Experimentelle Physik

Separation of Decay Photons and Prompt Photons Using a Neural Network

Diplomarbeit

von

Thomas Keuter

Institut für Kernphysik

Westfälische Wilhelms-Universität Münster

-Januar 2012-

Contents

1. Introduction

Presumably until 10^{-5} s after the Big Bang, the 'quark-gluon plasma' (QGP) existed. The QGP is a state of matter in which quarks and gluons are not confined within hadrons. Its study will help to understand the evolution of the universe from the Big Bang to the creation of hadrons and atoms. Nowadays, the QGP can be created under laboratory conditions in heavy-ion collisions, for example at the Large Hadron Collider (LHC). The produced particles are measured by the detectors of the ALICE experiment. Not only heavy-ion collisions, but also proton-proton collisions are investigated by ALICE to get reference data and to compare them to the results of heavy-ion collisions. In the experiment the quark-gluon plasma cannot be directly investigated because only colorless particles can be measured and therefore no single quarks and gluons. Due to this fact, special probes are needed which can be used to study the properties of the quark-gluon plasma. One of these probes are 'prompt photons', which will be the topic of this thesis. High energy photons from particle decays form a large background for prompt photons. The aim of this thesis is to distinguish between prompt photons and decay photons using a neural network on an event-by-event basis. Since no initial analysis of this type exists, only proton-proton collisions have been investigated. Chapter 2 introduces the theoretical basics of particle physics and the QGP. After a short description of the LHC and ALICE in chapter 3, the theory of artificial neural networks is presented in chapter 4. In this thesis, special simulated data are needed to train the neural network. Chapter 5 summarizes the principles of data simulation and the settings of the used simulations. In chapter 6, the separation of decay photons and prompt photons in the Monte-Carlo simulations is presented. For this part of the analysis, the prompt photons and the decay photons are identified by using the 'MC truth'. After identifying the photons, the input variables for the neural network can be calculated and the network can be trained. In the experiment, the MC truth is not available and therefore a different possibility to identify the photons and train the network is needed. In chapter 7, the identification of decay photons using an invariant mass analysis is described. Since prompt photons cannot be identified in this way, a

method to calculate the input variables for prompt photons in 'minimum bias events' is presented in chapter 8. Using the results of chapters 7 and 8, a neural network can be trained with measured data. This network is able to allocate every measured photon to the set of prompt photons or decay photons.

2. Theoretical Background

In this chapter, the theoretical background which is necessary for the understanding of this thesis will be presented. After summarizing the Standard Model of Particle Physics, the creation of a quark-gluon plasma, its properties and signatures will be briefly explained.

2.1. The Standard Model

In 1808, John Dalton formulated the hypothesis that matter consists of small indivisible units called 'atoms[1]'[Dal08]. Nowadays, the hypothesis of atoms is generally accepted because of many experiments like scanning tunnel microscopy, in which atoms can be seen indirectly. The atomic substructure was discovered by Ernest Rutherford in 1911 [Rut11]. He explained the results of his scattering experiment by the existence of a positively charged, massive nucleus inside the atom and negatively charged, light electrons around the nucleus. The nucleus is composed of protons and neutrons. These were discovered by Rutherford in 1919 and by James Chadwick in 1932 [Cha32], respectively. In the middle of the 20th century, a lot of other particles were discovered, like π^0 and Δ^+. In particle accelerator and cosmic ray experiments, particles were observed which interact strongly with protons and neutrons, but have different masses. Together with protons and neutrons, these strongly interacting particles are called 'hadrons'. In 1964, Gell-Mann and Zweig explained the existence of the hadrons by new fundamental particles, called quarks [Gel64, Zwe64]. According to Gell-Mann and Zweig, all hadrons are composed of quarks and anti-quarks representing different combinations of them. Nowadays, the Standard Model of Particle Physics describes the fundamental particles of which matter[2] is composed and the forces between the particles.

[1] From the Greek $\alpha\tau o\mu o\varsigma$, which means uncutable or indivisible.

[2] In this context, 'matter' means visible matter. Dark Matter is not discussed in this thesis.

Force	**Applies to**	**Gauge Boson**	**Mass**	**Range**
Strong	Color	8 gluons (g)	0	10^{-15} m
Electromagnetic	Electric charge	Photon (γ)	0	∞
Weak	Weak charge	$W^{\pm}$	$80.4\,\mathrm{GeV/c^2}$	10^{-18} m
		Z^0	$91.2\,\mathrm{GeV/c^2}$	

Table 2.1.: Fundamental forces of the Standard Model [PDG11].

Fermions	**Name**	**Electric Charge**	**Interaction with**	**Mass**
Leptons	e^-	$-$ 1 e	Weak and e.m.	0.511 MeV/c^2
	ν_e	0	Weak	$<$ 2 eV/c^2
	μ^-	$-$ 1 e	Weak and e.m.	105 MeV/c^2
	ν_μ	0	Weak	$<$ 0.19 MeV/c^2
	τ^-	$-$ 1 e	Weak and e.m.	1.78 GeV/c^2
	ν_τ	0	Weak	$<$ 18.2 MeV/c^2
Quarks	u	+ 2/3 e	Strong, weak	1.7 - 3.1 MeV/c^2
	d	$-$ 1/3 e	and e.m.	4.1 - 5.7 MeV/c^2
	c	+ 2/3 e		1.2 - 1.3 GeV/c^2
	s	$-$ 1/3 e		80 - 130 MeV/c^2
	t	+ 2/3 e		173 - 174 GeV/c^2
	b	$-$ 1/3 e		4.1 - 4.9 GeV/c^2

Table 2.2.: Fundamental particles of the Standard Model [PDG11]. For each particle, a corresponding anti-particle exists with same multiplicative quantum numbers (e.g. same spin), but opposite additive quantum numbers (e.g. the electric charge for electron (e^-) and positron (e^+)).

As listed in tables 2.1 and 2.2, the Standard Model contains 12 fermions (particles with half-integer spin) as well as the gauge bosons photon, $W^{\pm}$, Z^0, and gluons (bosons are particles with integer spin). The fermions form the matter we are surrounded by whereas the gauge bosons act as mediators of the forces.

