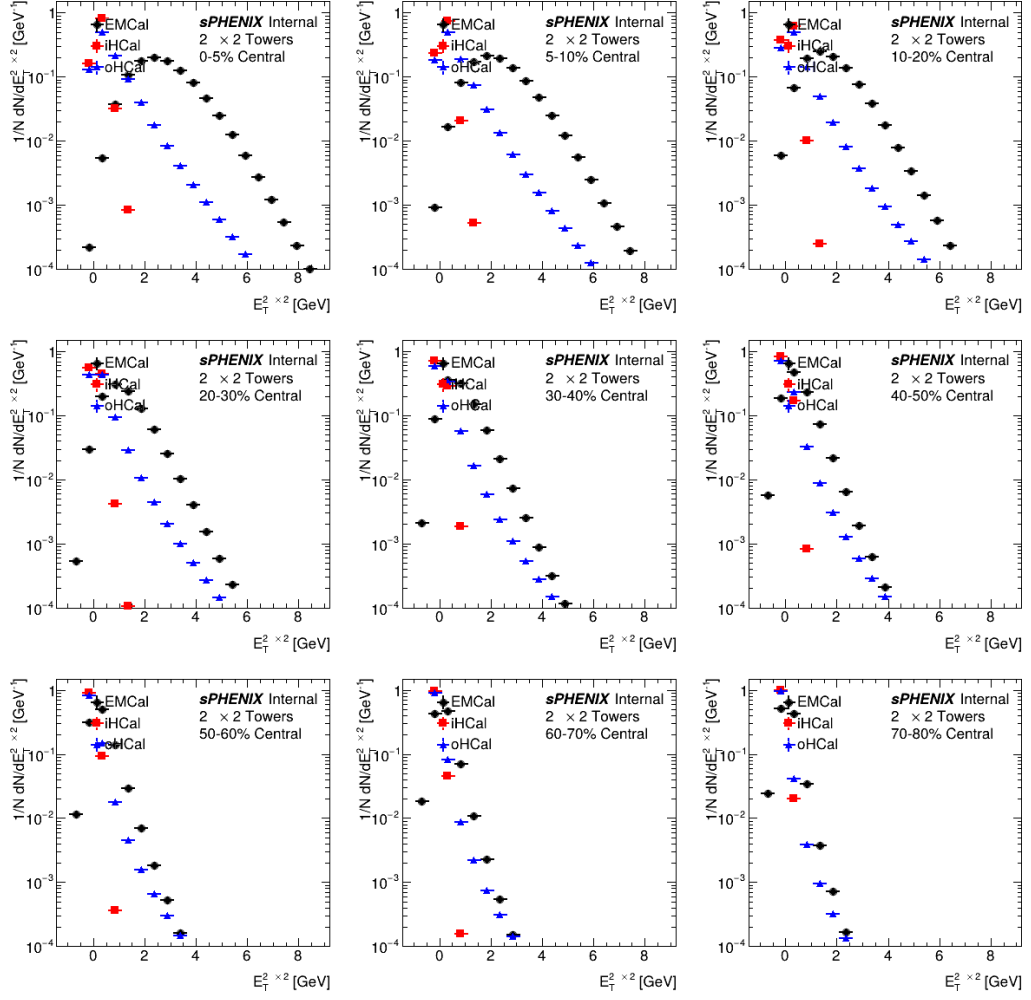


Figure 21: Distribution of E_T^2 for each sub-system (EMCAL, iHCal, and oHCal) calorimeter windows for all centralities.

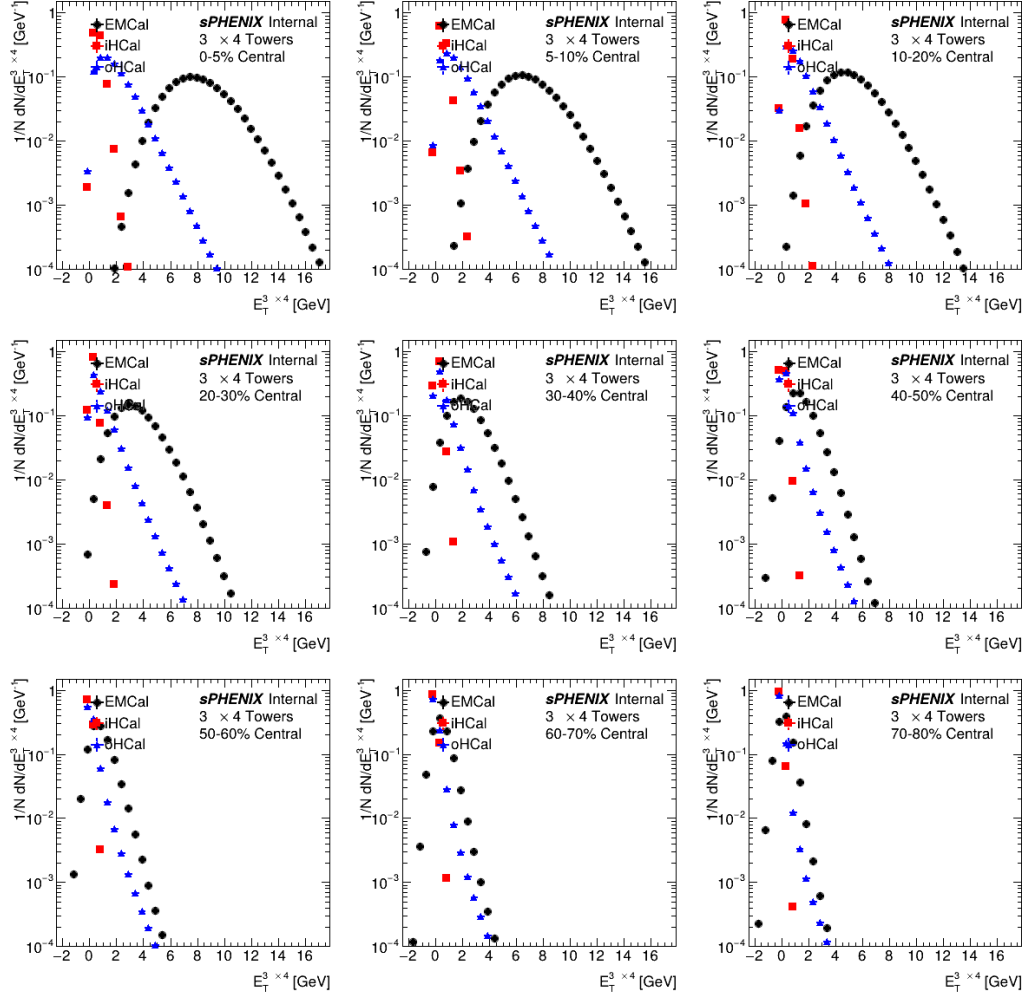


Figure 22: Distribution of E_T^{3x4} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

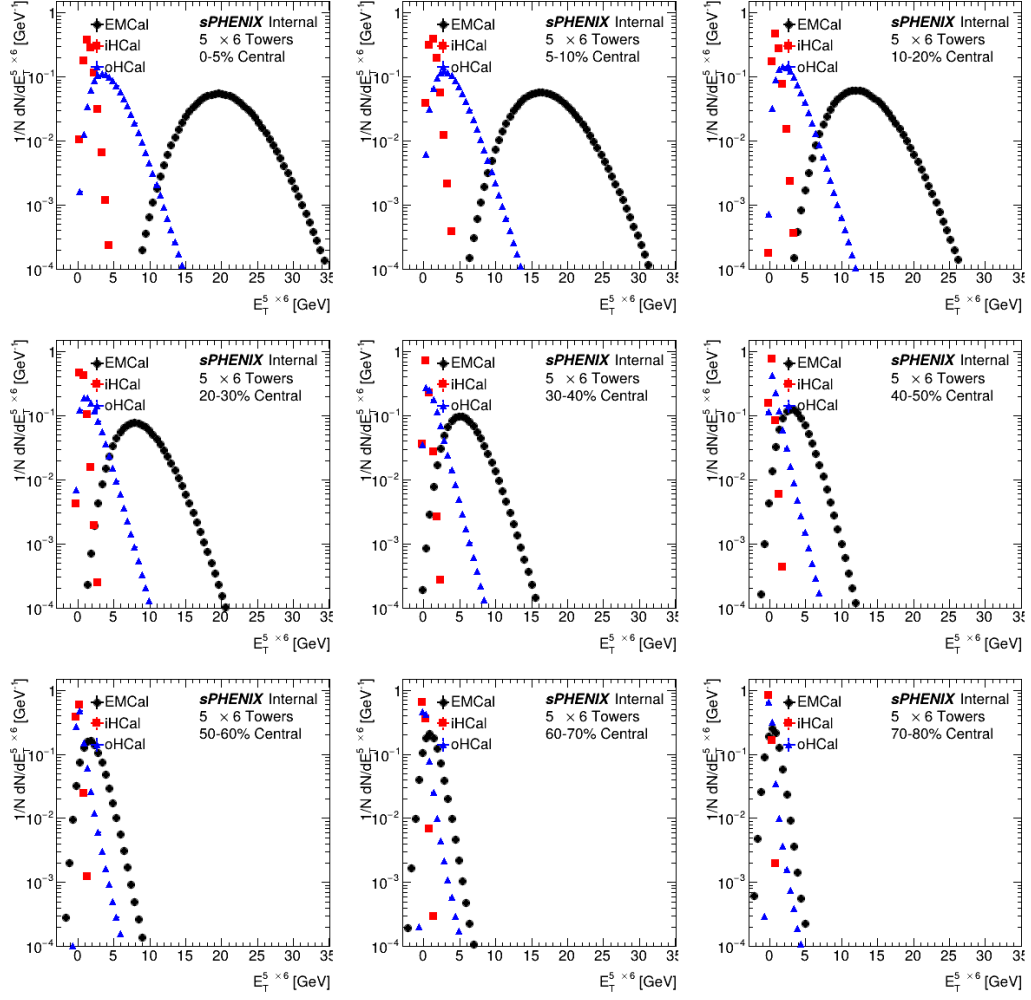


Figure 23: Distribution of $E_T^{5 \times 6}$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

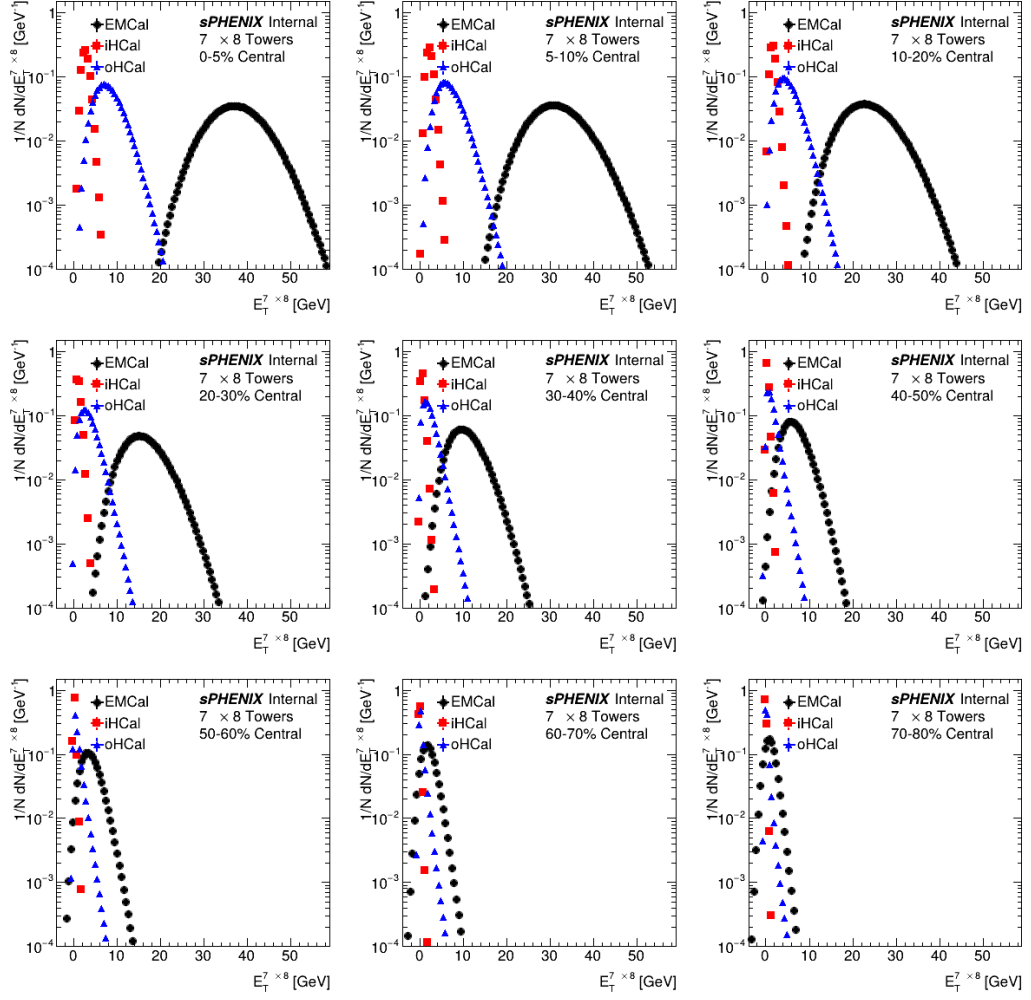


Figure 24: Distribution of E_T^{7x8} for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

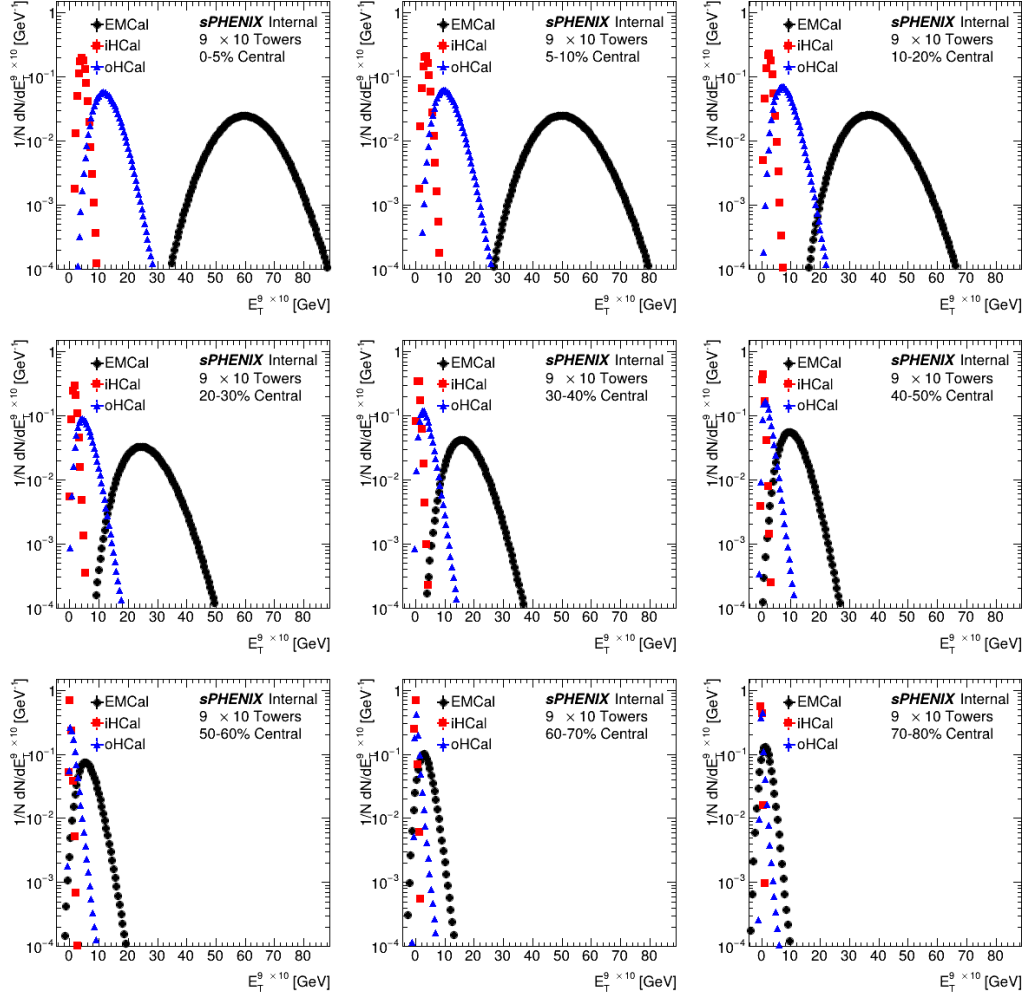


Figure 25: Distribution of $E_T^{9 \times 10}$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

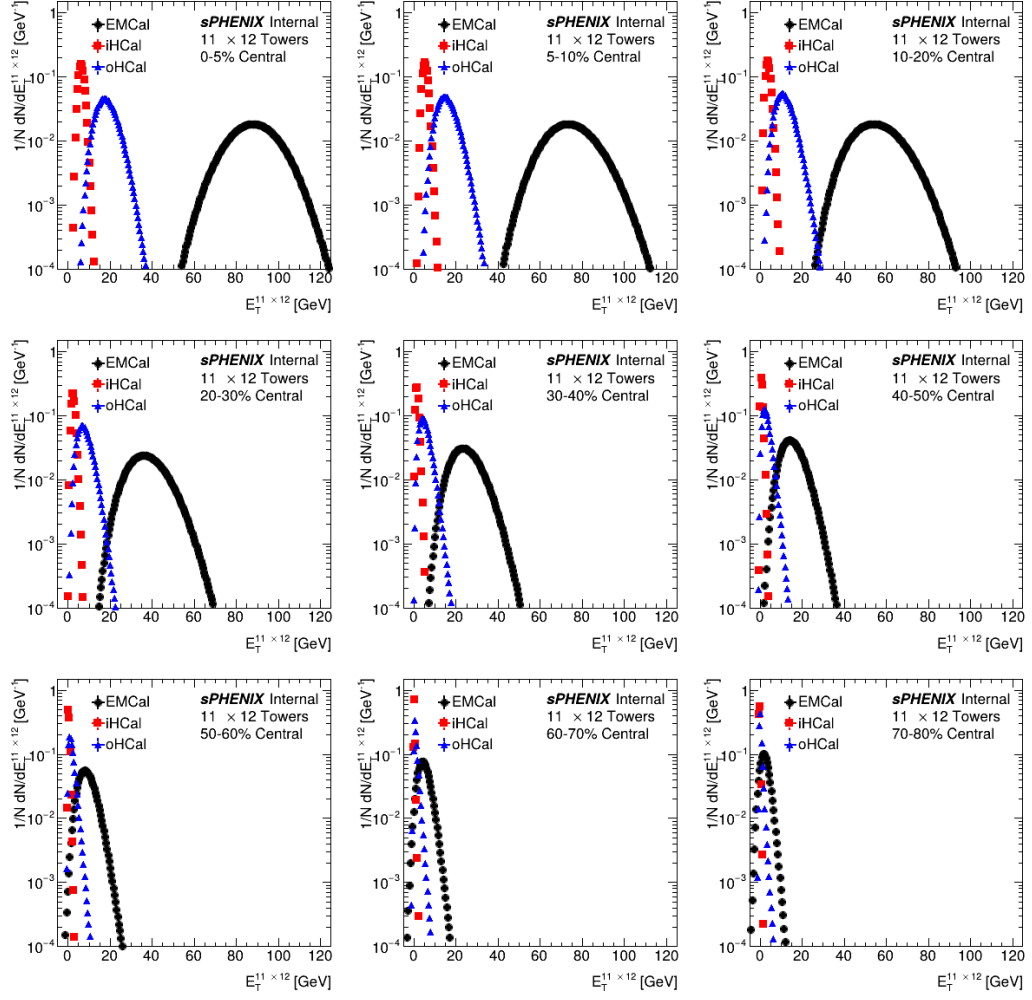


Figure 26: Distribution of $E_T^{11 \times 12}$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

