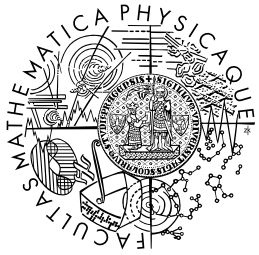

FACULTY
OF MATHEMATICS
AND PHYSICS
Charles University

MASTER THESIS

Matej Haviernik

Neutral pion identification at Future Circular Collider

Institute of Particle and Nuclear Physics

Supervisor of the master thesis: Mgr. Jana Faltová, Ph.D.

Study programme: Physics

Study branch: Particle and Nuclear Physics

Prague 2023

I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources. It has not been used to obtain another or the same degree.

I understand that my work relates to the rights and obligations under the Act No. 121/2000 Sb., the Copyright Act, as amended, in particular the fact that the Charles University has the right to conclude a license agreement on the use of this work as a school work pursuant to Section 60 subsection 1 of the Copyright Act.

In ..... date ....

Author's signature

I would like to thank my supervisor, Jana, for the immense support provided during writing the thesis. Thanks to her solid advice, patient consultations and understanding I was able to finish the thesis with motivation and enthusiasm and push my knowledge of programming and multivariate methods further. I credit her with introducing me to popularisation of science during my two years of master's degree, which led me to look at physics with newfound excitement.

I would further like to thank my partner David; he was a neverending source of empathy, love and understanding during times I was facing stress and nerves from impending deadlines or after heated arguments with lines of code that refused to work.

Title: Neutral pion identification at Future Circular Collider

Author: Matej Haviernik

Institute: Institute of Particle and Nuclear Physics

Supervisor: Mgr. Jana Faltová, Ph.D., Institute of Particle and Nuclear Physics

Abstract: Future Circular Collider (FCC) is a 100 km long particle collider to be built around the year 2040 in the CERN laboratory. The first stage of operation will be a lepton collider FCC-ee aiming to test the Standard model with unprecedented precision at maximal central energies of 365 GeV. Neutral pions originating from such collisions are crucial for reconstruction of particles such as the τ lepton and their identification poses a challenge for detectors. Neutral pions decay almost immediately into a pair of photons separated by a small angle and can be easily misidentified as a single photon. We should be able to distinguish the signal of a neutral pion from the signal of a single photon with a fine segmented calorimeter. In the thesis we will work with the FCC-ee noble liquid calorimeter design. The main goal of the thesis is to understand the geometry of the calorimeter planned for FCC-ee experiment and investigate the options offered by multivariate analysis methods for the reconstruction and identification of neutral pions against a single photon background.

Keywords: calorimeter FCC pions

Název práce: Identifikace neutrálních pionů na Future Circular Collider

Autor: Matej Haviernik

Ústav: Ústav částicové a jaderné fyziky

Vedoucí diplomové práce: Mgr. Jana Faltová, Ph.D., Ústav částicové a jaderné fyziky

Abstrakt: Future Circular Collider (FCC) je 100 km dlhý urýchľovač častíc, ktorého výstavba sa plánuje okolo roku 2040 v laboratóriu CERN. Prvú etapu bude tvoriť leptónový urýchľovač FCC-ee, ktorého cieľom bude otestovať Štandardný model s nebývalou presnosťou pri najvyššej centrálnej energii 365 GeV. Neutrálne pióny pochádzajúce z takýchto zrážok sú dôležité na rekonštrukciu častíc ako napríklad τ leptóny a ich identifikácia predstavuje výzvu pre detektory. Neutrálne pióny sa rozpadajú na pár fotónov s veľmi malým rozletovým uhlom, kvôli čomu sa dajú ľahko zameniť za jeden fotón. Mali by sme byť schopní rozlíšiť signál z neutrálneho piónu od signálu z jediného fotónu v kalorimetri s dostatočne jemnou segmentáciou. V tejto práci budeme pracovať s návrhom kalorimetru pre FCC-ee, ktorý využíva kvapalné vzácne plyny. Hlavným cieľom našej práce je porozumieť stavbe kalorimetra plánovaného pre experiment FCC-ee a preskúmať možnosti identifikácie neutrálnych piónov na fotónovom pozadí pomocou "multivariate analysis" metód.

Klíčová slova: kalorimetr FCC piony

Contents

Introduction

As the High Luminosity LHC project continues to be the predominant program at CERN, projects aiming to continue improving the state of physics after the HL-LHC program are being considered. One of the proposals is a new collider built with a 100 km circumference. The operation of the collider would be divided into several stages, with the first stage being an electron-positron collider FCC-ee aiming to improve accuracy of measured properties of heavy vector bosons, the Higgs boson and the top quark. The following phase FCC-hh would run proton-proton collisions with maximal luminosity reaching $L = 3.10^{35}$ cm^{-2}s^{-1}. The collider would be at the frontier of probing for physics beyond Standard model and measuring currently known couplings with heretofore unreached precision.

The neutral pions created in collision decay predominantly into two photons, contributing significantly to electromagnetic showers in the hadronic and electromagnetic calorimeter. They are an abundant final state product of numerous processes and their identification is important in subsequent reconstruction of τ leptons and various hadrons. Due to the Lorentz boost of the pion frame, the decay products are observed to be highly collimated with the decay angle rapidly decreasing with growing energy. The resultant showers in the calorimeter largely overlap and due to this it then becomes challenging to distinguish di-photon clusters originating from π^0 from background of single photons.

Because clustering algorithms tend to group the two decay photons into a single cluster at higher energies, it becomes necessary to turn to other methods when seeking satisfactory π^0 detection. Aside from cluster information, we can observe the information provided by calorimetric cells directly that reveal the internal structure of the cluster and the profile of energy deposition in layers. The variables characterizing the structure of clusters and energy deposition can be subsequently analyzed via multivariate analysis methods. In this thesis, we will work with data containing Monte Carlo simulations of passage of particles through calorimetric environment at the FCC-ee. Two data samples in the energy range of 0-100 GeV were used for the study — one consisting of single π^0, the other of single photons — containing information on reconstructed clusters and cells. We will attempt to define the pions resolved by the clustering algorithm. Next we will compare two multivariate analysis methods — the Rectangular Cuts method and Boosted Decision Trees — to determine the reconstruction efficiency of pions against single photon background.

1. Neutral pion decay

1.1 Standard Model

The Standard Model is a spontaneously broken non-Abelian gauge theory defined by the local SU(3)×SU(2)×U(1) symmetry. It describes interactions between elementary spin $\frac{1}{2}$ fermions (distinguished by flavor and ordered into three generations), spin 1 boson fields ($W^{\pm}, Z, \gamma, G$) and one Higgs scalar field H. A non-Abelian field theory is a theory, where generators of the appropriate internal symmetry do not commute. Explicitly, for a transformation of a field Ψ [1]

$$\begin{aligned} \Psi' &= S\Psi \\ \overline{\Psi}' &= \overline{\Psi}S^{-1} \end{aligned} \tag{1.1}$$

we can write the S as an exponential in the form

$$S = exp(i\alpha^a T^a) \tag{1.2}$$

where generators τ^a satisfy the commutation relation

$$[T^a; T^b] = if^{abc}T^c \tag{1.3}$$

f^{abc} being the appropriate structure constant of the corresponding Lie group. For SU(2) group the generators of the corresponding multiplet are conventionally chosen as normalized Pauli matrices

$$T^a = \frac{1}{2}\sigma^a$$

and for the SU(3) group the generators become (again normalized) Gell-Mann matrices

$$T^a = \frac{1}{2}\lambda^a$$

Whether the parameters α^a depend on x or not defines if the theory is local or global. The imposition of local gauge invariance is the reason behind the introduction of the covariant derivative D_μ into the Standard Model Lagrangian

$$D_\mu = \partial_\mu - igA^a_\mu T^a \tag{1.4}$$

which in turn results in an interaction term, turning the otherwise free theory into an interacting one. Using a perturbation expansion of an S matrix element into an infinite series of terms containing the interaction term, we can calculate the corresponding matrix elements for various processes up to an arbitrary order. The constant g denotes a coupling constant. A^a_μ is a multiplet of vector fields, which are then responsible for mediating the respective force. These fields transform as [1]

$$A^{a\prime}_\mu = A^a_\mu - f^{abc}\alpha^b A^c_\mu + \frac{1}{g}\partial_\mu \alpha^a \tag{1.5}$$

In a basic gauge theory, the vector bosons are by default massless — the electroweak mediating bosons W and Z acquire mass via interaction with a scalar Higgs field, which also generates masses of elementary fermions. As the Higgs field does not directly couple to the photon or the gluon, they keep their zero masses. A direct consequence of this is that while the weak force mediated by heavy bosons is a short-range force, the electromagnetic field is long-range, being mediated by charge neutral massless photons. However, this principle does not apply to the strong interaction, mediated by massless gluons with non-zero color charge, which display a phenomenon called "color confinement" at low energies — confining quarks into hadrons. This phenomenon results in quantum chromodynamics being unable to perturbatively describe bound states of quarks.

1.2 Properties of the neutral pion

The strong force binding quarks together is mediated by an octet of gluon fields carrying color charge. Neutral pions (which we will also denote as π^0) belong to the octet of light pseudoscalar mesons and are described in the static quark model as the bound state of $u\overline{u}$ and $d\overline{d}$ quarks, being a superposition [2]

$$\frac{u\overline{u} - d\overline{d}}{\sqrt{2}}$$

with the mass of $m_{\pi^0} = (134.9768 \pm 0.0005)$ MeV and the lifetime $\tau_{\pi^0} = (8.43 \pm 0.13)\times10^{-17}$ s [3]. Lifetime of the charged pions $\pi^{\pm}$ is several orders higher, at $(2.6033 \pm 0.0005)\times10^{-8}$ s. The reason for this is, while the decay of charged pions is mediated via the weak force (primary decay products being μ leptons and corresponding ν_μ neutrinos) the neutral pions decay predominantly via the electromagnetic interaction. The main decay channels of π^0 are

$$\pi^0 \to \gamma + \gamma;\ \mathrm{BR} = 0.98823 \pm 0.00034$$
$$\pi^0 \to \gamma + e^+ + e^-;\ \mathrm{BR} = 0.01174 \pm 0.00035$$
$$\pi^0 \to e^+ + e^- + e^+ + e^-;\ \mathrm{BR} = (3.34 \pm 0.16)\times10^{-5}$$

where the two-photon decay mode represents the main contribution to the neutral pion decay width, followed by cases when one of the photon (or both) undergoes a conversion into an electron-positron pair. Such decays are also known as single and double Dalitz decays.

π mesons - both charged and neutral pions - play an important role in modeling strong nuclear force at low energies. Their existence was first predicted by Hideki Yukawa and their role was understood to be that of mediators of residual strong force between nucleons. Nonetheless, pions play an important role at high energies as well. First direct observations of charged pions came from studying cosmic radiation in the upper atmosphere, where they are produced in collisions of high energy protons with the atmosphere. Neutral pions are produced in all modern hadron colliders, where they form an important part of hadronic showers in calorimeters and contribute to their electromagnetic portion via the important double-photon decay. They also play an important role in τ lepton reconstruction, being a part of multiple τ decay modes.

1.2.1 Perturbative model of π^0 decay via the triangle loop

The quarks making up π^0 are confined according to principles of quantum chromodynamics below the scale at which the running strong coupling constant diverges at the leading order. Due to their inherent non-perturbative nature, neutral pion decays cannot be fully and accurately described via an infinite perturbative expansion. However, there is still a number of phenomenological models, which attempt to calculate its decay width with varying degrees of precision. An example of such model is based on the interaction Lagrangian

$$\mathcal{L} = ig\overline{\Psi}_p\gamma_5\Psi_p\pi \tag{1.6}$$

describing a pseudoscalar coupling of proton and π^0. Proton is considered in this model to be a point-like fermion, which directly couples to the photon field. The decay $\pi^0 \to \gamma\gamma$ then proceeds via a closed proton loop, as displayed on figure 1.1

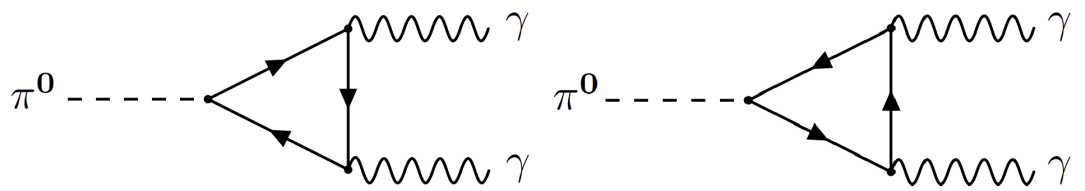

Figure 1.1: A decay of π^0 via closed loop with Bose symmetrization explicitly pictured

As the loop integral is convergent, the final result for the decay width is finite. In the limit $q^2 << m_p^2$, where q is the pion momentum, we get a result

$$\Gamma(\pi^0 \to \gamma\gamma) = \frac{1}{64\pi} m_{\pi^0}^3 \frac{\alpha^2}{\pi^2} \frac{1}{f_\pi^2} \tag{1.7}$$

where $f_\pi \doteq 93$ MeV is an experimentally determined constant also present in decays of charged pions. Numerically, the above expression is equivalent to lifetime of approximately 8.58×10^{-17} s. The result is in a good agreement with the experimental value, predicting well the order and the approximate value.

1.2.2 Kinematics of the π^0 decay

When observing the $\pi^0 \to \gamma\gamma$ decay, it is often the case that the angle between the two photons is very small, which poses a challenge for most modern detectors. The reason behind such small angle and its dependence on π^0 energy can be easily described by a simple boost from center-of-mass (CMS) frame to the laboratory frame. In the CMS frame, the angle between the decay products is equal to $\theta' = \pi$; the direction of flight of both photons is oriented opposite to each other alongside a randomly oriented axis. The energy of the decay products is $E_\gamma^{CMS} = m_\pi/2$ in order to conserve total energy. The photon momenta then form a sphere with a diameter equal to m_{pi^0}. If the z axis is taken to be identical to the direction of π^0 movement, then working in spherical coordinates we can write the four-momenta of the decay products in the CMS frame as

$$P_{\pm}^{CMS} = \frac{m_\pi}{2}(1, \pm\sin\theta\cos\phi, \pm\sin\theta\sin\phi, \pm\cos\theta) \tag{1.8}$$

Where we set $E_\gamma = p$ and $\phi = \arctan(y/x), \theta = \arccos(z/r), r = \sqrt{x^2+y^2+z^2} = m_\pi/2$. The number of photons per space angle Ω displays a constant distribution for ϕ and θ given by

$$\frac{\mathrm{d}N_\gamma}{\mathrm{d}\Omega} = \frac{\mathrm{d}N_\gamma}{\mathrm{d}(\cos\theta)\mathrm{d}\phi} = C \tag{1.9}$$

Via Lorentz transformation we then boost the $P_{\pm}^{CMS}$ four-momenta into the laboratory frame, which then results in the observed decay angle α being the angle between the transformed momenta $\vec{p_\pm}'$. The minimal decay angle α_{min} corresponds to a decay symmetric around the z axis in the CMS frame — when $\theta = \pi/2$. The angle is then equal to

$$\sin\frac{\alpha_{min}}{2} = \frac{m_\pi}{E_\pi} \tag{1.10}$$

Dependence of the minimal decay angle on π^0 energy is explicitly displayed on the graph 1.2. As per (1.9) the distribution of N_γ reaches a maximum for $\theta = \pi/2$, a majority of the decay angles observed will be in the immediate vicinity of α_{min}.

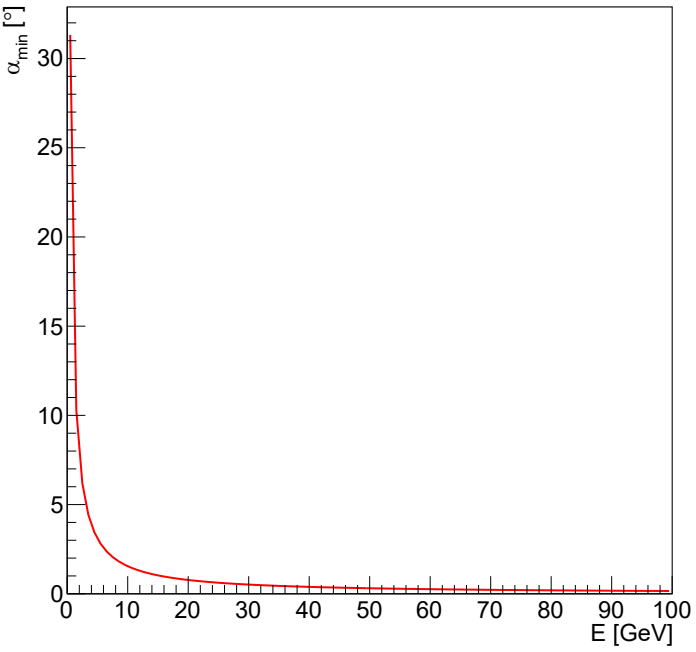

Figure 1.2: Dependence of α_{min} on pion energy E

2. Future Circular Collider

Future Circular Collider - FCC - is an option being considered to take over after the operation of HL-LHC at CERN comes to an end. It offers significant improvement on results and bigger available phase space for processes both investigated currently, as well as being part of yet unprobed physics. The circular collider with a circumference of 100 km is projected to operate at maximal center of mass energy of p-p collisions reaching $\sqrt{s} = 100$ TeV and a maximal luminosity of $L = 3.10^{35}$ cm^{-2}s^{-1}. This corresponds to O(20) ab^{-}1 per experiment. The initial stage of operation is projected to be a lepton collider FCC-ee [4] expected to gather data on Z and H bosons and the strong interaction at maximal $\sqrt{s} = 365$ GeV with a possible intermediate step of a lepton-hadron FCC-eh collider [5]. The final stage of the FCC operation consists of a hadron collider FCC-hh [6] aiming to study new potential interactions, rare decay modes and hypothetical heavy yet undiscovered particles.

2.1 Coordinate system and unit definition

When working in the calorimetric environment it is useful to use the (r, ϕ, η) coordinate system and identify the collision vertex as the coordinate zero point. The ϕ and r coordinates determine the transverse plane, defined as the plane orthogonal to the beam axis, which we identify as the z axis, when working in Cartesian or cylindrical coordinates. The third coordinate, η, is then defined by the equation

$$\eta = -\ln\tan\left(\frac{\theta}{2}\right) \tag{2.1}$$

where θ is the angle between the positive z axis and a momentum vector p, for $\theta \in [\pi/2; 0]$ is $\eta \in [0; \infty]$. In the $(\eta \times \phi)$ space, a distance is then given by the relation

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.2}$$

When working with a particle momentum p we also define the transverse momentum

$$p_T = \sqrt{p_x^2 + p_y^2} = |p| \sin\theta \tag{2.3}$$

During an experiment we measure the transverse momentum and the (η, ϕ) coordinates. To obtain the Cartesian momenta we can transform the measured physical quantities according to the set of equations

$$\begin{aligned} p_x &= p_T \sin\phi \\ p_y &= p_T \cos\phi \\ p_z &= p_T \sinh\eta \\ |p| &= p_T \cosh\eta \end{aligned} \tag{2.4}$$

2.1.1 Resolution of a calorimeter and sampling parameter

For the sake of energy reconstruction and particle identification, a calorimeter is segmented in $(\Delta\eta, \Delta\phi)$ that characterize its cell size (and subsequently granularity). Granularity can be homogeneous throughout the entire volume of the calorimeter or vary through several layers. Along with granularity we use the energy resolution of the calorimeter, defined as

$$\frac{\sigma_E}{E} \approx \frac{a}{\sqrt{E}} \oplus \frac{b}{E} \oplus c \tag{2.5}$$

where a is called the stochastic term that determines the contribution of statistical fluctuations, b is the noise term accounting for electronic noise and c is a constant term, that counts in any other random contributions to the measurement uncertainty (such as material inhomogeneities). $\oplus$ represents a quadratic summation. [7].
When working with sampling calorimeters, deposited energy is registered in the active layers. Due to design of the calorimeter, however, this represents only a fraction of the true shower energy, denoted as f_{sampl}; thus the calorimeter needs to be calibrated to account for this partial loss. The basic equation for the reconstructed cell energy E_{rec} reads

$$E_{rec} = \frac{E_{dep}}{f_{sampl}} \tag{2.6}$$

where E_{dep} is the energy deposited in the active material.

2.2 Physics at the FCC-ee

As a first stage of the proposed FCC program, the main objective of the circular e^+e^- collider is to measure properties of the Standard Model with previously unattained precision and probe for deviations that would point to either very rare couplings or high energy scales. The center of mass energy will be in the range of 88-365 GeV to cover the electroweak sector along with the Higgs boson and the top quark [4]. The FCC-ee will particularly focus on processes at center-of-mass energies around the Z pole (91 GeV), the WW pair threshold (161 GeV), ZH threshold (240 GeV) and $t\bar{t}$ threshold (340-365 GeV) with the CMS energy calibration precision reaching 100 keV level at the Z and WW scales. The current state of measurements points to observable quantities in the electroweak sector agreeing with the Standard model predictions within current uncertainties. Therefore, it is of interest to reduce those uncertainties as much as possible to uncover potential deviations from the SM predictions that would lead to new physics. At the Z pole, luminosity will be increased by a factor of $\mathcal{O}(10^5)$ when compared to LEP, which would decrease statistical uncertainties by a factor of $\mathcal{O}(300)$. Aside from quantities such as the Z mass and decay width, we would improve measurement of $\sin^2\theta_W^{eff}$ by measuring forward-backward asymmetry A_{FB} in the $e^+e^- \to Z \to \mu^+\mu^-$. Last, but not least, we would be able to reduce experimental uncertainty on $\alpha_S(m_Z)$ by a factor of $\mathcal{O}(10)$ by measuring ratio R_l of the Z hadronic width to the Z leptonic width. The Z also figures in $e^+e^- \to Z\gamma$ process that will be used to study the decay of Z to neutrinos by measuring its

invisible decay width, enabling us to directly test the unitarity of the neutrino mixing matrix and search for right-handed quasi-sterile neutrinos.

At the WW and $t\bar{t}$ production threshold, it is imperative to reduce the uncertainties in measuring the W and t mass, as well as their electroweak coupling values, as they play part in electroweak radiative corrections. Finally in the Higgs sector, FCC-ee is expected to improve significantly on the data yet gathered by the LHC - because higher order corrections to Higgs couplings are at the order of few percent, measurements need to reach precision at least a few per mil. To achieve such precision, at least 10^6 H bosons need to be produced with the most productive channels being $e^+e^- \to HZ$ and $e^+e^- \to WW \to H$. Finally, the FCC-ee is expected to reach precision of $\pm 12\%$ when measuring the Higgs trilinear coupling (combined with measurements from HL-LHC and when only κ_λ is allowed to vary). The luminosity parameters expected for working points of the FCC-ee are displayed in the table 2.1.

Table 2.1: Design parameters of the CLD detector

Working point	Tot. lum./year [ab^{-1}/year]	Run [years]	Goal [ab^{-1}]
Z pole (88-94 GeV)	24	2	150
Z pole (88-94 GeV)	48	2	
WW threshold ($\sim$ 161 GeV)	6	1-2	10
ZH threshold ($\sim$ 240 GeV)	1.7	3	5
$t\bar{t}$ threshold (340-350 GeV)	0.20	1	0.2
$t\bar{t}$ threshold ($\sim$ 365 GeV)	0.34	4	1.5

2.3 Detectors at the FCC-ee

To satisfy the requirements put onto the FCC-ee program regarding angular and energy resolution, particle identification, missing energy resolution and tracking, strict constraints are put on criteria the detectors must satisfy. The colliding electron and positron bunches cross at an angle of 30 mrad via crab waist collision scheme with the time between bunch crossings ranging from a minimum of 20 ns to a maximum of 7 μs. High production of synchrotron photons and electron-positron pairs in the general vicinity of the interaction region is to be expected due to high luminosities, the first of these is to be addressed with tungsten masks (see 2.1). Regarding the pair production, simulations indicate that most of the resulting particles do not reach the detector and the background is therefore moderate enough to be remedied using high readout electronics. Several detector designs have been proposed and studied for FCC-ee, which I will present in the subsections below.

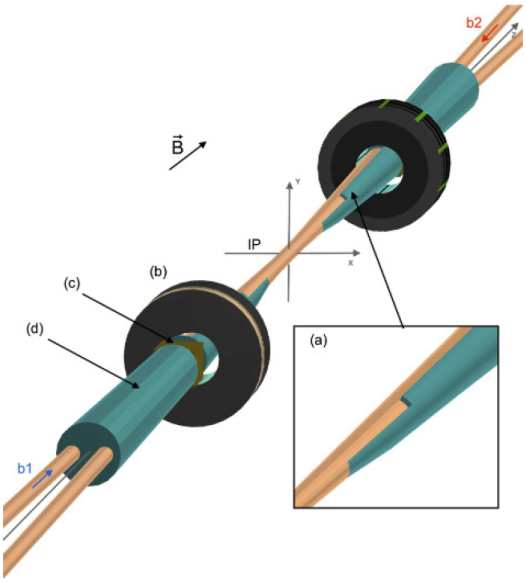

Figure 2.1: Scheme of the interaction point with the tungsten masks displayed in blue. From point (a) to (b) the shielding is 0.1 mm thick, being replaced by a 15 mm thick tungsten cone (d) behind absorber (c) [4]

2.4 CLD detector

CLD, meaning "CLIC-Like Detector", is a design using a Si tracker and a 3D-imaging calorimeter with high granularity. The dimensions of the CLD detector are displayed in a table 2.2 with the layout displayed in figure 2.2

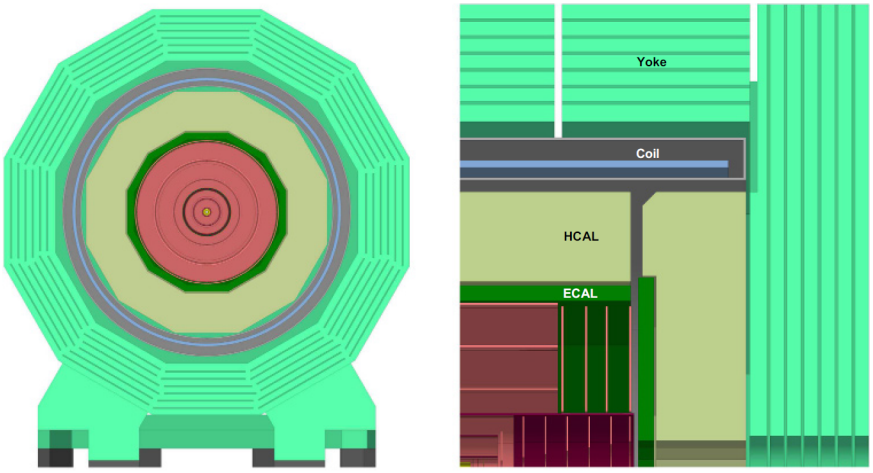

Figure 2.2: CLD detector layout with the transversal plane displayed on the left and the longitudinal cut of the upper right quadrant pictured on the right, with hadronic and electromagnetic calorimeters shown around the central tracking region [4]

Table 2.2: Design parameters of the CLD detector

Vertex inner radius [mm]	17
Tracker outer radius [m]	2.1
Tracker half length [m]	2.2
ECAL thickness [X_0]	22
HCAL thickness [λ_I]	5.5
Overall height [m]	12.0
Overall length [m]	10.6
Solenoid field [T]	2

The CLD detector follows the model designed for CLIC with the inner vertex region consisting of a cylindrical silicon pixel vertex detector and a silicon tracker followed by highly granular calorimeters. The vertex detector itself is divided into a cylindrical barrel consisting of three double layers and disks covering the forward regions, also divided into three double layers. The silicon tracker is divided into the inner tracker built out of three barrel layers and sever forward disks. Beyond that is located the outer tracker which completes the tracking region with three barrel layer and four forward disks. Single-point resolution of the tracking system is assumed to be 3×3 μm^2 for the vertex detector, 5×5 μm^2 for the first disk of the inner tracker and 7×90 μm^2 for the remaining layers of the inner and outer tracker. Tracking efficiency for such configuration has been determined to be 100 % for muons with $p_T > 1$ GeV, remaining high even for lower energies (eg. 96 % at $p_T > 0.1$ GeV). The transverse momentum resolution $\sigma(1/p_T)$ is expected to be lower than 5×10^{-5} GeV^{-1} for high-momentum muons.

2.4.1 Calorimetry at the CLD detector

As pictured on 2.2, detection beyond the tracking region will be covered by an electromagnetic calorimeter ECAL and a hadronic calorimeter HCAL. The baseline option for the ECAL is a silicon-tungsten sampling calorimeter. The choice is motivated by the particle flow method that is to be used at CLD, which individually reconstructs each particle and thus optimizes jet energy resolution [8]. The ECAL is divided into 40 identical Si-W layers with segmentation of 5×5 mm^2. The segmentation has been chosen as to adequately resolve energy deposited by particles in neighboring jets. The chosen number of layers has been found to give the best γ energy resolution. Depth of the ECAL of 22 X_0 has been then determined to limit leakage into the HCAL.

Simulations have shown that when combining silicon tracker with high granularity calorimetry and particle flow reconstruction, jet energy resolution reaches 4.5 % for energies at 50 GeV, decreasing further down to 4 % at 100 GeV and higher [8]. For single photons, stochastic term in the formula 2.5 has a value of $15\%/\sqrt{E(GeV)}$ for energies of 5-100 GeV.