The fundamental fermionic particles can be subdivided into quarks and leptons. In this model, 6 quarks with different 'flavors' exist: up-quark (u), down-quark (d), charm-quark (c), strange-quark (s), top-quark (t), and bottom-(or beauty-)quark (b). In addition, 6 corresponding anti-quarks exist: anti-up-quark ($\overline{u}$), anti-down-quark ($\overline{d}$), anti-

charm-quark ($\overline{c}$), anti-strange-quark ($\overline{s}$), anti-top-quark ($\overline{t}$), and anti-bottom-quark ($\overline{b}$). The quarks carry 'color', which is in the strong force the analogon to the electric charge in the electromagnetic force. Three different colors (blue, red, and green) exist in the strong force and together, they cancel each other out like positive and negative electric charges. Particles that carry color cannot be isolated (they are confined), only particles without color can be observed directly and therefore only hadrons, which are neutral in color, can be measured in experiments. Hadrons are classified into baryons and mesons. Baryons consist of three quarks, like the proton $p = uud$, and mesons consist of a quark and an anti-quark, like the pion $\pi^+ = u\overline{d}$. In baryons, the three quarks carry different colors (red, green, and blue) which sum up to no color (red + green + blue = white = colorless); in mesons, the quark carries color and the anti-quark carries anti-color, which also sum up to no color (e.g. red + anti-red = white).
The leptons are electron (e^-), muon (μ^-), and tau (τ^-) and the corresponding neutrinos ν_e (electron-neutrino), ν_μ (muon-neutrino), and ν_τ (tau-neutrino). The anti-leptons are positron (e^+), anti-muon (μ^+), anti-tau (τ^+), anti-electron-neutrino ($\overline{\nu}_e$), anti-muon-neutrino ($\overline{\nu}_\mu$), and anti-tau-neutrino ($\overline{\nu}_\tau$). Contrary to quarks, leptons do not interact strongly; they are only affected by the weak force, and the charged ones additionally by the electromagnetic force.
The gauge boson of the electromagnetic force is the photon (γ), the quantization of the electromagnetic field, which is described by quantum electrodynamics (QED). Using the condition that the Lagrangian of this theory has to be invariant under a local U(1) transformation, it can be seen that the photon carries no electric charge and has no mass. Therefore, the electromagnetic force has an infinite range as in Coulomb's law.
The gauge bosons of the weak force are the W^+-, W^--, and the Z^0-boson. They can change the flavor of a fermion (e.g. the β-decay: $d \to u + e^- + \overline{\nu}_e$) and have large masses resulting in a short range of the weak force. An explicit mass term in the Lagrangian of the (electro-)weak theory would violate the gauge invariance, therefore, the masses of the gauge bosons are created by the 'Higgs mechanism' by adding a 'Higgs field' to the Lagrangian. At the LHC, the experiments ATLAS and CMS are searching for the corresponding particle of the Higgs field, the Higgs boson.
The gauge bosons of the strong force are the gluons, which are massless. The gluons interact via color and are described by quantum chromodynamics (QCD, chroma $\hat{=}$ color). The underlying symmetry group of QCD is SU(3), which is resulting in 3 different colors and 3^2-1 = 8 gluons. Since SU(3) is a non-abelian group, the gluons carry color and anti-color and interact with each other. This limits the range of the strong force to

the size of a proton ($\approx$ 1 fm). The potential $V(r)$ between a quark and an anti-quark can be estimated by analyzing the charmonium, a bound system of a c- and a $\bar{c}$-quark. For small distances r of the c-quarks, the charmonium is similar to the positronium, a bound system of an electron and a positron, which results in a Coulomb-like summand in the potential, $V(r) \propto 1/r$. Up to now, no free, colored particle was observed. This confinement of the quarks within hadrons is taken into account by adding a summand $\propto r$ to the potential. The ansatz for the potential is given by [Pov09]

$$V(r) = -\frac{4}{3}\frac{\alpha_s(Q^2)\hbar c}{r} + k \cdot r, \qquad (2.1)$$

where k is a constant and $\alpha_s(Q^2)$ is the coupling constant of the strong force. Due to the polarization of the vacuum[3] by gluons, α_s depends on the momentum transfer Q^2. In first order of perturbation theory in QCD, $\alpha_s(Q^2)$ is given by [Hal84]

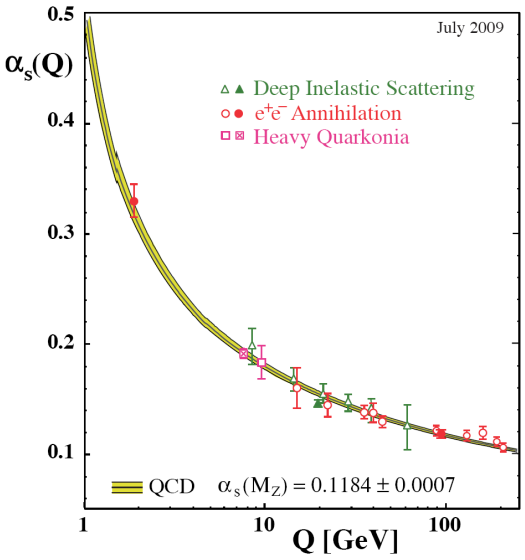


Figure 2.1.: Strong coupling constant α_s as a function of the momentum transfer Q [Bet09].

$$\alpha_s(Q^2) = \frac{12\pi}{(33 - 12n_f) \cdot \ln(Q^2/\Lambda^2)}, \qquad (2.2)$$

where n_f is the number of the involved types of quarks and Λ is a scale parameter of the QCD ($\Lambda \approx 250$ MeV [Pov09]). For momentum transfers $Q \gg \Lambda$, the coupling constant becomes $\ll 1$ and QCD processes can be calculated using perturbation theory (see figure 2.1). The decrease of α_s with increasing momentum transfer is called 'asymptotic freedom'.

2.2. The Quark-Gluon Plasma

As mentioned in section 2.1, quarks are confined within hadrons under normal conditions. In certain QCD calculations [Cab75], it can be shown that for high temperatures T or a high baryochemical potential μ_B, the quarks are not confined anymore. This

[3] See appendix A.3.

new state of matter is called quark-gluon plasma (QGP). It is assumed that the QGP exists in neutron stars nowadays and that it existed until approximately 10^{-5} s after the Big Bang [BM07].