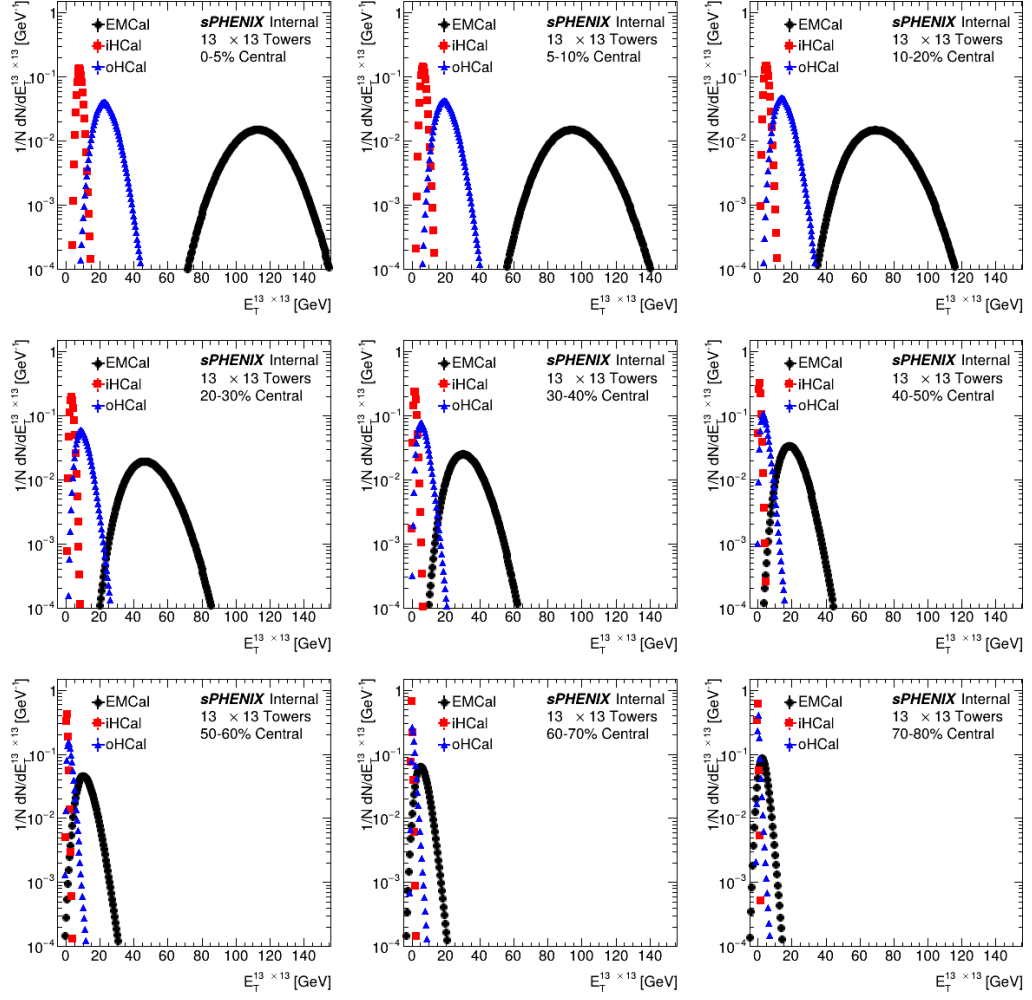


Figure 27: Distribution of $E_T^{13 \times 13}$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

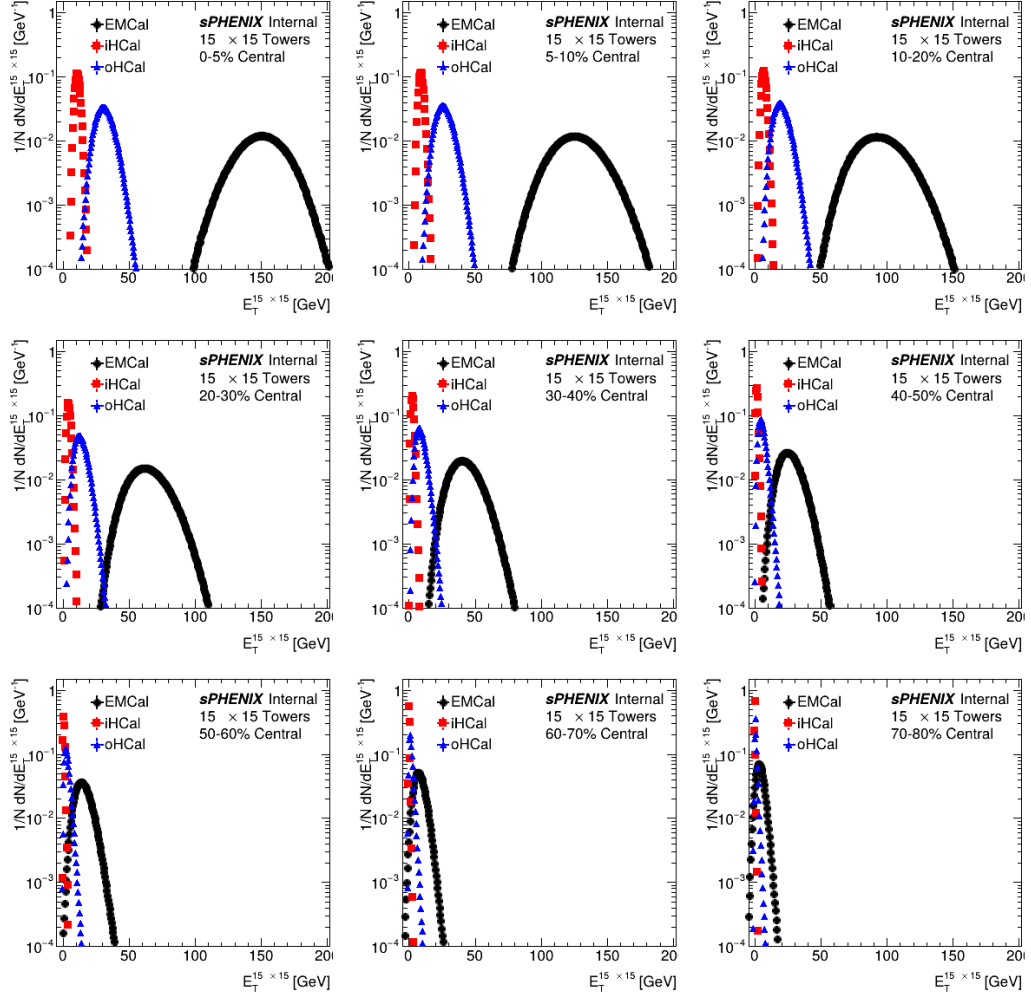


Figure 28: Distribution of $E_T^{15 \times 15}$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

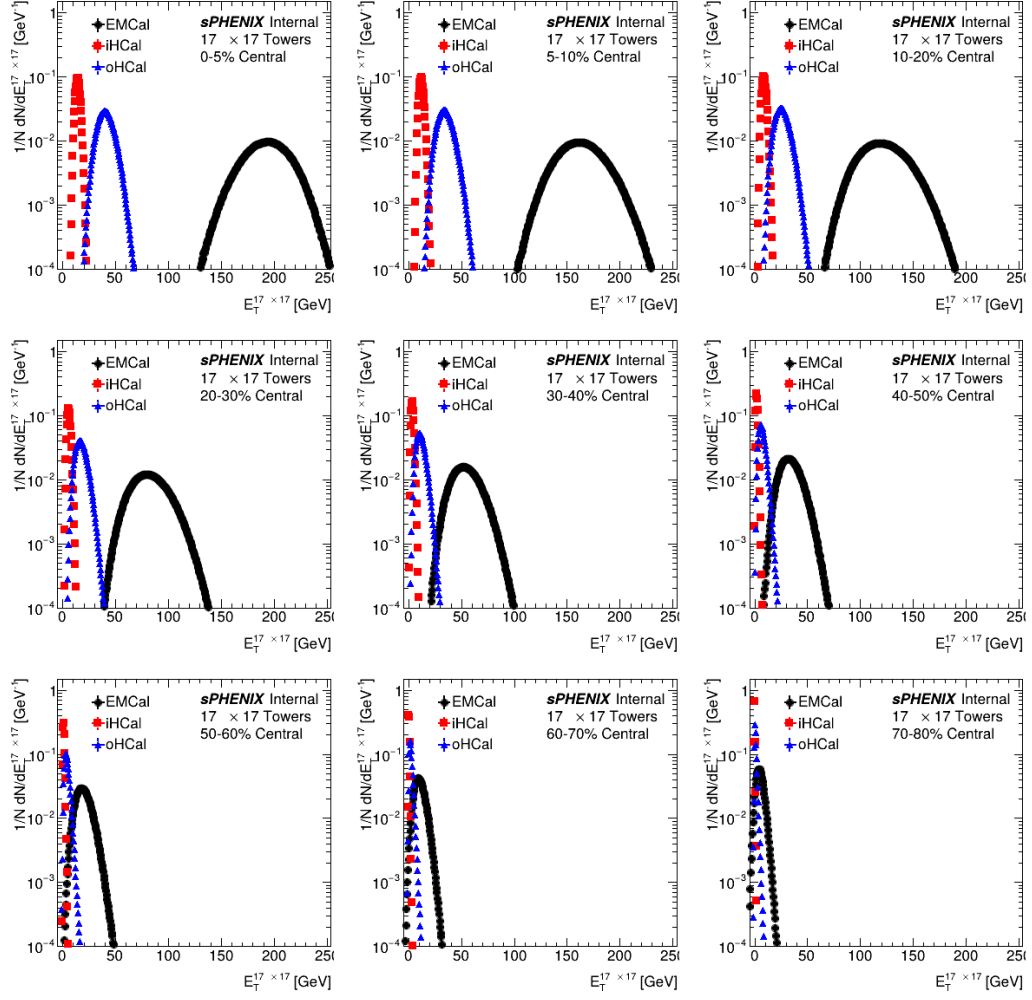


Figure 29: Distribution of $E_T^{17 \times 17}$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

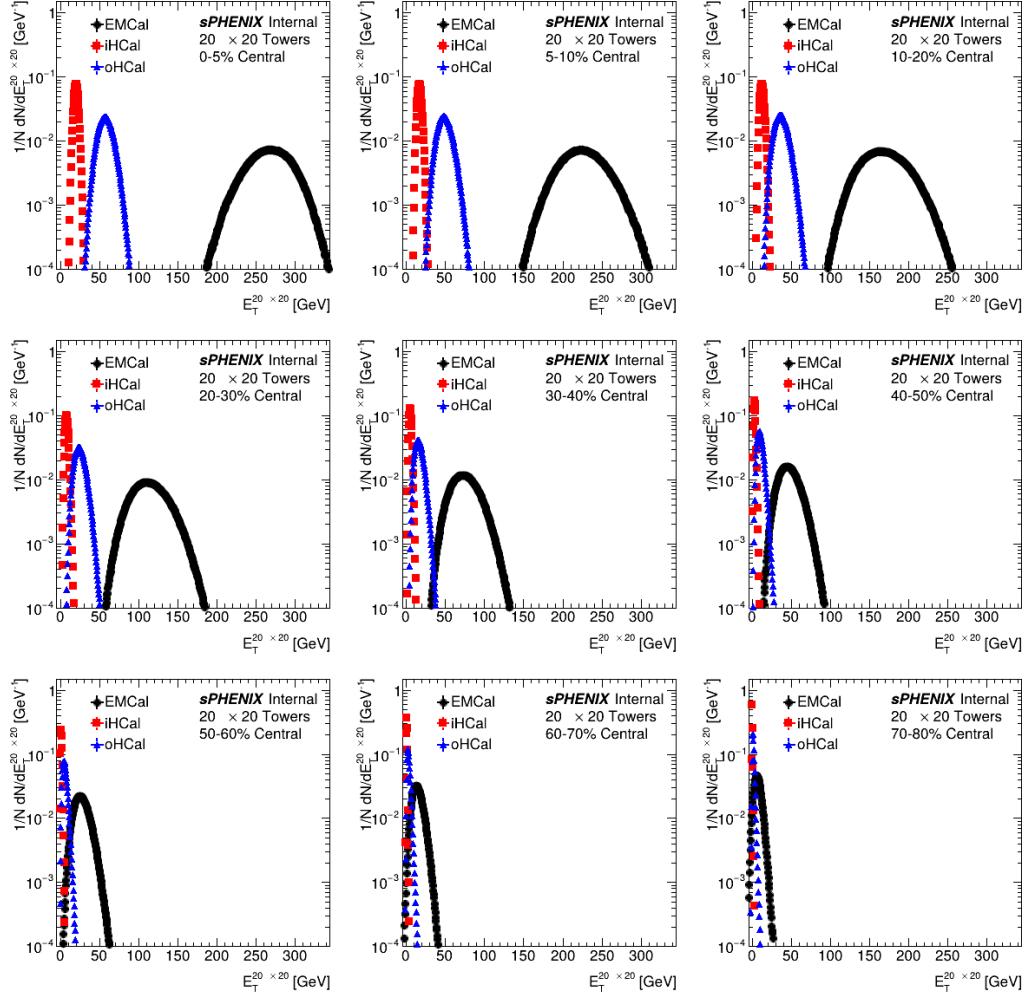


Figure 30: Distribution of $E_T^{20 \times 20}$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

B.2 Window Energy Fluctuations

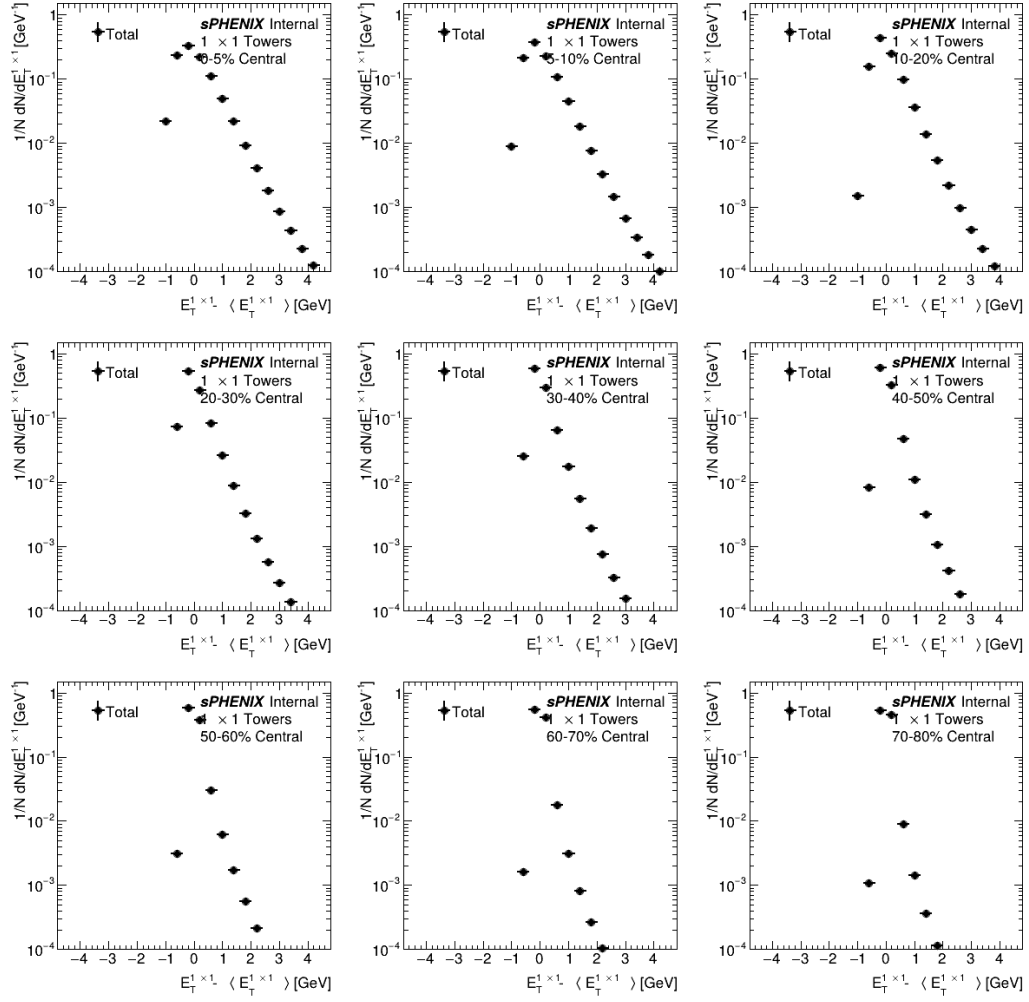


Figure 31: Distribution of $E_T^{1x1} - \langle E_T^{1x1} \rangle$ full calorimeter windows for all centralities.

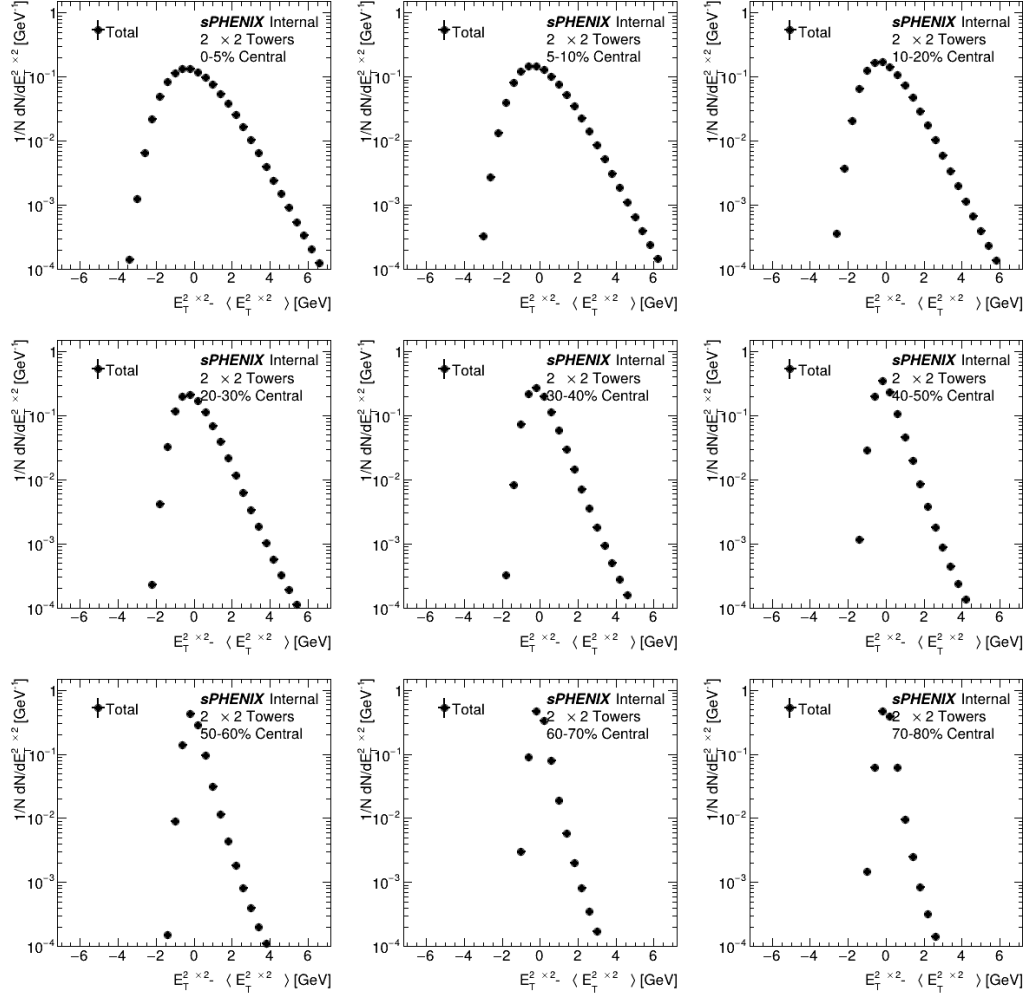


Figure 32: Distribution of $E_T^{2x2} - \langle E_T^{2x2} \rangle$ full calorimeter windows for all centralities.