2.5 IDEA detector

IDEA, meaning "Innovative Detector for Electron-positron Accelerators", is another proposition for a detector at FCC-ee. The proposal approaches detector design differently than CLD and consists of a silicon pixel vertex detector with the tracking region being a large volume short-drift wire chamber surrounded by a layer of silicon micro-strip detectors. The vertex detector is based on a very light design planned for the ALICE ITS upgrade, which features 5 μm resolution and low dark-noise rate. Beyond that a pre-shower detector is located, which ultimately leads to a dual-readout calorimeter. The relevant parameters of the IDEA detector design are located in the table 2.3 along with a layout displayed at 2.3.

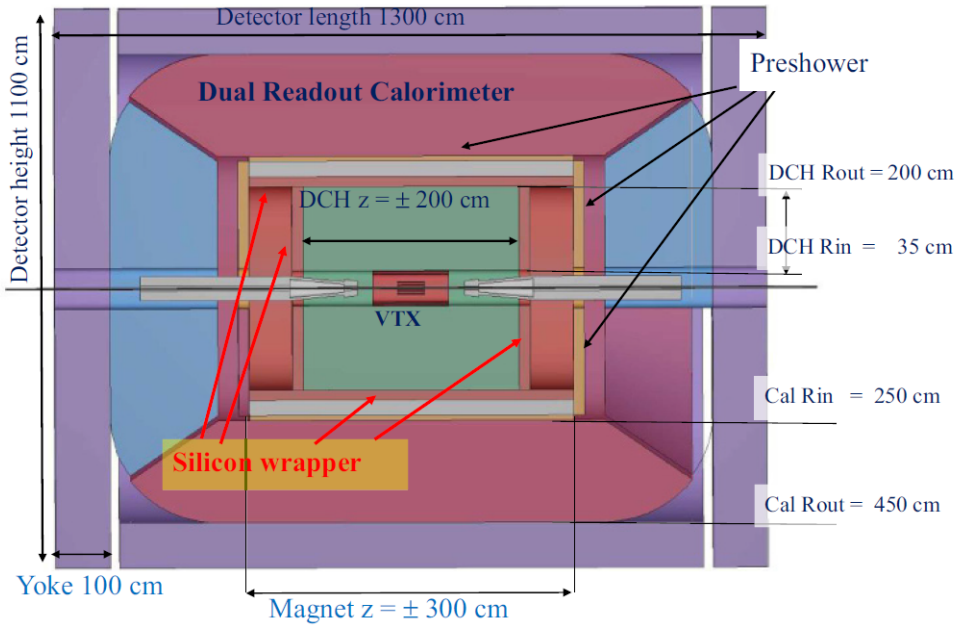

Figure 2.3: IDEA detector layout [4]

Table 2.3: Design parameters of the IDEA detector

Vertex inner radius [mm]	17
Tracker outer radius [m]	2.0
Tracker half length [m]	2.0
DR calorimeter inner radius [m]	2.5
DR calorimeter outer radius [m]	4.5
DR calorimeter thickness [λ_I]	7
Overall height [m]	11
Overall length [m]	13
Solenoid field [T]	2

The motivation behind using a wire drift chamber (DCH), as opposed to a silicon tracker, is the low material budget of the DCH. The magnetic field to be used in the wire chamber has a value of 2T and the chamber extends from inner radius of 0.35 m to an outer radius of 2 m with length along the beam axis of

4 m. The DCH consists of 112 co-axial layers arranged into 24 azimuthal sectors with alternating sign stereo angles. This arrangement results in an approximately square-shaped cell with an edge of 12.0-14.5 mm. The gas mixture within the chamber is 90%He and 10%$i\mathrm{C}_4\mathrm{H}_{10}$ which together with the cell size amounts to a maximal drift time of ≈400 ns. The ancestor of this design, MEG2 drift chamber for the KLOE experiment, has reached a spatial resolution of 100 μm, which is the conservative estimate also for the IDEA drift chamber. Together with cluster counting/timing techniques, we expect improvement of the spatial resolution — also resulting from longer drift distances, when compared to the MEG2. The drift chamber is surrounded by a layer of silicon micro-strip detector to provide additional position measurement. Coefficients in transverse momentum resolution $\sigma(1/p_T) = a \oplus b/p_T$ are expected to reach $a \simeq 3 \times 10^{-5}$ GeV^{-1} and $b \simeq 0.6 \times 10^{-3}$.

The drift chamber is followed by a preshower detector consisting of two alternating layers of Micro Pattern Gas Detector (MPGD) chambers and absorbers. In the barrel region, the magnetic coil acts as the first absorber and the second absorber is made from lead. In the forward region, both absorber layers are lead ones. The preshower detector further improve tracking resolution and accurately determine impact point of charged particles and photons. They also aid in π^0 identification by identifying the corresponding decay photon pairs.

2.5.1 Calorimetry at the IDEA detector

The calorimeter design for the IDEA detector is a lead-fibre dual-readout (DR) environment with a thickness of approximately $7\lambda_I$. A dual-readout design offers many advantages over a sampling calorimeter design proposed for the CLD, as it boasts excellent electromagnetic and hadronic shower energy resolution and particle discrimination. The calorimeter collects signals from scintillators (S) and Cherenkov detectors (C) and combines them to reach resolution, which simulations estimate at $10\%/\sqrt{E}$ for electrons and $30\%/\sqrt{E}$ for pions, constant terms being negligibly small. For isolated particles, the DR calorimeter displays good discrimination between muons, electrons, photons and hadrons using variables such as the C/S ratio, lateral shower profile, charge-to-amplitude ratio and starting time of the signal. The intrinsic discrimination ability will be paired with fine transverse granularity to allow for good shower separation and aid particle-flow reconstruction by matching showers to tracks from the inner region and to muon tracks. Longitudinal segmentation of the calorimeter is an open problem as of now, with multiple proposals being studied to optimize resolution of signals produced by overlapping electromagnetic and hadronic showers.

2.6 Detector concept with noble liquid calorimetry at FCC-ee

Noble liquid calorimetry has been adapted to many experiments, such as NA48 [9], ATLAS [10], SLD [11] and others, with the ATLAS experiment at the LHC being the latest example. It has been chosen for its energy resolution, linearity of response and radiation hardness, among other advantages. A noble liquid LAr calorimeter with high granularity is being studied for the FCC-hh detector due to the harsh radiation environment expected at collisions with center-of-mass energy reaching 100 TeV. For the purpose of a lepton collider, radiation hardness is not a concern — however, the requirements of uniform and linear response, stability and excellent resolution have led to proposals for a noble liquid calorimeter to be used at the FCC-ee as well. A design of such detector is displayed on figure 2.4.

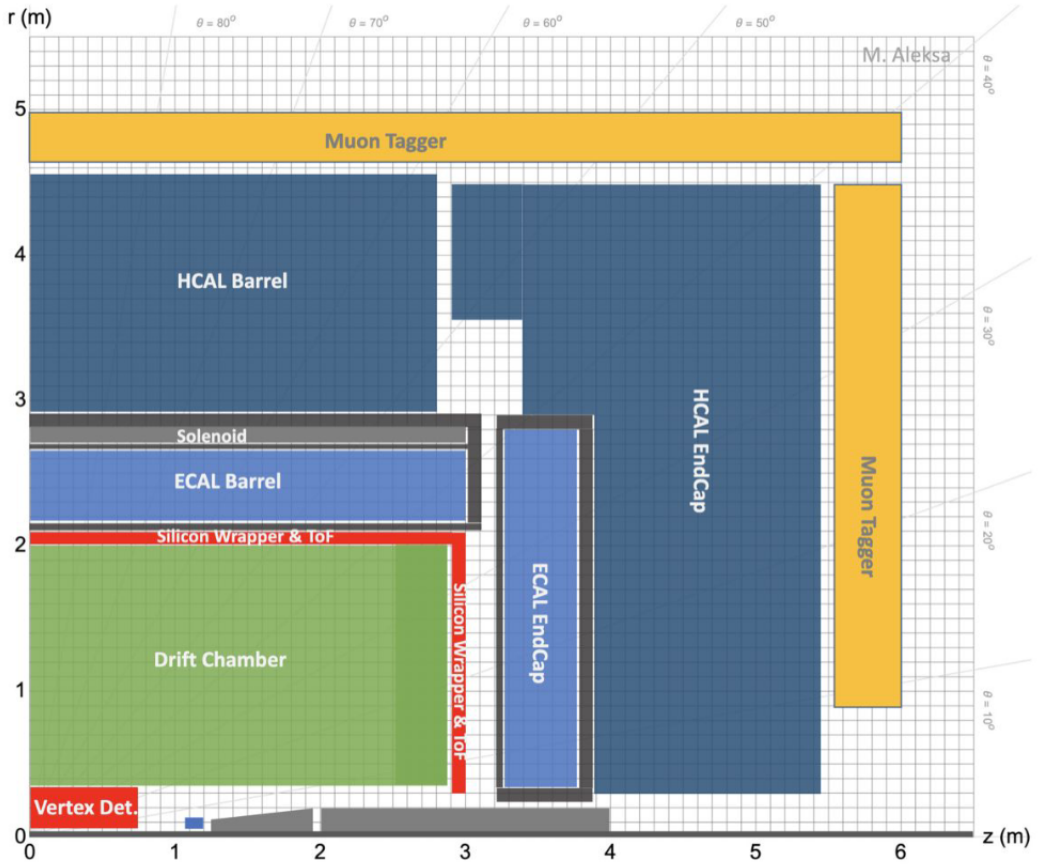

Figure 2.4: Detector concept for the FCC-ee with a LAr calorimeter [12]

The full detector concept is a modified design of the IDEA detector, with a common drift chamber concept that together with a silicon pixel vertex detector makes up the inner tracking region. The surrounding and forward regions are covered by ECAL and HCAL barrels together with respective endcaps and a HCAL extended barrel calorimeter HEB. The proposed detector combines a liquid argon (LAr) sampling electromagnetic calorimeter (ECAL) with an ATLAS TileCal-like hadronic calorimeter made of scintillating tiles and steel absorbers, with signal read out by silicon photomultipliers [8].The baseline idea for the absorber material of ECAL is lead, with alternative proposals for the absorber being tungsten and

liquid krypton or xenon considered as alternative active media[13]. The advantage of such alternatives is smaller radiation length and Molière radius, which will lead to more contained showers and better separation of close-laying clusters. The baseline idea of a LAr calorimeter with 1.2 mm thickness of sensitive gaps near the inner radius and 1.8 mm thick straight lead absorbers would have a thickness of $\sim 22\ X_0$ and a Molière radius of $R_M \approx 4$ cm. The absorbers are azimuthally inclined by approximately 50°. A cross-section of a sampling-calorimeter concept for the FCC-ee is displayed on figure 2.5.

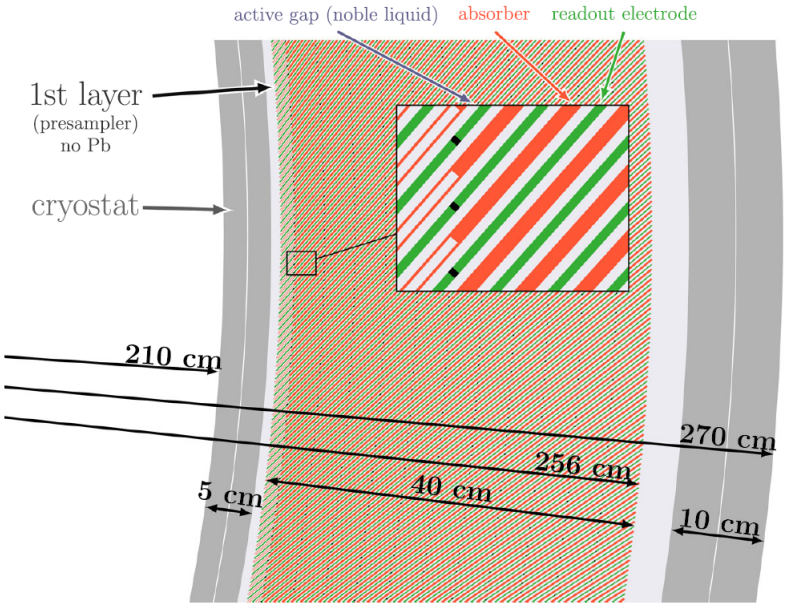

Figure 2.5: LAr calorimeter transversal cross-section [14]

The calorimeter is to be divided into twelve longitudinal layers with granularity of the cells equal to $\Delta\theta \times \Delta\phi = 0.57° \times 0.47°$ [13] which will result in a typical cell size of about $2 \times 2 \times 3$ cm^3 [14]. Due to the inclination of the layers (as seen on figure 2.5), sampling fraction f_{sampl} of the calorimeter will be increasing with depth. The effect of this will be combated by the longitudinal segmentation and energy calibration changing with depth. The first layer will be built without absorber plates and will act as a presampler to correct for energy loss upstream. A second strip layer has granularity brought further down to $\Delta\theta \times \Delta\phi = 0.14° \times 0.47°$ (corresponding to a cell size of 5.4 mm $\times$ 17.7 mm) and is aimed at reconstruction of π^0 and other particles decaying into a pair with very small angle between the decay products. A sampling term of 8 % and a constant term of 0.68 % has been achieved in performance studies for electrons and photons, when considering leas absorbers and LAr active material. The sampling term has been brought further down to 7 % when studying configuration with LKr or LXe, or tungsten absorbers.

2.7 Clustering at the FCC

Reconstructed cells are grouped together to create clusters that are used for particle reconstruction and identification. There are two methods of clustering used at the FCC, which we will introduce below.

2.7.1 Sliding window algorithm

The sliding window works with cell granularity in the $\eta \times \phi$, without taking longitudinal layers into consideration. It is used mostly for reconstruction of electrons and photons. The starting point is building a tower in the radial direction, while the dimensions in the $\eta \times \phi$ space are kept constant — for illustration, see figure 2.6. In the simplest case, the tower size in $\eta \times \phi$ is 1×1. The energy of the tower is equal to the sum of cell energies in the tower [6].

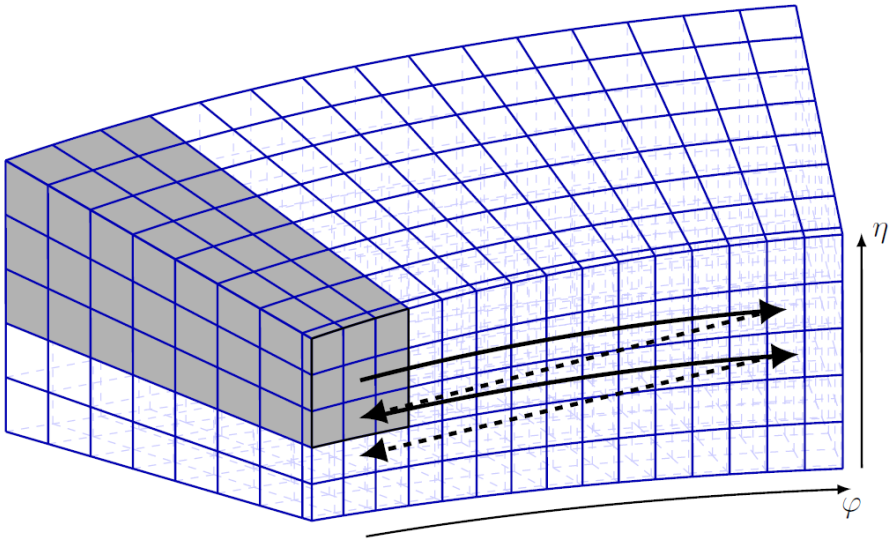

Figure 2.6: Depiction of towers with fixed $\eta \times \phi$ dimensions in the sliding window algorithm. The shaded area is a window that scans the towers for local energy maxima [6]

The towers are then scanned by a window of a fixed size $N_\eta \times N_\phi$ centered symmetrically around one tower that searches for local maxima. If the sum of energies of towers in the window is above a preselected cut E_T^{cut}, a pre-cluster is created with position being determined as a weighted average of cells in the window [15]. It is possible for multiple pre-clusters to overlap, out of them only the one containing the most energy is kept. Finally, the tower containing the pre-cluster coordinates is selected as the central tower, around which is the final cluster built. The cluster is located again within a fixed window with constant $\eta \times \phi$ dimensions; the size of the final window is selected to limit lateral shower leakage and suppress noise contribution at the same time.

2.7.2 Topological clustering

The topological clustering works instead with variable cluster size by grouping together cells containing significant energies, when compared to the overall noise levels. It is an iterative procedure that grows clusters procedurally in the outward direction from the seed [15]. In contrast to the sliding window method, topoclustering works in the radial dimension as well. The cells are topologically connected together to form clusters while signals originating from noise contribution are rejected. Cell significance ξ_{cell} is then defined as [6]

$$\xi_{cell} = \left| \frac{E_{cell}}{\sigma_{cell}^{noise}} \right| \tag{2.7}$$

where σ_{cell}^{noise} is the expected noise level in the cell (determined as the standard deviation of a Gaussian noise distribution). The first step is searching for seeds, which are defined as cells with significances above a cut $\xi_{cell} \geq S$. The seeds are then ordered by energy and every seed is assigned a protocluster. For each seed, the neighboring cells are added to the protocluster if their significance is above a set threshold T and if they haven't been selected as seed cells. After being included in the protocluster, they are considered as seeds and their neighbors are considered next. If the neighboring cell is already assigned to another cluster and its significance is above T or P, the two protoclusters are merged. Alternatively, if T or P is set to 0, the cell will be assigned to the more energetic cluster of the two, without merging. This step is repeated until there are no more neighbors with significance above T. Finally, the cluster growing process is finished by adding all neighboring cells on the immediate outer perimeter of the protocluster with significance above a third cut P. Thus a cluster is created with irregular shape, that is characterized typically by a core of cells containing highly significant signals and a surrounding shell of cell with smaller energy deposits. In the case of clusters being merged, it is possible to observe a cluster with more seeds. The topological clustering is used for reconstruction of jets, for example.

2.8 Event simulations

The simulation of events in our thesis was carried out using Geant4 within FCCSW, which is a set of packages, tools and standards to coordinate various FCC studies, compare their results and avoid duplicate works [16]. Geant4 is an open toolkit for simulating passage of particles through matter that includes all aspects of the simulation process, such as tracking, system geometry, physics models and processes governing the interactions, storage of events, hit management, visualization [17] etc. Scales covered by the physics models offered by Geant4 vary from 250 eV up to several PeV. The programming language used in Geant4 is C++ and the simulation of processes is based on the Monte Carlo method of random sampling. The interface for the generation of primary particles that define a physics event is provided by the *event* category, which contains the primary particles and vertices. The *geometry* category then describes the detector geometry. Transportation of particles and representation of physical processes by characteristics such as "at rest", "along step" and "post step" is handled by the *tracking* category. The physical properties of particles and

materials are implemented in the *particles* and *materials* categories necessary for the simulation of particle-matter interaction. Finally, the *physics* category describes the physical processes, divided into several subcategories: particle decay, electromagnetic physics, hadronic physics, transportation, optical physics, photolepton-hadron physics and parametrization. When the simulation propagates a particle through matter, the transportation is carried out in steps (in units of time or length, depending on whether the process happens at rest or not) determined by physics processes or by the detector geometry. While traversing the detector, hits are recorded as snapshots of physical interactions in a sensitive detector component. The hits are created using information provided in the current step (for example energy loss), while the detector response is given by the user. For the sake of the performance, cuts are applied on particle generation in interactions (mainly to suppress the generation of large numbers of soft electrons and photons when simulating passage of charged particles through matter). When simulating a calorimeter, hits located inside a cell are summed with the result being the whole energy deposited inside the cell, cell position and cell ID. After simulating particle passage and energy deposition, the cells become input for clustering algorithm. In this way, we obtain a full set of data that contains information on the initial particle state along with the final detector output.

3. Multivariate analysis

Multivariate analysis can be defined as a statistical study of data with simultaneous observation and analysis of multiple variables. It is used in experiments when working with multiple measurements made on each experimental unit [18]. In physical experiments, this usually means using multiple variables (referred to as discriminating variables) to test hypotheses about events present in the experiment. Discriminating variables are provided by detectors (or calculated from data provided) and are selected for their discriminating power for specific analyses. The discriminating power is determined by the overlap of their respective probability distribution functions (or "pdf"). In our thesis, the hypothesis being tested is a question, whether the signature is a signal (a neutral pion) or a background (photon). The first case will be denoted as a null hypothesis H_0, the latter case then an alternate hypothesis H_1.

A success of an analysis test can be characterized by quantities ϵ_S and r_B. The first quantity is the signal efficiency defined as the probability to accept a signal event. It is related to the type I error α of the test as [19]

$$\epsilon_S = 1 - \alpha \tag{3.1}$$

The latter quantity is the background rejection — the probability to reject a background event, equal to

$$r_B = 1 - \beta \tag{3.2}$$

where β is the type II error of the test. A successful analysis displays the highest possible r_B for ϵ_S, which is the goal of multivariate analysis. The dependence of the background rejection on signal efficiency constitutes a receiver operating characteristic, or an ROC curve. A typical shape of an ROC curve is displayed on figure 3.1.

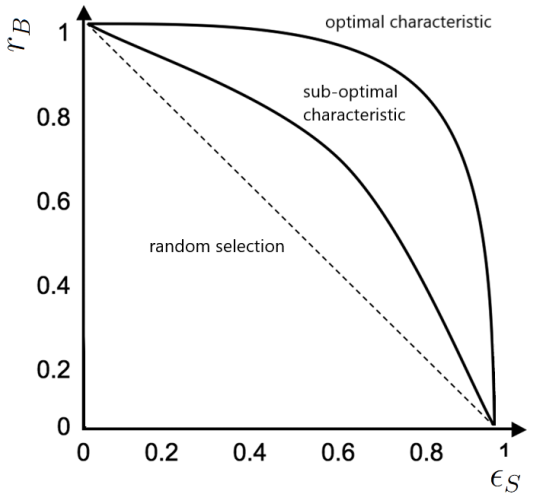

Figure 3.1: An ROC curve displaying multiple tests with varying performance [20] (*edited for the purposes of this thesis*)

The final value of ϵ_S and r_B is obtained by choosing a particular working point (also known as WP) on the ROC curve. The conditions determining the choice of the working point vary between experiments; we can work with ϵ_S at a required value of r_B or we can demand that the working point maximizes the signal significance.

An integral of the ROC curve (known also as AUROC, meaning "area under the ROC") is often used to compare multiple analyses; the optimal test usually being the one with the highest AUROC. This is connected to the fact, that a test with the highest power is a test that maximizes likelihood $\lambda(\vec{x})$, defined as [21]

$$\lambda(\vec{x}) = \frac{S(\vec{x})}{B(\vec{x})} > c_\alpha \tag{3.3}$$

where $\vec{x}$ is a set of discriminating variables and S and B are the probability distribution functions for signal and background respectively. c_α is then a constant, that determines the final signal efficiency (and size of the test α). As the shape of a probability distribution function corresponding to a discriminating variable in a real experiment is not exactly known, they are approximated by numerical approaches that encompass various MVA methods.

3.1 Rectangular cuts

A simple MVA method is the method of rectangular cuts. The method consists of applying a cut on each discriminating variable separately and identifying all events excluded by the cut as background. The cuts are optimized by maximizing the expected significance, which is done by fitting using a specific fitting method. The fitters used most commonly for rectangular cuts are eg. Monte Carlo, Genetic Algorithm and Simulated Annealing.

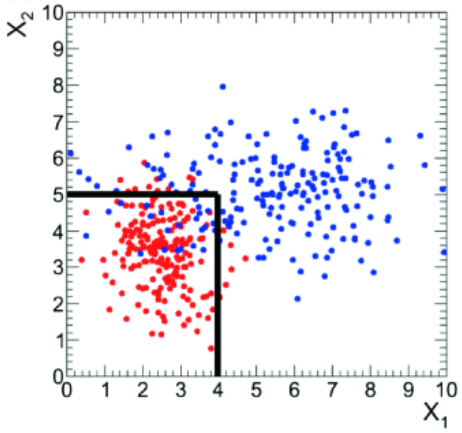

Figure 3.2: A simple visualization of rectangular cuts imposing a boundary between signal and background events [22]

3.1.1 Monte Carlo

Monte Carlo approach consists of randomly sampling the fit parameters and choosing the cuts that maximise the r_B value for a given ϵ_S. The drawback of this method is that for a large value of discriminating variables n_{var}, smoothness of the ROC curve drops and the required number of MC samples to correctly analyze data grows with powers of $2n_{var}$. The main parameter influencing the MC performance is the number of events in the toy sample.

3.1.2 Genetic Algorithm

The Genetic Algorithm finds approximate solutions to optimization problems by modelling a population of genomes (abstract representations) of individuals (possible solutions). The population then evolves towards an optimal solution of the relevant problem by selecting individuals with good fitness to produce the next generation and randomly changing (mutating) their parameters. The process is repeated in a preselected number of steps, until the optimization converges. The fitness of an individual is determined by a fitness function, that either returns a value representing the fitness of an individual, or compares two individuals and returns the better performing one. The parameters involved in GA algorithm are the population size, number of steps for convergence and number of independent cycles of fitting. Of these, population size is the most important value when it comes to determining the quality of the results, at the cost of increasing the optimization time.

3.1.3 Simulated Annealing

Simulated annealing aims to solve an optimization problem with several local or global minima. The algorithm evolves the problem slowly to reach a convergence with multiple solutions in a process inspired by a slow cooling (annealing) of metals, wherein atoms move slowly towards the state of the lowest energy. To avoid possible local minima, parameters are perturbed with a probability proportional to exp $(-\Delta E/T)$, where ΔE is an energy shift and T is an ambient temperature.

3.2 Boosted Decision Trees

A more advanced method of multivariate analysis is via Boosted Decision Trees (BDT). Decision tree is, similarly to a rectangular cut, a binary classifier analyzing events and assigning them as a signal or a background by a sequence of decisions on single variables, until the phase space is split into separate hypercubes of signal/background events. The data used in training BDT algorithm is weighted such that the sums of background and signal events respectively are normalized to one. When building a decision tree, multiple cut values are tested on every discriminating variable and then optimal splitting is selected by defining a splitting index — the most common (and the one used in our analysis) being the Gini index defined as [19]

$$G = p(1-p) \tag{3.4}$$

where p is the signal purity of a phase space Φ defined as

$$p = \frac{N_S}{N_S + N_B}; \quad S, B \in \Phi \tag{3.5}$$

N_i being the number of background/signal events contained in Φ. The optimal cut is then determined as the one corresponding to maximum of the difference C:

$$C = N_0 G_0 - N_1 G_1 - N_2 G_2 \tag{3.6}$$

Indices 1 and 2 designate two nodes created by a single branching of one root node with index 0, the Gini indices in the equation are weighted by the total number of events in each corresponding phase space. The final nodes are called leaves, which are then collectively designated as signal or background — the decision tree can be then after training applied on a test sample to check its performance, when events pass through the tree and are classified as signal or background according to their location among the final nodes. With enough branchings during the training phase, we could split the phase space of events and background into two sets of hypercubes containing pure signal and background — the result of such process would be an extremely overtrained decision tree with very little testing power. The boundary between background and signal would encompass any statistical fluctuations present in the distribution of the two types of events, to which decision trees are already sensitive. To reduce the effect of this sensitivity on performance and to enhance classification performance, decision trees undergo boosting or bagging. Boosting involves growing a large number of decision trees in order to counter statistical fluctuations. The number of trees being grown during boosting is a basic free parameter of the algorithm which determines the precision of classification, however with higher number of trees we risk quickly overtraining the boosted decision tree; other free parameters are maximum depth of the tree that limits number of branchings and leaf size that puts upper limit on the percentage of the original sample to be contained in the final node. Below we will introduce two most common boosting algorithms.

3.2.1 Adaptive Boost

Adaptive Boost, or AdaBoost, is an algorithm that works by modifying weights of events misclassified by a preceding decision tree. The misclassified events are reweighted by a higher weights (with the rest of the events being reweighted so that the sum of weights remains constant) and the modified sample is used to train the next decision tree. The magnitude of reweighting is determined by parameters α and β with α being given by the misclassification rate ϵ of a previous decision tree

$$\alpha = \frac{1 - \epsilon}{\epsilon} \tag{3.7}$$

The β parameter is the learning rate of the AdaBoost algorithm that works by modyfing the boost parameter as $\alpha \to \alpha^\beta$ and it is a free parameter of the BDT algorithm. By lowering β we can reduce overtraining at the expense of lower sensitivity to event misclassification during the training period. The modified weight of the misclassified j-th event in the i-th tree is then equal to

$$w_{i,j} = w_{i-1,j}\alpha_{j-1}\frac{\sum_i w_{i-1,j}}{\sum_i w_{i,j}} \tag{3.8}$$

The final output is the BDT score $t(\vec{x})$ and it is given as the sum [23]

$$t(\vec{x}) = \frac{\sum_i \ln(\alpha_i) t_i(\vec{x})}{N} \tag{3.9}$$

where $t_i(\vec{x}) = 1$ for an event identified as signal and $t_i(\vec{x}) = -1$ for an event identified as background. To classify events as signal or background (and determine final signal efficiency or background rejection), the algorithm applies a cut on $t(\vec{x})$, classifying events below (above) the cut as background (signal). Adaptive boost performs well when combined with weak classifiers characterized by small maximum depth and large leaf size to prevent overtraining, boosting their performance significantly. Combined with β chosen suitably for a specific sample, the AdaBoost algorithm makes for a powerful boosting tool.

3.2.2 Gradient Boost

Gradient boost is an algorithm that works by minimizing the classification error from the previous tree. When growing trees, each tree is assigned a base function $f(x; a_i)$, the final output being characterized by a function

$$F(\vec{x}) = \sum_i b_i f_i(x; a_i) \tag{3.10}$$

The boosting algorithm then adjusts parameters a_i, b_i such that the deviation between the model function $F(\vec{x})$ and the value y obtained from training is minimized. The deviation is characterized by a loss function $L(F, y)$, defined for classification by the Gradient Boost algorithm as [23]

$$L(F, y) = \ln(1 + exp(-2yF(\vec{x}))) \tag{3.11}$$

Similarly to AdaBoost algorithm, Gradient Boost performs best for shallow decision trees to prevent overtraining. In analogy to the β parameter of AdaBoost, the resistance to overtraining can be enhanced by controlling the Shrinkage parameter, which determines the weight of individual trees and controls the learning rate of the algorithm.