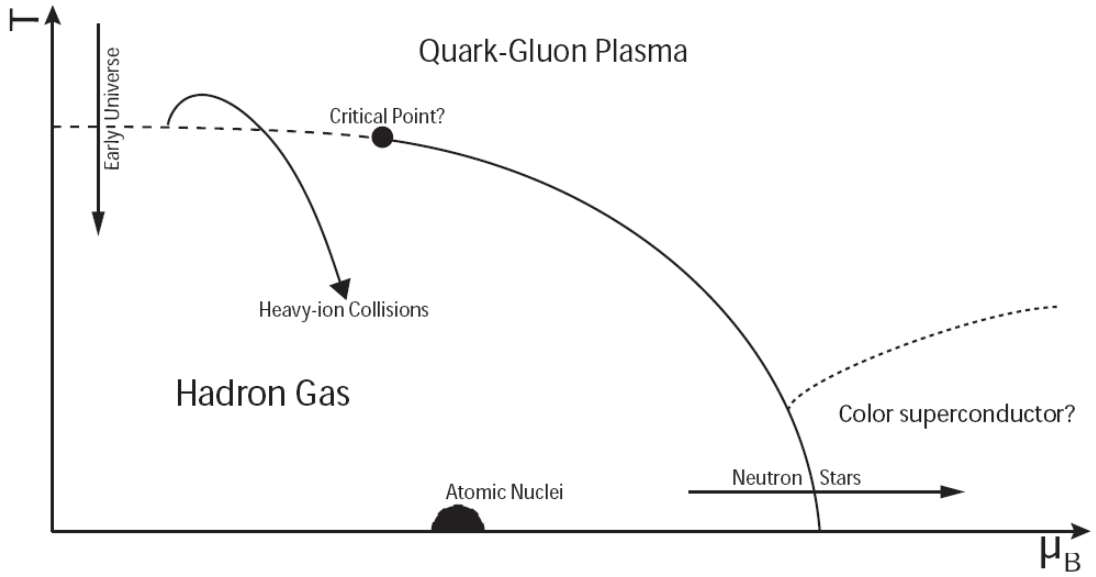


Figure 2.2.: The phase diagram of QCD. The solid line represents a first order phase transition, the dashed lines indicate crossover transitions. Neither the exact position of the critical point nor its existence has yet been measured [Sah10].

The study of the QGP will help to understand the evolution of the universe from the Big Bang to the creation of hadrons and atoms. It is also a test for some non-perturbative QCD calculations, so-called 'lattice QCD'. In figure 2.2, the phase diagram of strongly interacting matter is shown. For different temperatures T and baryochemical potentials μ_B, different states of matter exist. For small temperatures and baryochemical potentials, the quarks and gluons (also called partons) are confined within hadrons and form a hadron gas. The semi-circle at zero temperature represents the atomic nucleus, the ground state of nuclear matter. For a small baryochemical potential, the transition to the QGP is a smooth crossover, while it is a first order phase transition for a high baryochemical potential. The full circle in the diagram is the critical point, where the first order phase transition ends and the smooth crossover begins. Neither the exact position of the critical point nor its existence has yet been measured. In figure 2.3, the reduced energy density $\epsilon_{red} = \frac{\epsilon}{T^4}$ is plotted against the temperature of a hadron gas or QGP, calculated with lattice QCD. The calculations were done for different number of flavors, 2 (3) flavors means calculations by taking into consideration 2 (3) degen-

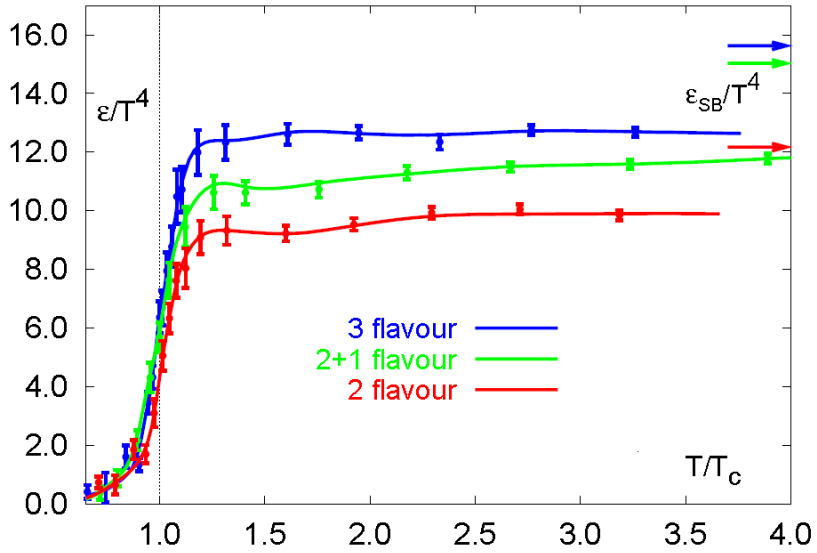


Figure 2.3.: The reduced energy density $\epsilon_{red} = \frac{\epsilon}{T^4}$ plotted against the temperature T. For $T = T_c$, a phase transition of the hadron gas occurred and a QGP is formed [Kar02].

erated quark masses and 2 + 1 flavors means calculations with 2 degenerated quark masses and 1 heavier quark. For a critical temperature T_c, the reduced energy density increases without increase of the temperature, which is characteristic of a phase transition. For an ideal gas, the energy density ϵ can be calculated by the Stefan-Boltzmann law, $\epsilon_{SB} \propto T^4$. The reduced energy density ϵ_{red} of an ideal gas does not depend on the temperature and is shown in figure 2.3 in the upper right corner. It can be seen that also for high temperatures, the reduced energy density does not reach the Stefan-Boltzmann limit and therefore, the QGP is not an ideal gas.

2.3. The Quark-Gluon Plasma in the Laboratory

The quark-gluon plasma can be produced in the laboratory in ultra-relativistic heavy-ion collisions, for example at the LHC, where lead ions are accelerated up to a center-of-mass energy per nucleon pair of $\sqrt{s_{NN}} = 5.5$ TeV, which is much higher than the rest mass of a nucleon. Due to Lorentz contraction, an ultra-relativistic lead ion is deformed into a thin disc in the laboratory frame. The collisions can be classified in different centralities described by the impact parameter b, which is small for very central collisions and large for peripheral collisions (compare figure 2.4). In central collisions, the two nuclei overlap very much and a lot of protons and neutrons participate in the collision.