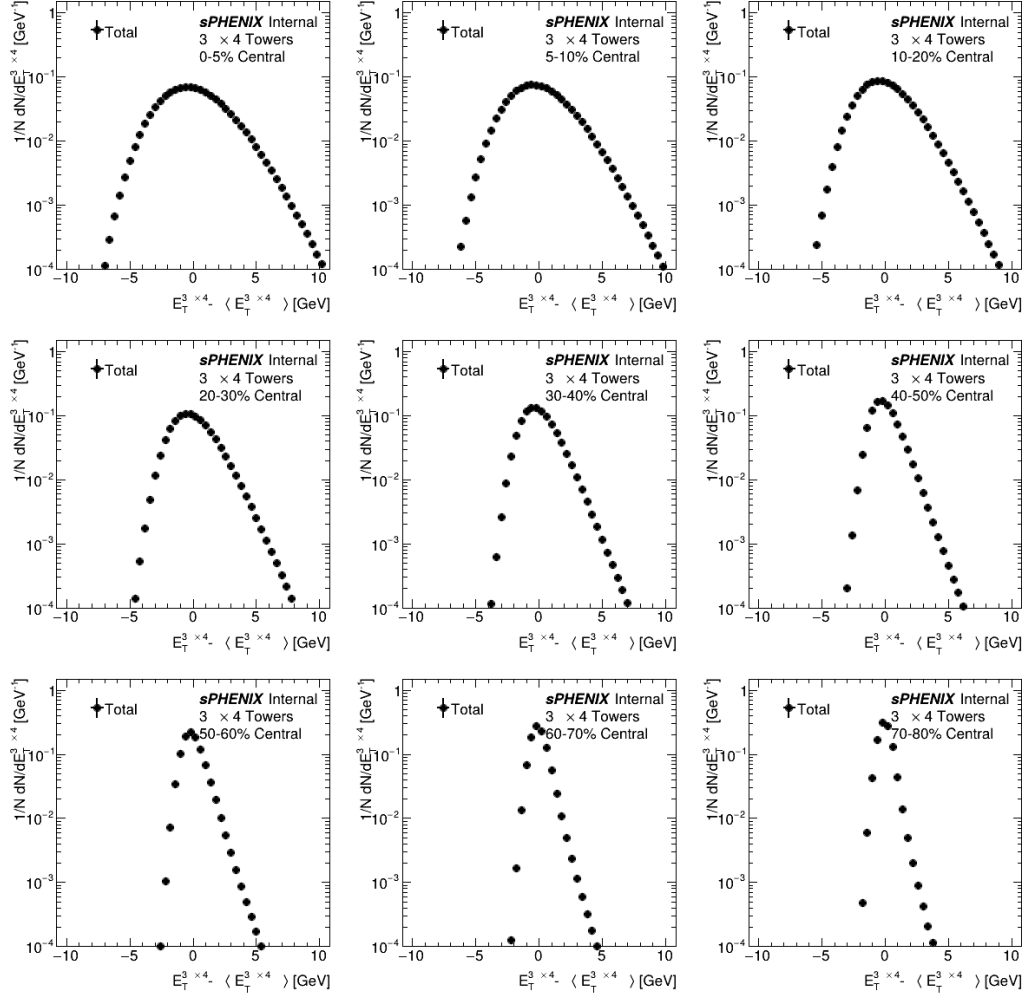


Figure 33: Distribution of $E_T^{3x4} - \langle E_T^{3x4} \rangle$ full calorimeter windows for all centralities.

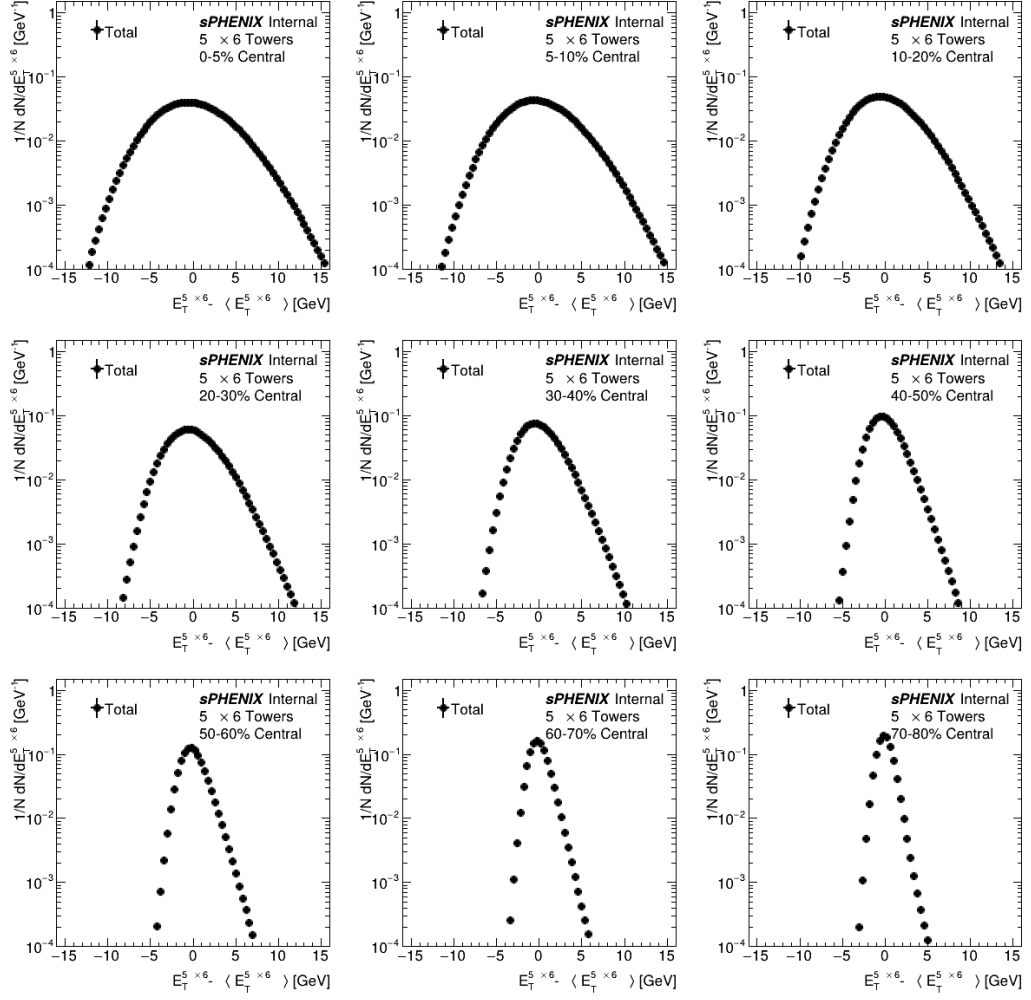


Figure 34: Distribution of $E_T^{5x6} - \langle E_T^{5x6} \rangle$ full calorimeter windows for all centralities.

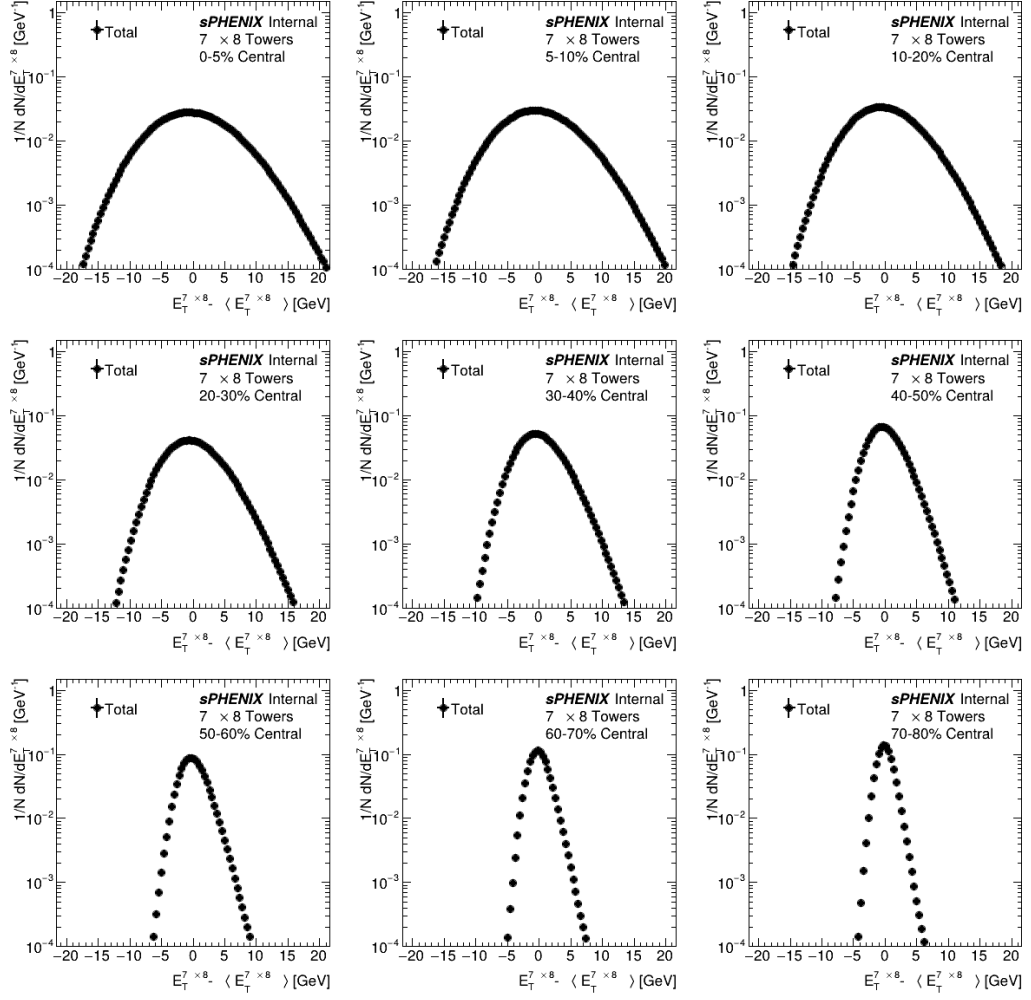


Figure 35: Distribution of $E_T^{7x8} - \langle E_T^{7x8} \rangle$ full calorimeter windows for all centralities.

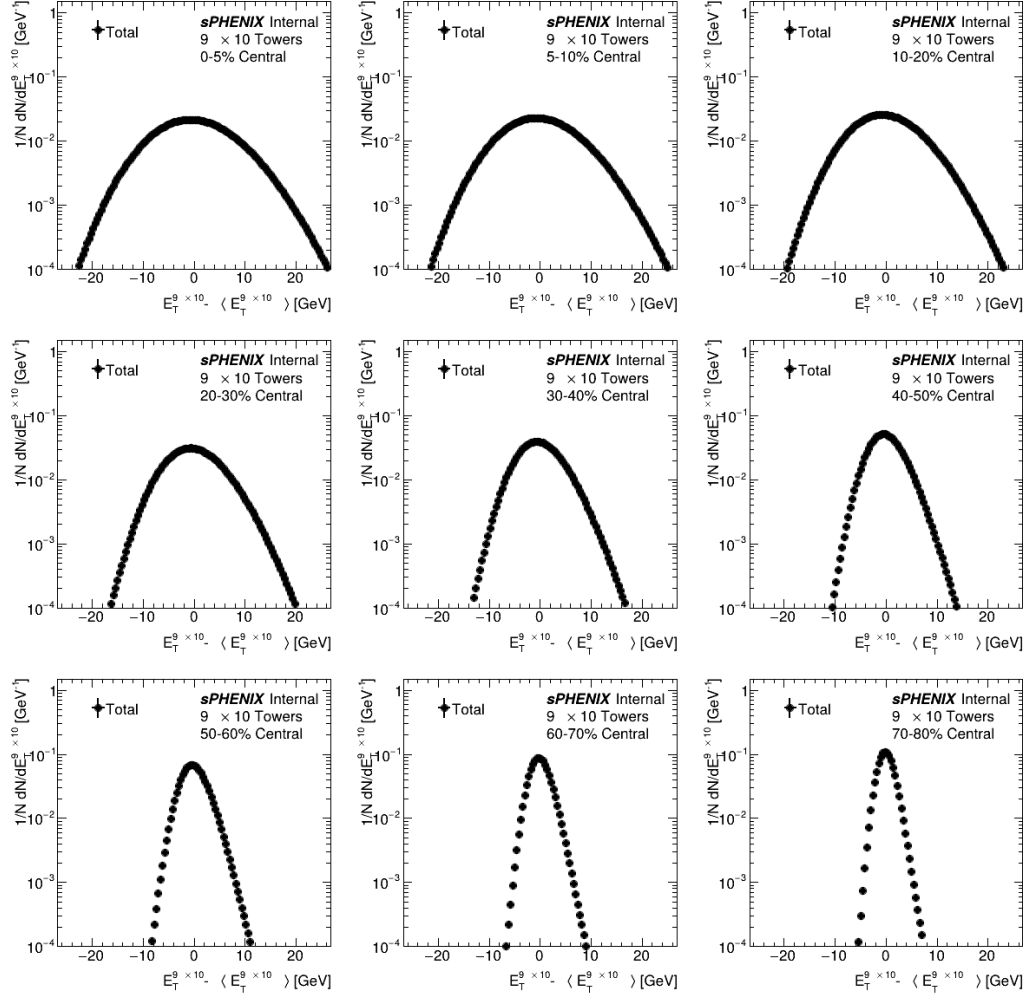


Figure 36: Distribution of $E_T^{9 \times 10} - \langle E_T^{9 \times 10} \rangle$ full calorimeter windows for all centralities.

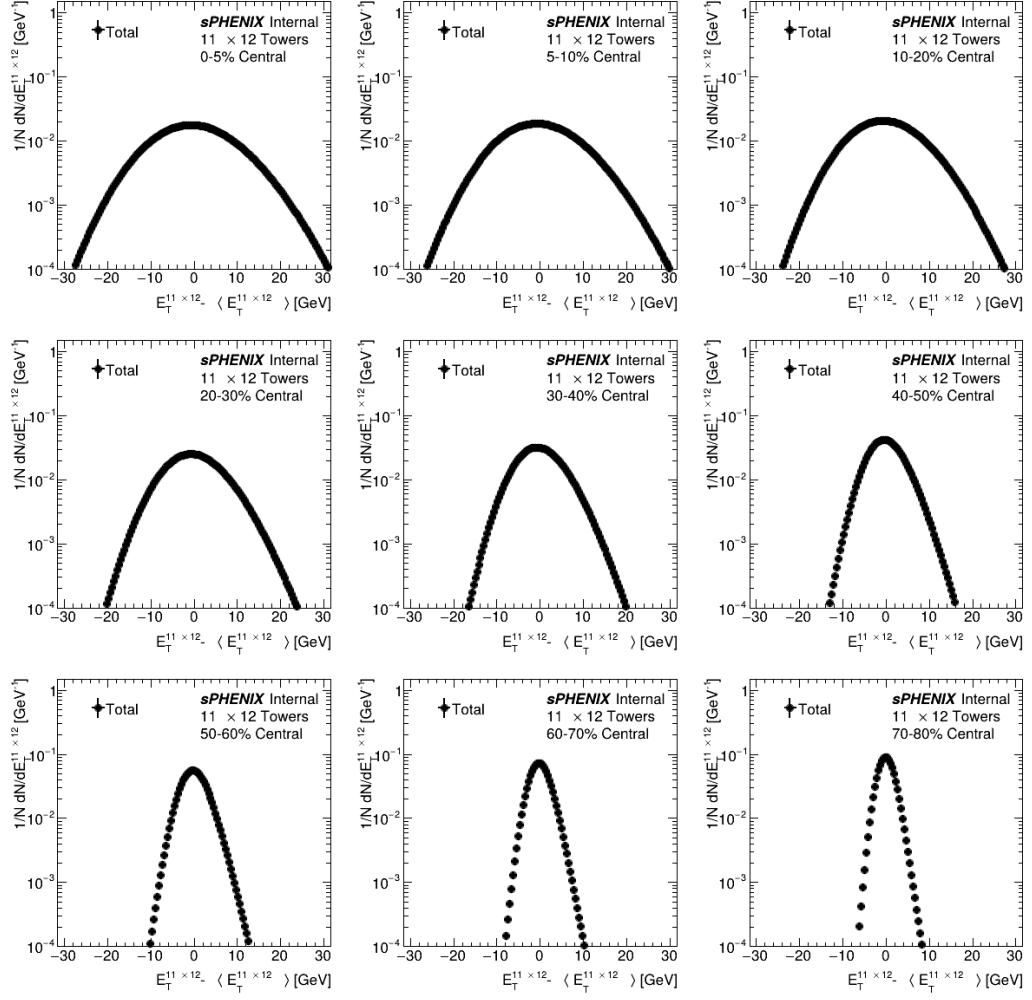


Figure 37: Distribution of $E_T^{11 \times 12} - \langle E_T^{11 \times 12} \rangle$ full calorimeter windows for all centralities.

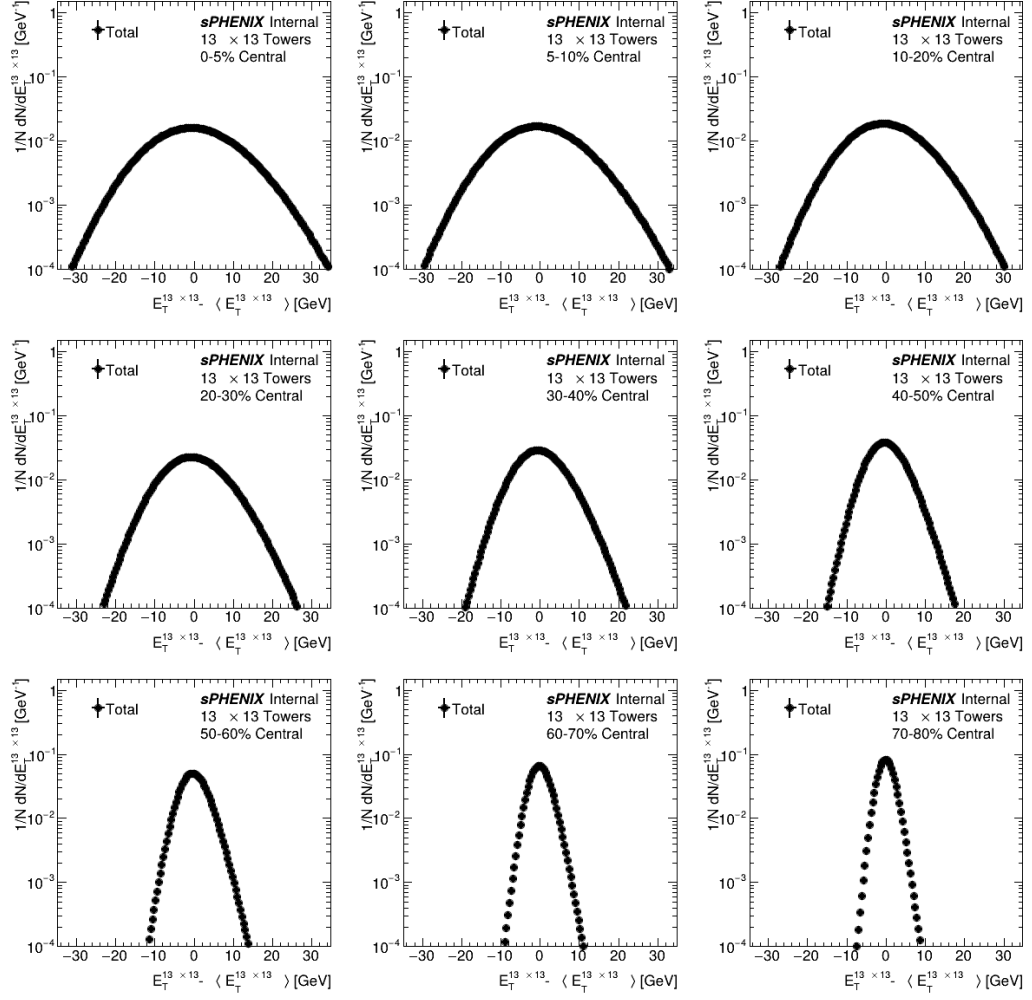


Figure 38: Distribution of $E_T^{13x13} - \langle E_T^{13x13} \rangle$ full calorimeter windows for all centralities.