3.2.3 Bagging

Aside from boosting, there is another technique to enhance the performance of decision trees and reduce overtraining. Bagging refers to a procedure, when the original training sample is resampled into a new training sample with possible replacement — a possibility that an event occurs multiple times in the new training sample. A classifier is then repeatedly trained using new samples, growing a large number of individual trees. The final combined classifier is a combination of individual trees. In contrast to boosting, bagging does not enhance the performance of weak trees, rather it aims to counter statistical fluctuations present in individual trees by cancelling and averaging them out by summing over a large number of trees with resampled training events.

3.3 TMVA Package in ROOT

The TMVA package in ROOT provides an environment for multivariate analysis — classification and regression — of data using numerous methods designed for use mainly in high-energy particle physics. The methods encompassed in TMVA package are [23]

1. Rectangular cuts
2. Projective likelihood method (PDE range-search, PDE-foam)
3. k-NN classifier
4. H-matrix discriminant
5. Linear discriminant analysis
6. Function discriminant analysis
7. Artifical neural networks
8. Deep learning methods (deep neural networks, convolutional neural networks, recurrent neural networks)
9. Boosted decision trees
10. Support vector machine
11. Predictive learning RuleFit
12. PyTMVA - Keras

Algorithms listed above all function by supervised learning, meaning that they are trained on data with known signal/background distribution to determine the phase space distribution which classifies events during testing as signal/background. During training the discriminating variables are analyzed to estimate their discriminating power, assess their linear correlation coefficients and for classification the package is able to decorrelate the variables by eg. transforming them into a normalized Gaussian shape. After classification, information is given along with calculated ROC curve and AUROC — which is also used by the TMVA to rate different MVA methods on their performance. The package provides tools that enable the user to optimize the parameters of MVA methods and in the case of decision trees provides information on overtraining via applying the Kolmogorov-Smirnov test onto a randomly selected subsample from the testing data.

In this thesis, we opted for using the Rectangular cuts method and Boosted decision trees to compare their performance and highlight the superior performance of BDT algorithm when used to discriminate between largely overlapping signal and background events. As a tool, the TMVA package is useful for practical analysis as well as an introduction to multivariate analysis due to its simple use and quick learnability for people already familiar with ROOT.

4. Pion identification

4.1 Source data statistics

For the study of neutral pion resolution the calorimetric environment was simulated in FCCSW software. The calorimeter was simulated using two separate granularity designs. The granularity settings of the designs were $\Delta\theta \times \Delta\phi = 0.57° \times 0.47°$ as the lower granularity setting and $\Delta\theta \times \Delta\phi = 0.14° \times 0.47°$ as higher granularity. The motivation behind studying higher granularity environment is the inclusion of a strip layer in the detector design described in section 2.6. Using Geant4 we simulated signal of 100 000 neutral pions. As a background we simulated 100 000 photons. For the purposes of this thesis we worked in the energy range of 0 to 100 GeV. The distribution of η, ϕ and energy E of the initial state photons is displayed on figures 4.1, 4.2 and 4.3. The entire statistic along with pions can be found in Appendix A.1.

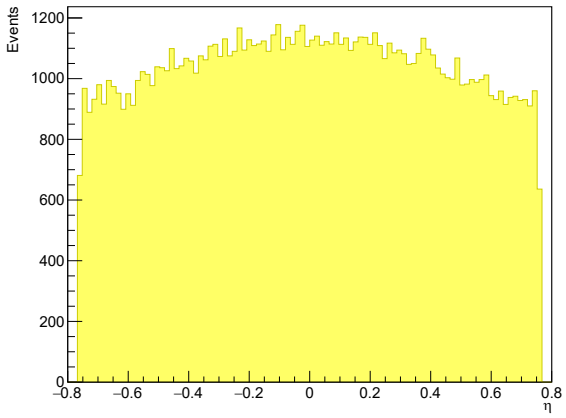

Figure 4.1: Distribution of η for photons

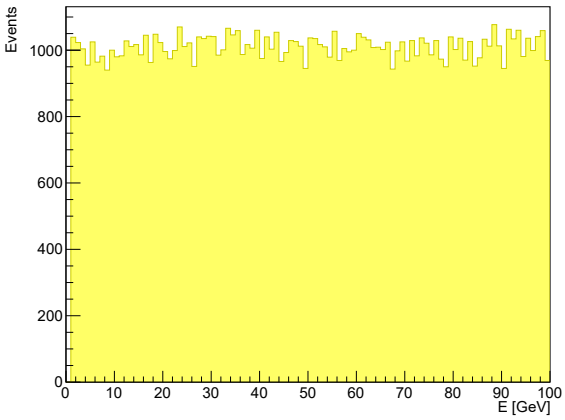

Figure 4.2: Distribution of E for photons

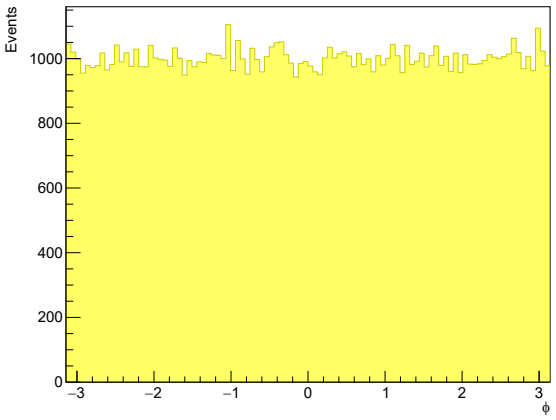

Figure 4.3: Distribution of ϕ for photons

4.2 Clustering

The clusters in the final state were reconstructed by both the sliding window algorithm and the topological clustering. The size of the sliding window used in our data was 9×17 cells in $\eta \times \phi$ and the energy threshold for soft particle production in showers was 0.05 GeV. Using the sliding window method, we typically obtained a low number of clusters for pions, with very few events containing more than 2 clusters. On the other hand, using the topological clustering we obtained a number of low-energy clusters for every event, in addition to a main cluster containing the majority of energy of the mother particle. The distribution of the number of clusters created by the respective methods is displayed on figures 4.4 and 4.5.

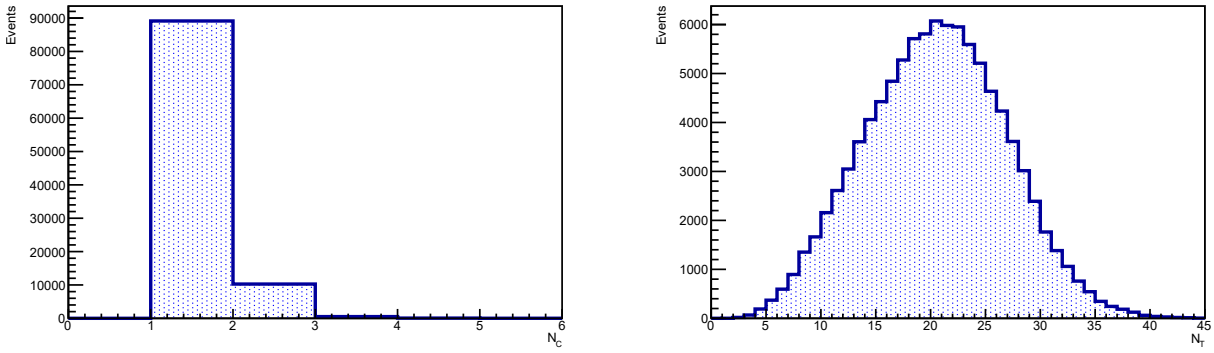

Figure 4.4: The number of clusters generated by the sliding window method N_C and topological clustering N_T for π^0 with lower granularity

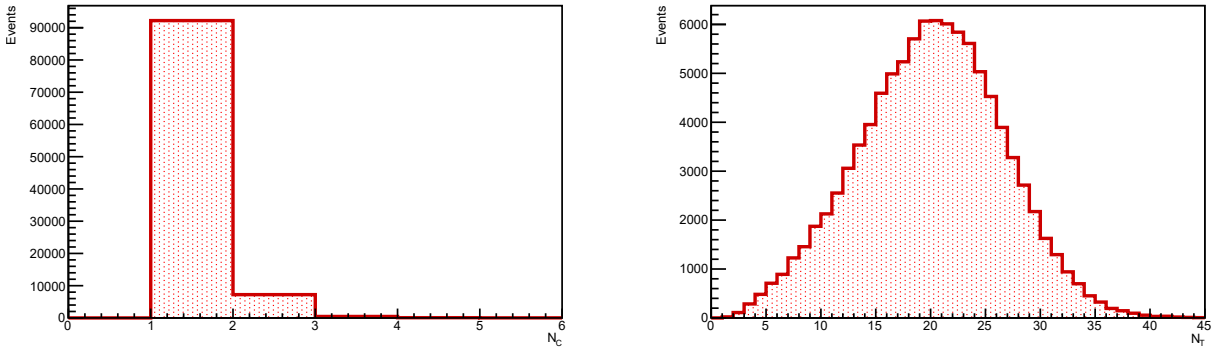

Figure 4.5: The number of clusters generated by the sliding window method N_C and topological clustering N_T for γ with lower granularity

The difference between sliding window and topological clustering is clearly visible on the histograms above. The number of clusters created by the sliding window algorithm is small and equal to one for an overwhelming majority of the events, with a small minority of events being reconstructed with two clusters in the final states. The two-cluster final states could be identified with pion decays; however in order to be precise we would have to reconstruct the invariant mass of those two clusters and check whether it corresponds to the pion rest mass.

However, even then the efficiency of neutral pion identification would be very low, as we observed in section 4.3. On the other hand, the topological clustering method created a large number of clusters in the final state. The difference between working with lower and higher granularity was negligible in terms of sliding window cluster distributions. For topological clustering the number of created clusters per event was increased by a factor of $\mathcal{O}(10)$ when working with smaller cell size.

4.3 Identification of resolved π^0s

Our first task was to identify events, where the clustering mechanism was able to reconstruct the photon pair as a pair of sliding window clusters. In order to correctly identify cluster pairs corresponding directly to the pion decay we computed the invariant mass m of two sliding window clusters. The full distribution of m is shown on figure 4.6, with what can be interpreted as the neutral pion mass peak visible at 0.149 GeV. To reject events beyond this peak, a cut was employed that accepted events with $m \leq 0.230$ GeV. We then calculated ratio of accepted events to all events in an energy bin to obtain a dependence displayed on figure 4.7.

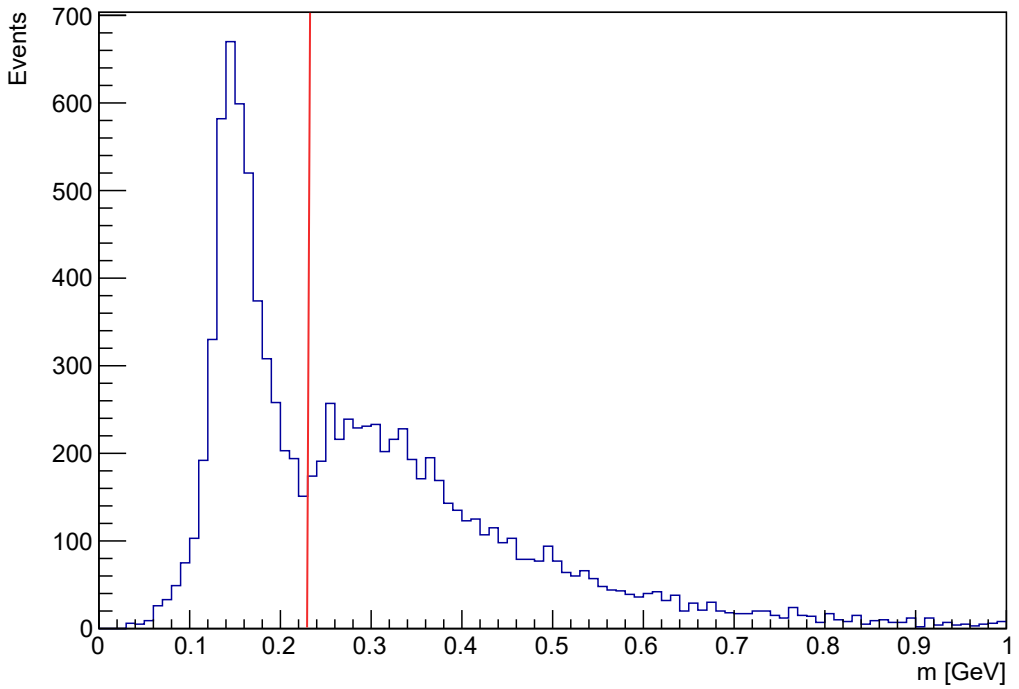

Figure 4.6: Sliding window cluster invariant mass distribution with the cut at E=0.230 GeV indicated

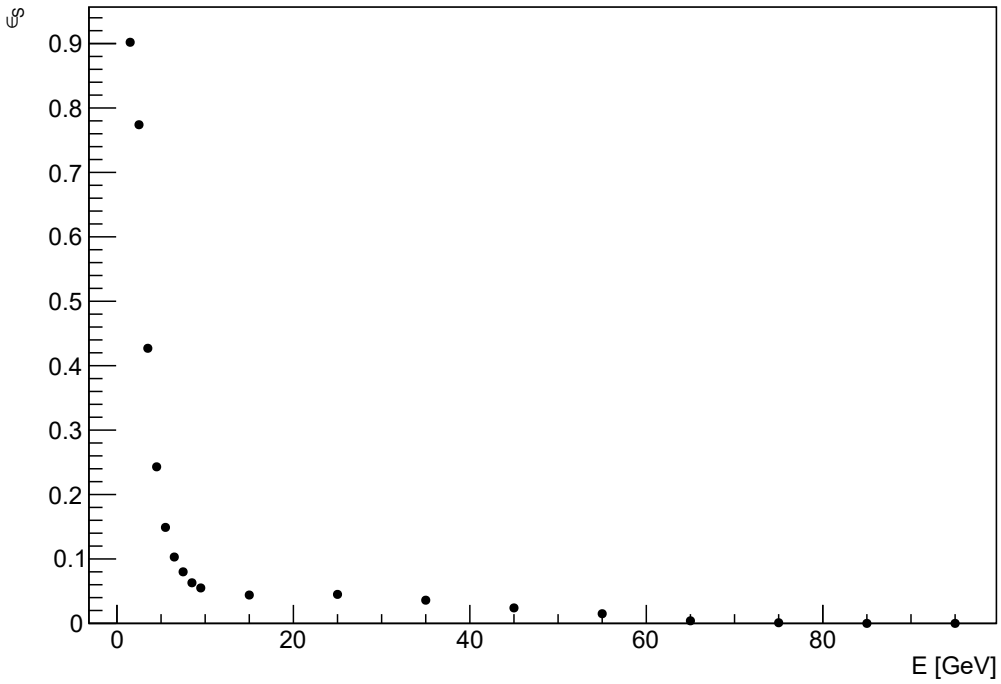

Figure 4.7: Dependence of resolved π^0 identification efficiency ϵ_S on pion energy E

We observed that regardless of the calorimeter granularity, the probability drops off sharply between 0-10 GeV, beyond which it slowly tends to zero at $E >$ 70 GeV. This is caused by the fact that our clustering algorithm was not optimized for π^0 identification. Minimal angle separating two decay photons at $E = 10$ GeV (beyond which the ϵ_S rapid descend slows and gradually tends to zero) is $\alpha_{min} = 1.59°$. To compare this with cell size of $\Delta\theta \times \Delta\phi = 0.57° \times 0.47°$, the separation of photons is only $\sim$3 cells at this energy for lower granularity and $\sim$12 cells for higher granularity. Compared with the size of our sliding window — 9×17 cells in $\eta \times \phi$ — we can immediately identify the reason behind cluster merging beyond $E \sim 10$ GeV.

4.4 Discriminating variables

In an ideal case, we would be able to reconstruct the π^0 decay as a pair of clusters, each corresponding to the relevant decay photon. In reality however, the electromagnetic showers caused by these photons largely overlap, a problem that becomes more pronounced when dealing with π^0 particles at higher energies (as discussed in section 1.2.2). To discriminate between π^0 and γ showers, we need to study the shower substructure using the energy deposition in cells. We chose a number of discriminating variables that represent the energy distribution in an electromagnetic shower. The variables are defined as:

1. E_{max} - The energy contained in a cell with the largest energy deposit in the second layer of the calorimeter

2. E_{2max} - The second largest energy deposit in the second layer of the calorimeter

3. E_{ocore} - Energy deposited in cells surrounding the shower centre defined as

$$E_{ocore} = \frac{E(3) - E(1)}{E(1)} \tag{4.1}$$

   where $E(n)$ is the total energy deposited in $\pm n$ cells surrounding the cell with the highest energy deposit

4. E_n as the sum of energy contained in the first n layers of the calorimeter

5. E_{i1} - Energy deposited in the i^{th} layer of the calorimeter divided by energy deposited in the first layer

$$E_{i1} = \frac{E_i}{E_1} \tag{4.2}$$

6. E_{iT} - Energy deposited in the i^{th} layer of the calorimeter divided by the total energy deposited in the calorimeter

$$E_{iT} = \frac{E_i}{E} \tag{4.3}$$

7. W_{nl} - variable determining shower width in a calorimeter layer l, defined as a normalized sum of energy over $\pm n$ cells in the η coordinate and ± 1 cells in ϕ, weighted by the distance in the $\eta \times \phi$ space

$$W_{nl} = \frac{\sum_{i=1}^{n} E_i \times \Delta R^2}{\sum_{i=1}^{n} E_i} \tag{4.4}$$

   where ΔR is defined as

$$\Delta R^2 = (\eta - \eta_{max})^2 + (\phi - \phi_{max})^2 \tag{4.5}$$

Variable E_{i1} was calculated for $i = 2$, $i = 3$. The variable E_{iT} was calculated for $i = 1$, $i = 1$ and $i = 3$. Finally, W_{nl} was calculated in the third layer for $n = 3$. The distribution was calculated for multiple energy bins in order to observe the effect of particle energy on the distribution shape. A selection of the variables is displayed on figure 4.8 for the energy range E = 30-50 GeV. The whole set of discriminating variables' distribution for various intervals of energy is displayed in Appendix A.2. The variables were also calculated separately for higher and lower granularity. In order to account for possible singularities, events with zero energy deposit for any layer were discarded, leading to a partial loss of events in the range of 0-15 GeV.

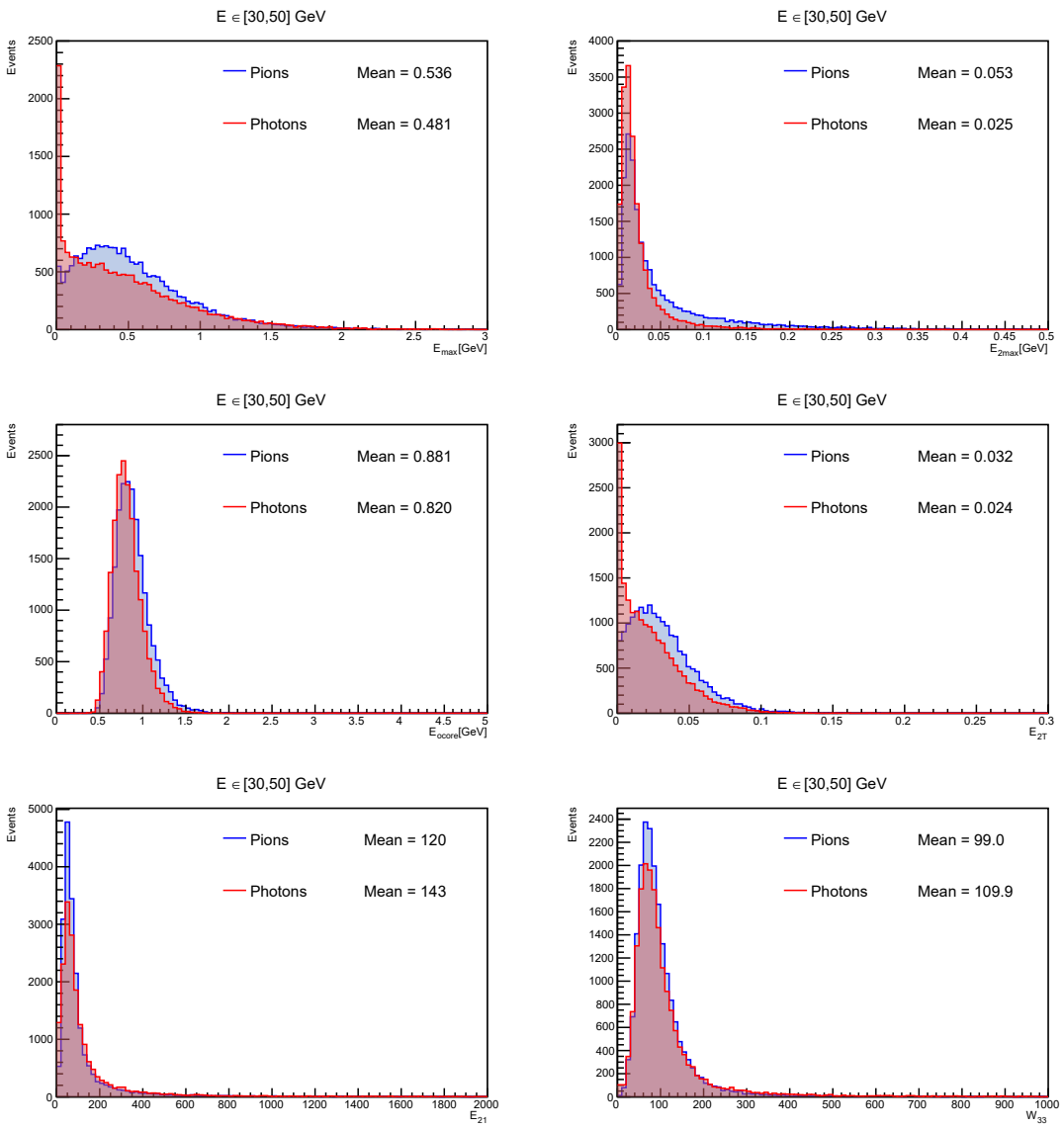

Figure 4.8: Distribution of discriminating variables for pion and photon showers at energies of E = 30-50 GeV in a lower granularity environment

4.5 Rectangular cuts

At the beginning the question was how to choose optimal cuts that would reject a significant portion of the background without loosing too much of the signal. The first simplest trial was to impose cuts by hand when using a small sample of discriminating variables — E_{max}, E_{2max} and E_{ocore}. We worked with data in a bin of 10-30 GeV. The variables were divided into a grid with points corresponding to cut values, for which we calculated ϵ_S, r_B and significance s defined as

$$s = \frac{S}{\sqrt{S+B}} \tag{4.6}$$

where $S(B)$ is the number of signal(background) events passing the cut. Out of the points we selected the optimal set of cuts as the set with maximal significance. This manual approach was compared with the TMVA Rectangular cuts

(henceforth also "Cuts") method, the main advantage of this method being its ease of use, transparency and robustness when it comes to overtraining. We also aimed to compare its performance to latter use of Boosted Decision Trees. For the multivariate parameter fitter we opted to use the Genetic Algorithm method. When compared to Monte Carlo method used for the Cuts MVA method, optimization by Genetic Algorithm exhibited higher AUROC values and significantly better ROC curve smoothness at the cost of longer run times. The parameters with which we configured the performance of the General Algorithm were population size for the GA *PopSize*, number of steps for convergence *Steps* and the number of independent cycles for fitting *Cycles*, the chosen values for our optimization are displayed in the table 4.1

Table 4.1: The selected configuration options for TMVA Cuts method

PopSize	800
Steps	40
Cycles	5 %

The cuts chosen by us and by the Cuts algorithm when finding a working point with maximal significance are compared in table 4.2. The agreement between the cut values was very good and the resultant value of cuts also hinted at the E_{2max} variable as being the weakest classifier out of the three. Nevertheless, the variable was kept, as the next phase of analysis included a wider set of discriminating variables.

Table 4.2: The selected configuration options for TMVA Cuts method

	Manual Cuts	**TMVA Cuts**
ϵ_S	0.87	0.87
r_B	0.50	0.50
E_{max}	0.020	0.020
E_{2max}	0.000	0.001
E_{ocore}	0.712	0.704

In the next analysis the full set of variables from the section 4.4 was used for training and testing. To evaluate performance at multiple energy scales the data has been split into discrete energy bins and the Cuts algorithm has been trained and tested on each bin to determine pion identification efficiency. We also analyzed separately the efficiency ϵ_S for configuration with higher granularity and configuration with lower granularity. The value of ϵ_S was extracted for $r_B = 0.8$. ϵ_S was normalized to the set of unresolved pions in an energy bin and the events. By unresolved pions we mean our statistic without events that were deemed accepted in section 4.3. The results of the analysis are displayed on the figure 4.9.

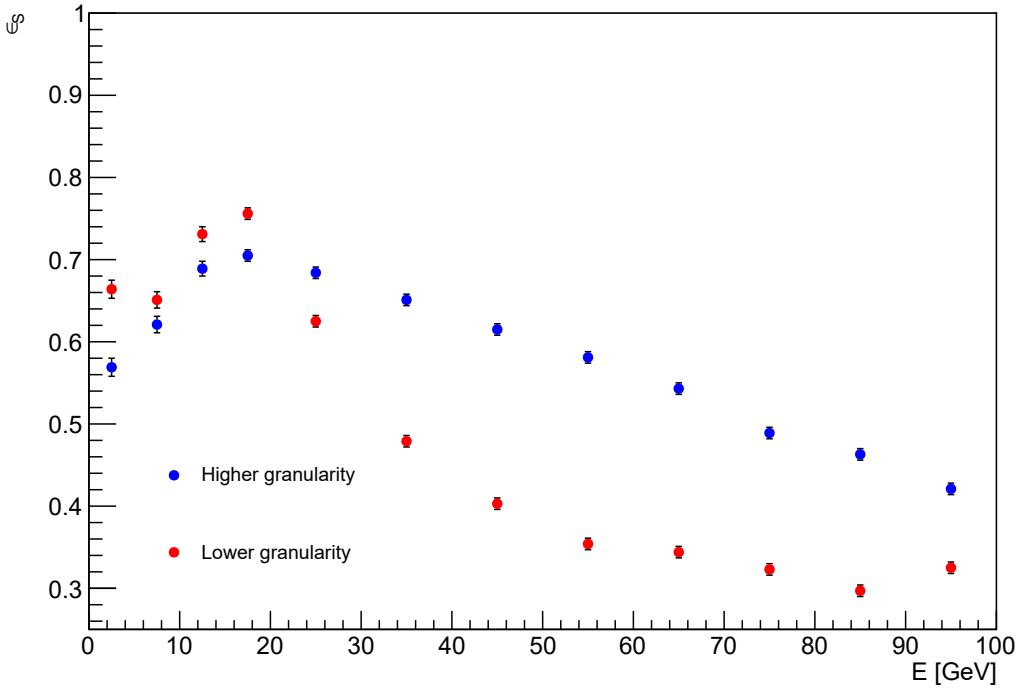

Figure 4.9: Dependence of π^0 identification efficiency ϵ_S on pion energy E

At lower energies, we observe the identification efficiency rising to a maximum located in the bin of $E = 15$-20 GeV after which efficiency drops gradually. The influence of higher granularity or ϵ_S is clearly visible for energies above $E \approx 25$ GeV, where the efficiency drops noticeably more slowly when working with smaller cells in the calorimeter. The dependence calculated for lower granularity drops more rapidly, reaching a presumably spurious plateau at energies corresponding to approximately 70 GeV. At these energies, the angle between two decay photons reaches merely 0.22 ° and the resulting electromagnetic showers overlap significantly. Considering the finite cell size, the efficiency is expected to continue falling for increasing neutral pion energies, instead of remaining constant.

4.6 Boosted Decision Trees

4.6.1 Optimization of BDT hyperparameters

To increase reconstruction efficiency and check whether the efficiency drop for low-energy pions persists, we turned to the BDT algorithm. In order to optimize BDT performance, we had to set a combination of parameters that maximized the area under the ROC curve (further referred to also as AUROC). In order to find a suitable combination we observed the dependence of AUROC on the number of trees in the BDT, or *NTrees*, β parameter of the AdaBoost algorithm *AdaBoostBeta*, minimum node size (or *MinNodeSize*) and maximal allowed depth of the decision tree *MaxDepth*. The parameters were first set to their default value and then successively optimized in that order. The default values provided by the TMVA package are displayed in the table 4.3 and the dependence of AUROC on

the respective hyperparameters on figures 4.10, 4.11, 4.12, 4.13. The dependence for all energy bins can be found in Appendix A.3.