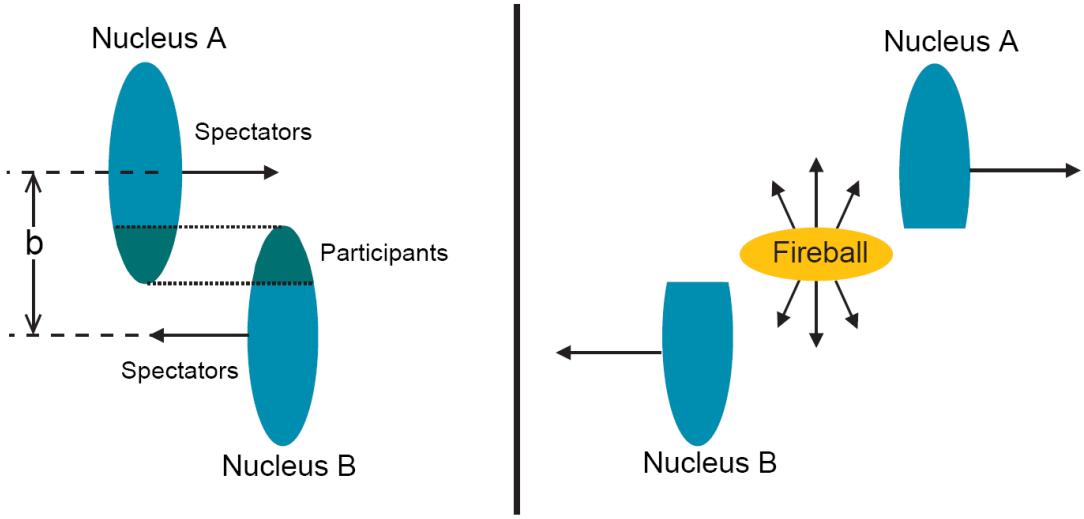


Figure 2.4.: Schematic view of the collision in the participant-spectator model [KB04].

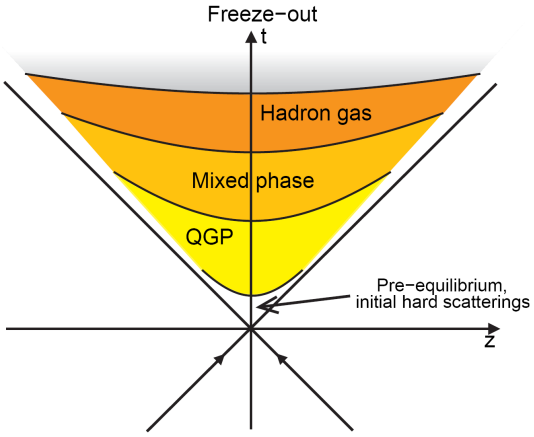


Figure 2.5.: Space-time evolution of an ultra-relativistic heavy-ion collision with a quark-gluon plasma [KB04].

These nucleons are called participants, the remaining nucleons are called spectators. In the collision, the quarks and gluons of the participants interact with each other and form the 'fireball'.

In figure 2.5, the space-time evolution of an ultra-relativistic heavy-ion collision is shown [Bjo83]. At the beginning (time $t = 0$) of the collision, the partons scatter hard and highly excited matter is formed. In the proceeding process, the fireball reaches a thermal equilibrium and it is possible that a quark-gluon plasma arises. The quark-gluon plasma cools down while expanding and after a possible mixed phase, the quarks and gluons are confined again within hadrons and a hadron gas is formed. These hadrons scatter inelastically and their composition changes until the chemical freeze-out is reached. Then, the composition of the hadron gas is fixed and the hadrons scatter only elastically. After the thermal freeze-out, also the elastic scattering stops and the momenta of the particles are fixed. These hadrons are measured in the detectors of the experiment.

2.4. Signatures of the Quark-Gluon Plasma

In the experiment, only colorless particles, like hadrons and leptons, can be measured, but no quarks and gluons. To get an evidence for a quark-gluon plasma, several signatures exist. Each of these signatures could come from another non-QGP scenario. In consequence, only combinations of the signatures can be an evidence for a quark-gluon plasma.

In collisions with protons instead of lead ions, no quark-gluon plasma is produced. To study the modification of the measurements due to a medium, the measured data in heavy-ion experiments are compared with data measured in proton-proton collisions. This comparison can be done directly by using the 'nuclear modification factor' R_{AA}. It is defined as

$$R_{AA} = \frac{\frac{d^2N}{dydp_T}|_{A+A}}{N_{coll} \cdot \frac{d^2N}{dydp_T}|_{p+p}}. \qquad (2.3)$$

$\frac{d^2N}{dydp_T}|_{A+A}$ is the number of measured particles in a certain rapidity and transverse momentum interval[4] in nucleus-nucleus collisions. $\frac{d^2N}{dydp_T}|_{p+p}$ is the number of measured particles in a certain rapidity and transverse momentum interval in proton-proton collisions. N_{coll} is the number of independent nucleon-nucleon collisions.

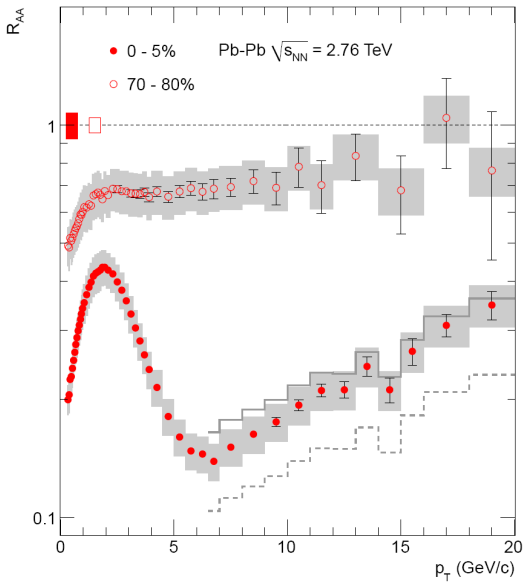


Figure 2.6.: Example of an R_{AA} measured by the ALICE experiment. Shown are central (0-5%) and peripheral (70-80%) lead-lead collisions with a center-of-mass energy per nucleon of $\sqrt{s_{NN}}$ = 2.76 TeV [ALI10].