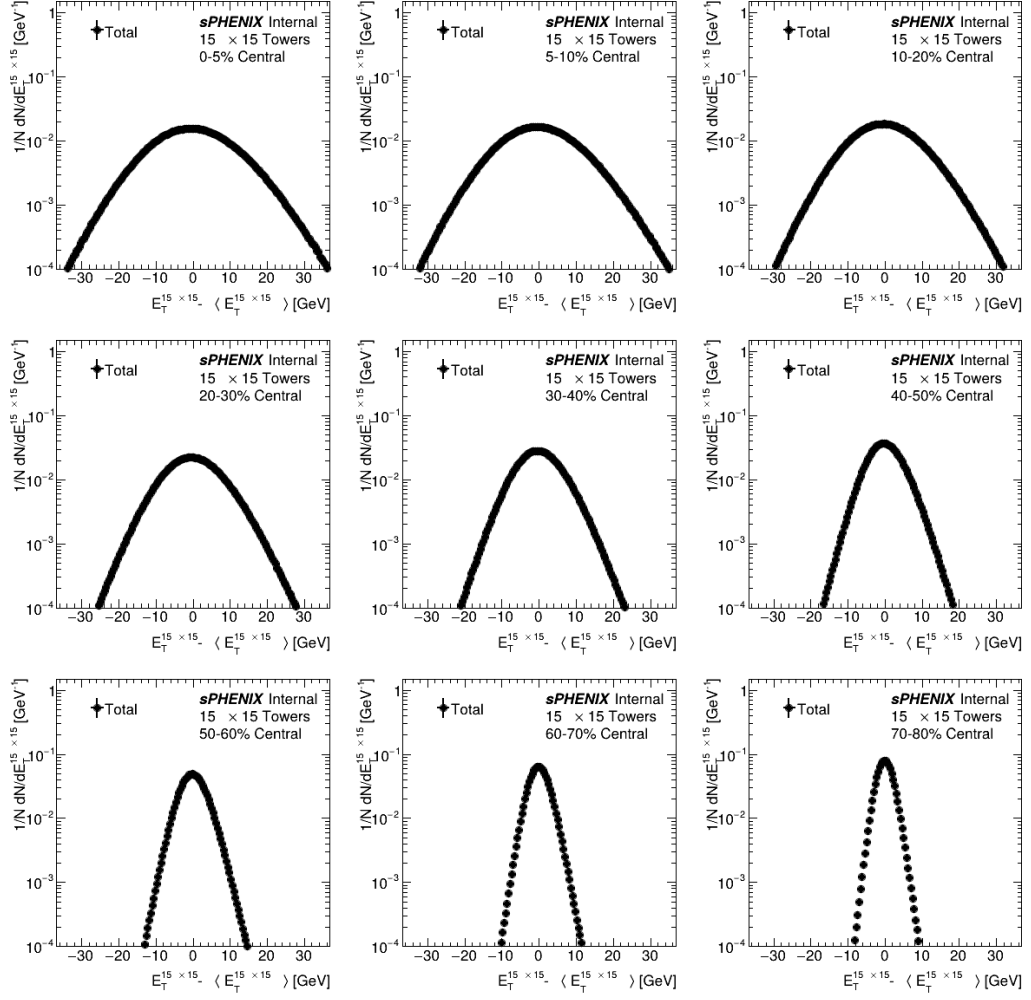


Figure 39: Distribution of $E_T^{15x15} - \langle E_T^{15x15} \rangle$ full calorimeter windows for all centralities.

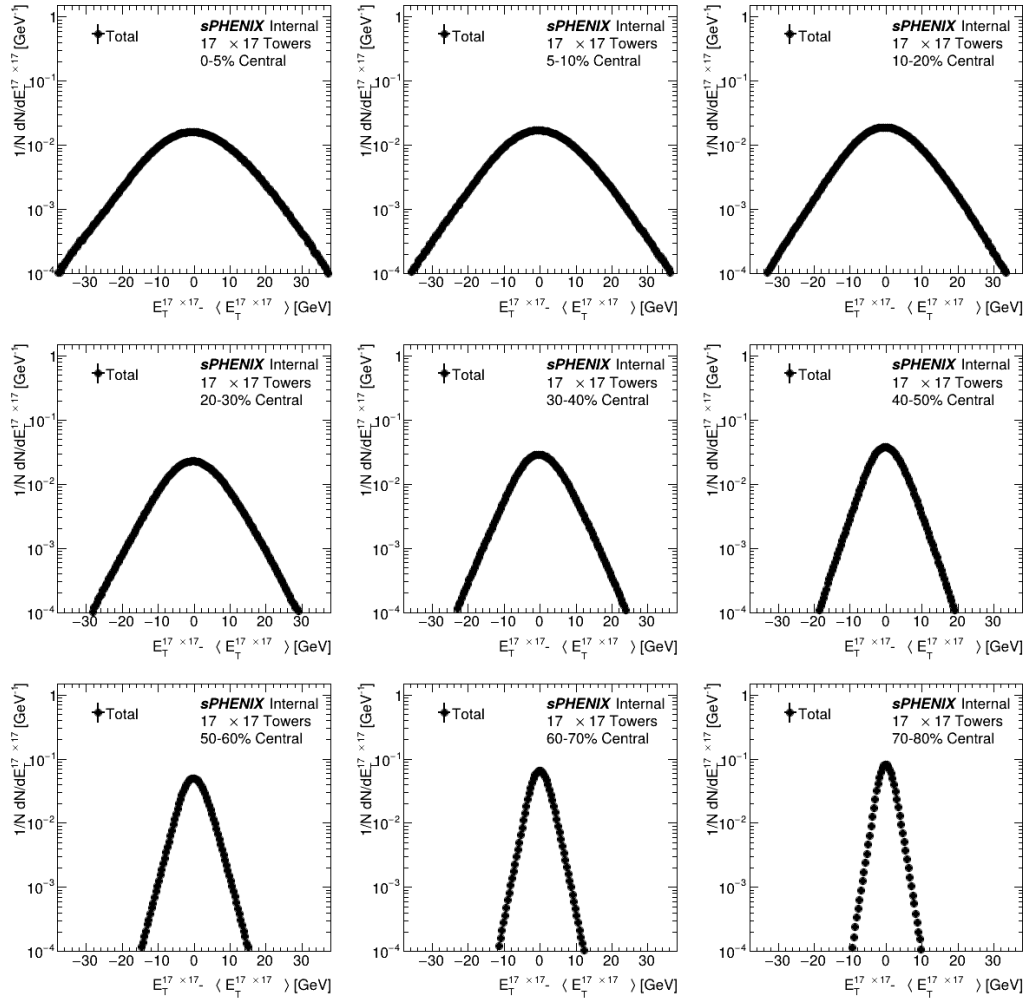


Figure 40: Distribution of $E_T^{17 \times 17} - \langle E_T^{17 \times 17} \rangle$ full calorimeter windows for all centralities.

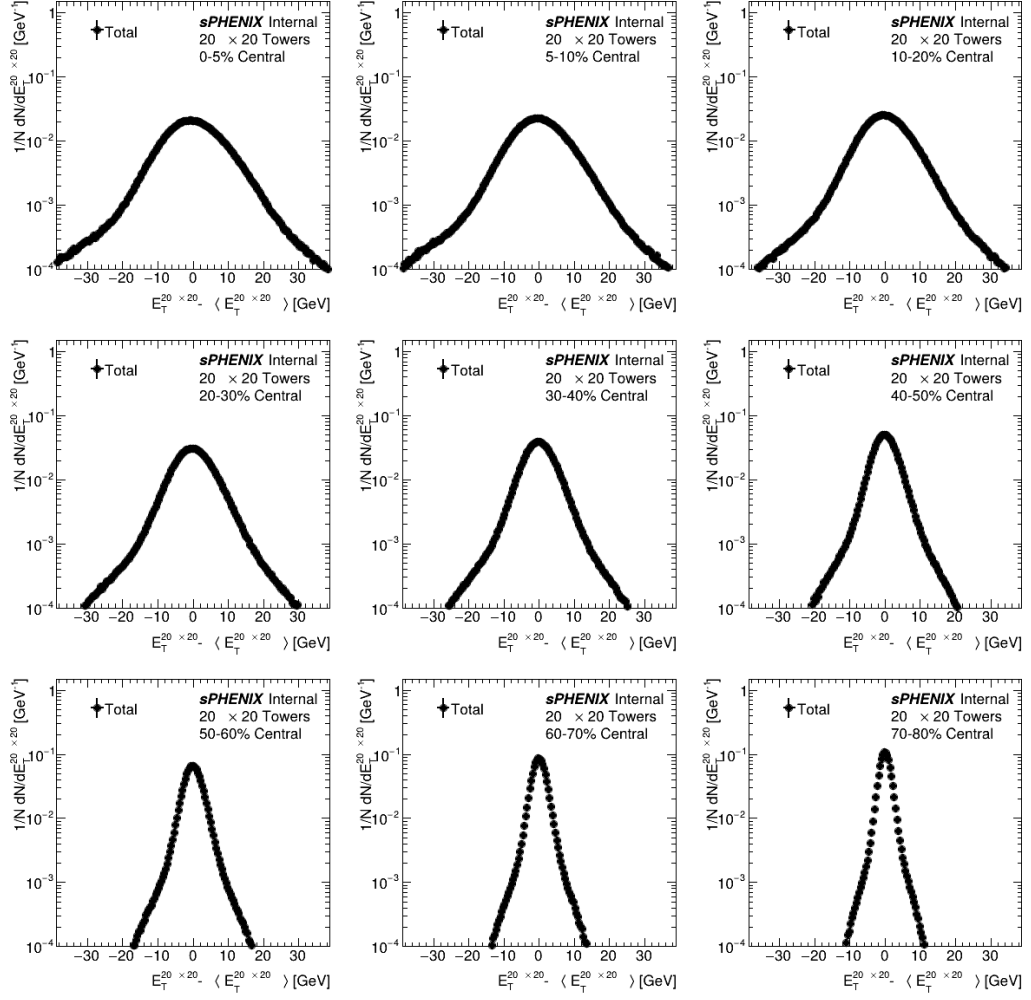


Figure 41: Distribution of $E_T^{20 \times 20} - \langle E_T^{20 \times 20} \rangle$ full calorimeter windows for all centralities.

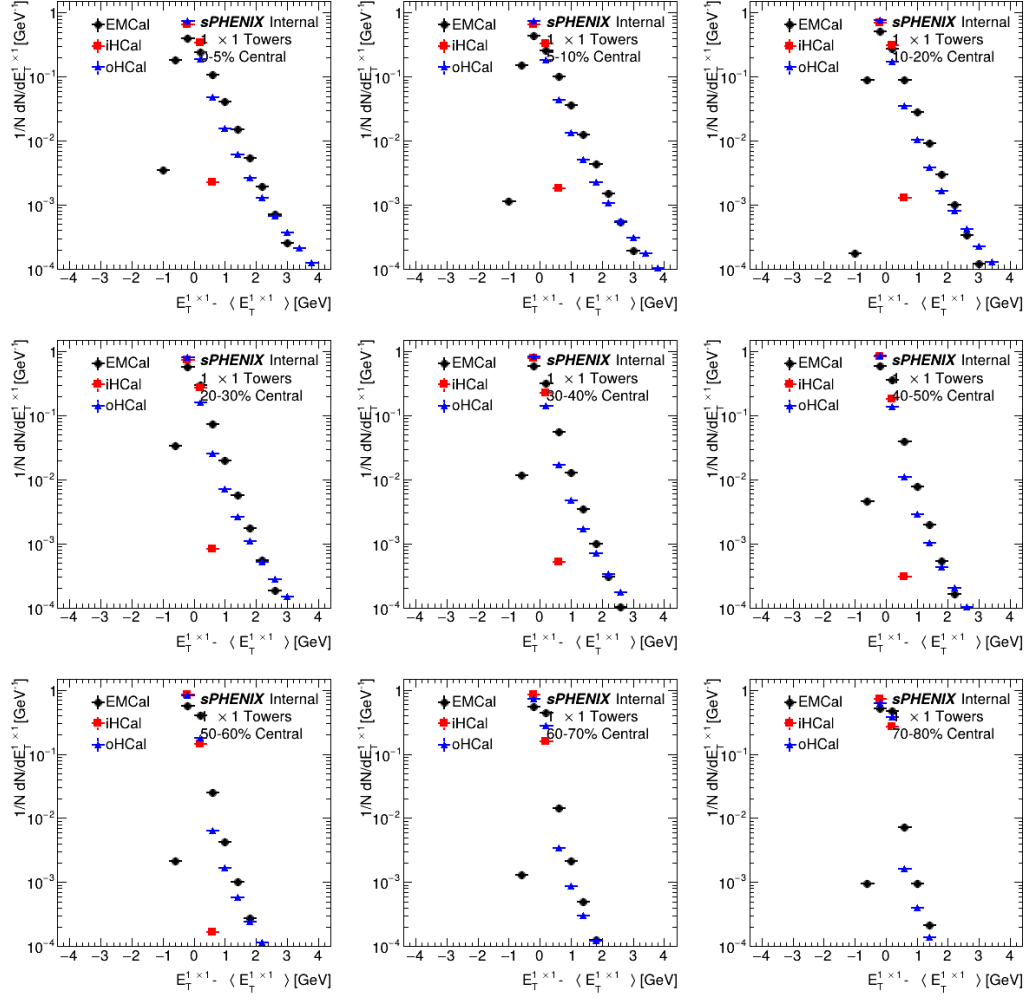


Figure 42: Distribution of $E_T^{1x1} - \langle E_T^{1x1} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

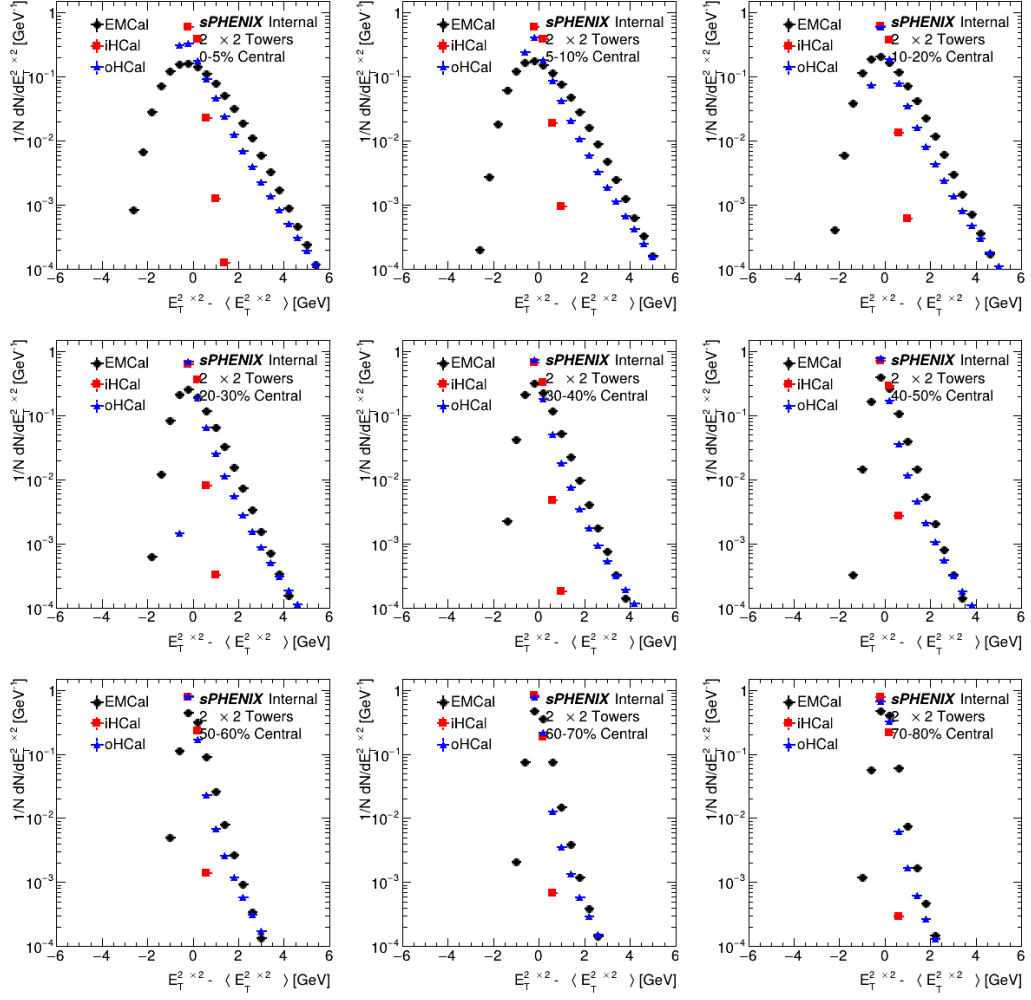


Figure 43: Distribution of $E_T^{2x2} - \langle E_T^{2x2} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

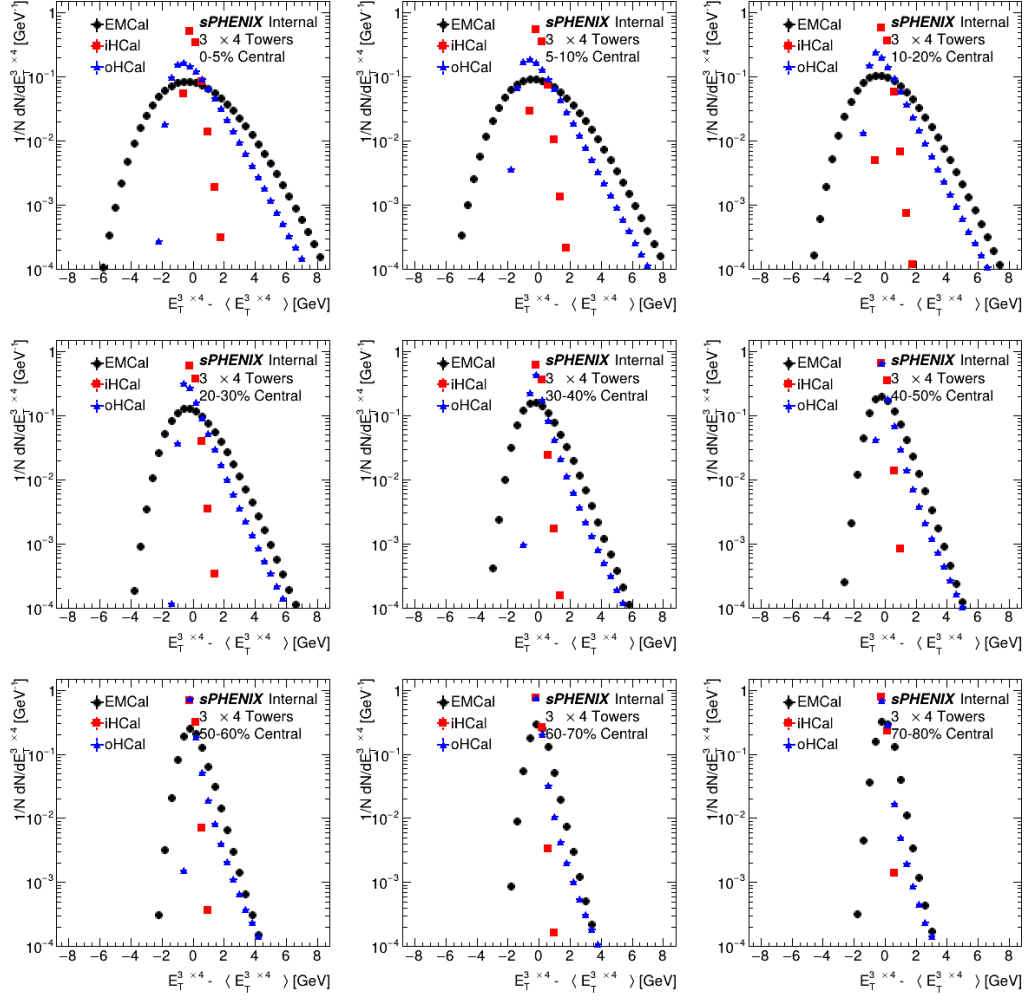


Figure 44: Distribution of $E_T^{3x4} - \langle E_T^{3x4} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