Table 4.3: The default values of BDT hyperparameters

NTrees	800
AdaBoostBeta	0.5
MinNodeSize	5 %
MaxDepth	3

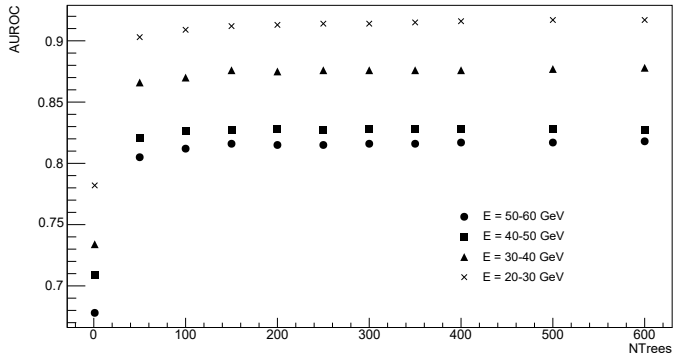

Figure 4.10: AUROC dependence on NTrees hyperparameter for E = 20-60 GeV

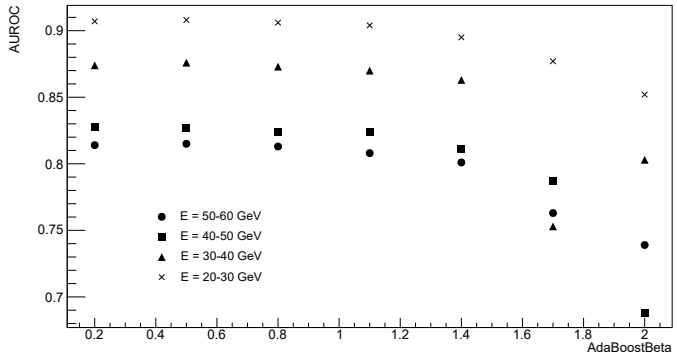

Figure 4.11: AUROC dependence on AdaBoostBeta hyperparameter for E = 20-60 GeV

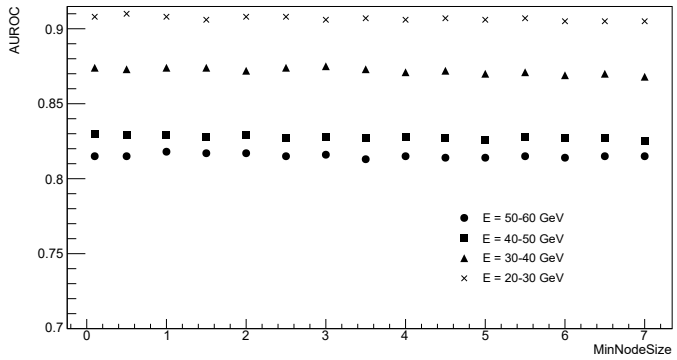

Figure 4.12: AUROC dependence on MinNodeSize hyperparameter for E = 20-60 GeV

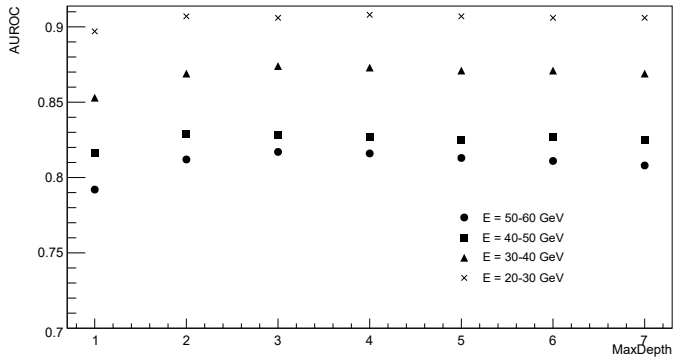

Figure 4.13: AUROC dependence on MaxDepth hyperparameter for E = 20-60 GeV

Along with the dependence of AUROC we also checked the value of the Kolmogorov-Smirnov test to prevent overtraining of the decision tree. Final parameters were chosen separately for data with lower granularity and data with higher granularity, as the parameters selected for one set of events led to overtraining, or underperformance in the other set. The final set of parameters is displayed in the table 4.4.

Table 4.4: The optimized values of BDT parameters

	Higher granularity	Lower granularity
NTrees	250	200
AdaBoostBeta	0.5	0.5
MinNodeSize	1.5 %	1.5 %
MaxDepth	3	2

The values were also compared to optimal values obtained by ROOT via using the *OptimizeTuningParameters* method. The parameters recommended by ROOT were identical for both lower granularity and higher granularity settings as shown in 4.5.

Table 4.5: Optimal hyperparameters chosen by the ROOT optimization algorithm

NTrees	1000
AdaBoostBeta	0.4
MinNodeSize	1 %
MaxDepth	4

However, the parameters were in reality suboptimal, as even though they maximized the available AUROC, they failed to account for overtraining. At the end we instead opted for parameters chosen previously by our method.

4.6.2 Training and testing

After the optimal hyperparameters were obtained we trained and tested the decision tree on our data. The data was randomly split into two separate groups to compare its performance and keep overtraining in check. We split the data into energy intervals and calculated signal efficiency for every bin to obtain dependence of efficiency on energy. The dependence is displayed on 4.14. The signal efficiency was obtained from a chosen working point at the ROC curve calculated for a fixed value of background rejection equal to 0.8.

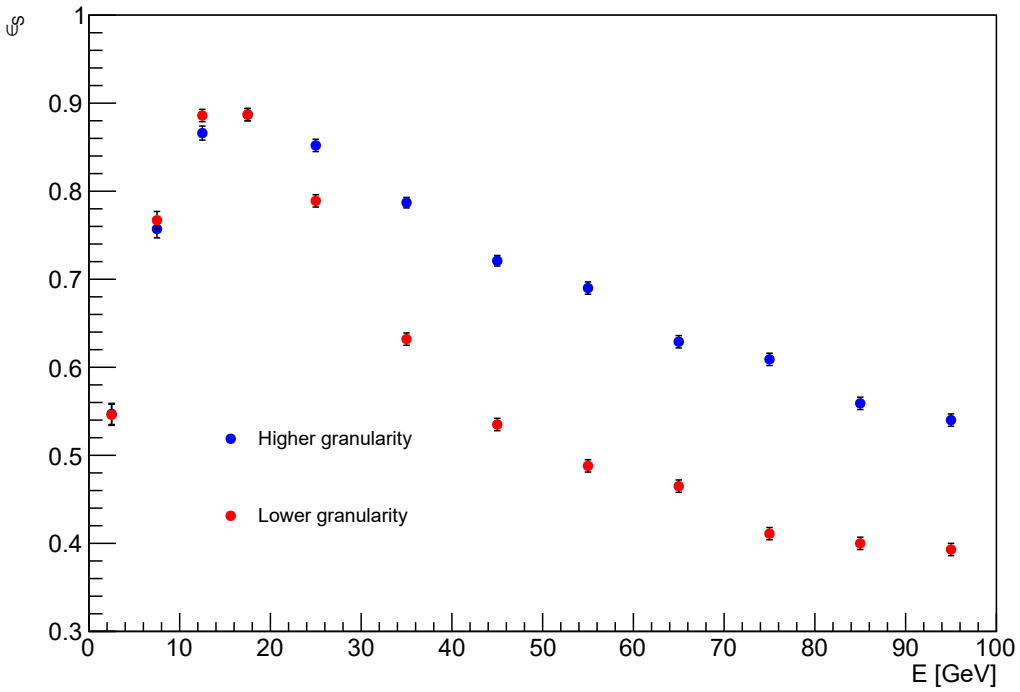

Figure 4.14: Dependence of π^0 identification efficiency ϵ_S on pion energy E

The identification efficiency displays a significant improvement over the efficiency obtained using the Cuts method, with the maximum located at $E \sim$ 15 GeV rising from ∼0.7 to ∼0.9 and the general efficiency also being higher. The behaviour at low energies, however still displays rapid dropping. This could be remedied by optimizing the clustering algorithm, which should be able to

identify neutral pions at low energies thanks to the large decay angle between the corresponding photons, however, it is still desirable to increase the performance of the method used. An option was to try invariant mass distribution reconstructed from the topological clusters, defined as:

1. m_{cc} as the invariant mass of the two clusters with the highest energy deposits reconstructed by topological clustering

The variable differs from the rest of the set by being obtained from clusters, rather than from the energy distribution in cells and layers of the calorimeter. The analysis was repeated with m_{cc} included. The results are displayed on 4.15.

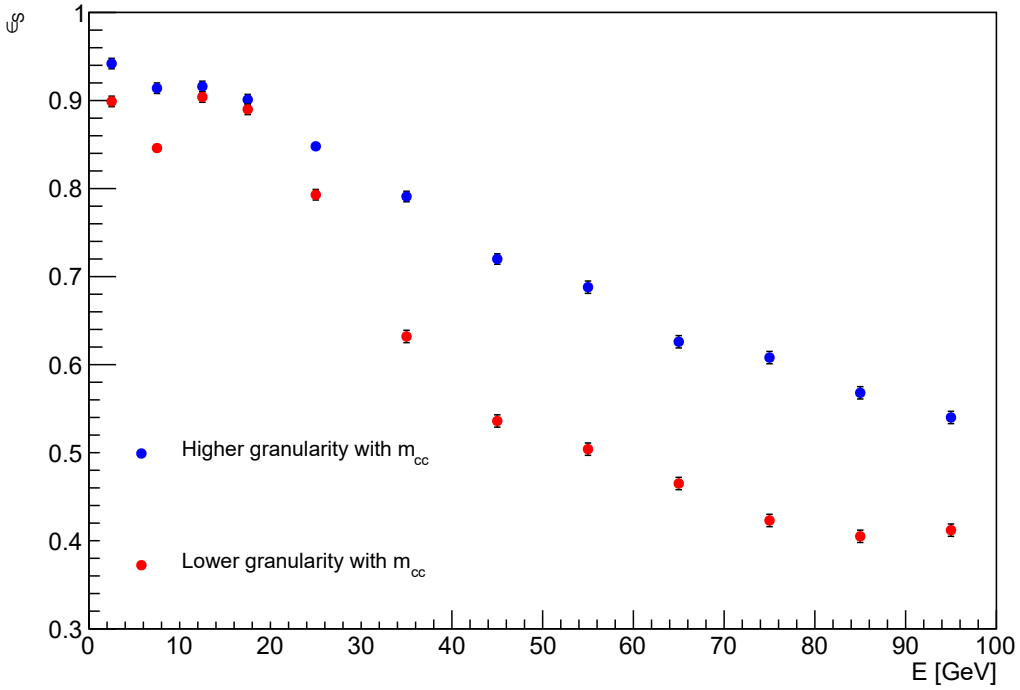

Figure 4.15: Dependence of π^0 identification efficiency ϵ_S on pion energy E with m_{cc}

When including m_{cc} in the analysis, there is a visible difference for low energy pions. The identification efficiency reaches a maximum instead the dropping behaviour exhibited by 4.14. This behaviour can be attributed to a fact that for low energy pions, the m_{cc} variable distribution contains a prominent peak, which is located around value $\mu = 0.130$ GeV for lower granularity and $\mu = 0.125$ GeV for higher granularity. The value can be interpreted as corresponding to the π^0 mass, which is well resolved from the photon m_{cc} distribution as seen on 4.16 and 4.17. For higher energies, the two distributions rapidly overlap and merge, so that the m_{cc} variable has very little discriminating power when it comes to high energy pions and the energy dependence of π^0 resolution efficiency above $E \sim 15$ GeV is identical regardless of m_{cc} inclusion in our analysis. The overall comparison of identification efficiency obtained by Cuts and BDT algorithm is summarized in table 4.6 by an average of the efficiency curve $\langle \epsilon_S \rangle$.

Table 4.6: $\langle \epsilon_S \rangle$ for the Cuts method and for BDT calculated for low granularity (LG) and high granularity (HG)

	Cuts	**BDT**	
		m_{cc} excluded	m_{cc} included
$\langle \epsilon_S^{LG} \rangle$	0.432	0.540	0.563
$\langle \epsilon_S^{HG} \rangle$	0.546	0.658	0.684

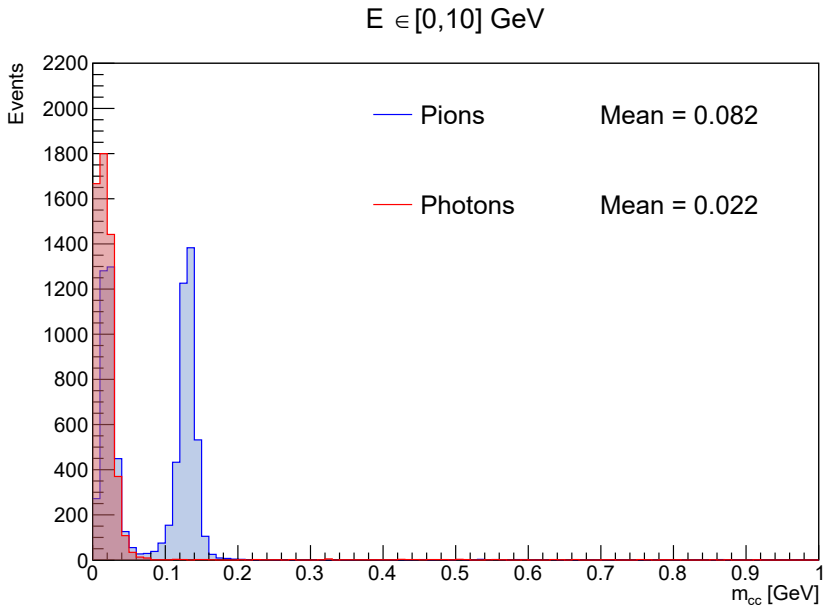

Figure 4.16: m_{cc} distribution for low-energy pions in low granularity environment

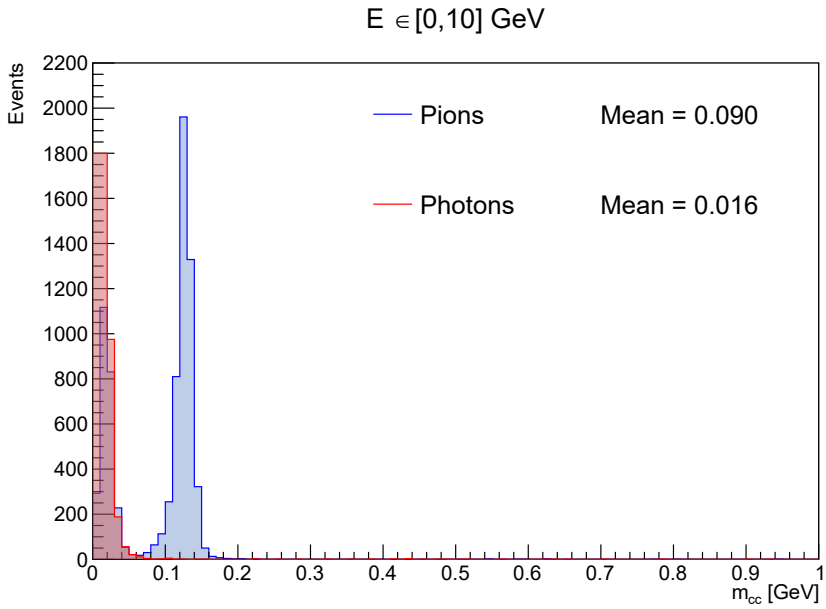

Figure 4.17: m_{cc} distribution for low-energy pions in high granularity environment

Conclusion

In this thesis our task was to explore methods of neutral pion identification in a noble liquid calorimeter deisgned for the FCC-ee. We worked with Monte Carlo simulations from Geant4 as part of FCCSW package that simulated passage of photons and neutral pions through the calorimeter. The energy range we worked with was 0-100 GeV and the calorimeter was simulated with lower and higher granularity. During the course of the thesis, we introduced mechanisms of the neutral pion decay ($\pi^0 \rightarrow \gamma\gamma$), described the FCC-ee project and introduced several detector designs and then explained basics of multivariate methods that were used in our analysis.

In our analysis we first explored the possibility of identifying neutral pions against a single photon background using information from sliding window clustering. We found that the identification efficiency decreased rapidly with rising energy of the pions irrespective of granularity. This we attributed to the clustering algorithm being unoptimized for shower resolution. The topological clustering method could be used for neutral pion identification when paired with a merging algorithm, that would combine the large number of low-energy clusters created during clustering.

Next we used information on energy deposition in cells of the calorimeter to characterise events using a set of discriminating variables. We worked with the Rectangular cuts method and Boosted Decision Trees. By choosing a working point at background rejection equal to 0.8 we calculated identification efficiency against pion energy. We observed a dropping of efficiency in the range of $E \sim 0 - 15$ GeV for both Rectangular cuts and BDT, although the BDT method displayed higher overall efficiency. We also observed that the higher granularity setting consistently led to a higher efficiency. This demonstrates the importance of including a strip layer with smaller segmentation in the real calorimeter that aims to identify precisely neutral pions and other particles decaying into particles with a very small angular separation. Note that for the purposes of our thesis, the granularity was constant across all layers. The real design would have varying segmentation because of the strip layer, thus to accurately simulate the detector response it is necessary to simulate the calorimeter with segmentation differing across layers as well.

To counter the drop in efficiency at lower energies, we introduced a new variable m_{cc}, obtained from topological clustering. The variable was shown to have a very good discriminating power at low energies, which successfully reversed the low-energy dropping. As the variable distribution for pions quickly merges with the distribution for photons, it becomes a rather weak classifier for pions at higher energies. It is therefore imperative to employ different variables at these energies, paired with a calorimeter design, that would enable us to study and resolve even highly overlapping electromagnetic showers such as those created by decaying high-energy neutral pions.

Bibliography

[1] J. Hořejší. *Fundamentals of Electroweak Theory.* First edition. The Karolinum Press, Praha, 2002.

[2] C. Amsler et al. Quark Model. *Physics Letters*, B667(3):173, 2008.

[3] R.L. Workman et al (PDG). *Progress of Theoretical and Experimental Physics*, 083C01 (2022).

[4] A. Abada et. al. FCC-ee:The Lepton Collider. *The European Physical Journal Special Topics*, 228:261–623, 2019.

[5] Future Circular Collider study. [online]: `https://fcc.web.cern.ch/`.

[6] M. Aleksa et. al. Calorimeters for the FCC-hh, 2019. [online]: `https://arxiv.org/abs/1912.09962`.

[7] J. Beringer et. al. (Particle Data Group). *Review of Particle Physics.* Physical Review D 86, 010001 (2012).

[8] M. Aleksa et. al. Calorimetry at FCC-ee. *The European Physical Journal Plus*, 136(1066), 2021.

[9] G. Unal. Performances of the NA48 liquid krypton calorimeter. *Frascati Phys. Ser.*, 21:361–375, 2001.

[10] ATLAS collaboration. ATLAS liquid-argon calorimeter: technical design report. *Technical design report ATLAS*, 1996.

[11] A. Benvenuti et al. The SLD calorimeter system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 289:463–467, 1990.

[12] M. Aleksa. Noble Liquid Calorimetry: Input proposals in Track 2. Presented at: *ECFA Detector RD Roadmap Task Force 6: 2nd Calorimetry Community Meeting*, [online]: `https://indico.cern.ch/event/1246381/`.

[13] F. Brieuc. Noble liquid calorimetry for a future FCC-ee experiment. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 1040:167035, 2022.

[14] N. Morange. Noble Liquid Calorimetry for FCC-ee. *Instruments*, 6(4), 2022.

[15] W Lampl et al. Calorimeter Clustering Algorithms: Description and Performance. Technical report, CERN, Geneva, 2008.

[16] FCCSW. [online]: `https://hep-fcc.github.io/FCCSW/`.

[17] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[18] I. Olkin and A. R. Sampson. Multivariate Analysis: Overview. *International Encyclopedia of the Social and Behavioural Sciences*, pages 10240–10247, 2001.

[19] O. Behnke, K. Kröninger, G. Schott, and T. Schörner Sadenius. *Data Analysis in High Energy Physics*. Wiley-VCH, Berlin, 2013.

[20] Matthew.E. Cross and Emma.V.E. Plankett. *Physics, Pharmacology and Physiology for Anaesthetists*. Second edition. Cambridge University Press, 2014.

[21] G. Cowan. *Statistical Data Analysis*. First edition. Oxford University Press, New York, 1998.

[22] R. Quagliani. *Study of Double Charm B Decays with the LHCb Experiment at CERN and Track Reconstruction for the LHCb Upgrade*. First edition. Springer Cham, 2018.

[23] A. Hoecker et. al. TMVA - Toolkit for Multivariate Data Analysis, 2009.

List of Figures

List of Tables

A. Attachments

A.1 Initial state distributions

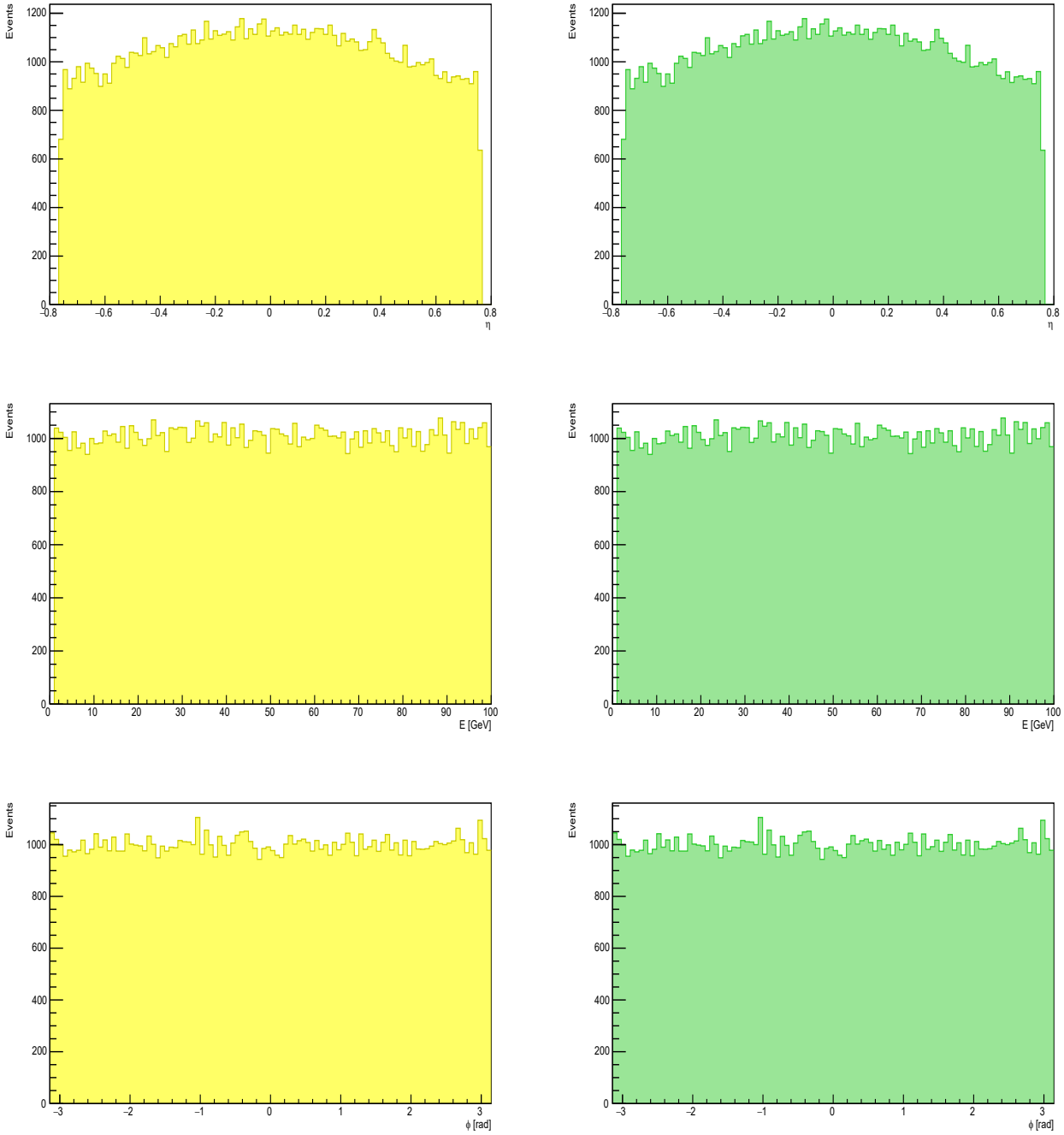

Figure A.1: Distribution of η, E and ϕ for photons (left) and pions (right) with lower granularity

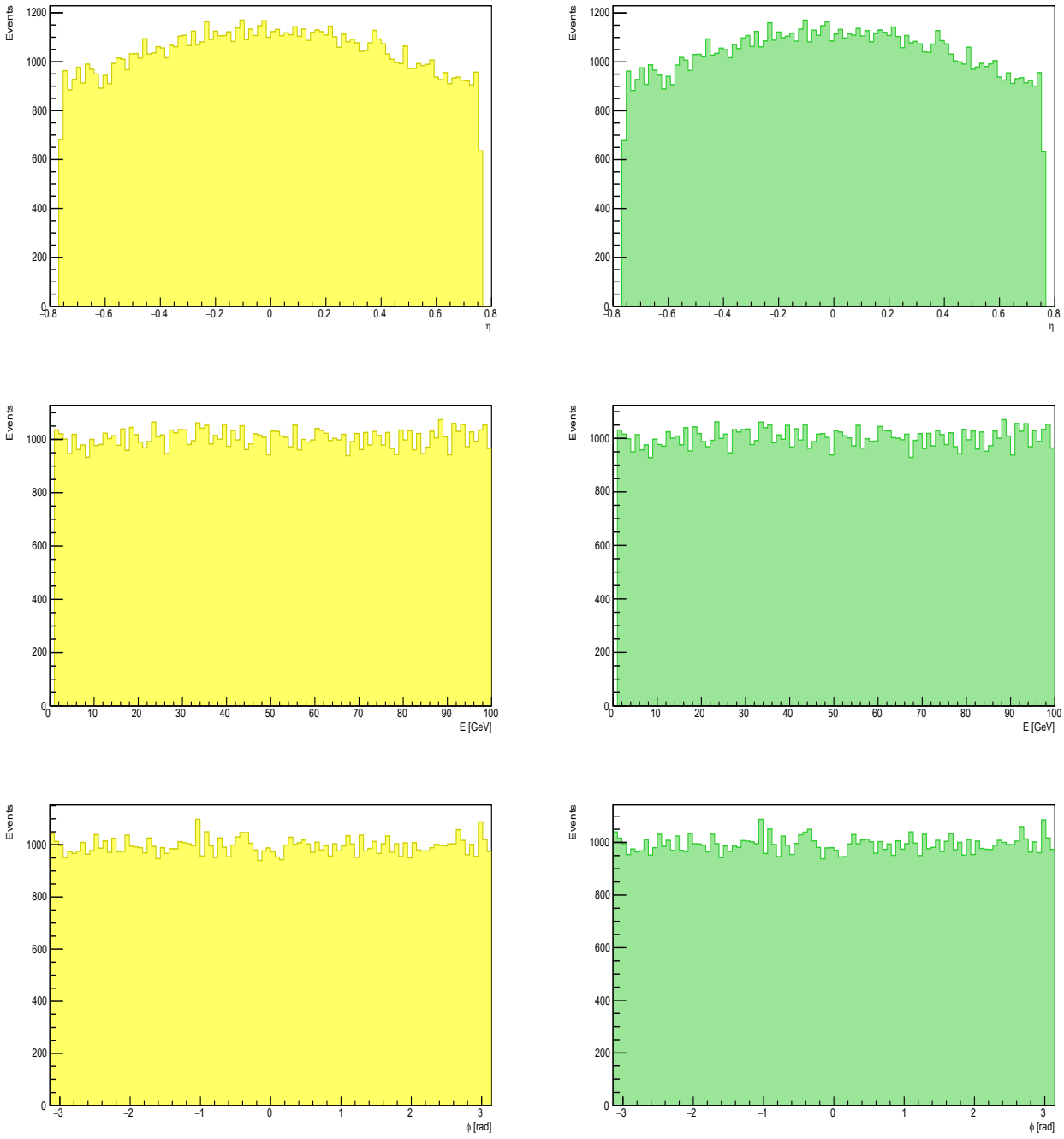

Figure A.2: Distribution of η, E and ϕ for photons (left) and pions (right) with higher granularity