In case the nuclear modification factor is equal to one, no medium effects influence the number of produced particles. For $R_{AA} < 1$, the number of produced particles per nucleon-nucleon collision is smaller in nucleus-nucleus collisions than in proton-proton collisions. In figure 2.6, the R_{AA}, measured by the ALICE experiment, is plotted as a function of the transverse momentum p_T. The R_{AA} is plotted for peripheral (70-80%) lead-lead collisions,

[4]Compare appendix A.1 for rapidity and transverse momentum.

expecting no quark-gluon plasma, and for central (0-5%) lead-lead collisions. A strong suppression of particles is measured in central collisions. This alone is not an evidence for a quark-gluon plasma, since the suppression could originate from the influence of e.g. 'Cold Nuclear Matter'. To estimate this influence, proton-nucleus collisions can be investigated.

The scaling of the number of produced particles with N_{coll} is not valid for low-p_T particles ($p_T \lesssim 2$ GeV/c). The particles with a small p_T scale with the number of participants of the collision N_{part}. Therefore, the R_{AA} is smaller than one for particles with $p_T \lesssim 2$ GeV/c.

The direct comparison of the different collision types by the nuclear modification factor implies that, for high p_T particles, a nucleus-nucleus collision can be described as a proton-proton collision scaled by the number of independent nucleon-nucleon collisions N_{coll}. To check this assumption, special probes are needed which do not interact strongly and are therefore not influenced by a medium. Two possible probes are photons and leptons. These probes travel through the medium without further collisions, therefore, they carry information about the stage of the collision in which they were produced. Some sources of leptons (l) are the Drell-Yan mechanism, a first hard scattering of two quarks and the creation of an $l\bar{l}$-pair via a virtual photon, semi-leptonic decay of charm or beauty mesons[5], and thermal leptons created in the plasma.

Photons can be divided into 'decay photons' and 'direct photons'. Direct photons are all photons not originating from decays and can be subdivided into 'thermal photons' and 'prompt photons'. The source of thermal photons is the hot medium, the prompt photons arise in hard parton-parton scatterings directly after the collision.

Besides thermal photons and the behavior of the reduced energy density ϵ_{red} as a function of the temperature T (see figure 2.3), the suppression and enhancement of J/Ψ is another signature of a quark-gluon plasma. The J/Ψ-meson is a charmonium ($J^{PC} = 1^{--}$) with a c-$\bar{c}$-potential $V(r)$ described by equation 2.1 under normal conditions. Inside a quark-gluon plasma, the color of the quarks is screened, similar to the electromagnetic screening in a normal plasma, and the potential $V(r)$ is changed to a Yukawa potential

$$V(r,T) = -\frac{4}{3}\frac{\alpha_s(Q^2)\hbar c}{r} e^{-\frac{r}{r_D(T)}}, \tag{2.4}$$

where $r_D(T)$ is the temperature-dependent Debye-radius, a parameter of the screening.

[5] Charm and beauty mesons are mesons with a charm-quark and a beauty-quark, respectively, and a light quark.

The Debye-radius decreases with increasing temperature and will be less than the J/Ψ binding radius for a certain temperature. The quarks are not confined anymore in this case, the mesons dissolve, and the number of J/Ψs is reduced.
This is only verified for low energies, like at RHIC. The Relativistic Heavy-Ion Collider (RHIC) is a particle accelerator at the Brookhaven National Laboratory (BNL) on Long Island. For high energies, like at the LHC, more c-$\bar{c}$-pairs are produced. Due to the Debye screening, the c-$\bar{c}$-pairs melt and the c ($\bar{c}$) can move independently through the plasma and recombine with other $\bar{c}$ (c) to J/Ψ. This results in an enhancement of J/Ψs, in case the charm-quarks are not confined and can travel freely for a certain distance. Therefore, the enhancement of J/Ψs is an indication for a quark-gluon plasma.

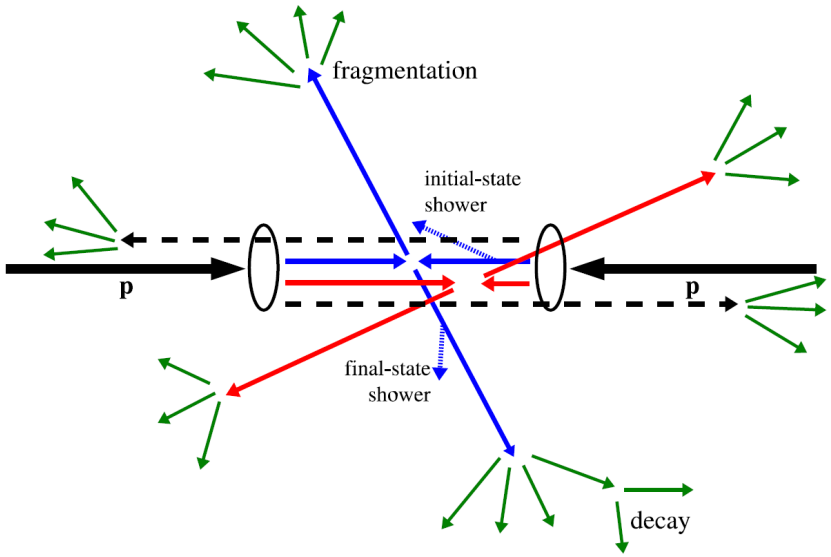


Figure 2.7.: Schematic view of a high-energy collision. Two protons collide (black), two partons scatter hard and two jets with high p_T are created (blue). In addition, two other jets with smaller p_T are produced (red). Due to gluon radiation, initial- and final-state showers arise. Colorless, principally measurable particles are drawn green [GO09].