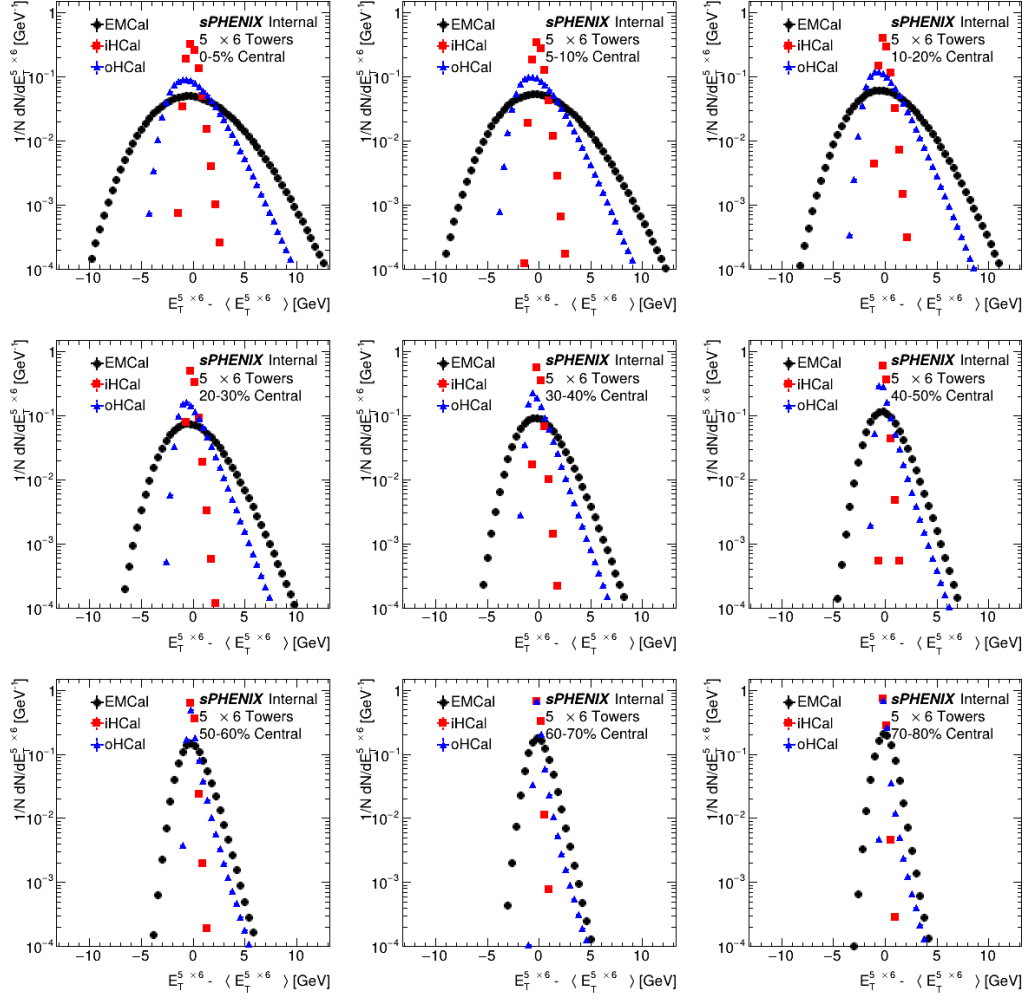


Figure 45: Distribution of $E_T^{5x6} - \langle E_T^{5x6} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

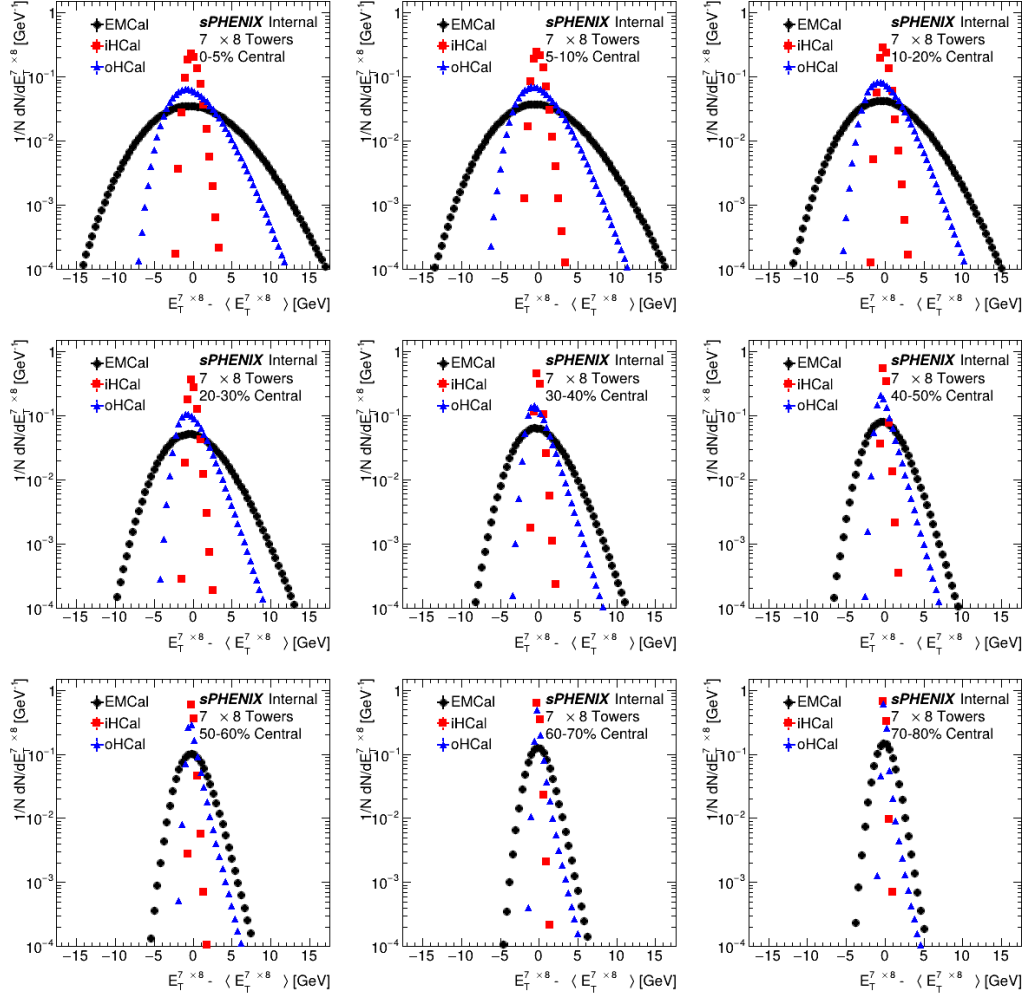


Figure 46: Distribution of $E_T^{7x8} - \langle E_T^{7x8} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

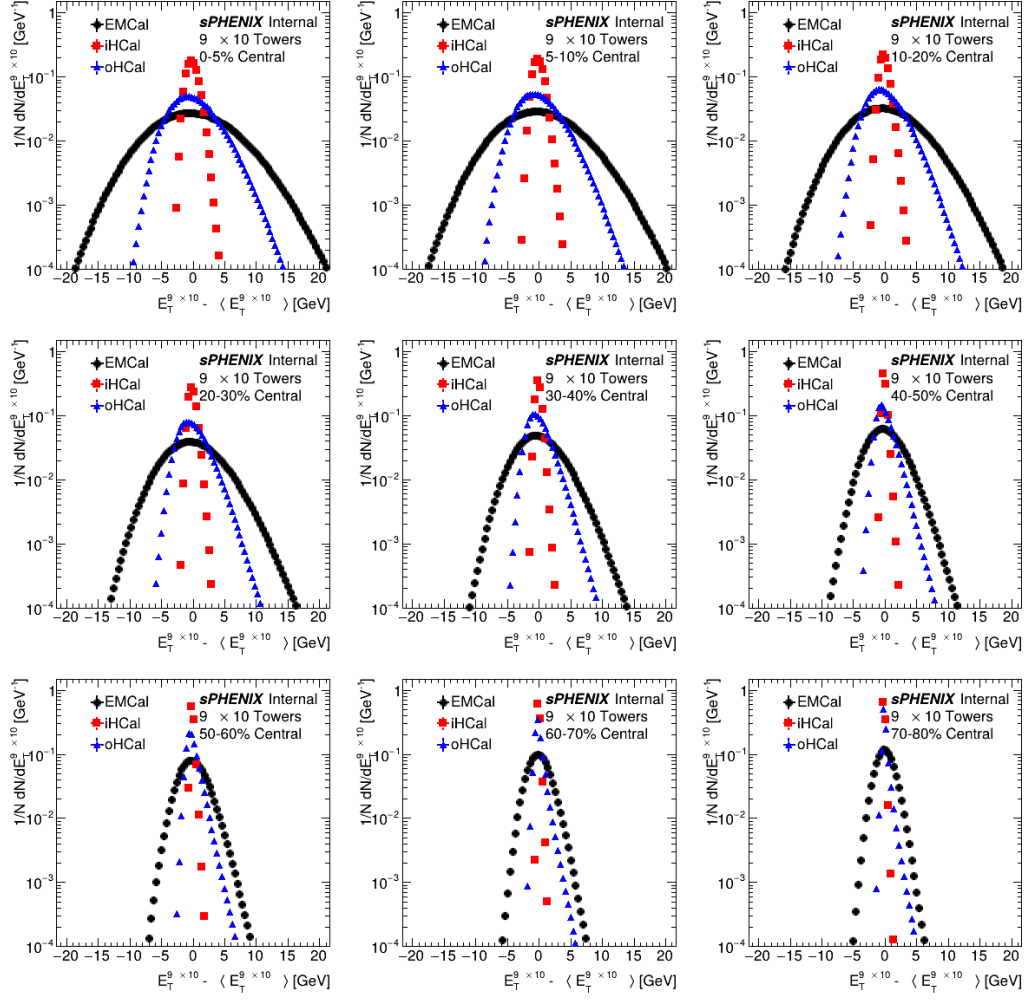


Figure 47: Distribution of $E_T^{9 \times 10} - \langle E_T^{9 \times 10} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

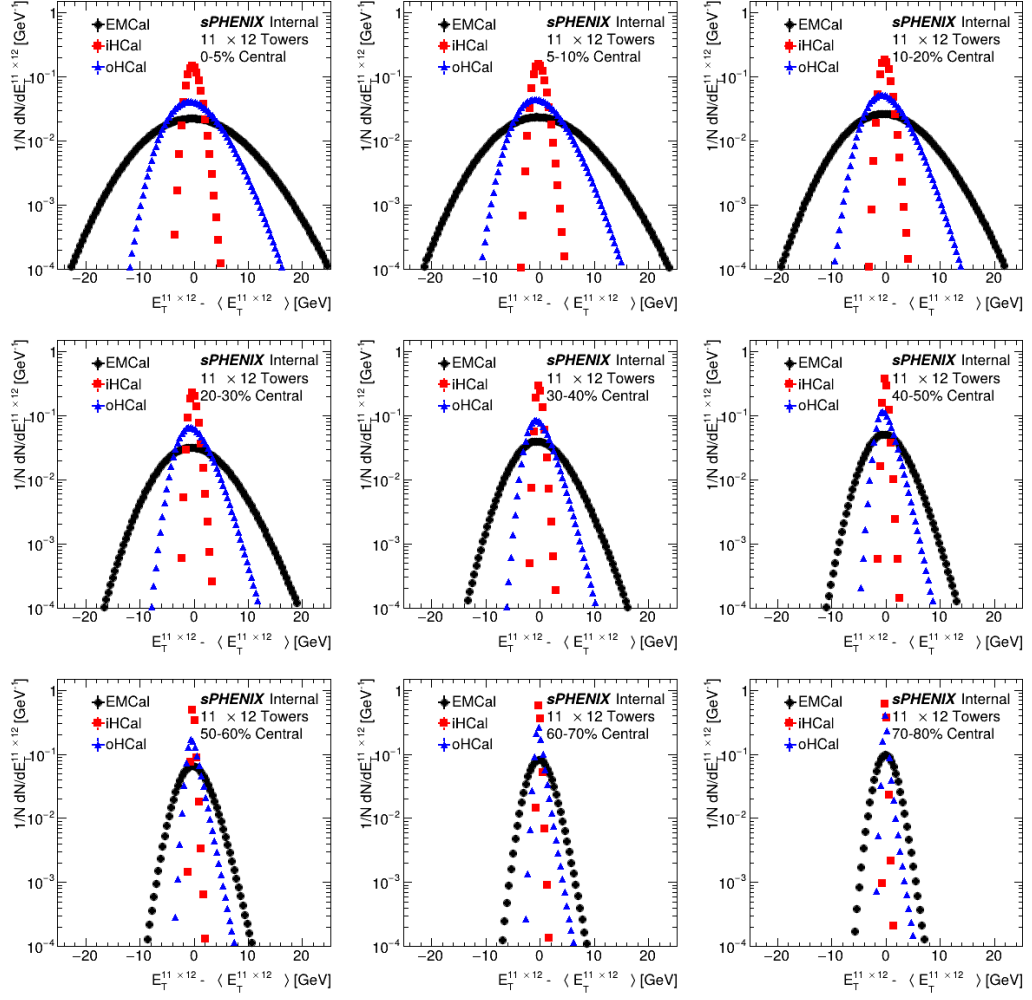


Figure 48: Distribution of $E_T^{11 \times 12} - \langle E_T^{11 \times 12} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

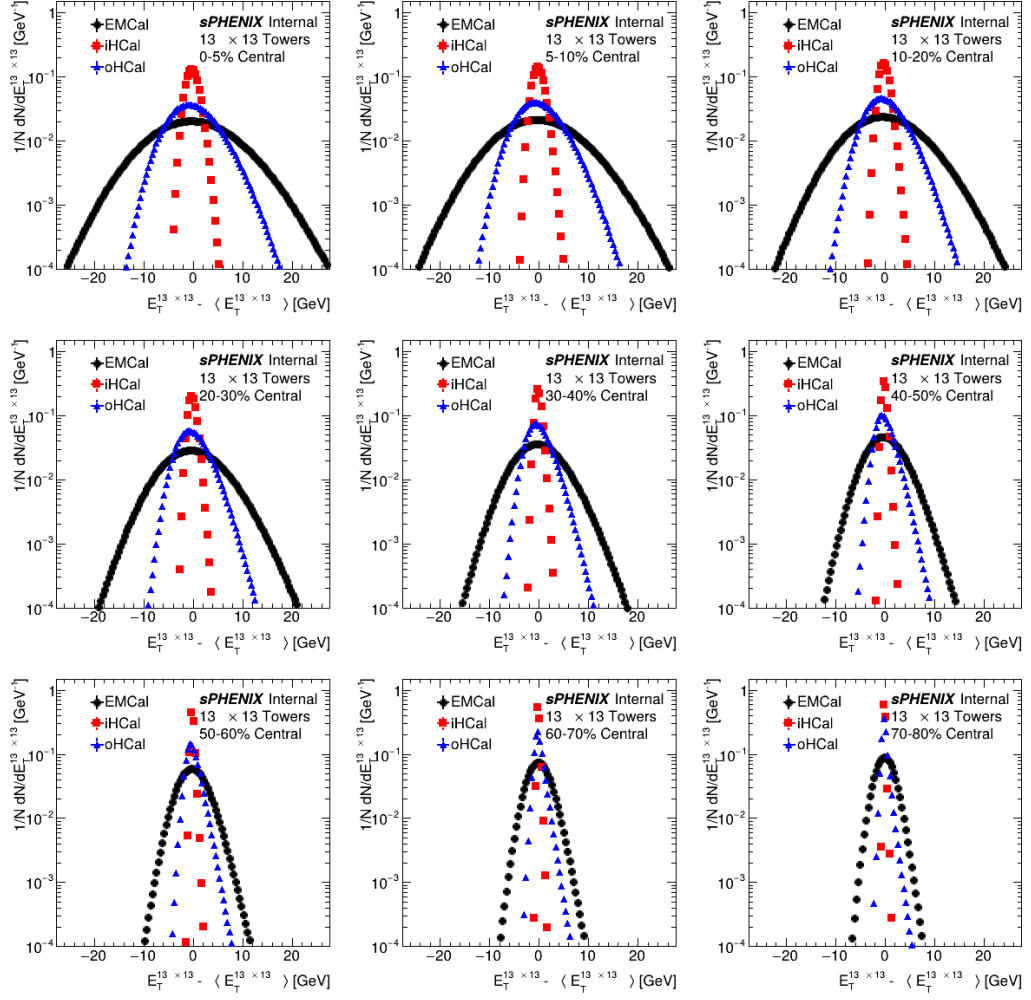


Figure 49: Distribution of $E_T^{13 \times 13} - \langle E_T^{13 \times 13} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

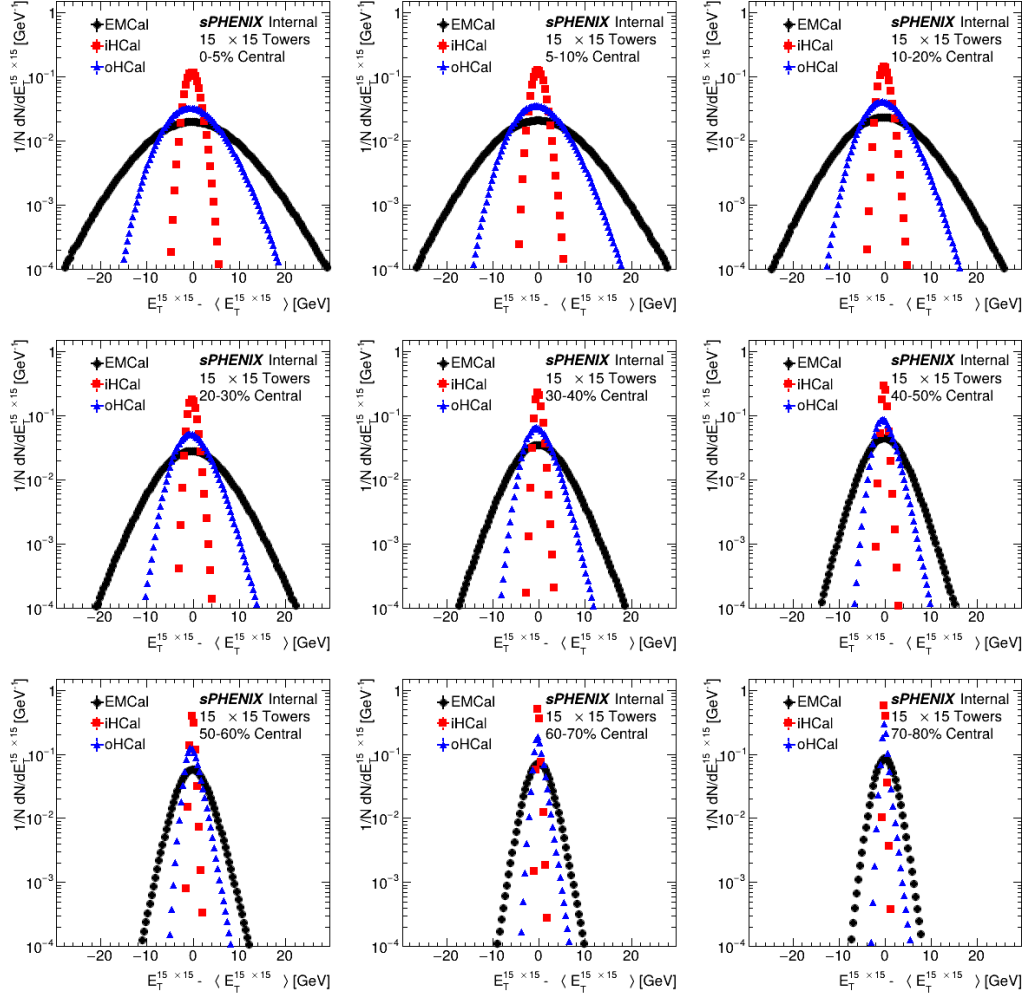


Figure 50: Distribution of $E_T^{15x15} - \langle E_T^{15x15} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