A.2 Discriminating Variables

A.2.1 Discriminating variables for lower granularity

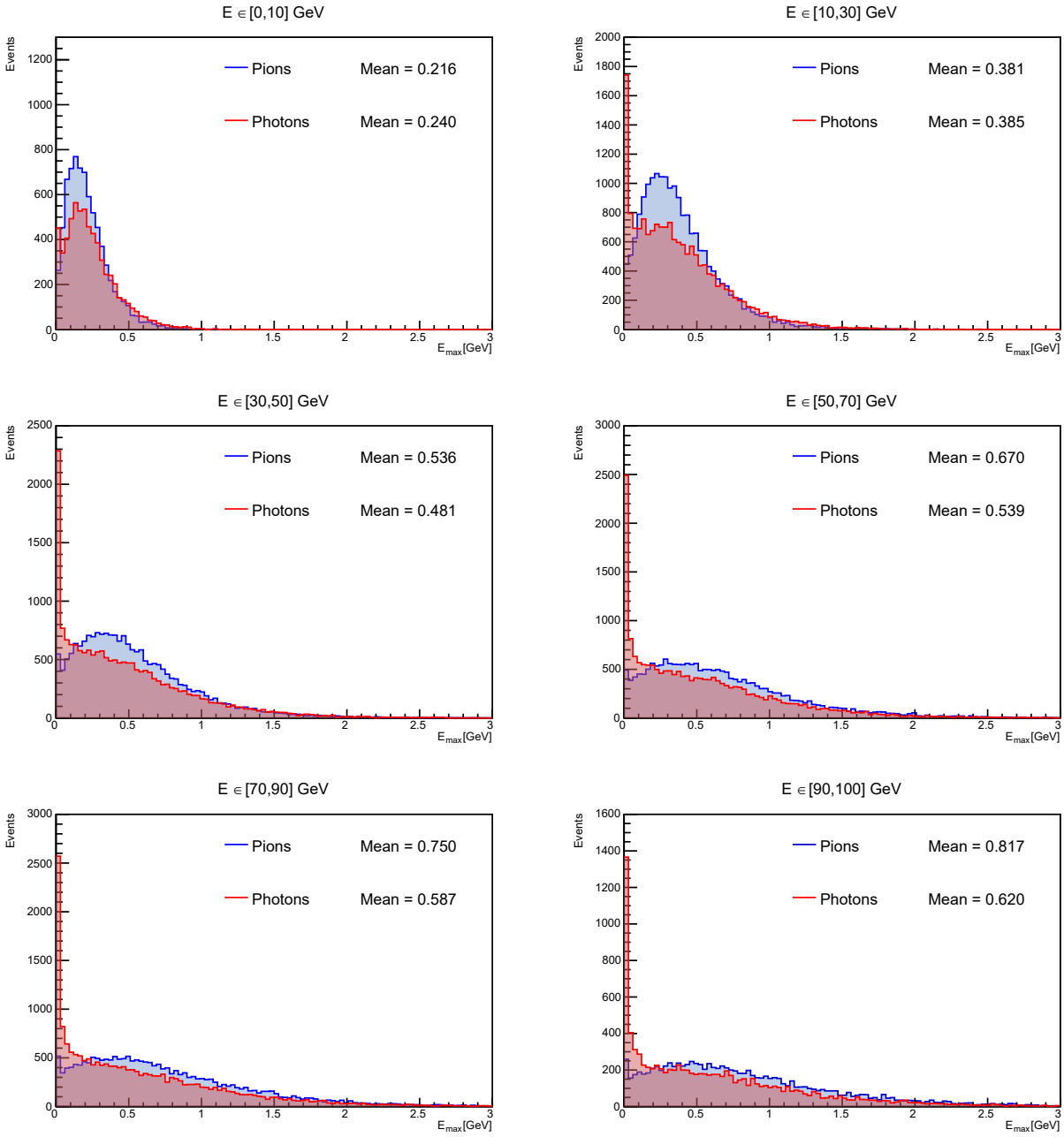

Figure A.3: Distribution of E_{max} for cells with lower granularity

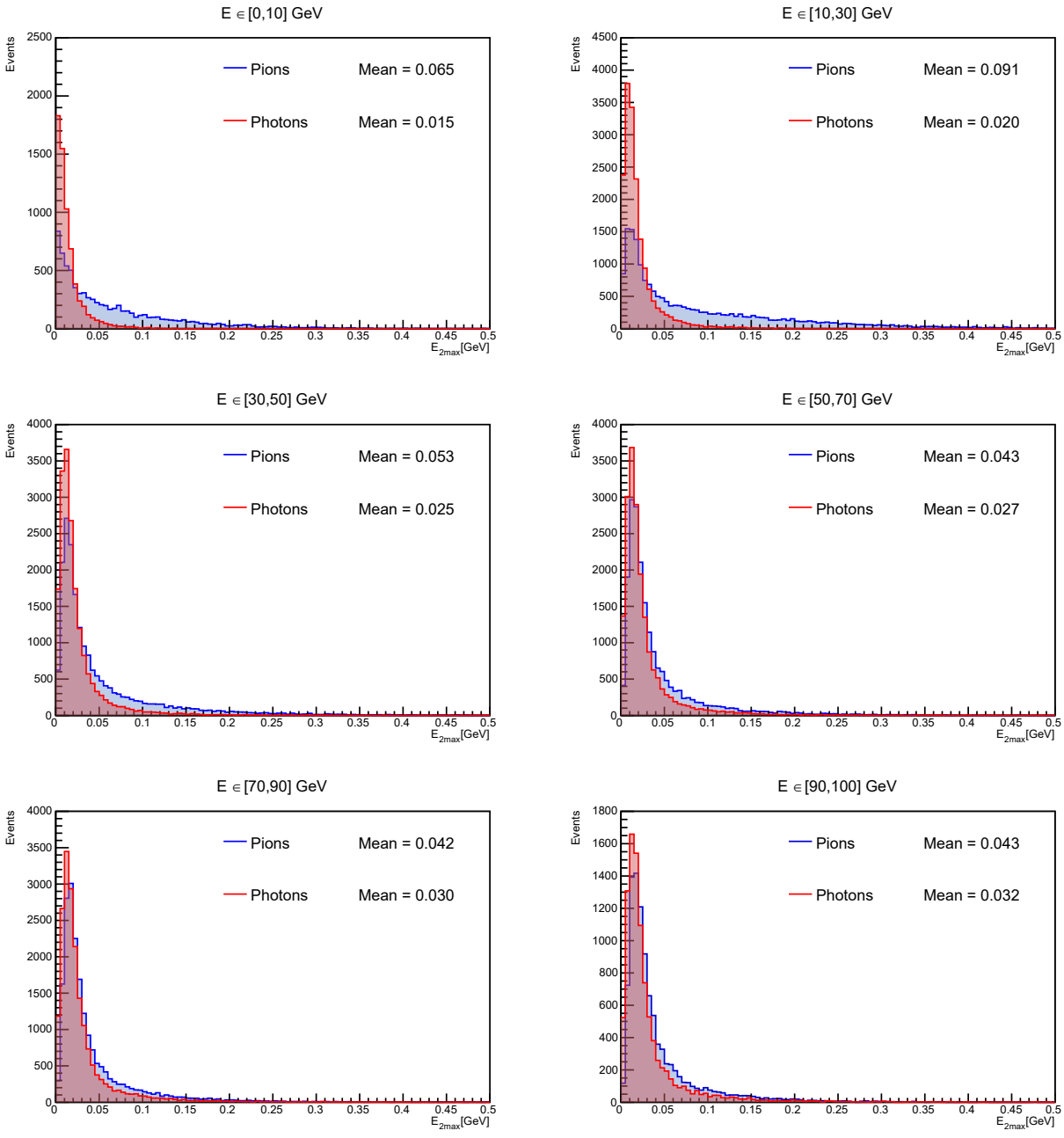

Figure A.4: Distribution of E_{2max} for cells with lower granularity

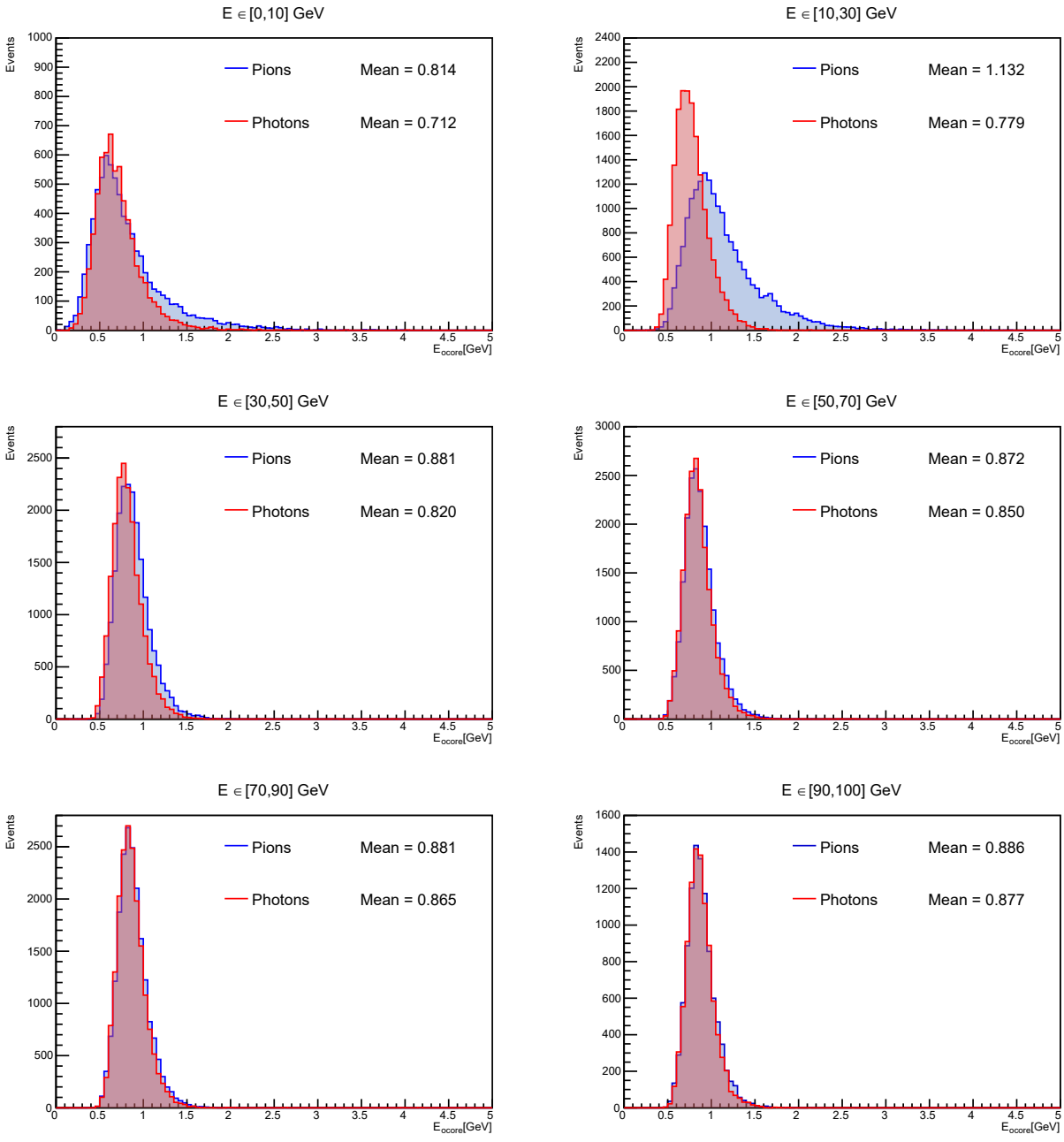

Figure A.5: Distribution of E_{ocore} for cells with lower granularity

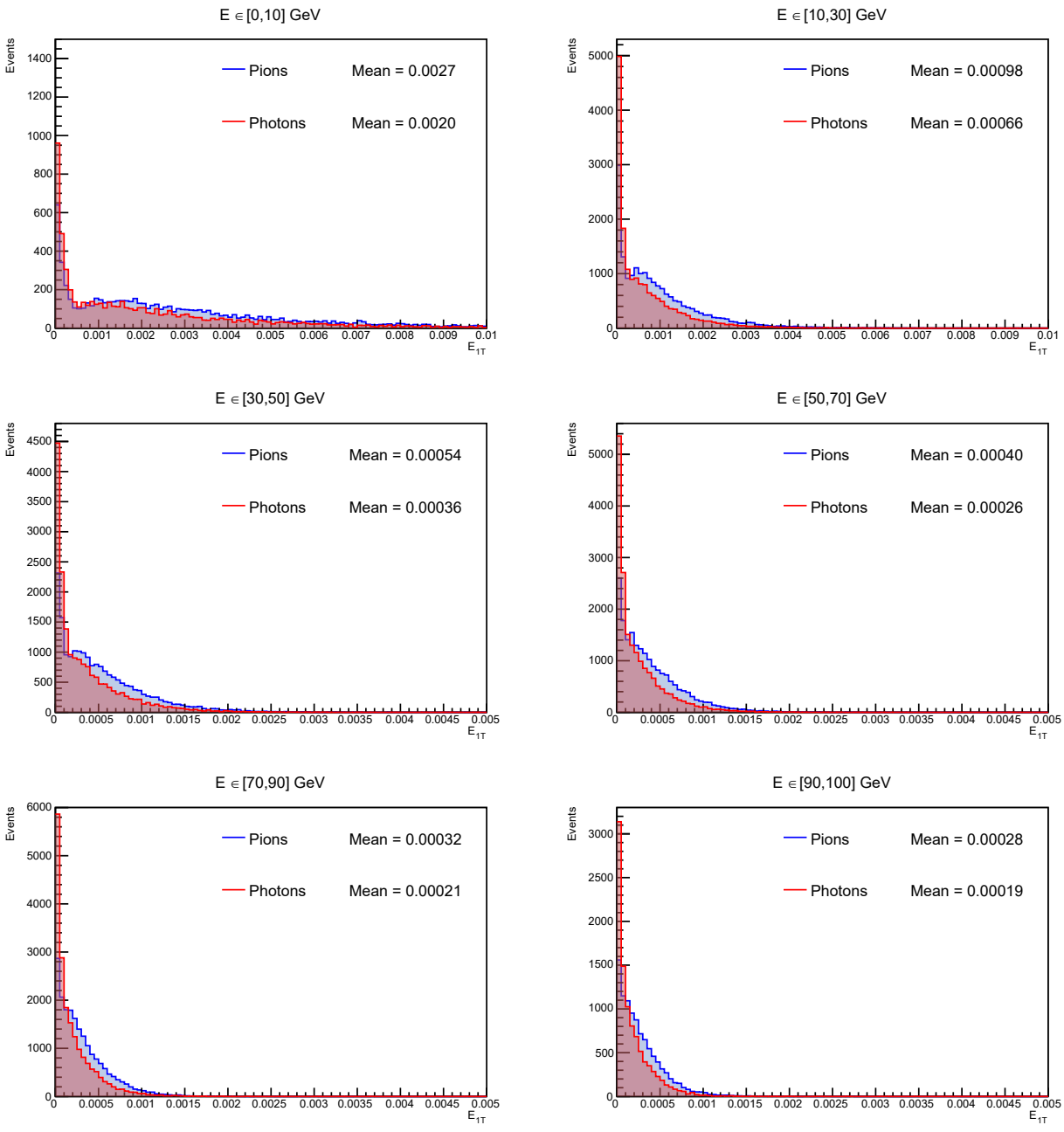

Figure A.6: Distribution of E_{1T} for cells with lower granularity

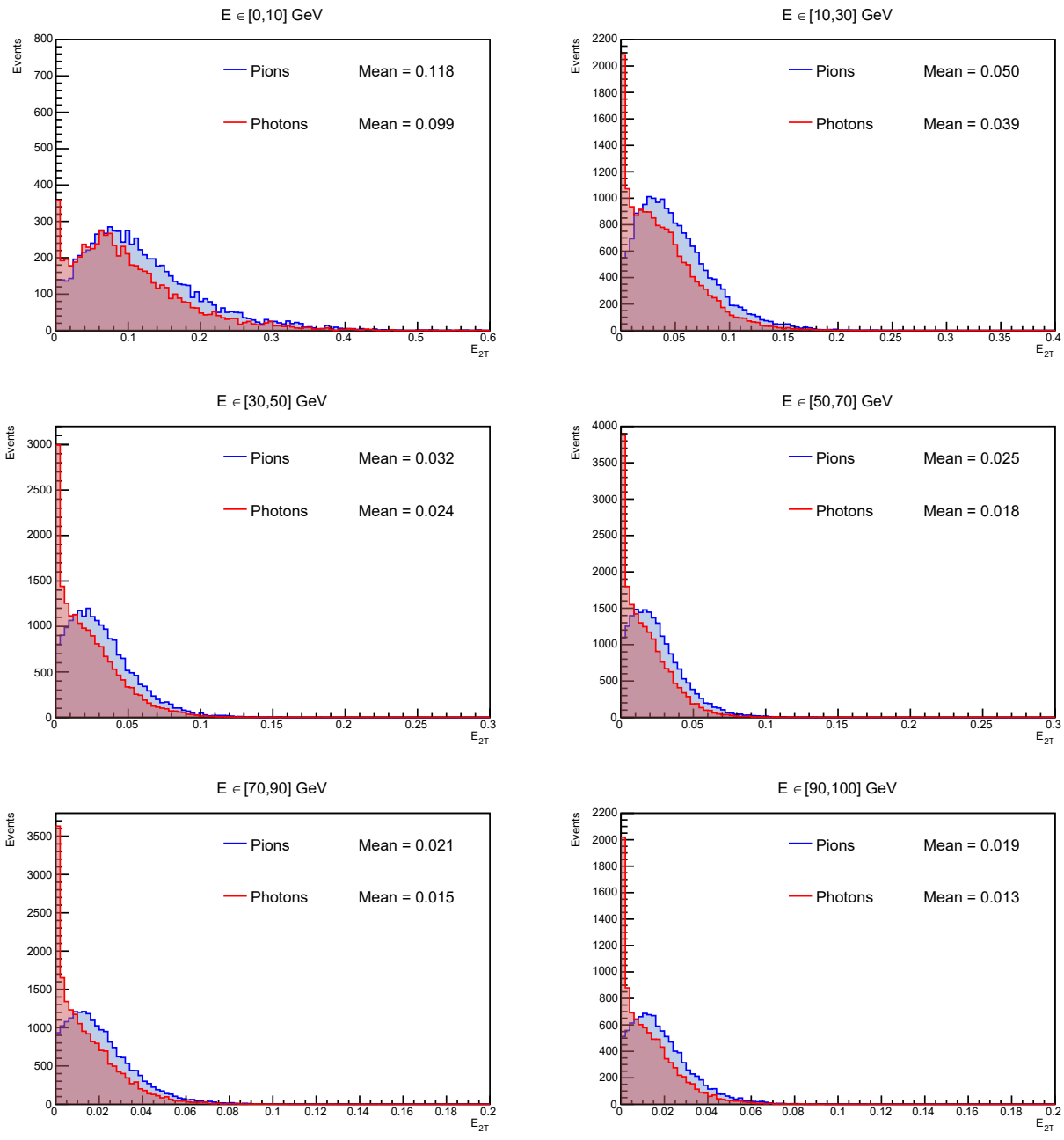

Figure A.7: Distribution of E_{2T} for cells with lower granularity

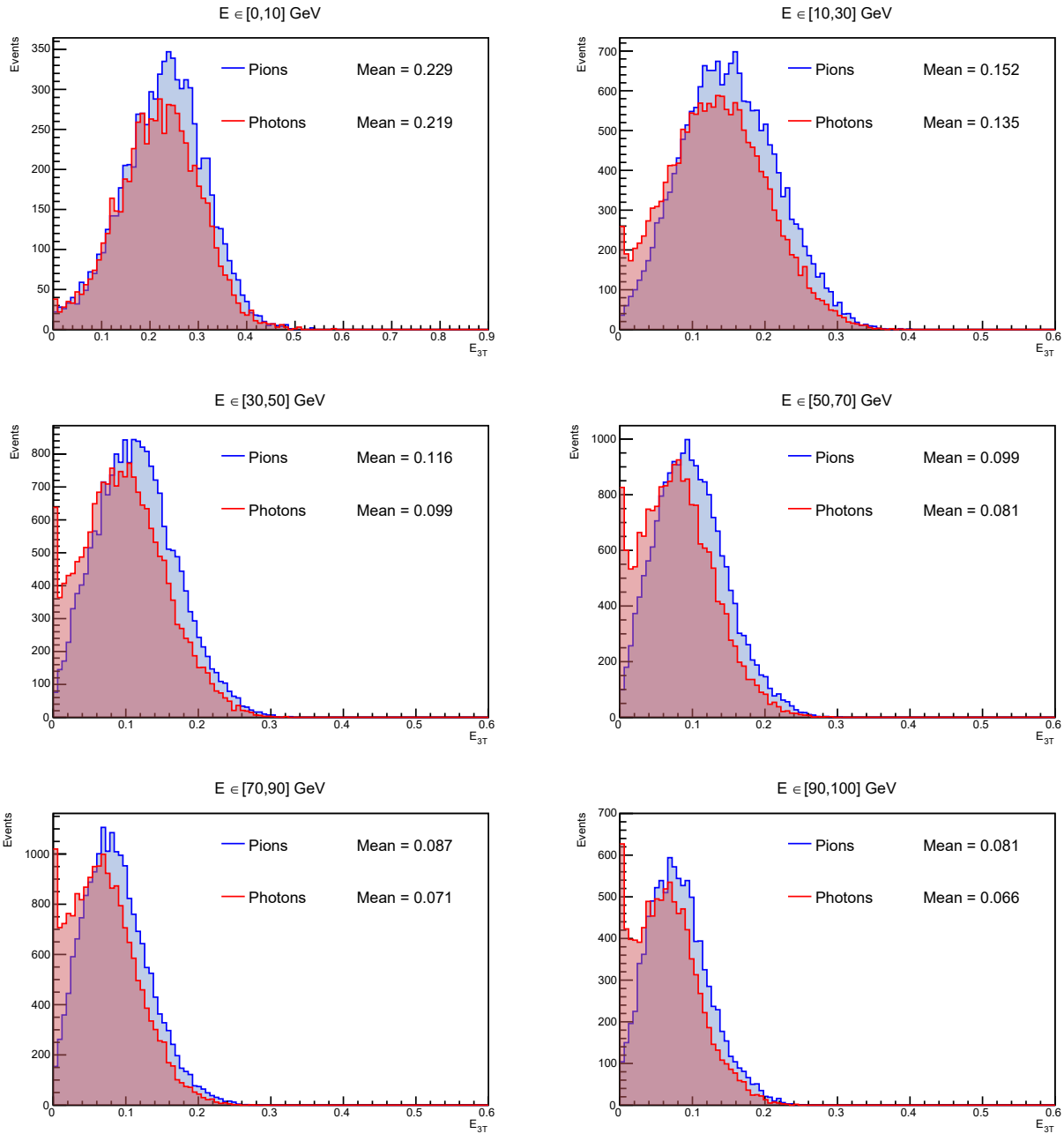

Figure A.8: Distribution of E_{3T} for cells with lower granularity

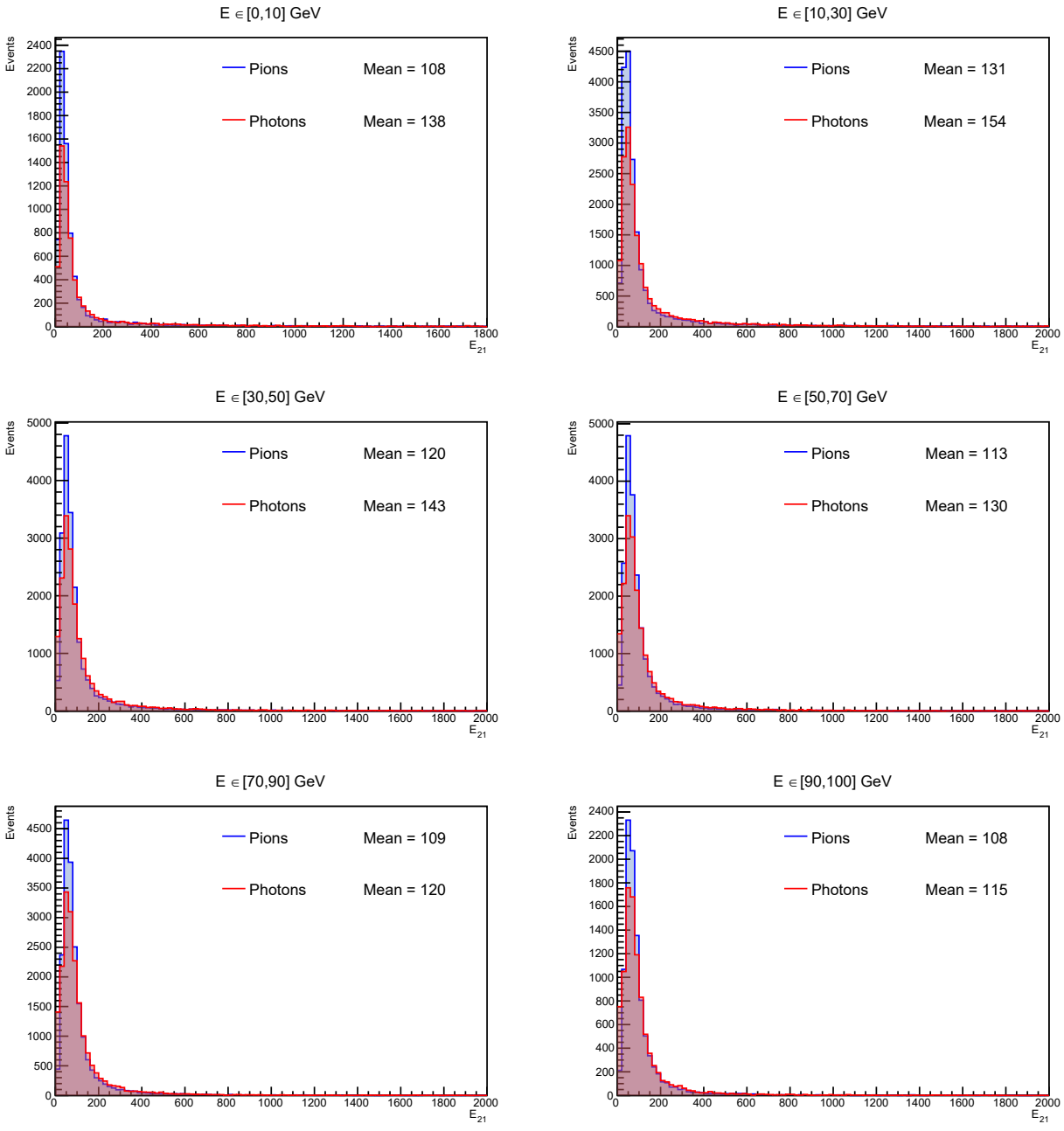

Figure A.9: Distribution of E_{21} for cells with lower granularity

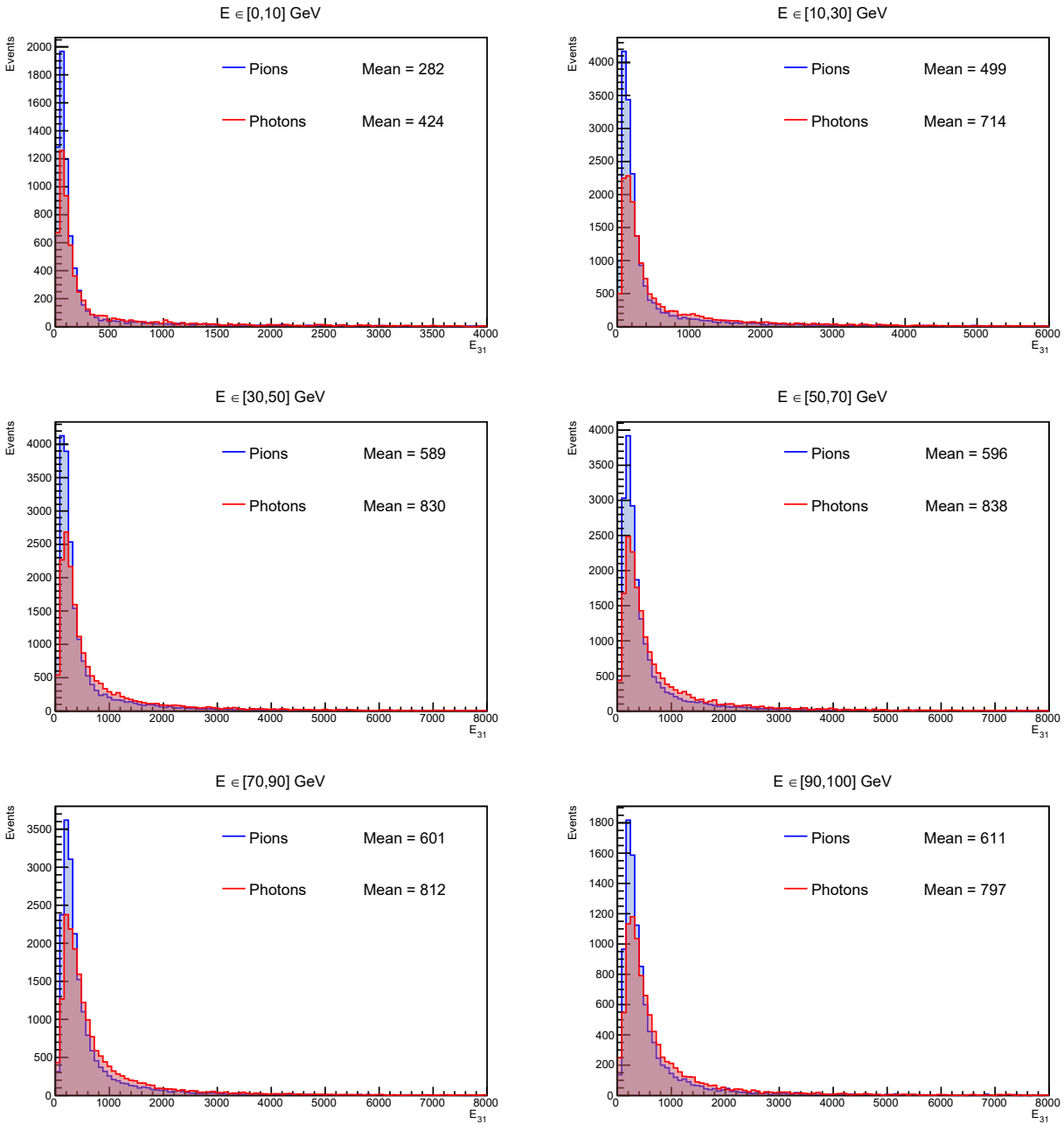

Figure A.10: Distribution of E_{31} for cells with lower granularity

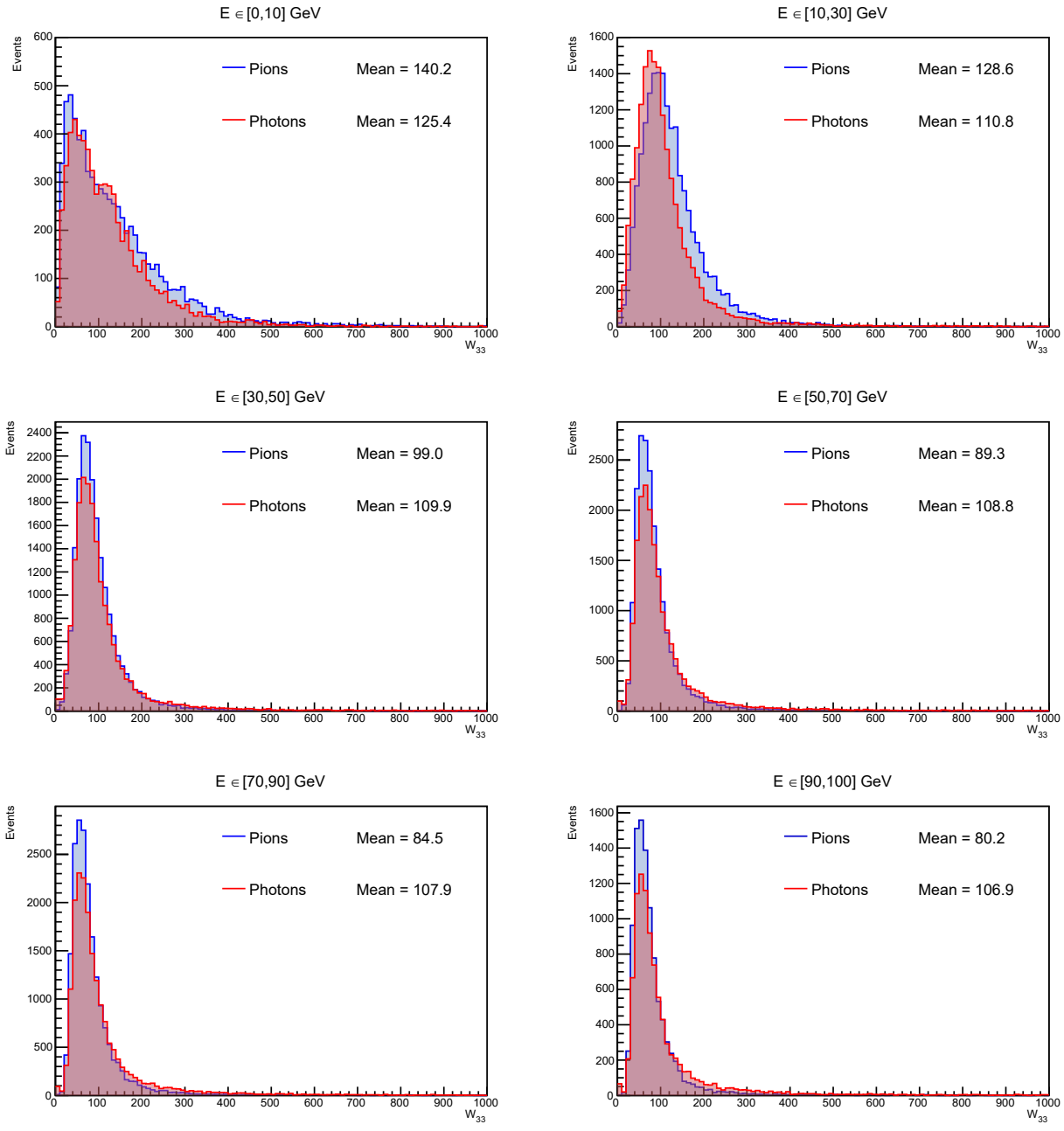

Figure A.11: Distribution of W_{33} for cells with lower granularity

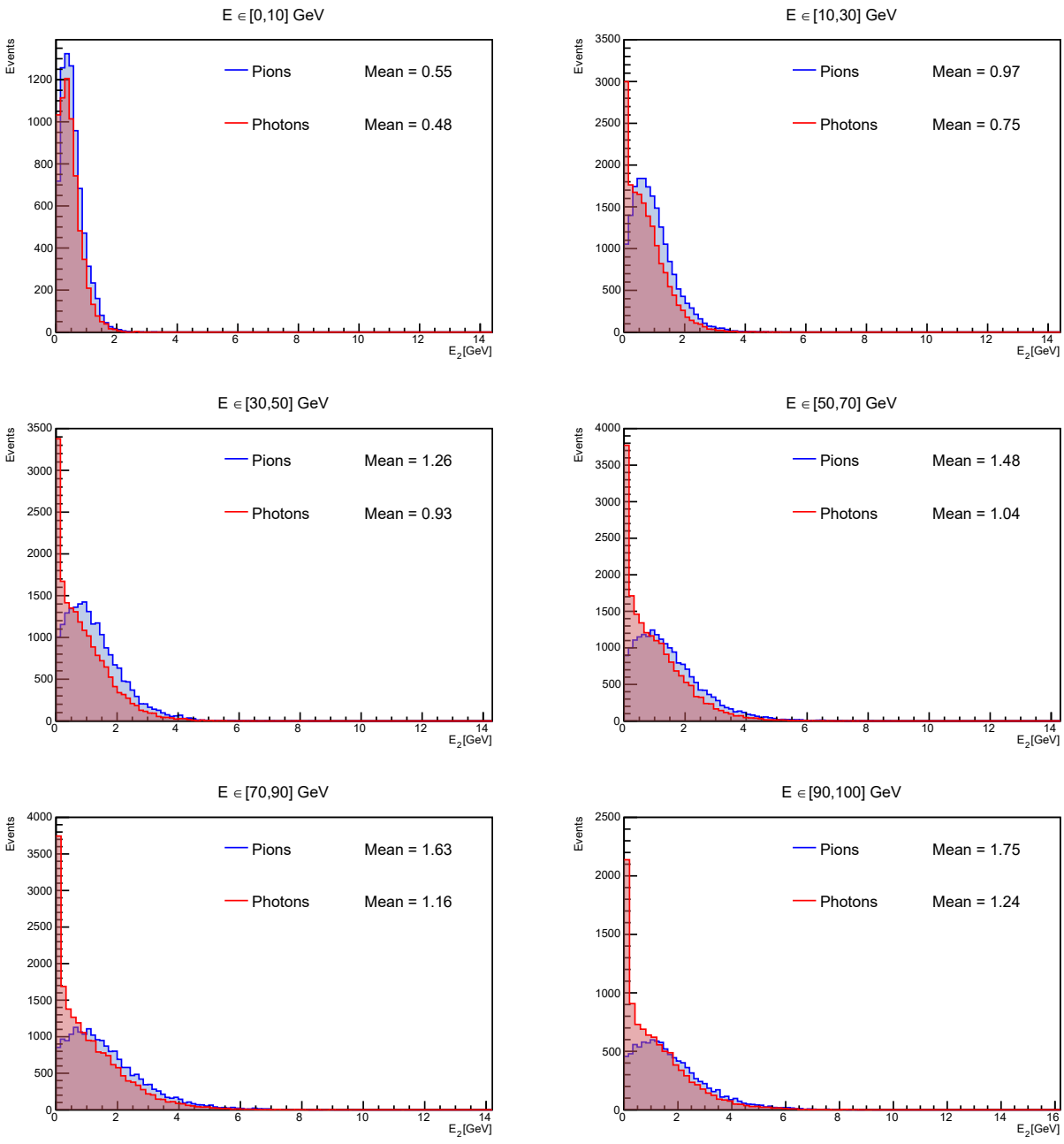

Figure A.12: Distribution of E_2 for cells with lower granularity