In the beginning of Pb-Pb collisions, as well as in the beginning of proton-proton collisions, the partons of the colliding nucleons scatter hard against each other and e.g. two partons with high transverse momentum can be produced, which are flying in opposite directions. The two partons (e.g. q and $\bar{q}$) carry color and, due to the potential between them (see equation 2.1), the energy which is necessary to separate them from each other increases with increasing distance. The potential between the two partons is described

by a 'color flux string' (q-$\bar{q}$), similar to an elastic band. At a certain energy, the string breaks and two additional partons (e.g. $\bar{q}'$ and q') are produced. Between the original and the additional partons, two new strings arise (q-$\bar{q}'$ and $\bar{q}$-q'). The two additional partons are flying approximately in the same direction as the original parton, but with a smaller momentum. Therefore, the distance between the original and the additional parton increases and the new string breaks at a certain distance. This results in a production of particles and therefore in a lot of high-energetic hadrons, which are flying approximately in the same direction as the original parton. This cluster of hadrons is called 'jet'. A jet is defined by the output of a special algorithm, called jet-finder. In figure 2.7, a schematic view of a high-energy proton-proton collision is shown. Among others, two high p_T jets are created.
In heavy-ion collisions, partons also scatter hard and partons with high transverse momentum are produced. These partons have to travel through a medium, interact with it and therefore they will lose energy. This results in a suppression of high p_T particles[6] and jets. This phenomenon is called 'jet-quenching'.

2.5. Photons and Jets

The sources of prompt photons are hard scatterings of partons at the beginning of the collision.

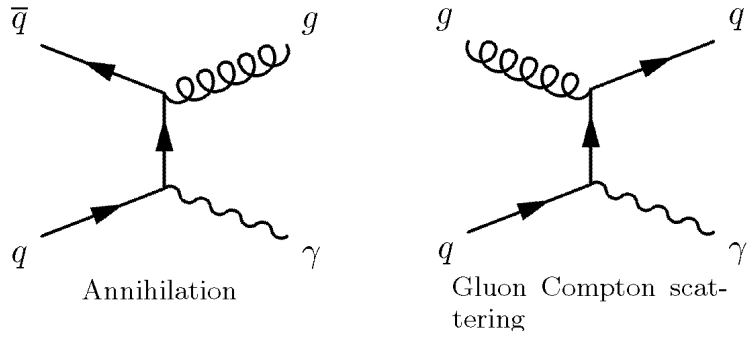


Figure 2.8.: Leading order Feynman diagrams of parton-parton scatterings which result in γ-jet events [Die06].

In figure 2.8, the Feynman diagrams of the origin of prompt photons are shown. Two partons of the colliding protons (or lead ions) scatter and due to annihilation or gluon

[6] Compare figure 2.6.

Compton scattering, a prompt photon and a parton are produced. The parton carries color and will therefore create a jet by fragmenting to hadrons. This fragmentation cannot be calculated in perturbative QCD because the momentum transfer Q is too small in these processes and hence the coupling constant of the strong force α_s is too large. To describe the creation of hadrons, a phenomenological fragmentation function is used. The cross section σ for the formation of a hadron h is given by [Boe02]

$$E\frac{d\sigma}{d^3\vec{p}}(p+p\to h+X) = \sum_{a,b,c,d}\int_0^1 dx_a \int_0^1 dx_b f_a^p(x_a,Q^2) f_b^p(x_b,Q^2)$$

$$\times \left(\frac{d\sigma}{dt}(ab\to cd)\right) D_h^c(z,Q^2)dz, \tag{2.5}$$

where E is the energy and $\vec{p}$ the three-momentum. X are the other produced particles except h. a,b,c, and d are different types of partons. x_a and x_b are the Bjorken scaling variables. Q^2 is the momentum transfer, $f_a^p(x_a,Q^2)$ and $f_b^p(x_b,Q^2)$ are the parton distribution functions of the colliding protons. t is a Mandelstam variable[7], $D_h^c(z,Q^2)$ is the fragmentation function, and z is the ratio of the hadron momentum p_h and the momentum of the original parton p_c, $z := \frac{p_h}{p_c}$.
According to equation 2.5, the fragmentation function can be interpreted as the probability to find a particle in the interval [z,z+dz]. The fragmentation function $D_h^c(z,Q^2)$ depends on the momentum fraction z and therefore the momentum of the measured hadron h and the momentum of the immeasurable parton c are needed to calculate $D_h^c(z,Q^2)$. To detect the momentum of the parton c, a prompt photon can be used. The transverse momentum of the scattering partons inside the two colliding protons is almost equal to zero and therefore the transverse momentum of the created parton is approximately the same as the transverse momentum of the prompt photon. Due to the conservation of momentum, the prompt photon and the parton are back-to-back correlated in the azimuthal angle φ. This means that the parton (the produced jet) has an azimuthal angle $\varphi_{jet} \approx \varphi_\gamma \pm \pi$ in case the prompt photon has an azimuthal angle φ_γ. The sum of the longitudinal momenta p_L of the scattering partons inside the two colliding protons is different from zero and therefore the jet and the prompt photon are not correlated in the pseudorapidity η; they can be boosted.

[7] See appendix A.1.

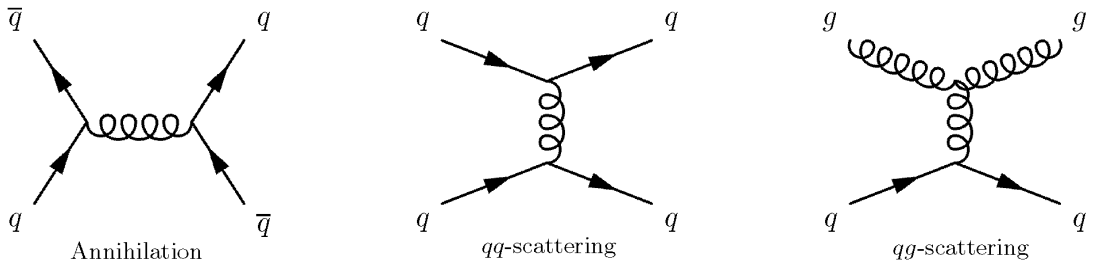


Figure 2.9.: Leading order Feynman diagrams of parton-parton scatterings which result in jet-jet events [Die06].