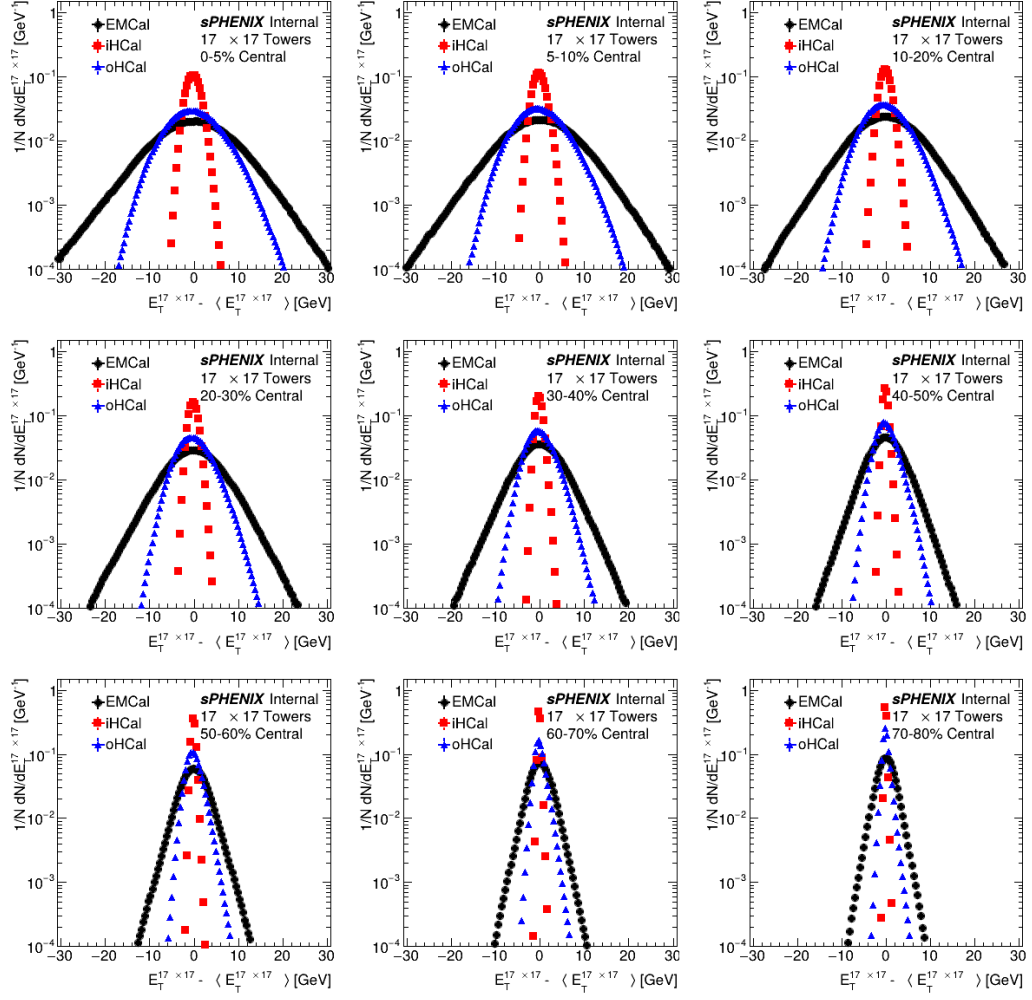


Figure 51: Distribution of $E_T^{17 \times 17} - \langle E_T^{17 \times 17} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

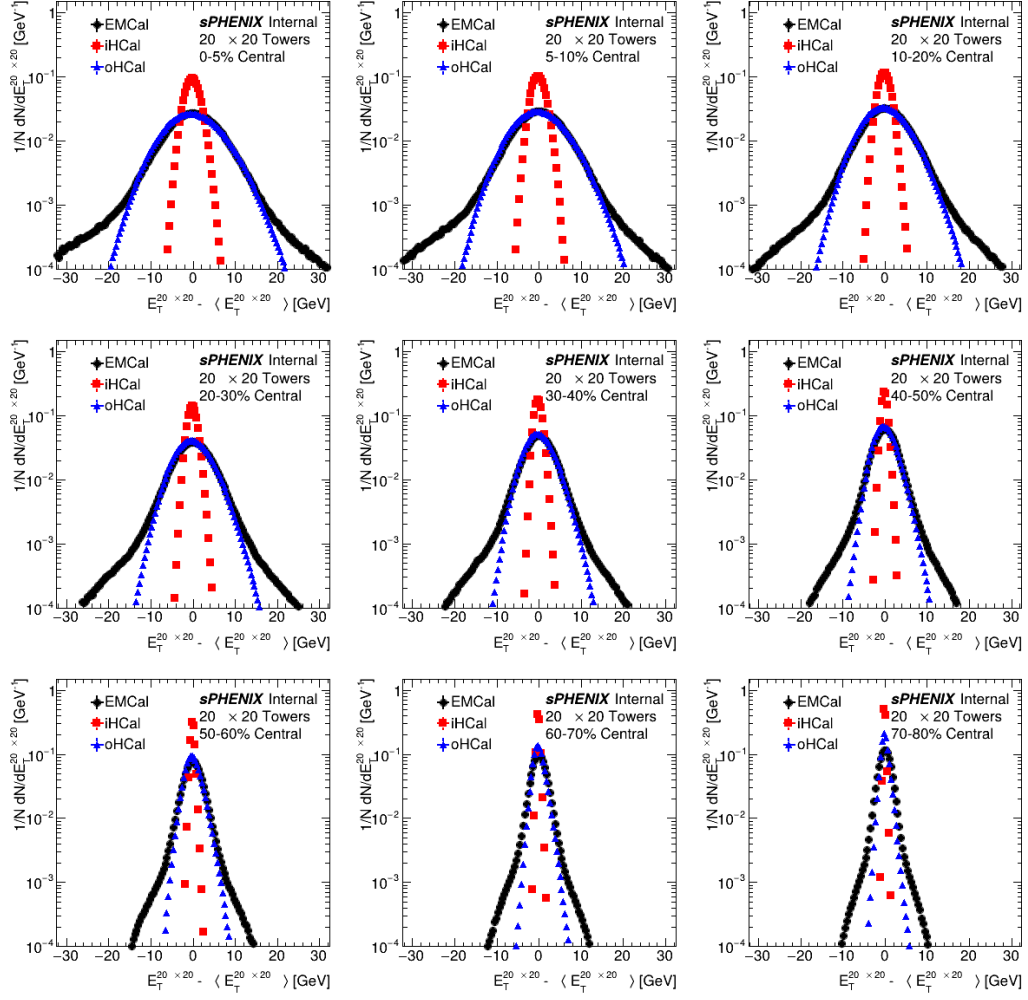


Figure 52: Distribution of $E_T^{20 \times 20} - \langle E_T^{20 \times 20} \rangle$ for each sub-system (EMCAL, iHCAL, and oHCAL) calorimeter windows for all centralities.

B.3 Window Energy Fractions

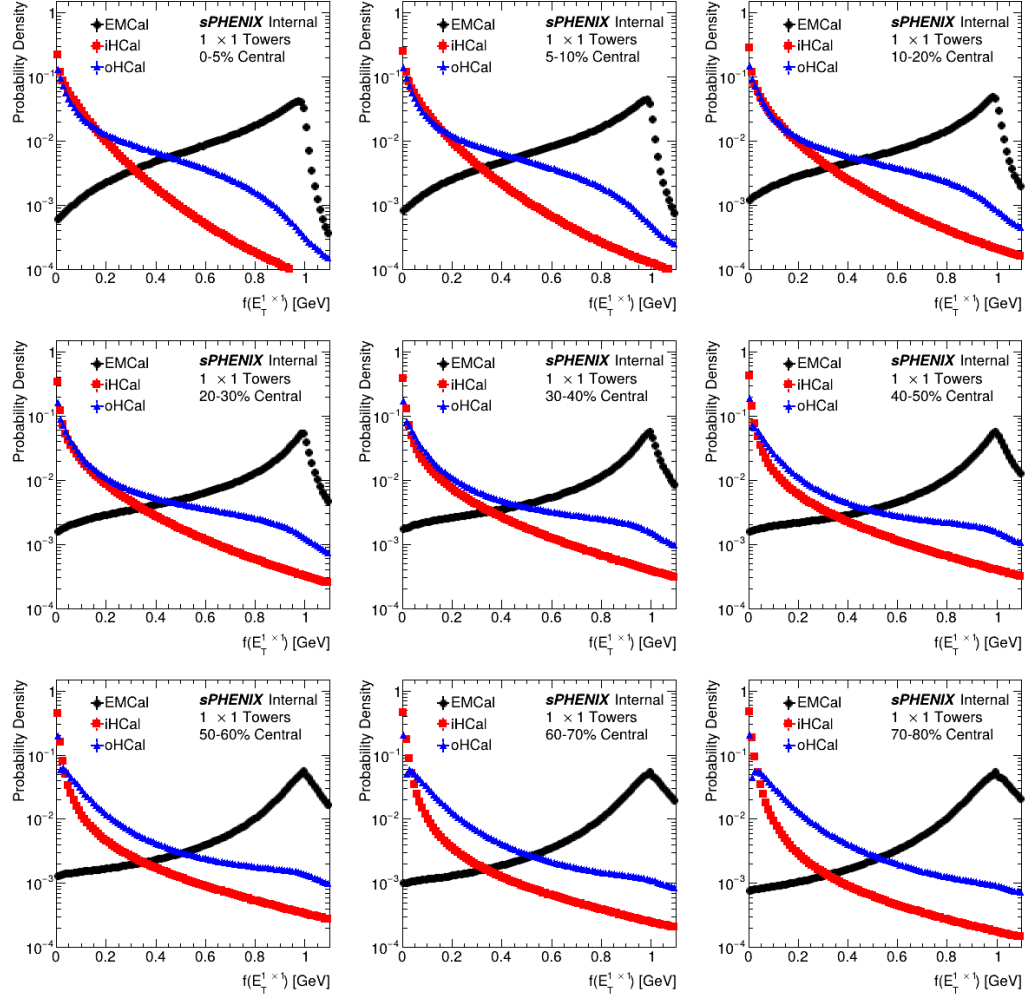


Figure 53: Distribution of fractional contribution $f(E_T^{1x1})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.

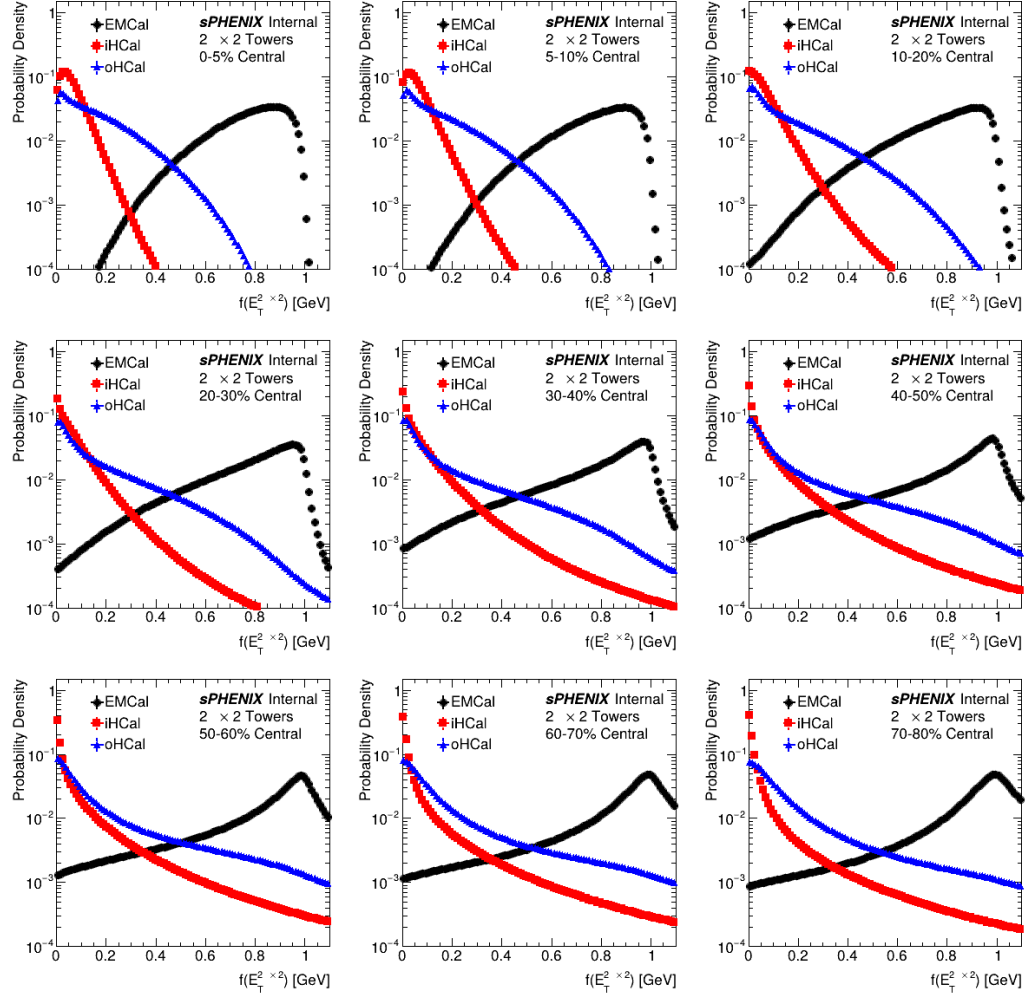


Figure 54: Distribution of fractional contribution $f(E_T^{2x2})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.

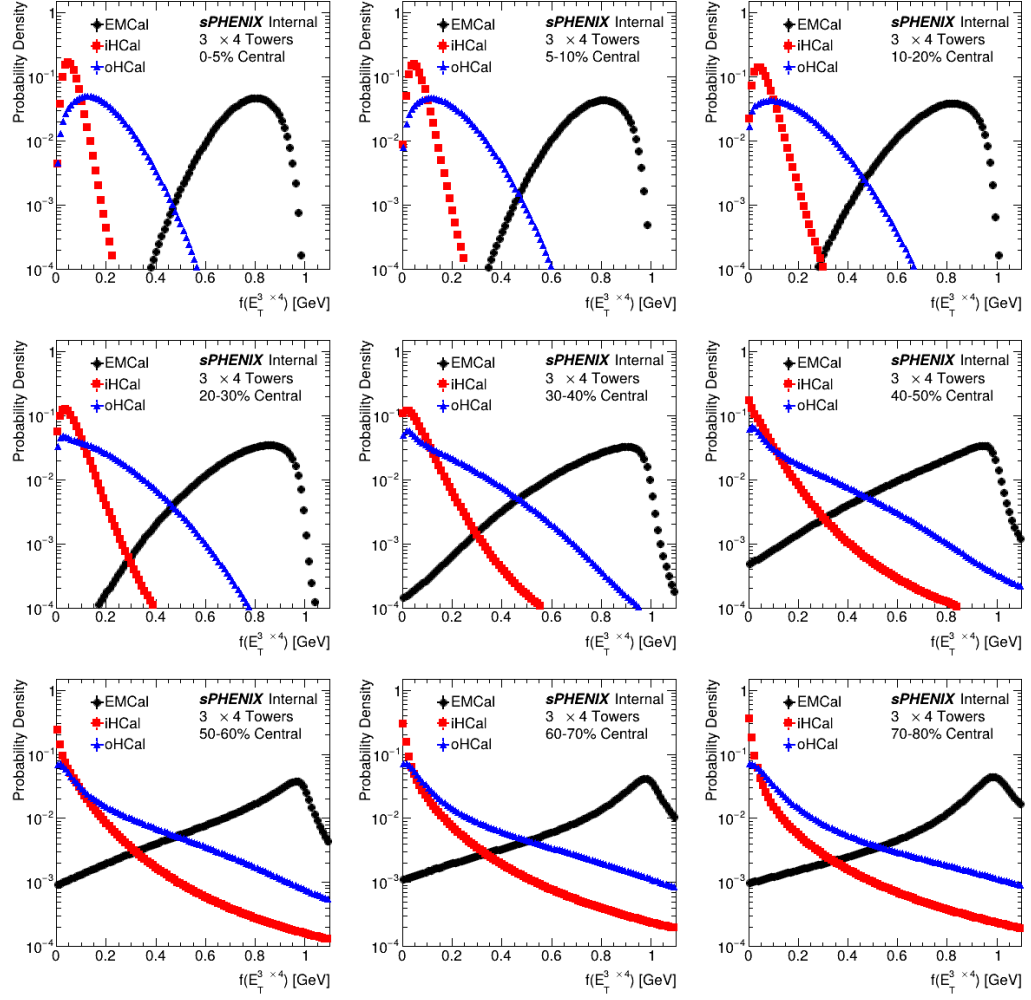


Figure 55: Distribution of fractional contribution $f(E_T^{3x4})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.