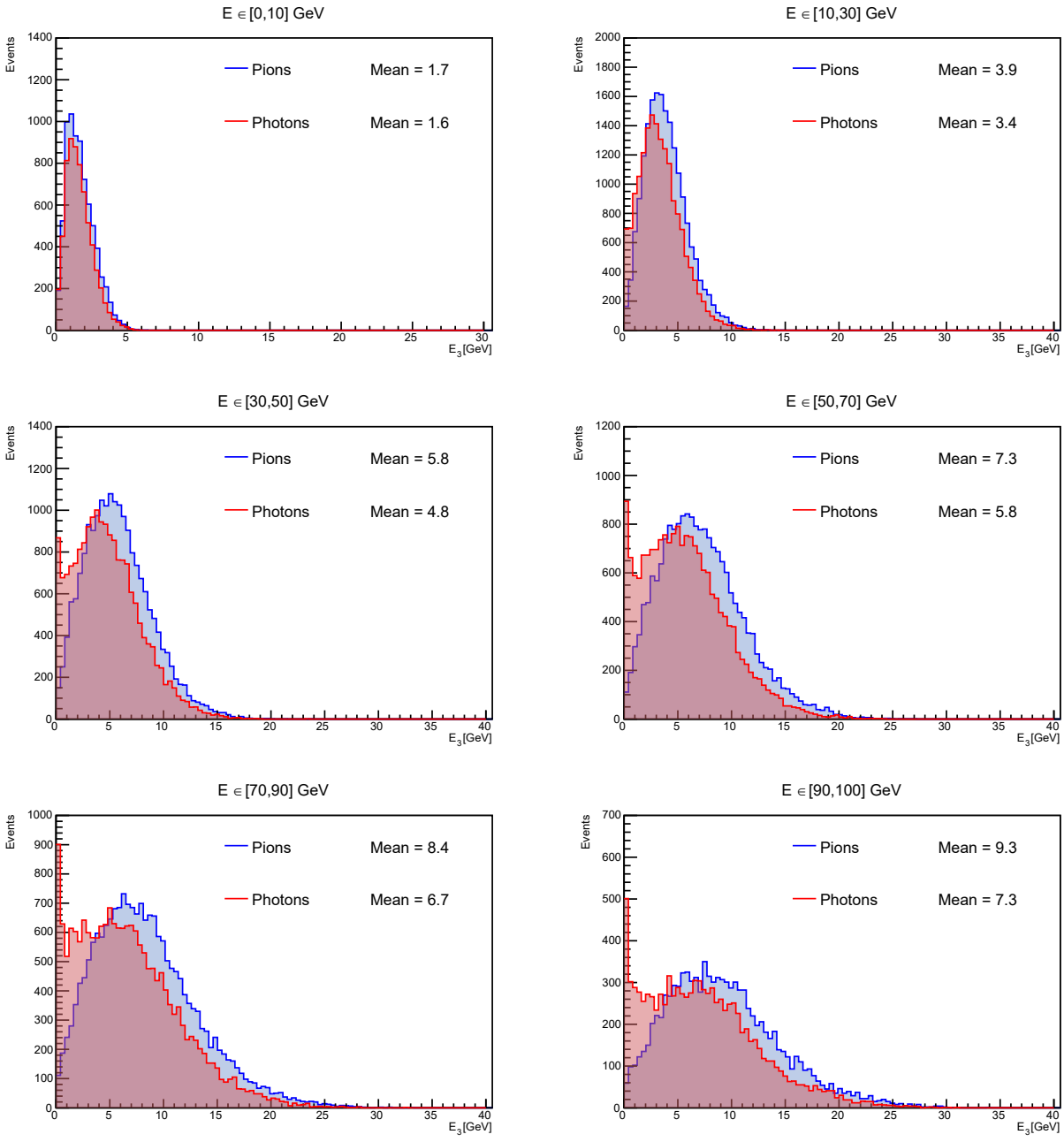

Figure A.13: Distribution of E_3 for cells with lower granularity

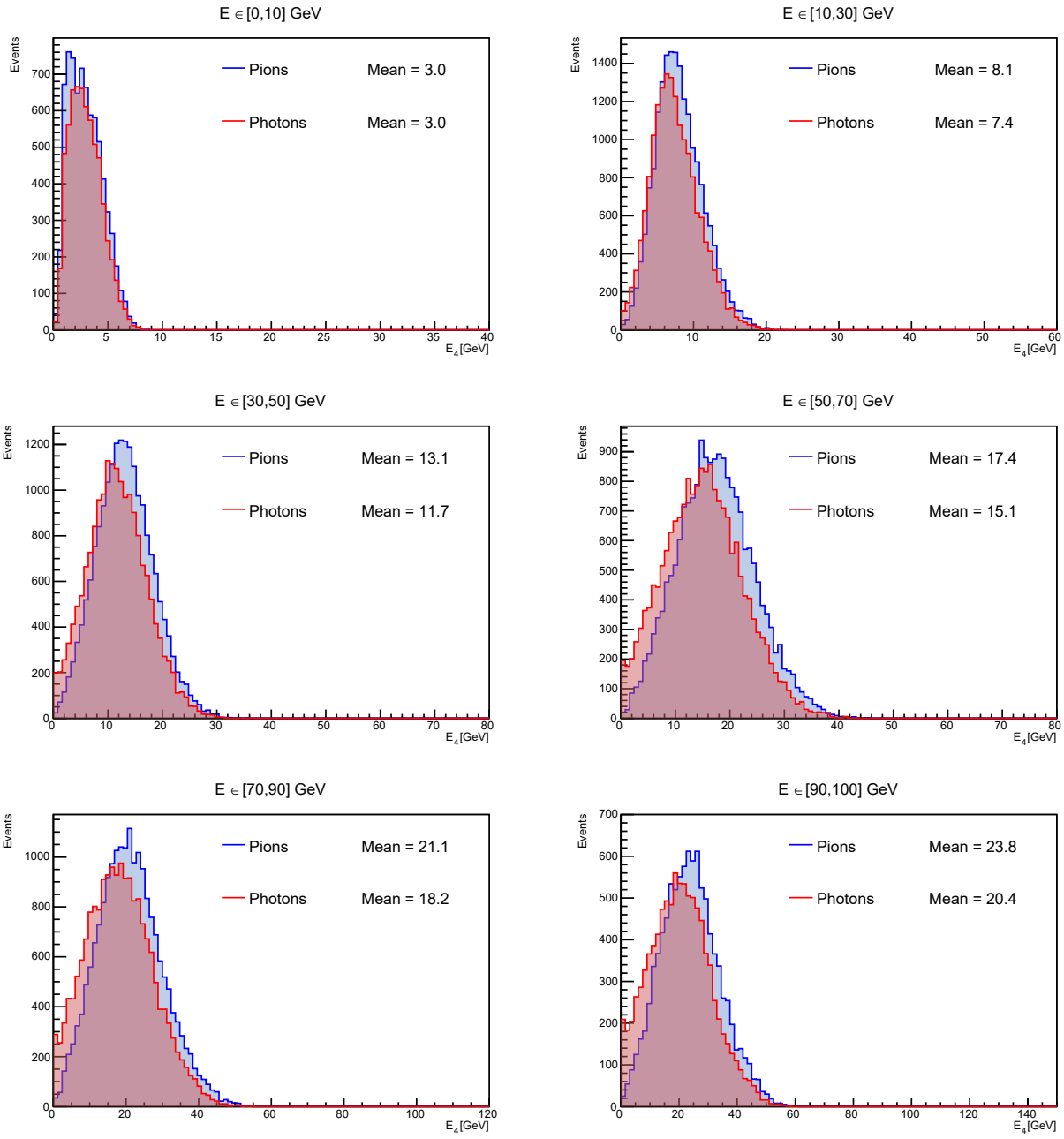

Figure A.14: Distribution of E_4 for cells with lower granularity

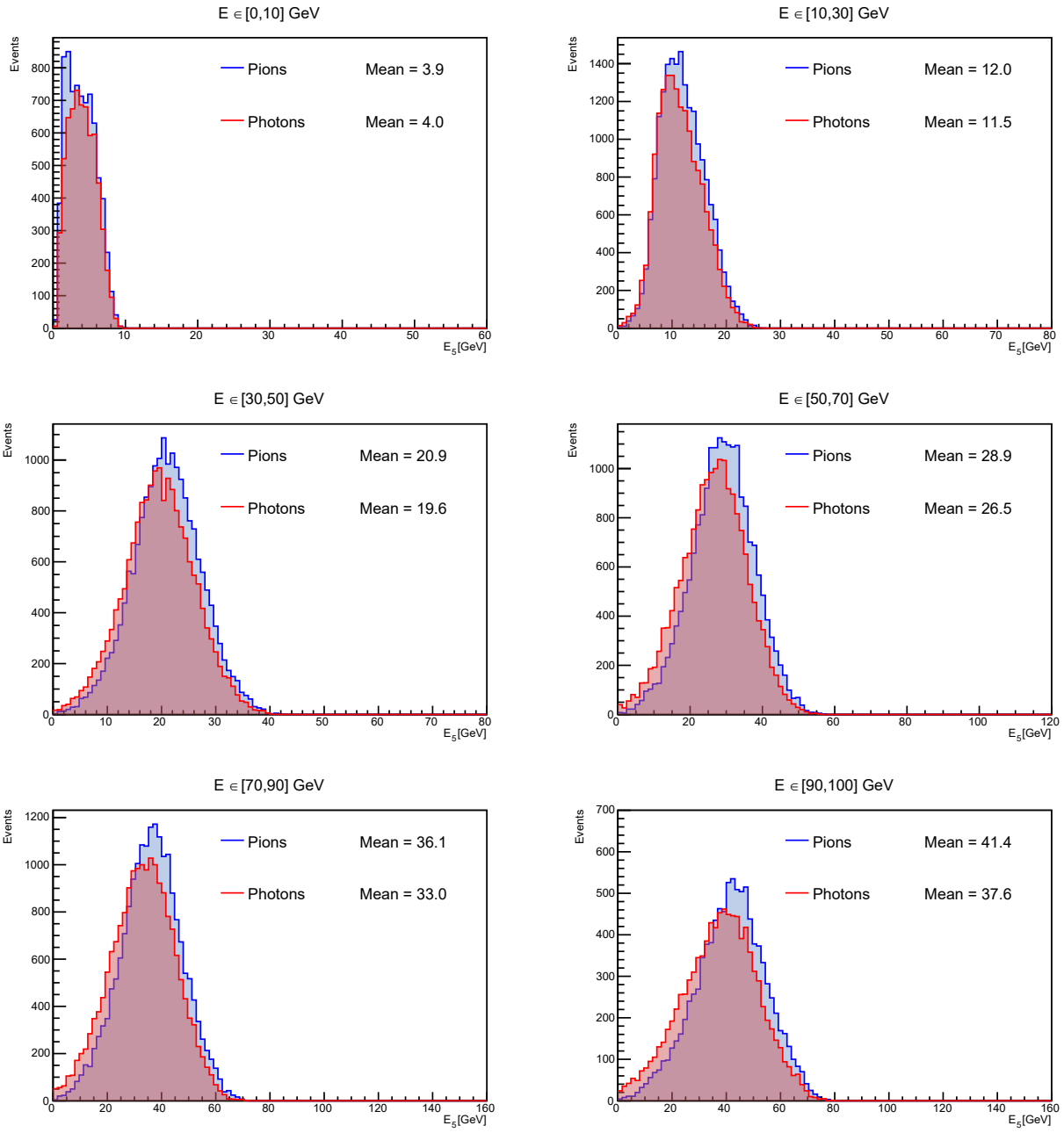

Figure A.15: Distribution of E_5 for cells with lower granularity

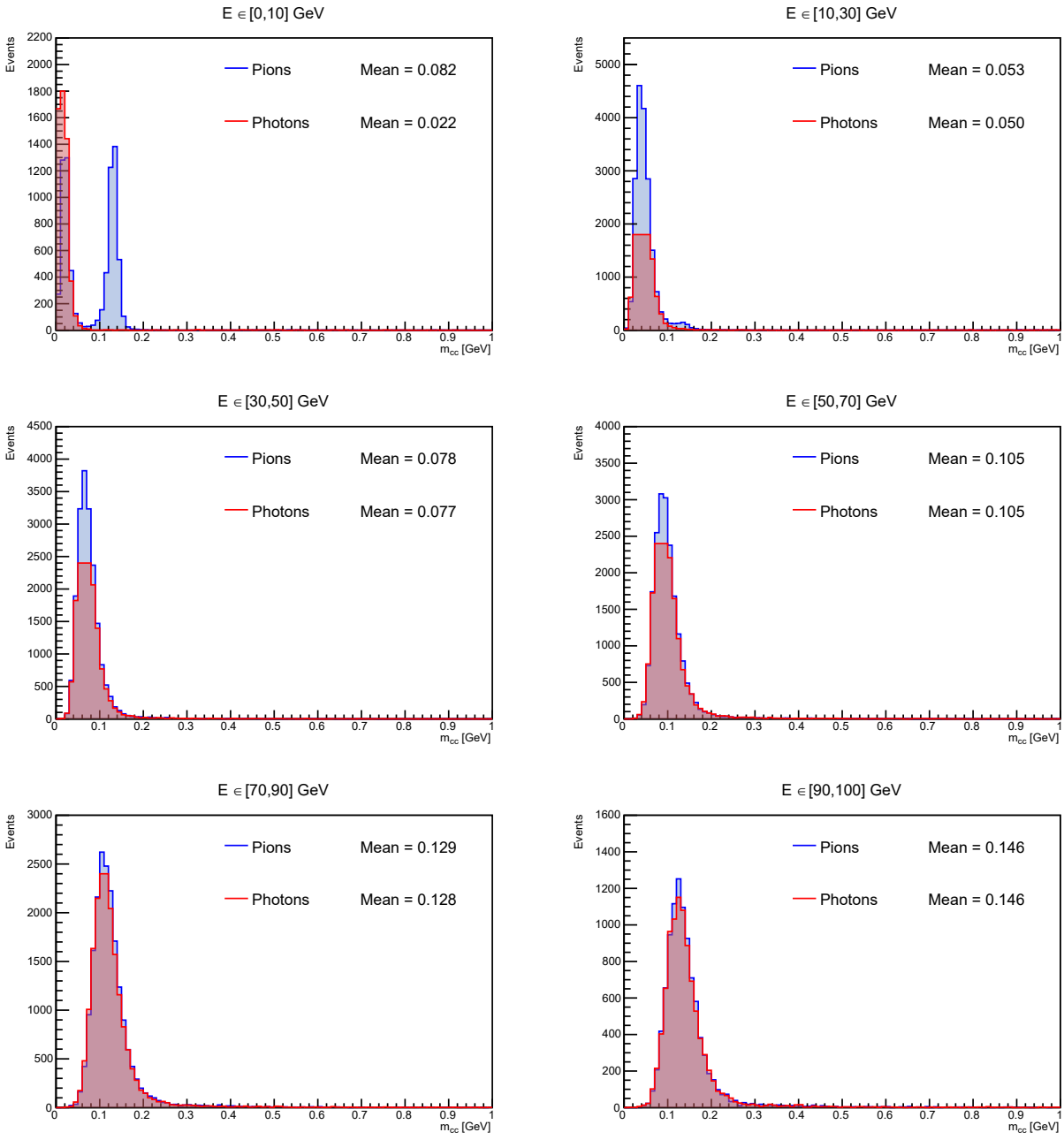

Figure A.16: Distribution of m_{cc} for cells with lower granularity

A.2.2 Discriminating variables for higher granularity

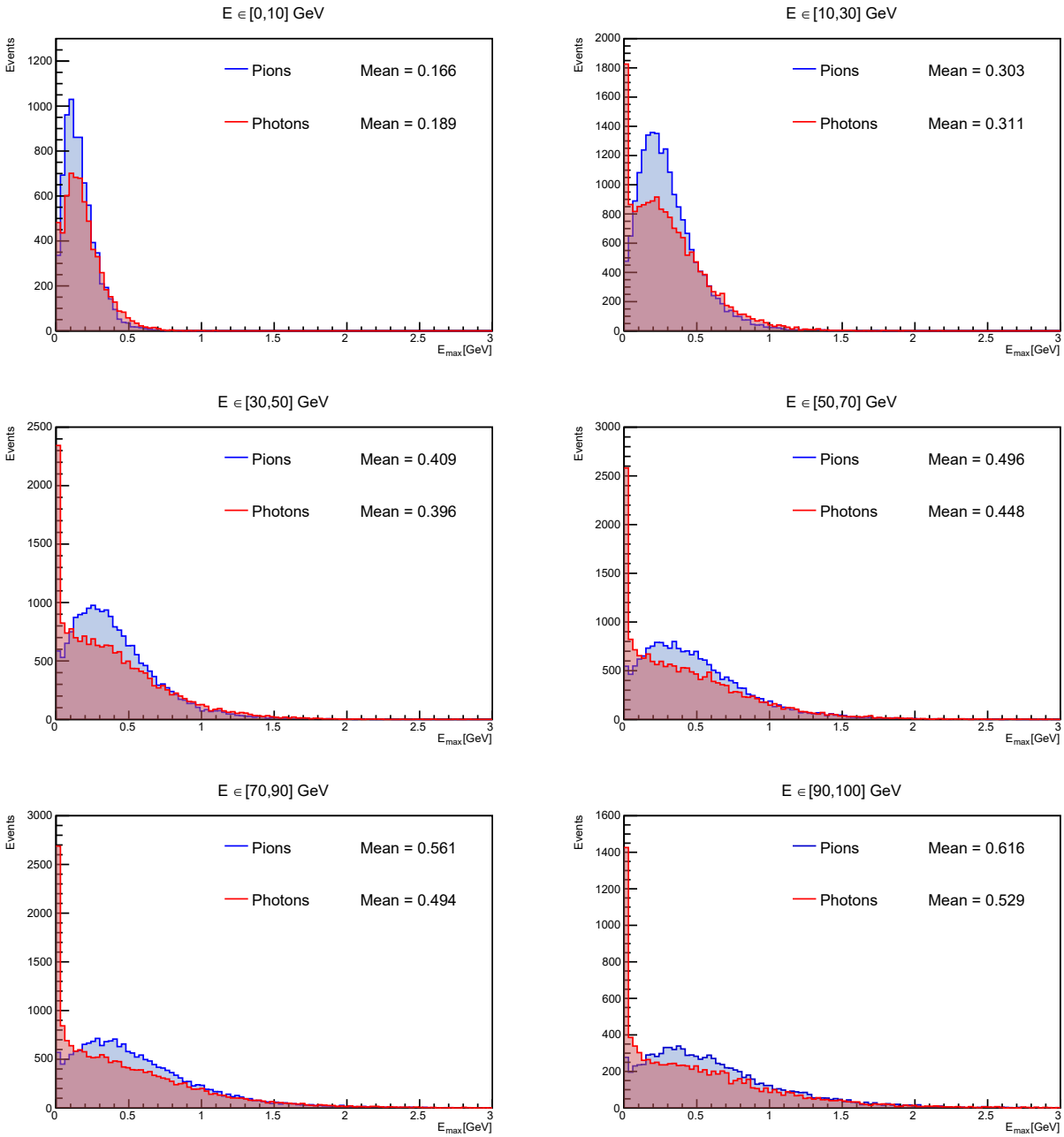

Figure A.17: Distribution of E_{max} for cells with higher granularity

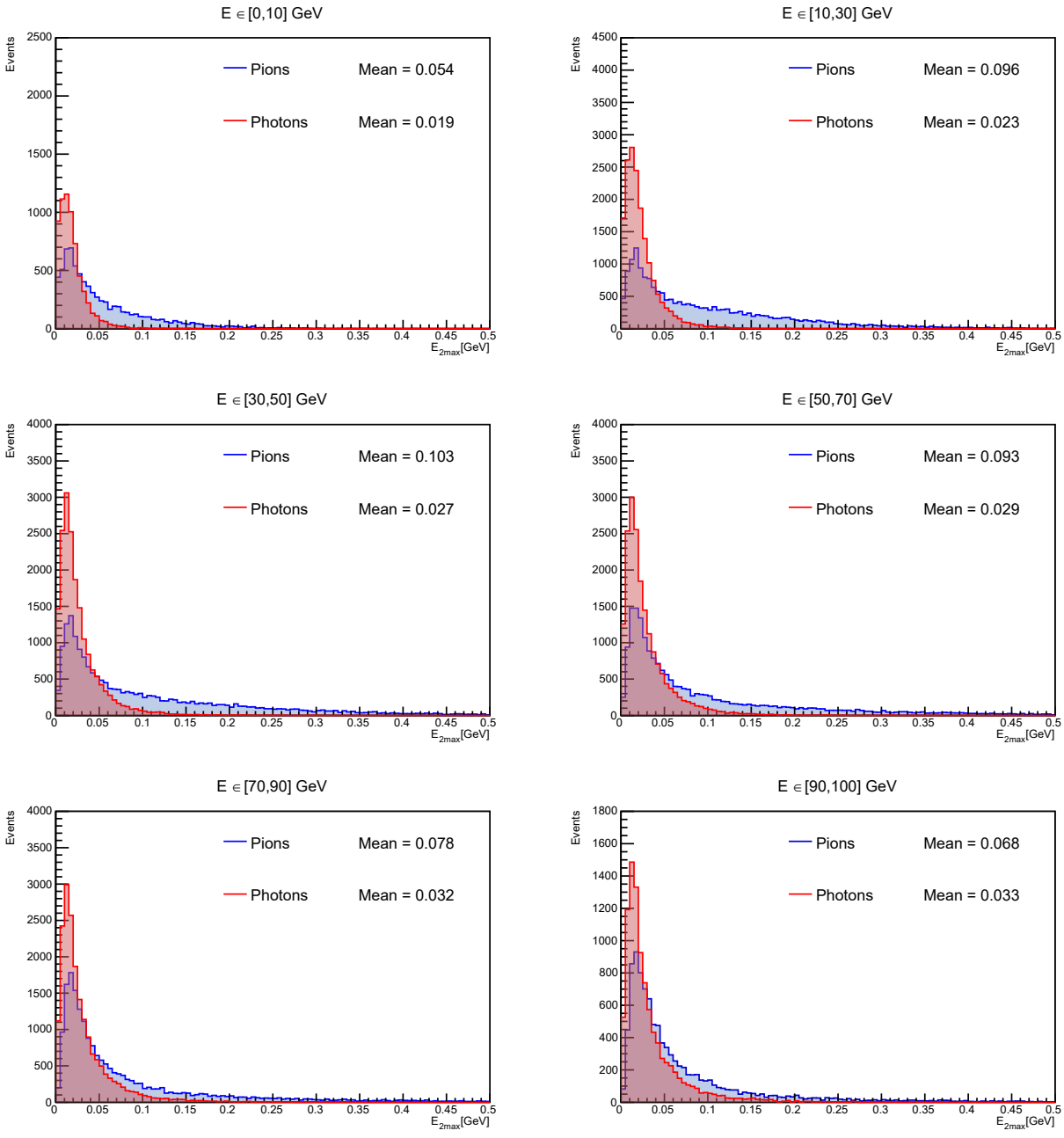

Figure A.18: Distribution of E_{2max} for cells with higher granularity

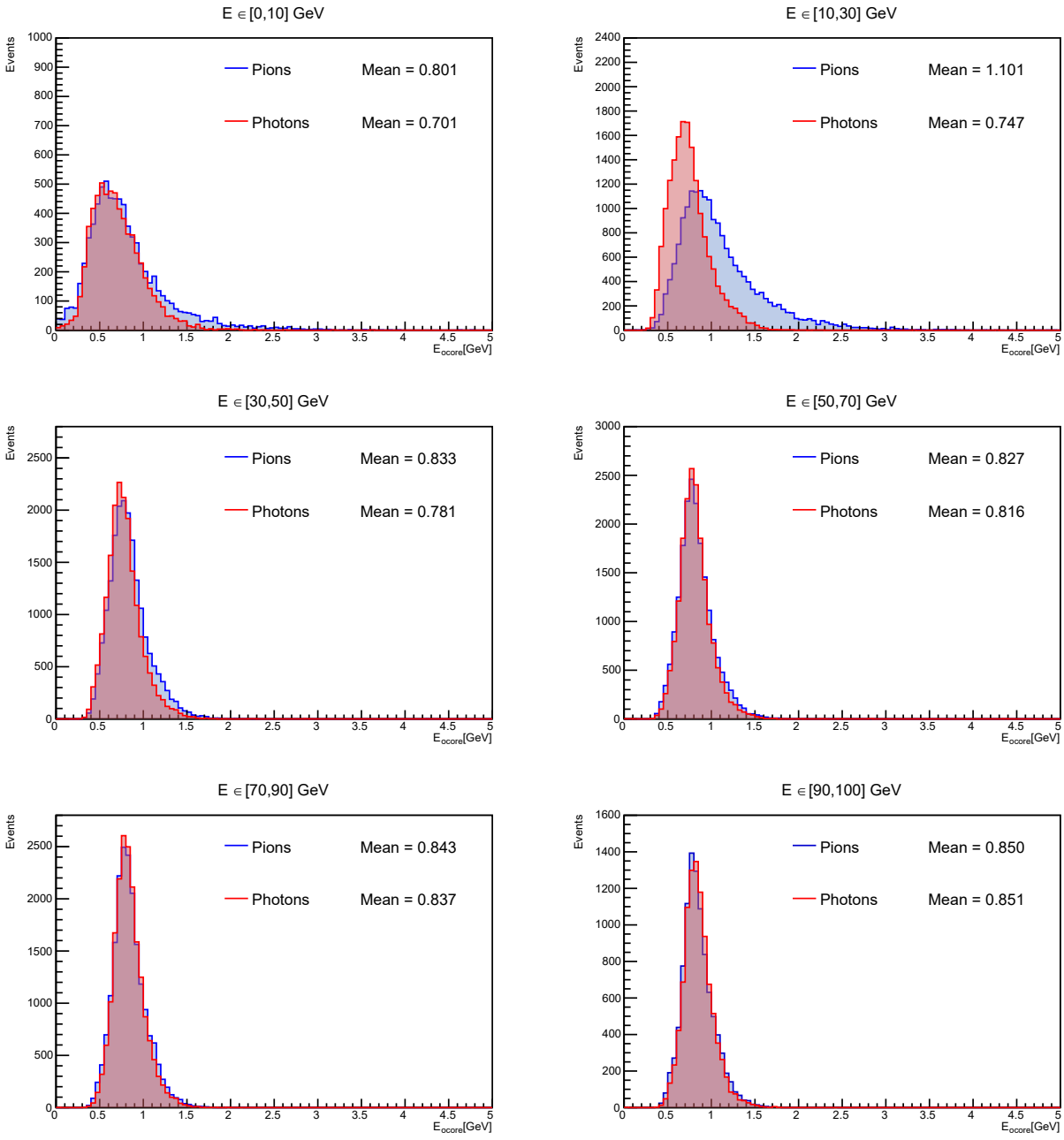

Figure A.19: Distribution of E_{ocore} for cells with higher granularity

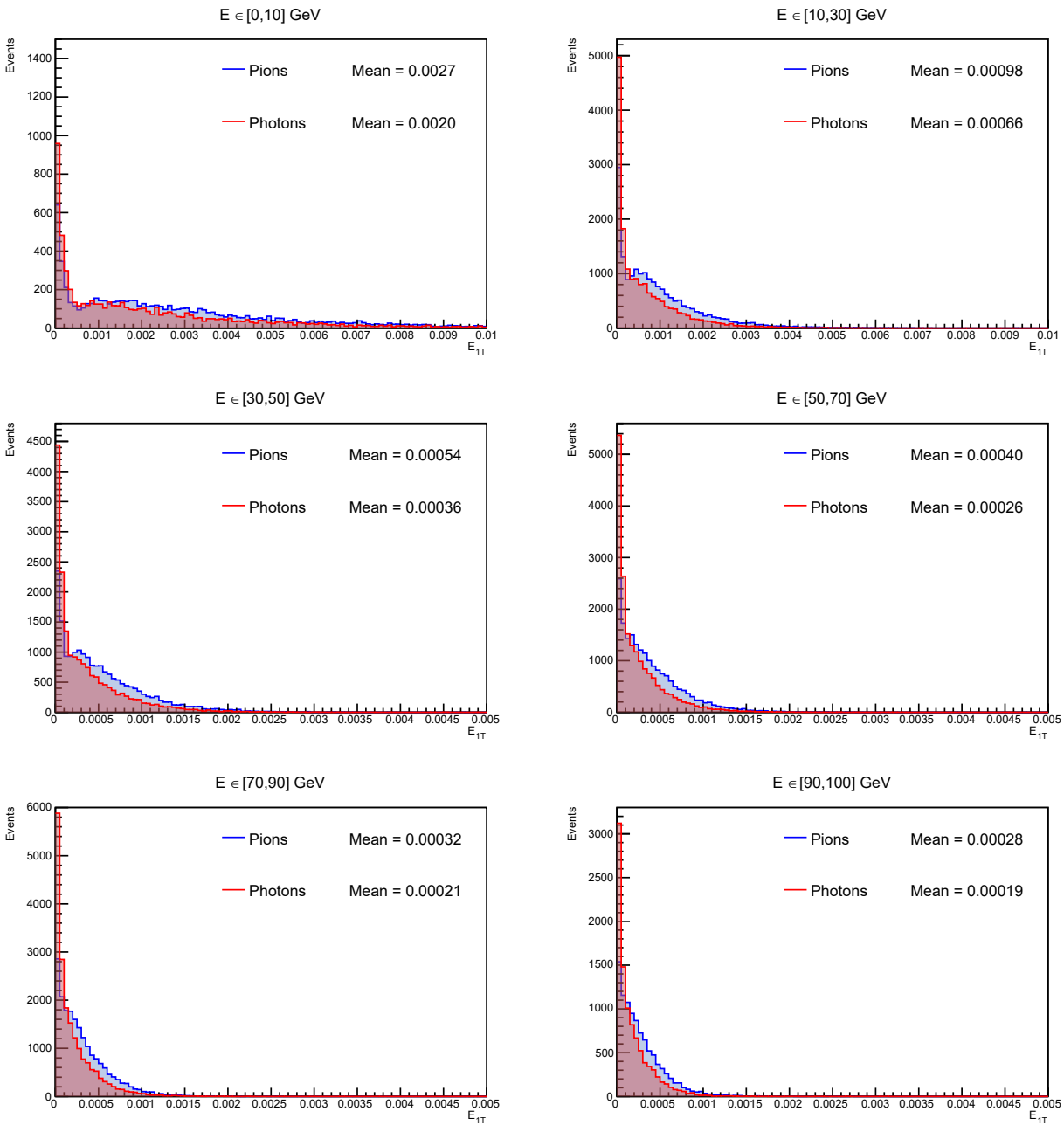

Figure A.20: Distribution of E_{1T} for cells with higher granularity

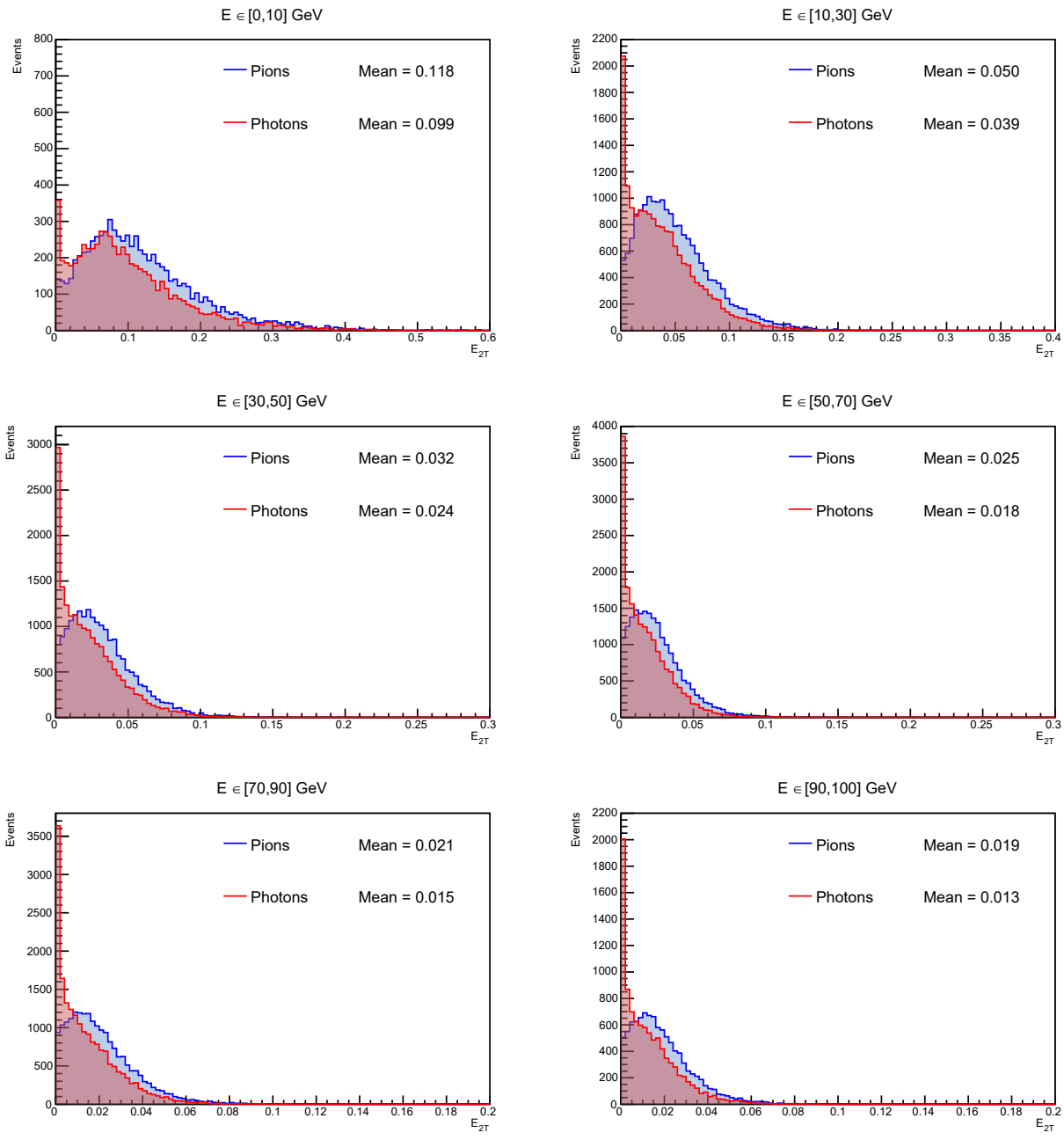

Figure A.21: Distribution of E_{2T} for cells with higher granularity

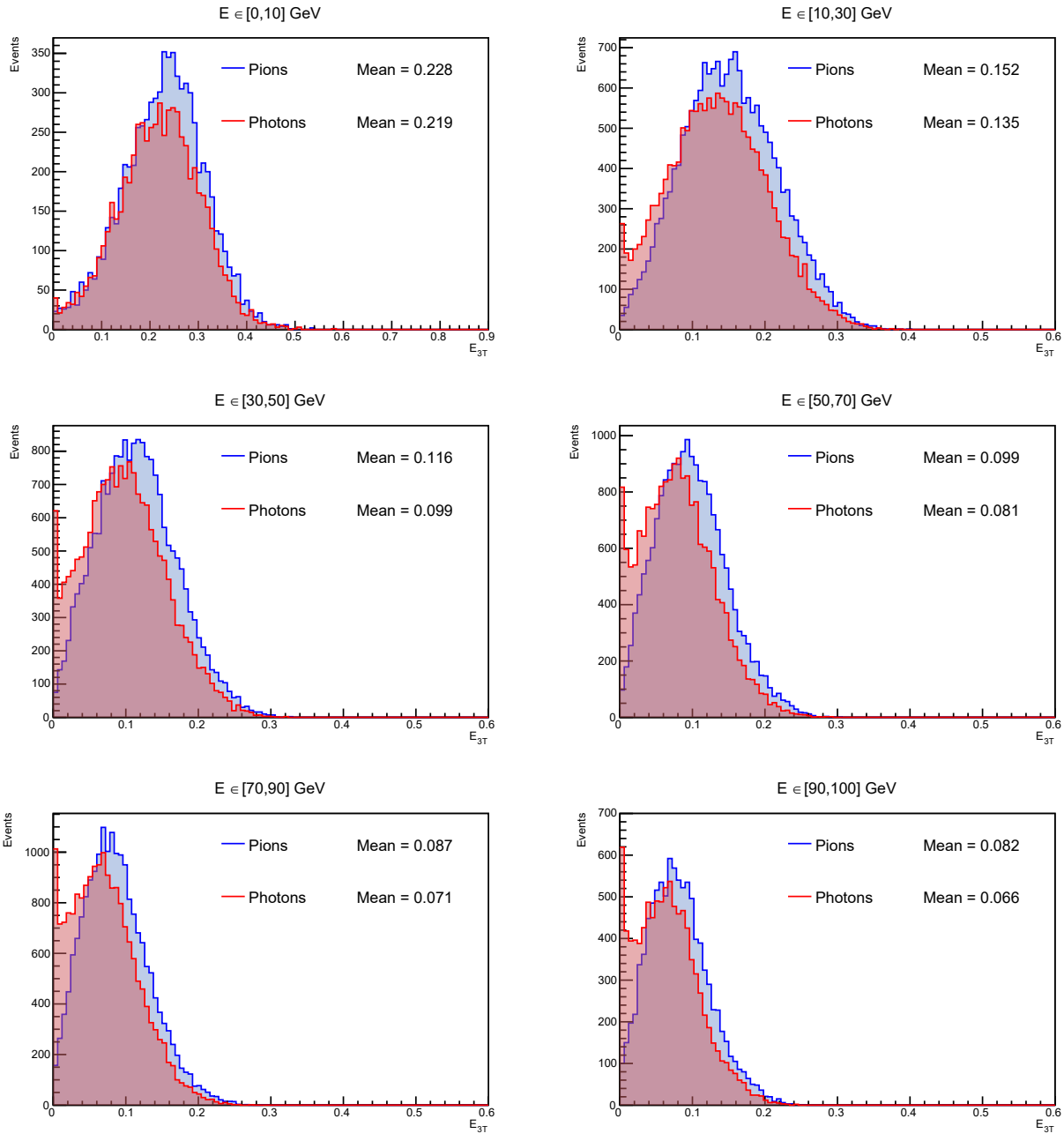

Figure A.22: Distribution of E_{3T} for cells with higher granularity