In figure 2.9, the Feynman diagrams of the origin of jet-jet events are shown. Two partons of the colliding protons (or Pb ions) scatter and due to annihilation, quark-quark scattering or quark-gluon scattering, two partons are produced which will fragment to hadrons and will produce two jets. These jets are also back-to-back correlated in φ and can be boosted in η. Two light neutral hadrons are the π^0-meson and the η-meson, which will be produced very frequently. The π^0 decays have a branching ratio of 98.8% into two photons and a branching ratio of 1.2% into an electron, a positron, and a photon. The η also decays into two photons (branching ratio = 39.3%), but also into 3 π^0s (branching ratio = 32.1%), which will most likely decay into photons. The decays of π^0s and ηs are the main source of decay photons.
For π^0s or ηs with a high momentum, the opening angle of the two photons is very small and the photons are detected close to each other. Due to the finite resolution of the detectors PHOS and EMCal[8], it is difficult to separate the two photons for very high momenta and it is possible that they are identified as a single high-energy photon which looks like a prompt photon. Using the transverse momentum of the wrongly identified photons as the parton momentum p_c in the momentum fraction z will lead to incorrect calculations of the fragmentation function. Therefore, a method to separate prompt photons and decay photons is needed. This method is the aim of this thesis and will be explained in detail in chapter 6.

[8] Compare section 3.2.1.

3. The LHC and ALICE

To produce a quark-gluon plasma in the laboratory, ultra-relativistic heavy-ion collisions are needed. This is realized at the LHC. The produced particles are measured by the detectors of the ALICE experiment. In this chapter, the LHC and the ALICE experiment will be shortly presented.

3.1. The Large Hadron Collider

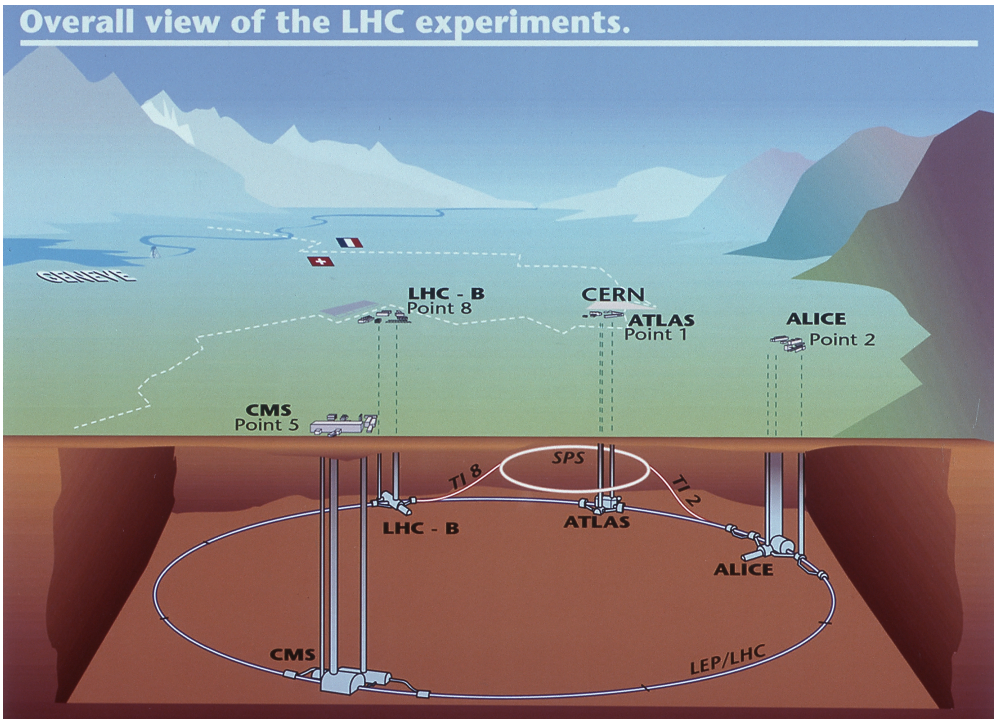


Figure 3.1.: Schematic view of the LHC and the experiments [CER11].

The Large Hadron Collider (LHC) at CERN is the world's biggest particle accelerator and is located under the Swiss-French border, 50 - 175 m below ground. It has a circumference of about 27 km and is designed to accelerate protons up to 7 TeV and lead ions up to 2.76 TeV per nucleon [LHC11]. The main goals of the LHC are [LHC11]

- finding the Higgs boson,
- testing the theory of supersymmetry,
- finding out what 'Dark Matter' (and 'Dark Energy') is,
- explaining the matter-antimatter imbalance in the universe, and
- studying the 'quark-gluon plasma'.

The LHC is a synchrotron with two beam pipes accelerating particles in opposite directions. Therefore, the total design collision energy is $\sqrt{s} = 14$ TeV for proton-proton collisions and $\sqrt{s_{NN}} = 5.5$ TeV for lead-lead collisions. The current total collision energy for proton-proton collisions is $\sqrt{s} = 7$ TeV. In each beam pipe, 2808 bunches with 1.1×10^{11} protons per bunch travel around the accelerator ring. The LHC has 1232 dipole magnets to bend the beams around the ring, 392 main quadrupole magnets to focus the bunches, and 8 cavities per beam to accelerate and focus the bunches. At four points, the bunches collide with a design luminosity[1] $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ for proton-proton collisions and $\mathcal{L} = 10^{27}\,\mathrm{cm}^{-2}\mathrm{s}^{-1}$ for lead-lead collisions. At these four points, six experiments are installed:

ALICE (A Large Ion Collider Experiment) [ALI95] studies the quark-gluon plasma, which is generated in Pb-Pb collisions, as well as proton-proton collisions to get reference data. The ALICE experiment will be described in detail in section 3.2.

ATLAS (A Toroidal LHC ApparatuS) [ATL99] and **CMS (Compact Muon Solenoid)** [CMS06] are experiments to explore physics at the TeV scale. The prime goals are the discovery of the Higgs boson and heavy particles which are postulated by the theory of supersymmetry. ATLAS is the largest experiment at the LHC and also the largest particle detector in the world. It consists of inner detectors, forward detectors, a hadronic and an electromagnetic calorimeter, a muon spectrometer, and a magnet system. ATLAS and CMS have the same goals, but use different techniques. Therefore, results can be checked against each other.