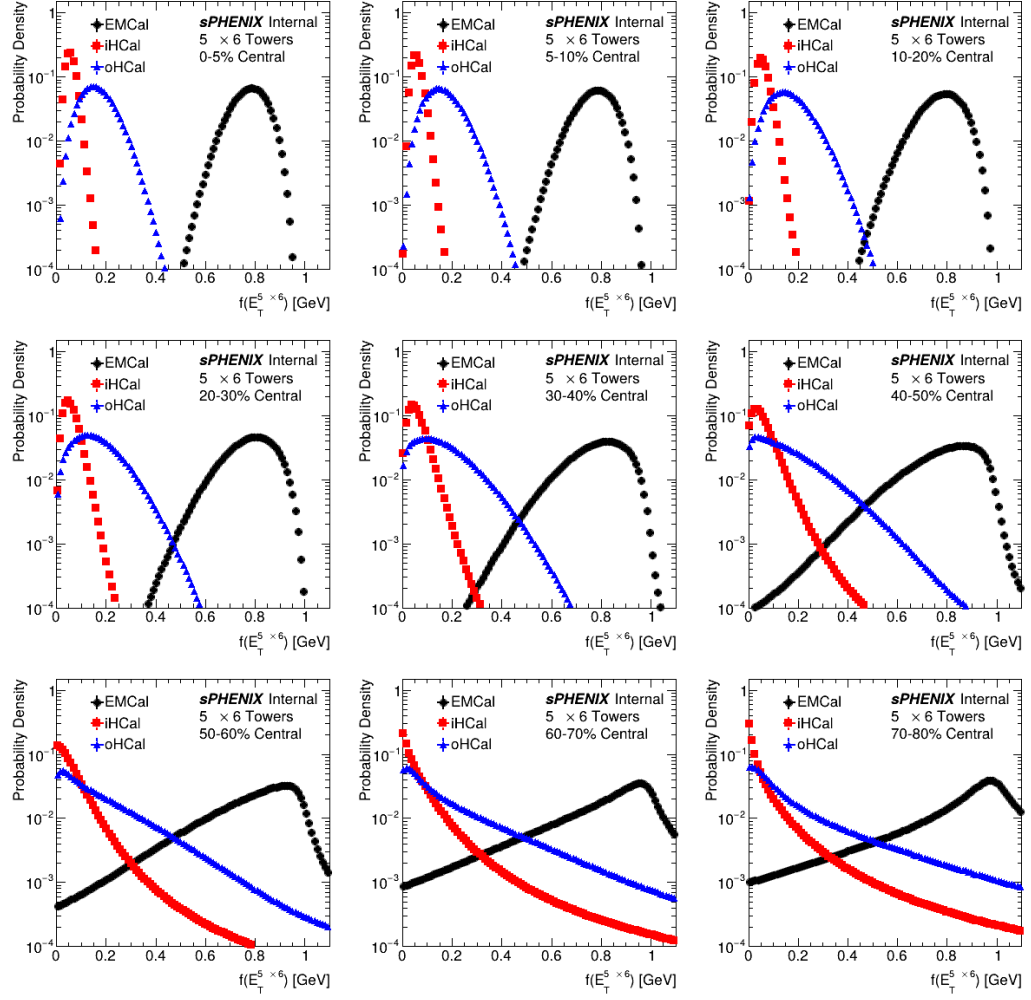


Figure 56: Distribution of fractional contribution $f(E_T^{5x6})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.

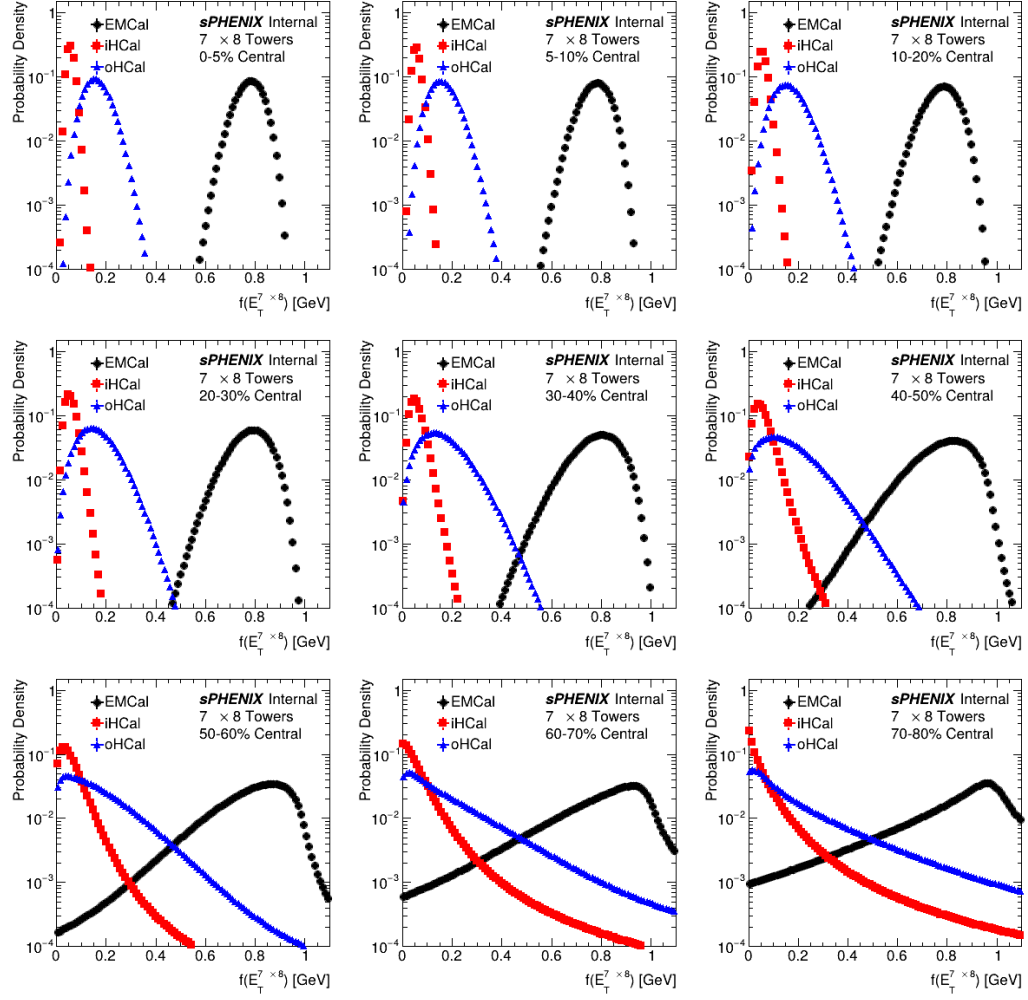


Figure 57: Distribution of fractional contribution $f(E_T^{7x8})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.

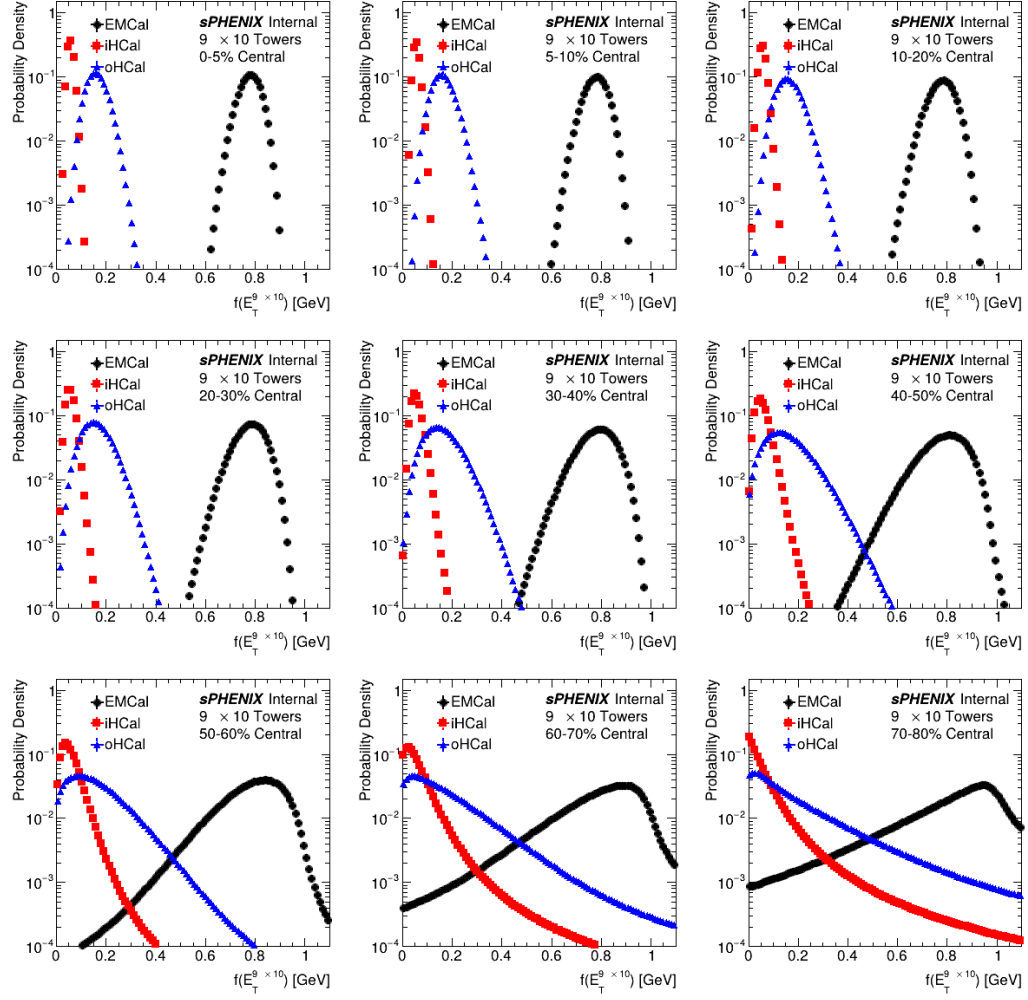


Figure 58: Distribution of fractional contribution $f(E_T^{9x10})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.

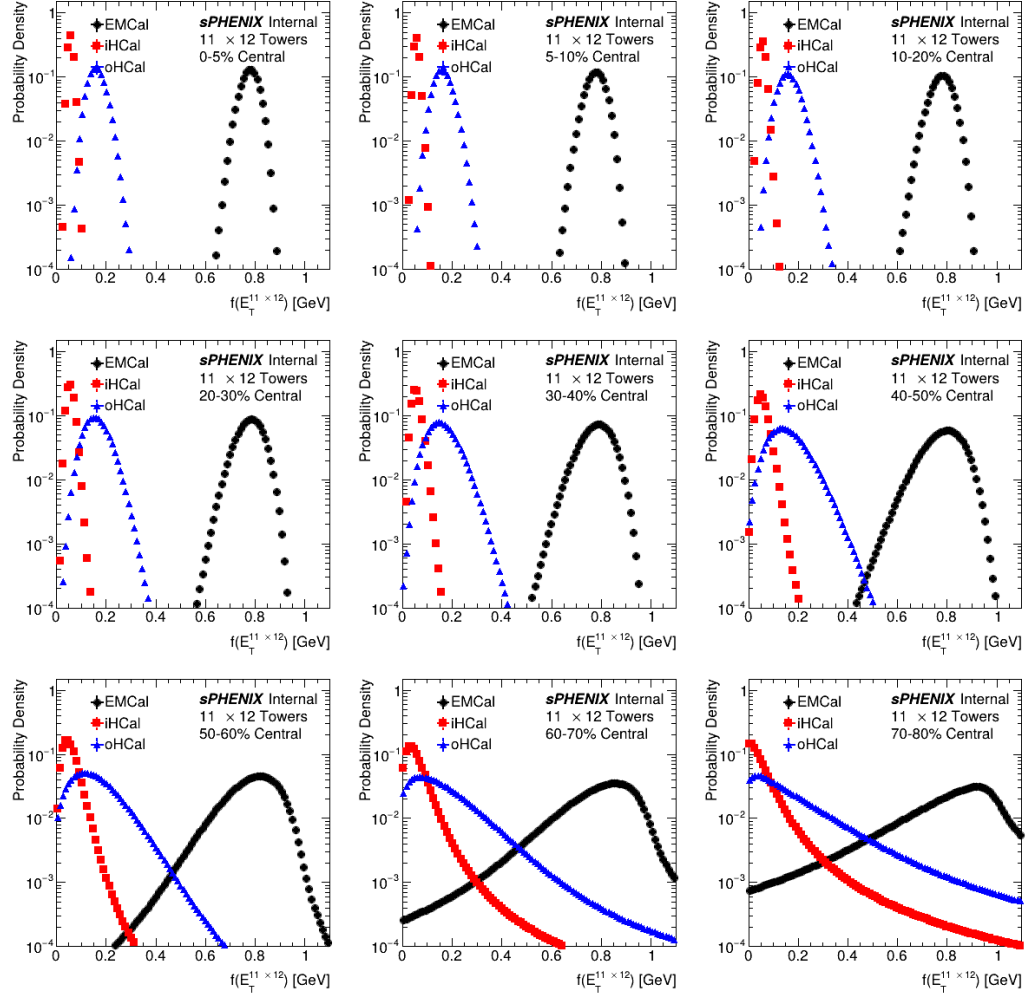


Figure 59: Distribution of fractional contribution $f(E_T^{11 \times 12})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.

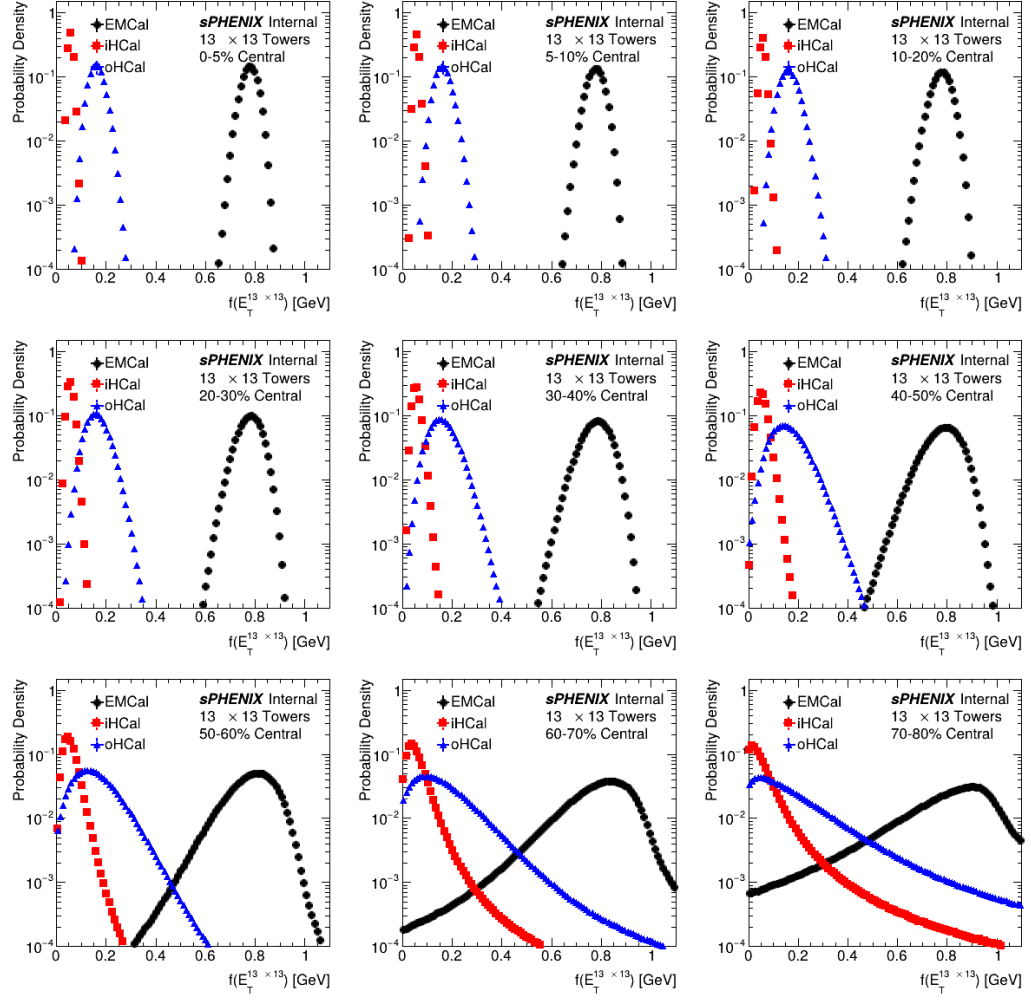


Figure 60: Distribution of fractional contribution $f(E_T^{13 \times 13})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.

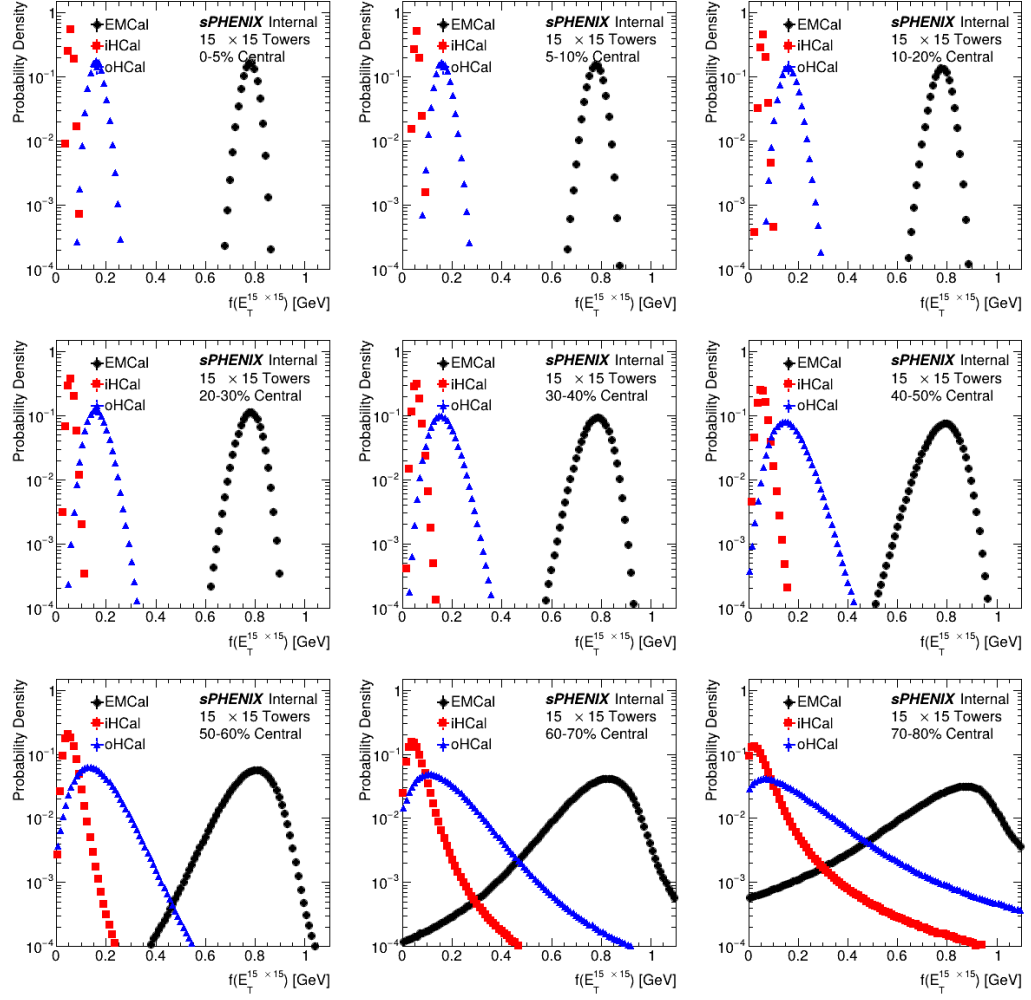


Figure 61: Distribution of fractional contribution $f(E_T^{15 \times 15})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.