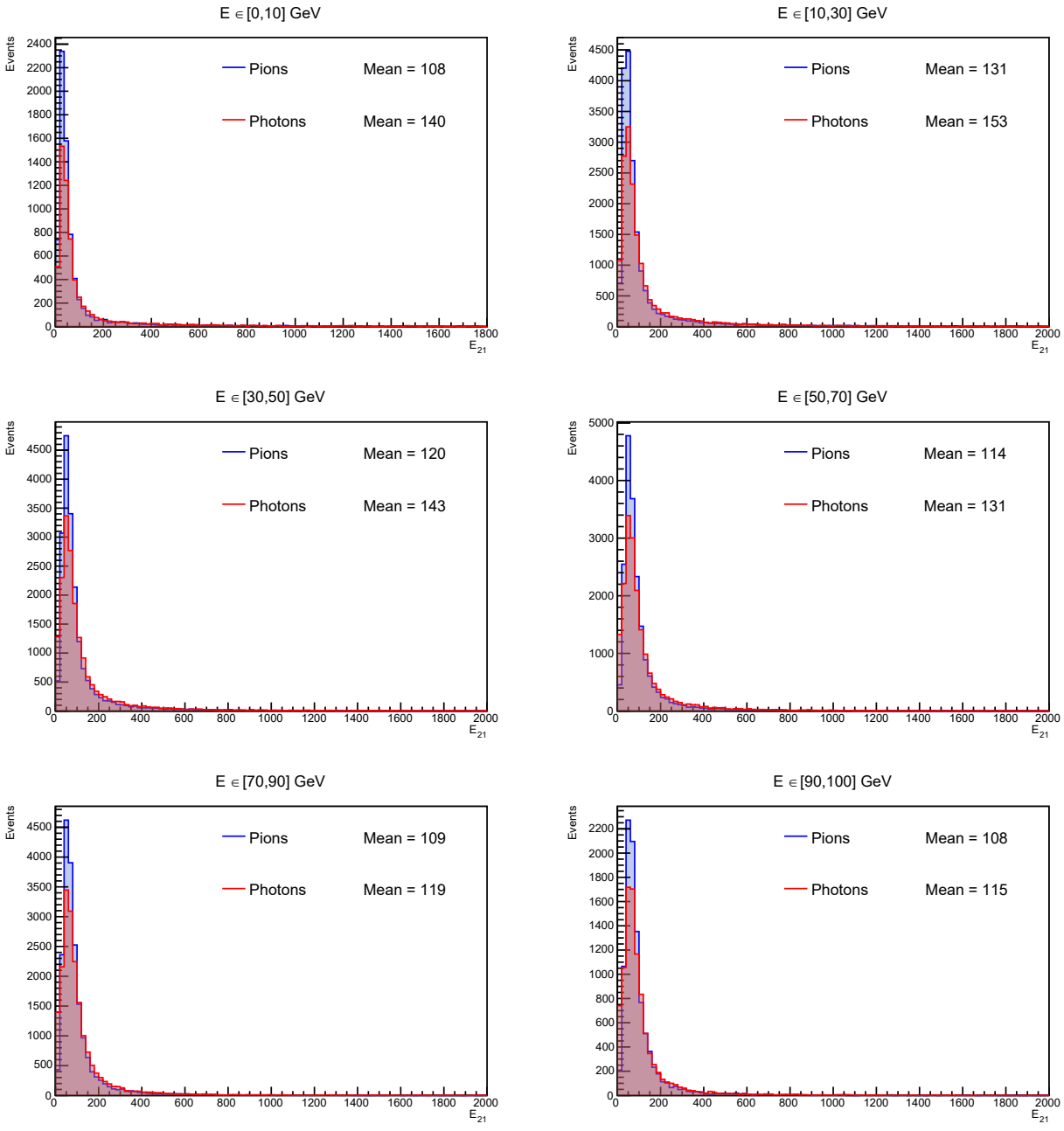

Figure A.23: Distribution of E_{21} for cells with higher granularity

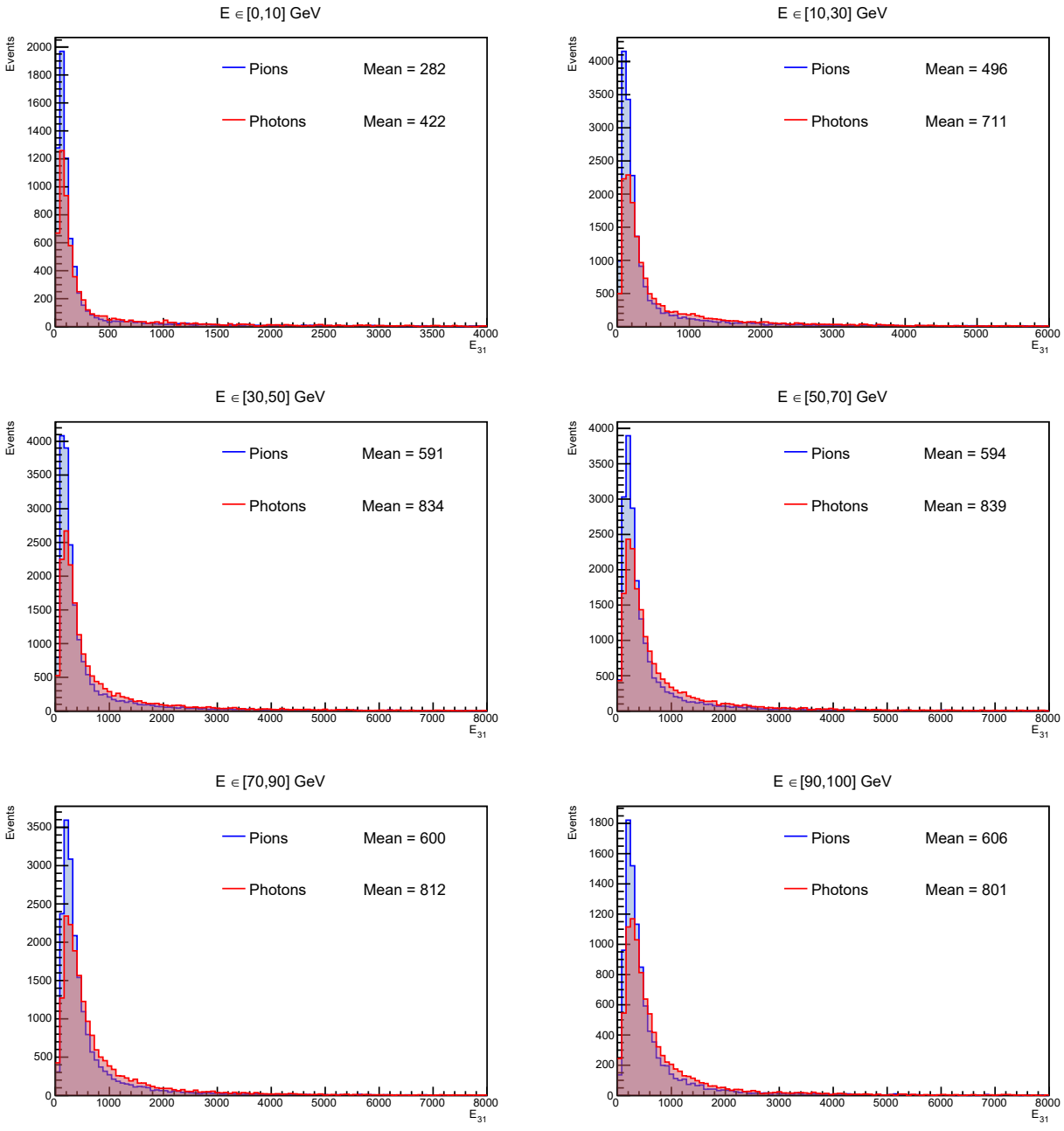

Figure A.24: Distribution of E_{31} for cells with higher granularity

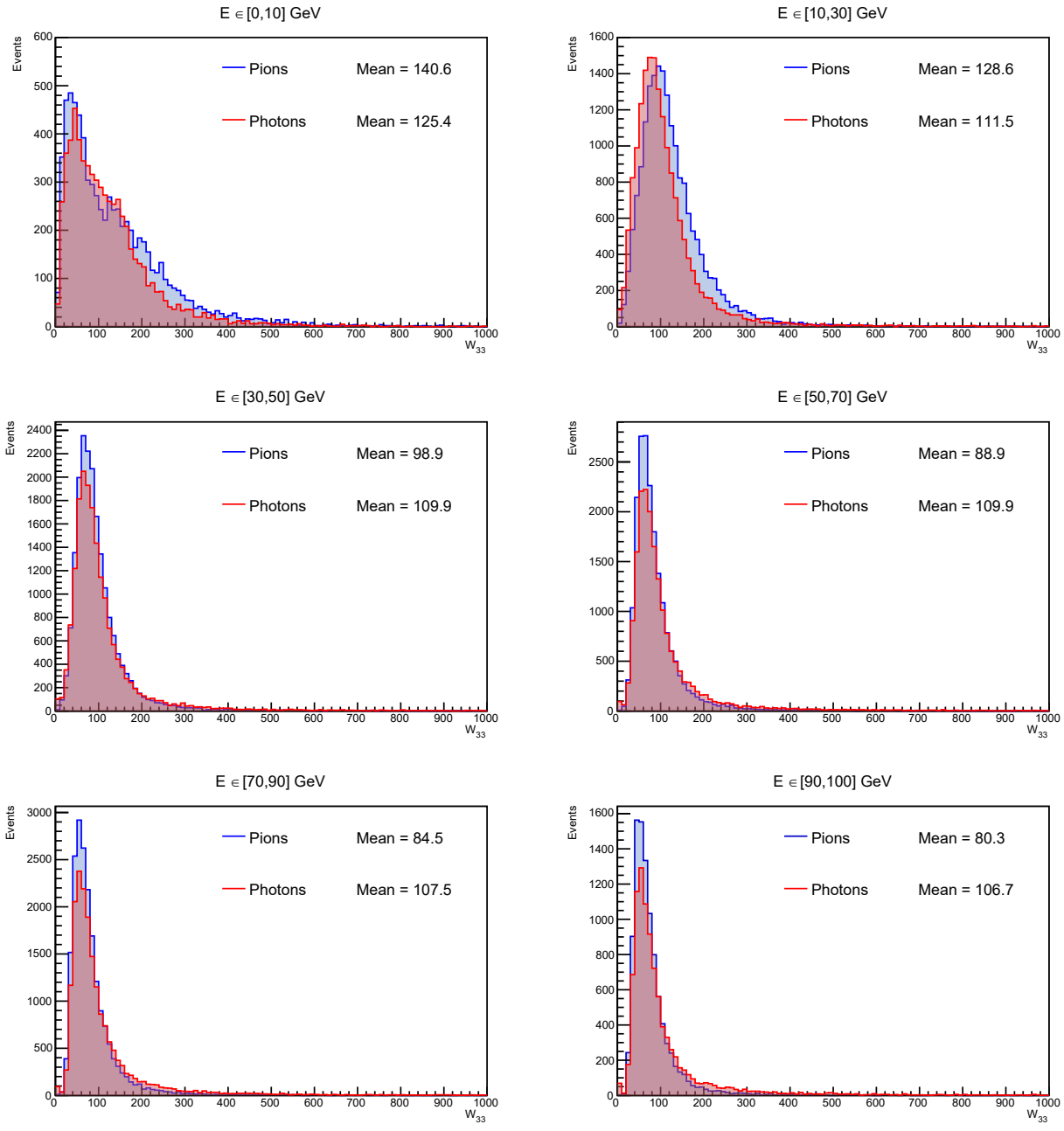

Figure A.25: Distribution of W_{33} for cells with higher granularity

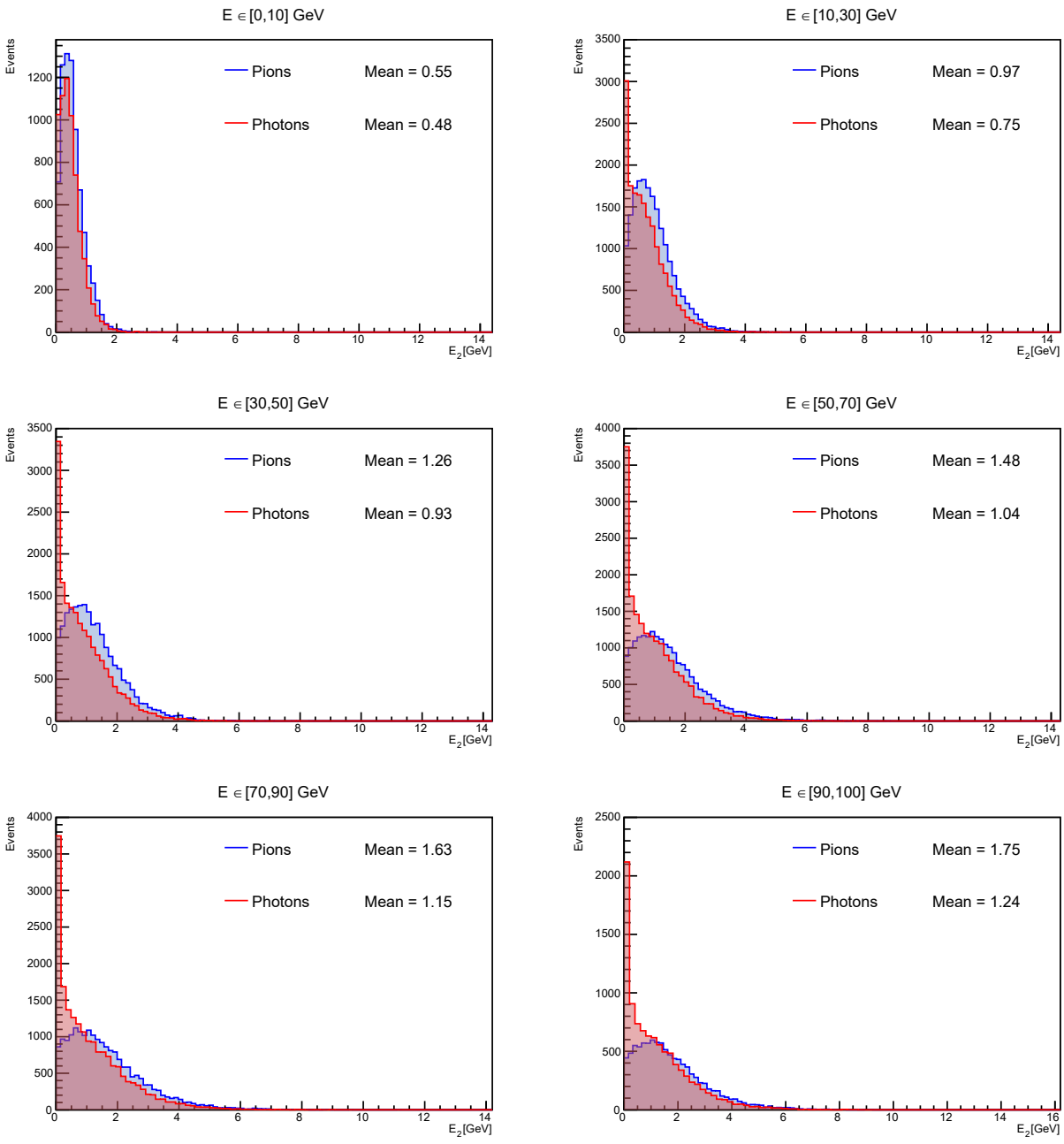

Figure A.26: Distribution of E_2 for cells with higher granularity

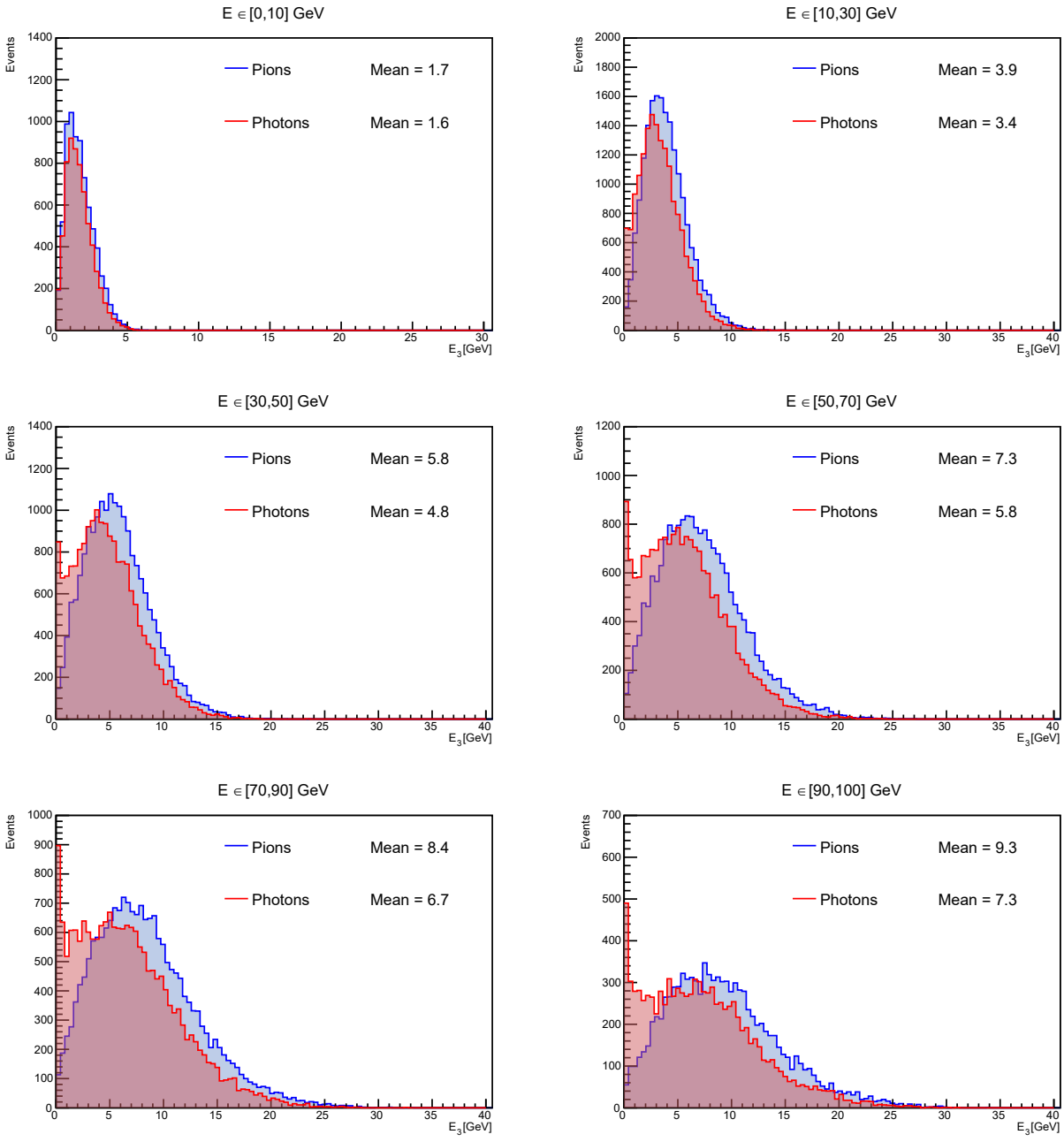

Figure A.27: Distribution of E_3 for cells with higher granularity

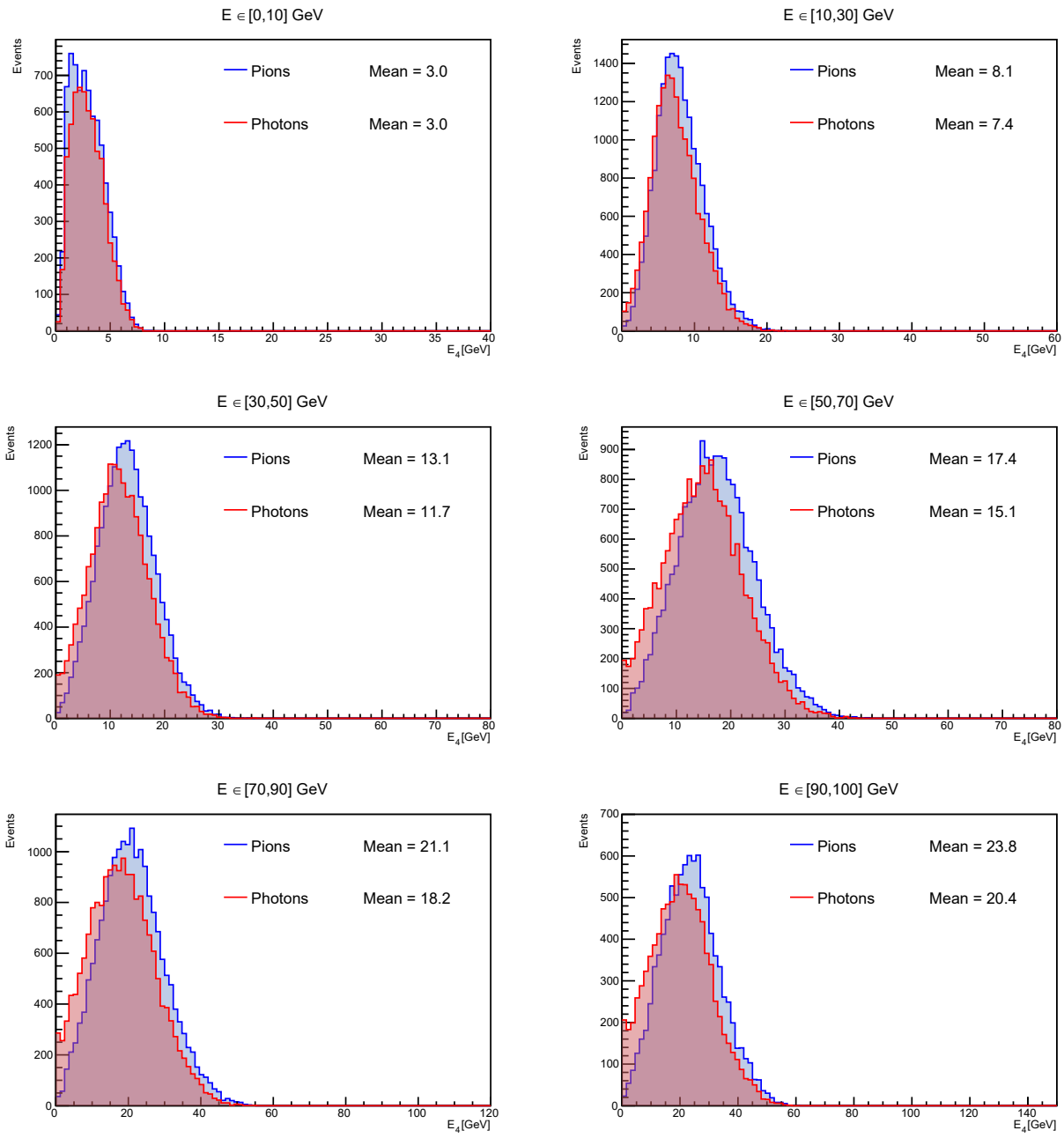

Figure A.28: Distribution of E_4 for cells with higher granularity

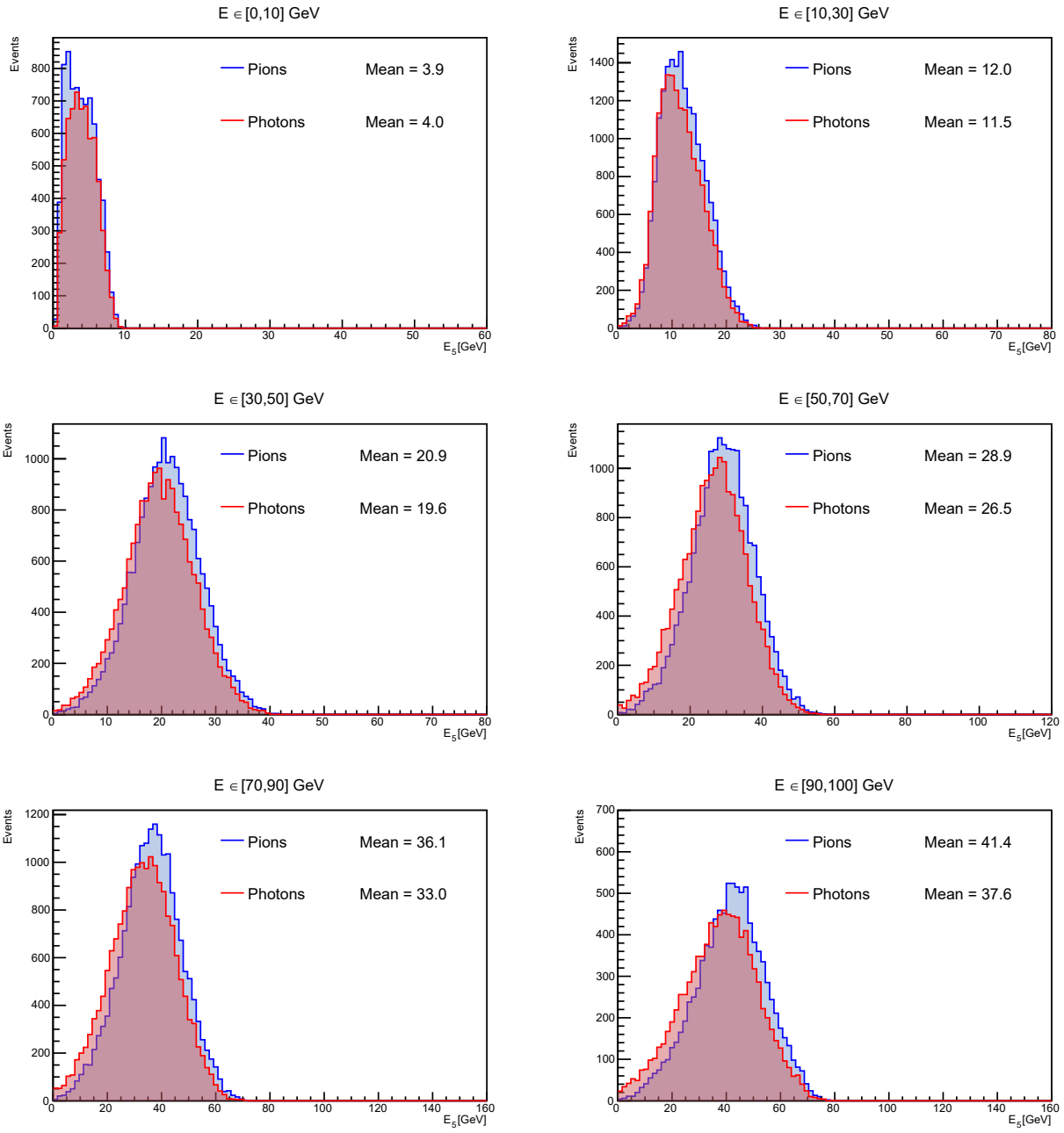

Figure A.29: Distribution of E_5 for cells with higher granularity

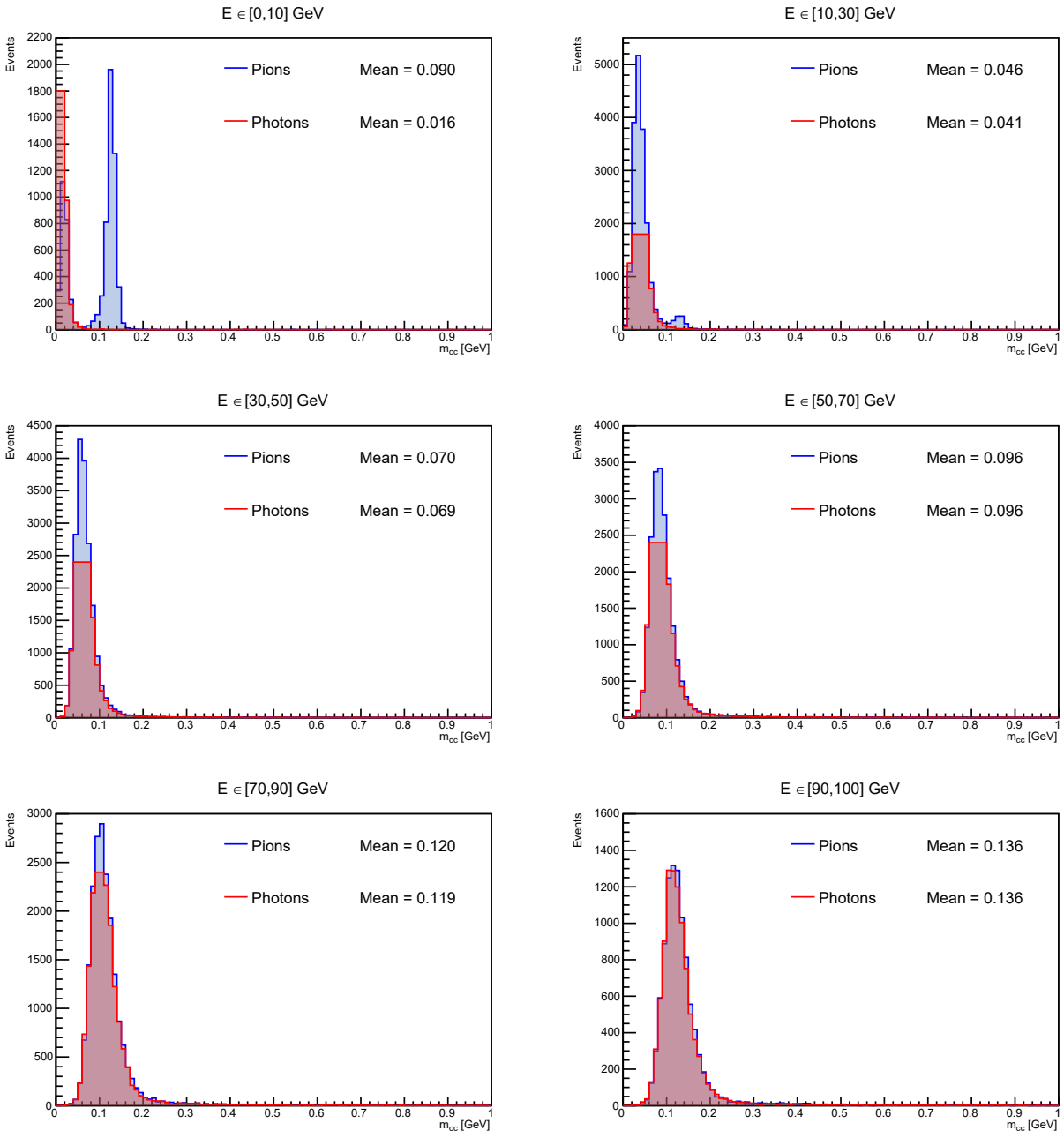

Figure A.30: Distribution of m_{cc} for cells with higher granularity

A.3 BDT hyperparameters optimization graphs

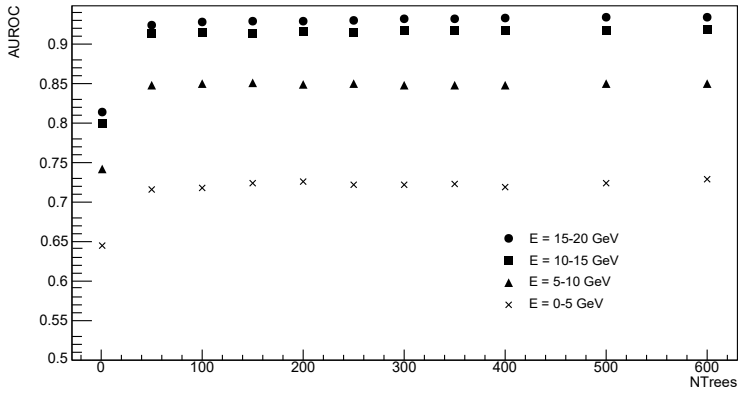

Figure A.31: AUROC dependence on NTrees hyperparameter for E = 0-20 GeV

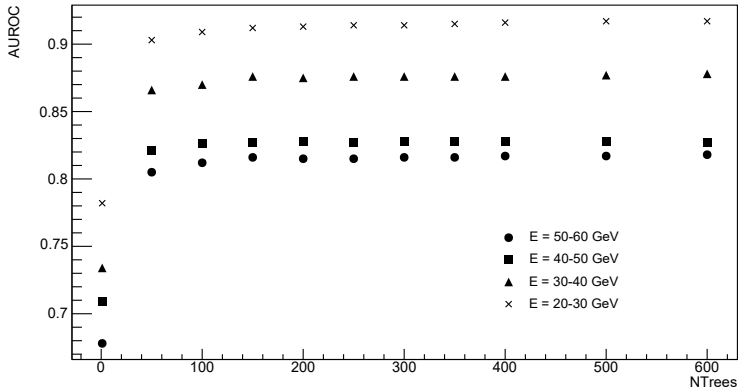

Figure A.32: AUROC dependence on NTrees hyperparameter for E = 20-60 GeV