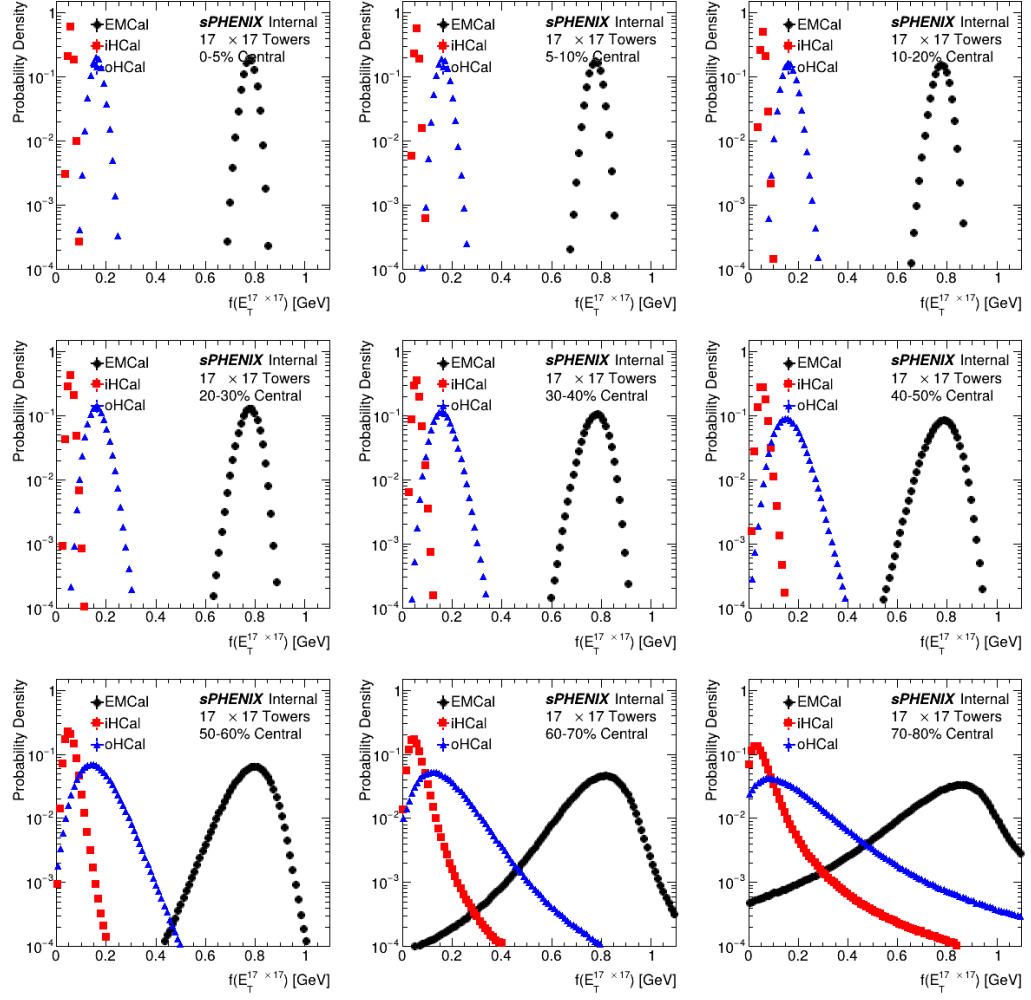


Figure 62: Distribution of fractional contribution $f(E_T^{17 \times 17})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.

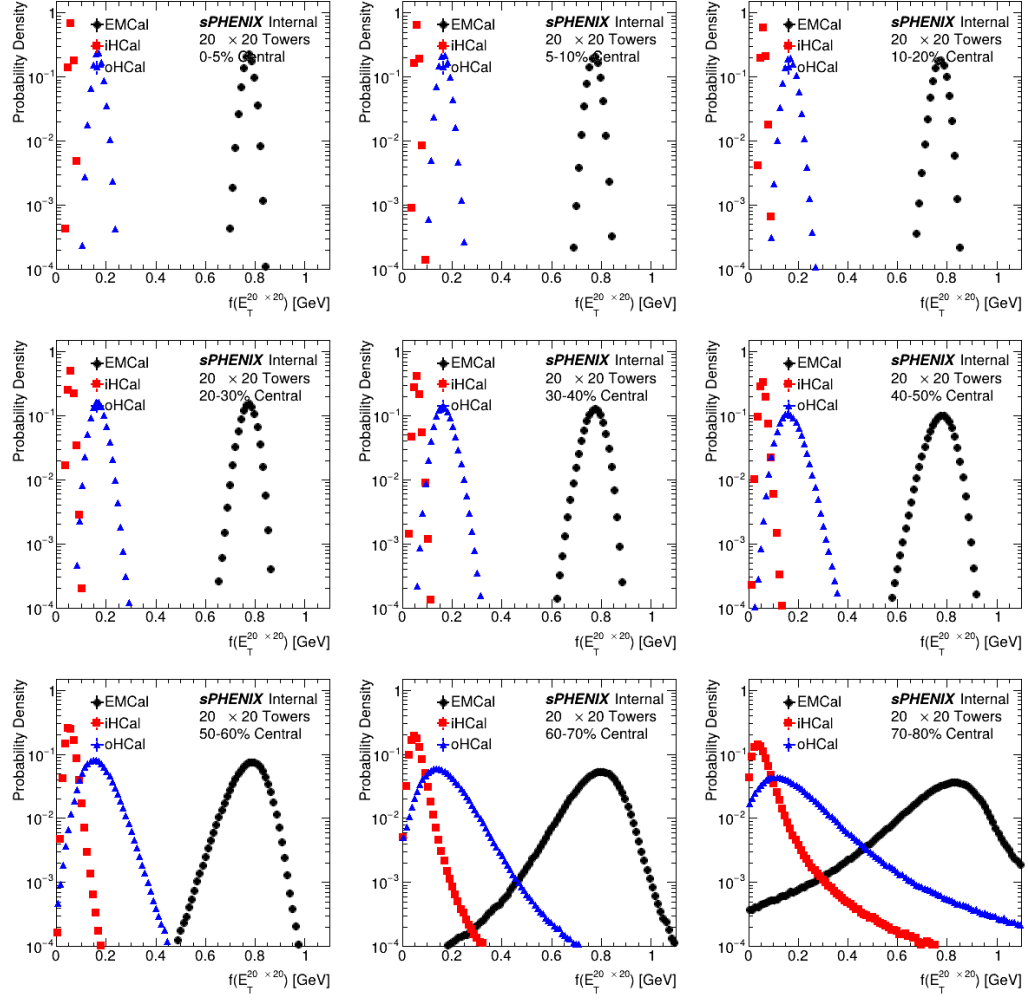


Figure 63: Distribution of fractional contribution $f(E_T^{20 \times 20})$ from each sub-system (EMCAL, iHCAL, and oHCAL) to full calorimeter windows.

B.4 Window Standard Deviations and Averages

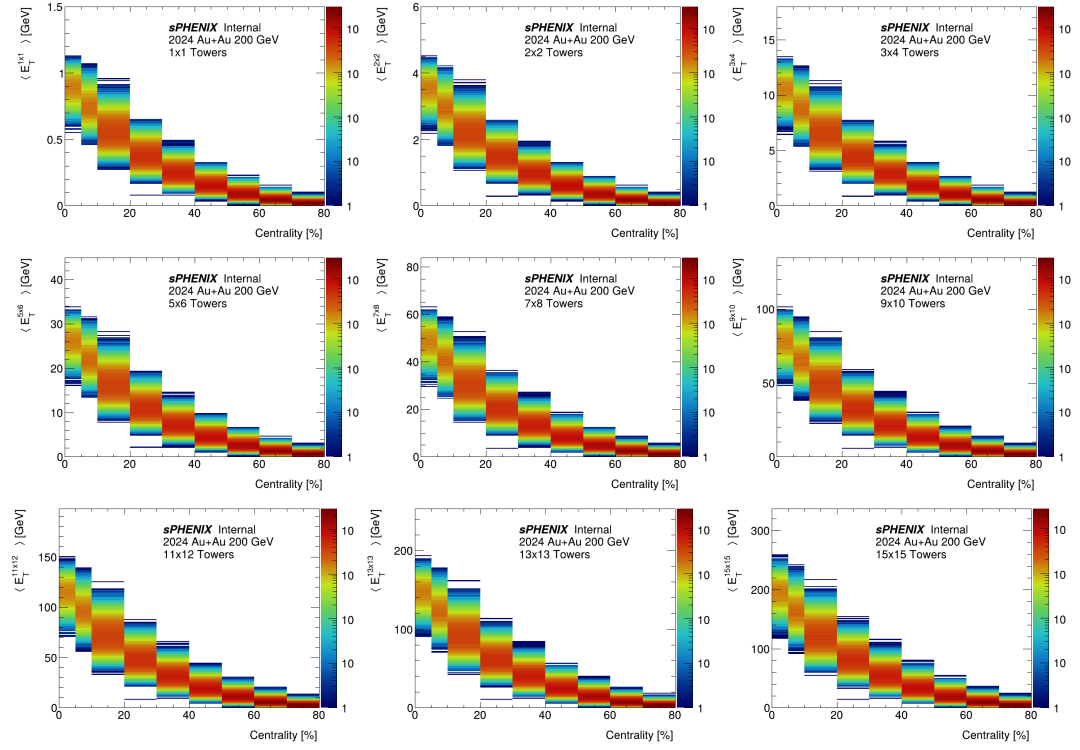


Figure 64: $\langle \delta E_T \rangle$ for all calorimeter window sizes as a function event centrality.

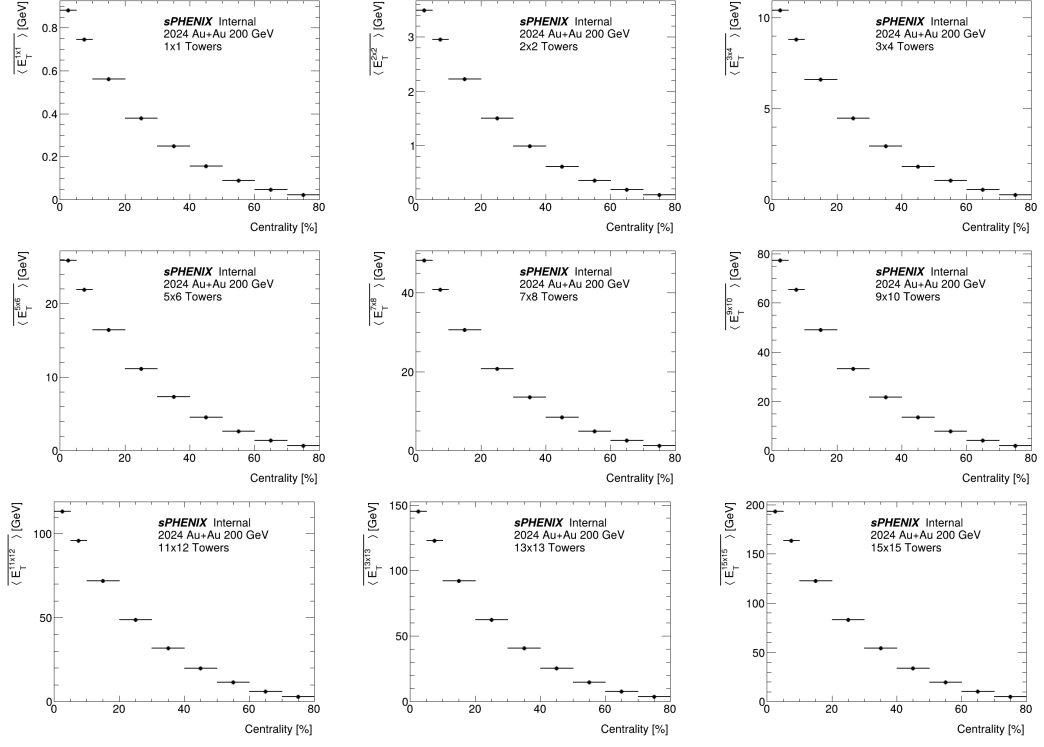


Figure 65: $\langle \delta E_T \rangle$ for all calorimeter window sizes as a function event centrality.

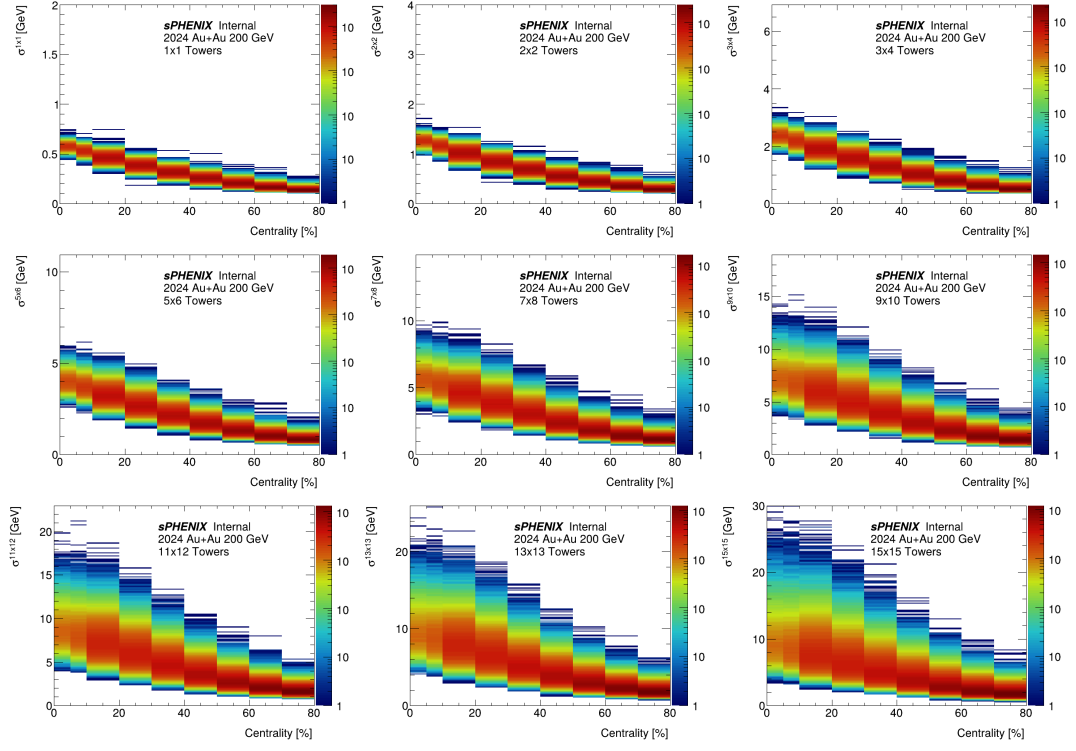


Figure 66: $\langle \delta E_T \rangle$ for all calorimeter window sizes as a function event centrality.

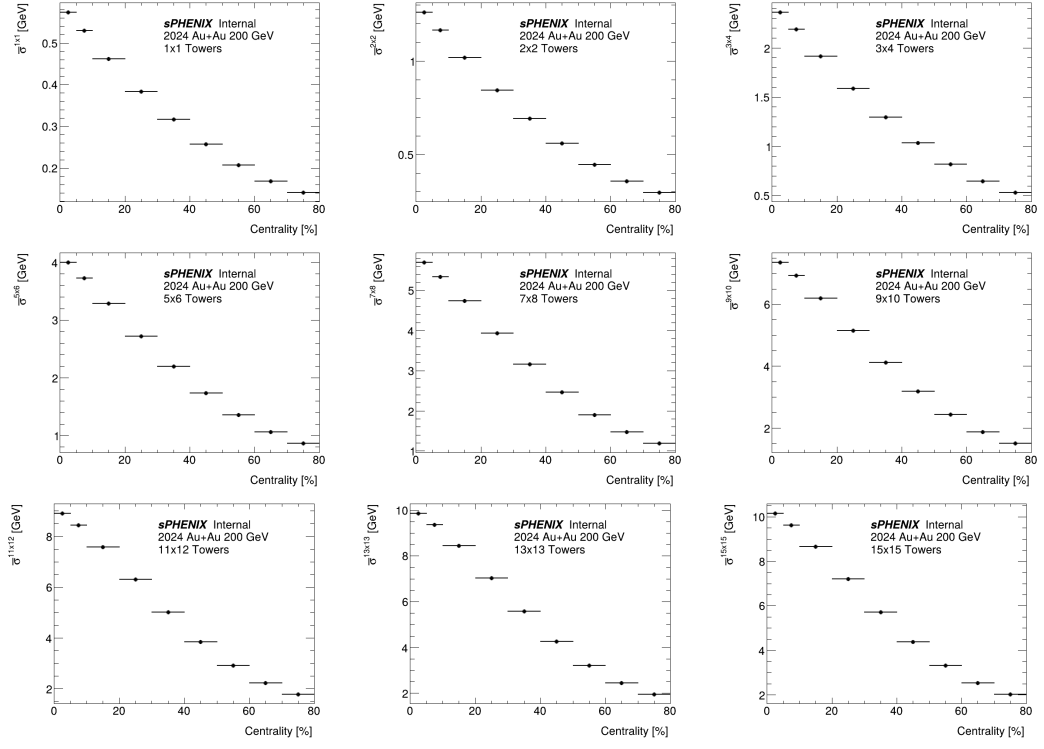


Figure 67: $\langle \delta E_T \rangle$ for all calorimeter window sizes as a function event centrality.

B.5 Window Area Correlation Fits

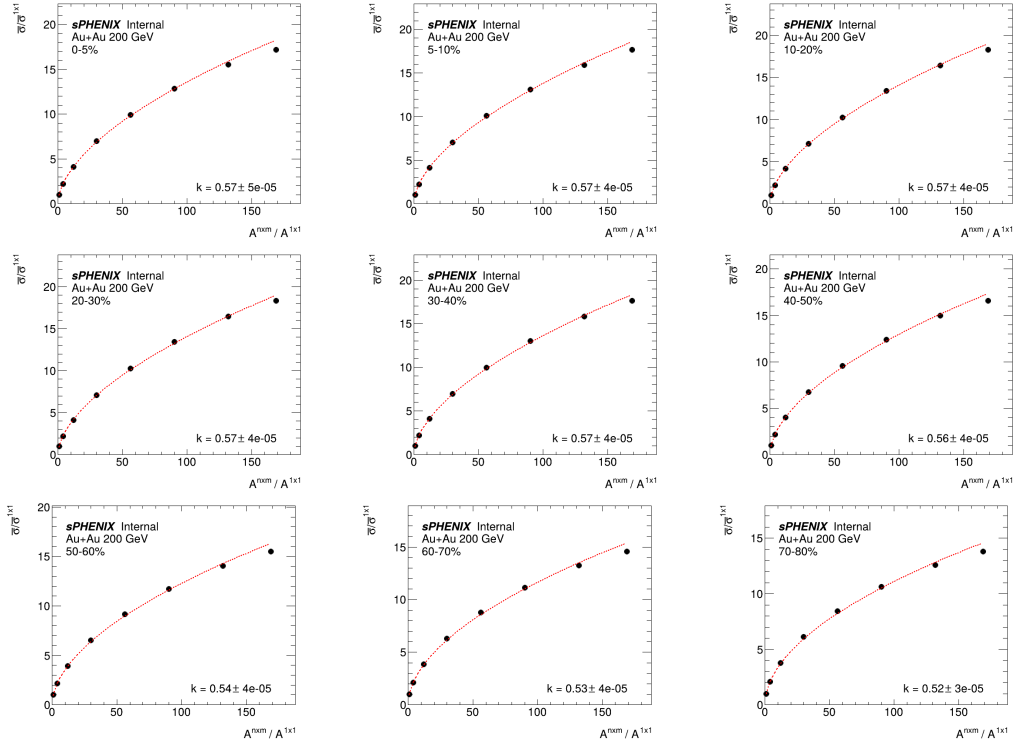


Figure 68: Fits on average standard deviation dependence on calorimeter window area to Eq. 5.1.

Vita

Tanner Austin Mengel was born in Harrisburg, Pennsylvania on September 16, 1997. He was raised in Allen, Texas, and graduated from Farragut High School in Knoxville, TN in 2016. He began his undergraduate studies in physics at the University of Tennessee, Knoxville. Tanner always had been interested in understanding the natural world. In his junior year, he began research in the field of relativistic heavy ion collisions, which is how he fell in love with the field. Tanner received his Bachelors in physics in May 2020. After graduation, he chose to remain at the University of Tennessee to continue his research and pursue a Doctor of Philosophy degree in physics.

Tanner's research interests include the study of jets in heavy ion collisions and small systems. His research brought him to work for the sPHENIX collaboration, which operates a detector on the Relativistic Heavy Ion Collider at Brookhaven National Lab in Upton, New York. His PhD research focused on underlying event characterizations and jet background subtraction methods. He also has worked on developing a method for measuring jet azimuthal anisotropy using cumulants and interpretable machine learning applications to nuclear physics. Tanner defended his PhD on February 26, 2025, and is scheduled to begin a post-doctoral research position focusing on jet analysis and detector operations for sPHENIX at the University of Colorado, Boulder on March 1, 2025.