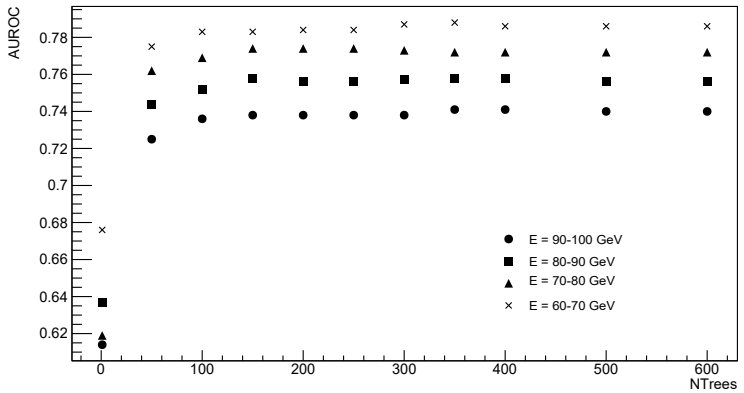

Figure A.33: AUROC dependence on NTrees hyperparameter for E = 60-100 GeV

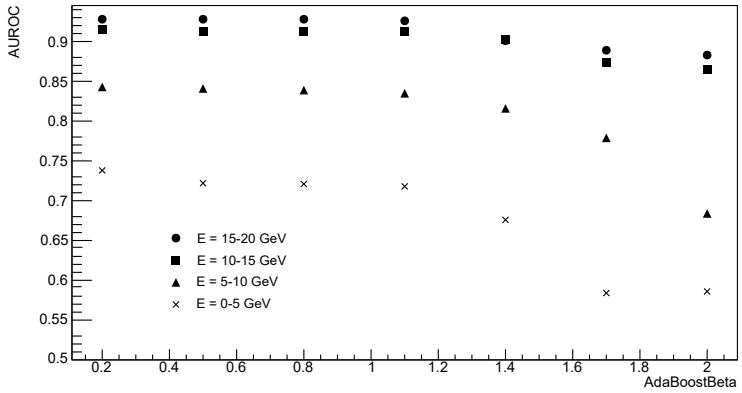

Figure A.34: AUROC dependence on AdaBoostBeta hyperparameter for E = 0-20 GeV

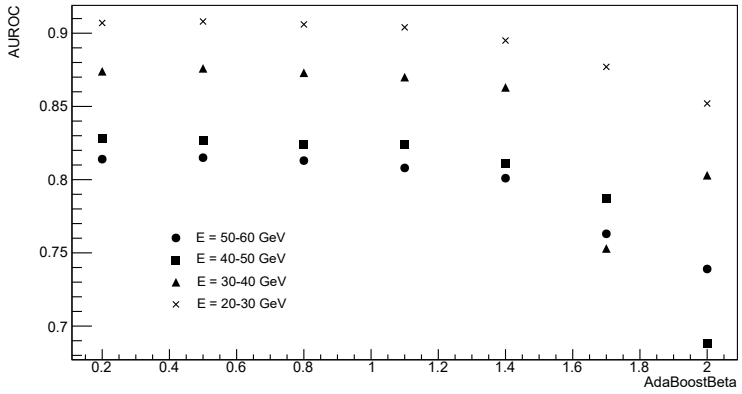

Figure A.35: AUROC dependence on AdaBoostBeta hyperparameter for E = 20-60 GeV

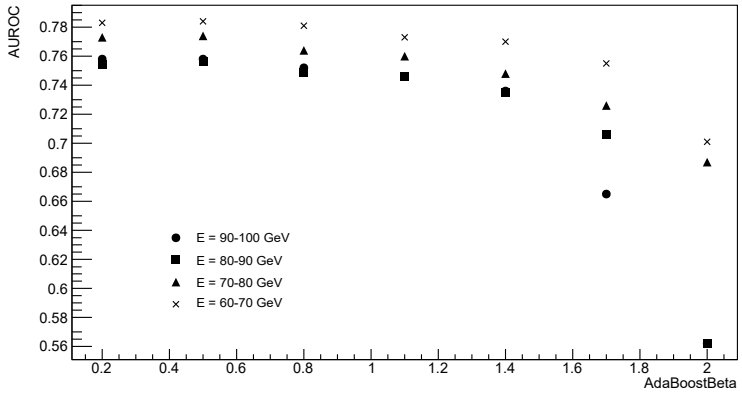

Figure A.36: AUROC dependence on AdaBoostBeta hyperparameter for E = 60-100 GeV

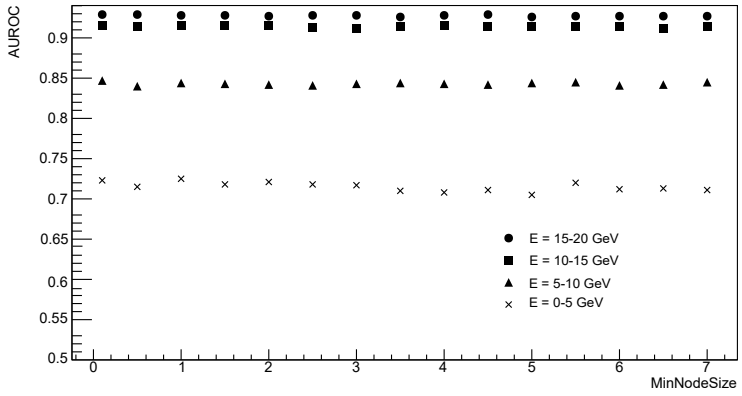

Figure A.37: AUROC dependence on MinNodeSize hyperparameter for E = 0-20 GeV

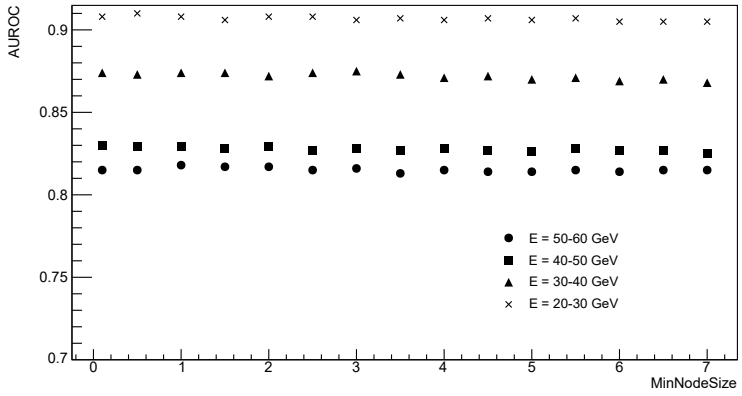

Figure A.38: AUROC dependence on MinNodeSize hyperparameter for E = 20-60 GeV

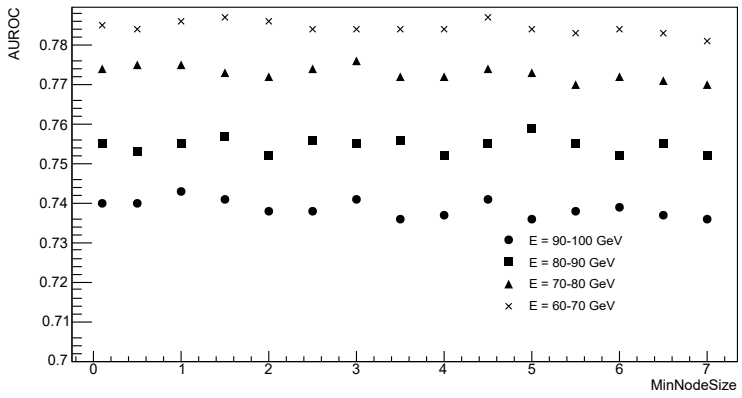

Figure A.39: AUROC dependence on MinNodeSize hyperparameter for E = 60-100 GeV

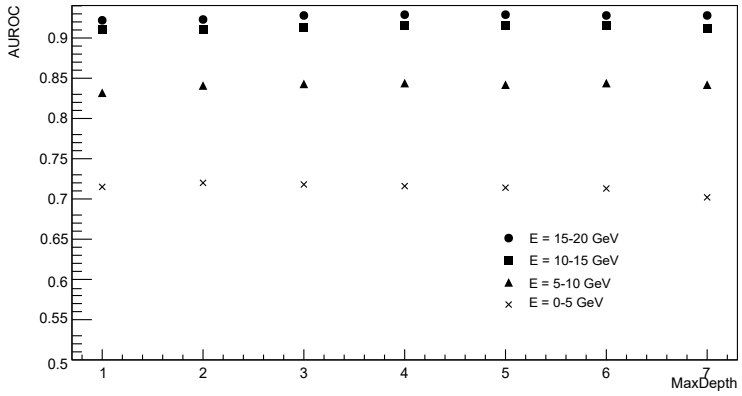

Figure A.40: AUROC dependence on MaxDepth hyperparameter for E = 0-20 GeV

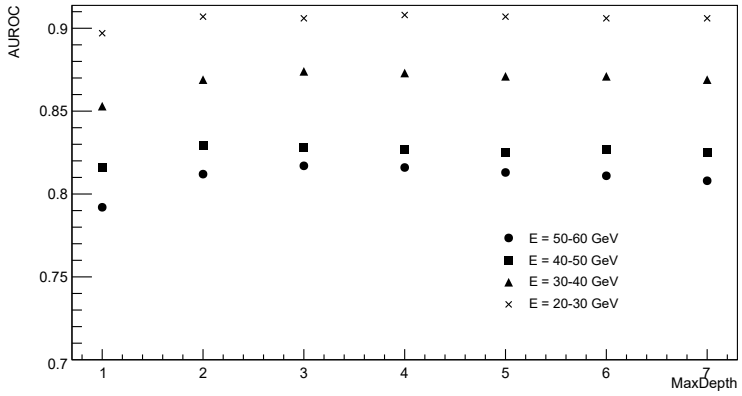

Figure A.41: AUROC dependence on MaxDepth hyperparameter for E = 20-60 GeV

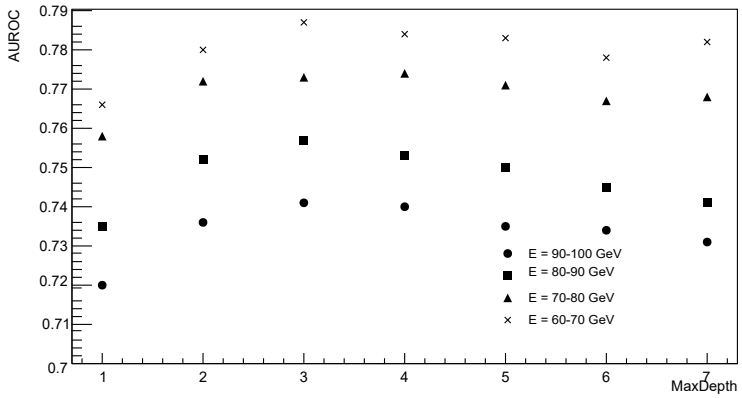

Figure A.42: AUROC dependence on MaxDepth hyperparameter for E = 60-100 GeV