[1] The luminosity is given by $\mathcal{L} = f \cdot N_1 \cdot N_2 \cdot n/A$ with n bunches in each beam, N_1 and N_2 particles per bunch, a cross-sectional area A of the beams and circulation frequency f. For a given cross section σ, the rate of a process dN/dt can be calculated with the luminosity: $dN/dt = \mathcal{L}\,\sigma$.

LHCb (Large Hadron Collider beauty experiment) [LHCb05] studies the CP-symmetry violation in b-quark hadrons. This is possible because of the large $b\bar{b}$ production cross section at the LHC energies.

LHCf (Large Hadron Collider forward experiment) [LHCf05] is close to ATLAS and measures forward particles, which are created in the collisions, to test models for the high energy region of cosmic rays.

TOTEM (TOTal Elastic and diffractive cross section Measurement) [TOT04] measures the total cross section of proton-proton collisions. It also studies elastic scattering and diffractive dissociation at the LHC.

3.2. The ALICE Experiment

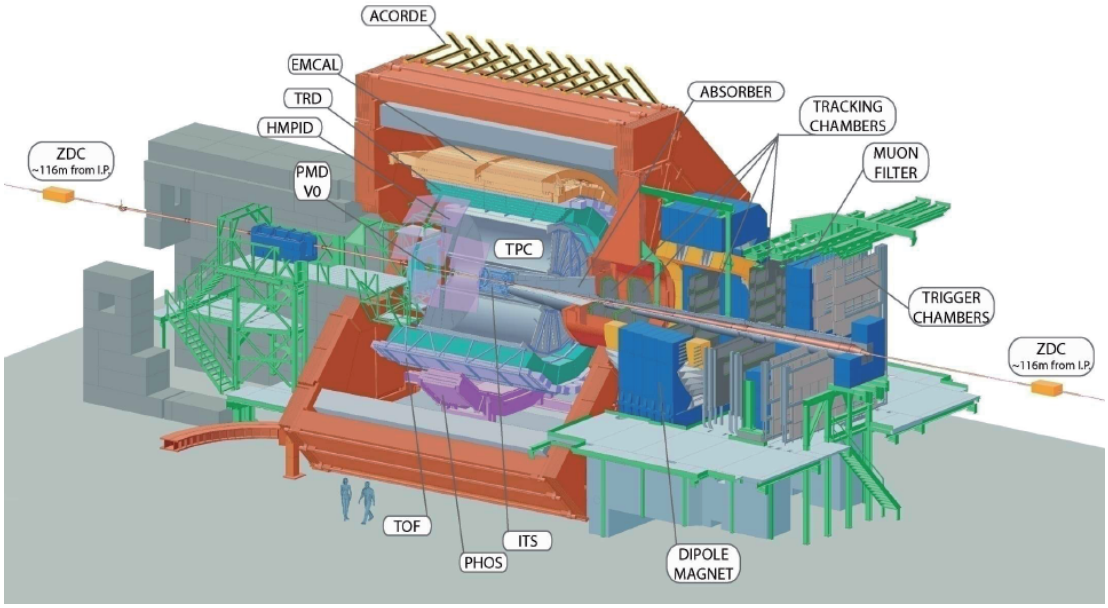


Figure 3.2.: ALICE detectors [Wil09]

A Large Ion Collider Experiment (ALICE) is a particle detector with the aim to explore the phase transition between normal nuclear matter and a quark-gluon plasma. It studies Pb-Pb collisions, in which the quark-gluon plasma is generated and also proton-proton collisions to compare the result with Pb-Pb data. The detector has a size of 16 m $\times$ 16 m $\times$ 26 m and it weights about 10,000 t.

ALICE is developed to measure 4000 charged particles per unit of η ($\mathrm{d}N_{ch}/\mathrm{d}\eta = 4000$), but it is able to handle events with a much higher multiplicity (up to $\mathrm{d}N_{ch}/\mathrm{d}\eta = 8000$). It has the special feature to detect different particles (hadrons, electrons, muons, and photons), reconstruct their tracks over a large range of momenta (from 100 MeV/c up to over 100 GeV/c), and identify them [ALI95].
The ALICE detector can be divided into the 'central barrel' located around the collision point from $\eta = -0.9$ to $\eta = 0.9$, the 'Muon Arm' at a pseudorapidity region $-4.0 \leq \eta \leq -2.5$, and other 'forward detectors'. The central barrel is surrounded by a solenoid magnet, called 'L3', which creates a homogeneous magnetic field of $B = \pm 0.5$ T at room temperature. The subdetectors of the central barrel are ITS, TPC, TRD, TOF, PHOS, EMCal, and HMPID; they are explained in section 3.2.1. The Muon Arm and the forward detectors are explained in section 3.2.2.

3.2.1. The Central Barrel

ITS (Inner Tracking System): The ITS is the innermost detector of ALICE. Its tasks are the reconstruction of the primary vertex of the collision and the reconstruction of the secondary decay vertices of D-, B-mesons, and hyperons. Using the specific energy loss ($\mathrm{d}E/\mathrm{d}x$), it is able to identify particles and their tracks for low momenta ($p \leq 100$ MeV/c) [ITS99]. The ITS consists of six layers of silicon detectors. The two innermost layers are Silicon Pixel Detectors (SPD) which are able to handle an expected track density of 80 tracks/cm^2 [ALI04]. The following two layers are Silicon Drift Detectors (SDD) and the outermost layers are Silicon Strip Detectors (SSD).

TPC (Time Projection Chamber): The TPC is a gas detector with a volume of 90 m^3 and a gas mixture of 90% Ne and 10% CO_2. The task of the TPC is the tracking of particles. It measures the specific energy loss $\mathrm{d}E/\mathrm{d}x$ (see figure 3.3) and has the capability to reconstruct and identify 20,000 tracks per event. The TPC is able to measure the momentum and the production vertex of a particle with a momentum between 100 MeV/c and 100 GeV/c and can identify it [TPC00, GO09]. The charged particles will ionize the gas and produce free electrons. These electrons drift in an electrostatic field to the end caps. Due to the drift time of about 90 μs, the TPC is the slowest detector in ALICE [ALI08].

TRD (Transition Radiation Detector): The tasks of the TRD are the separation

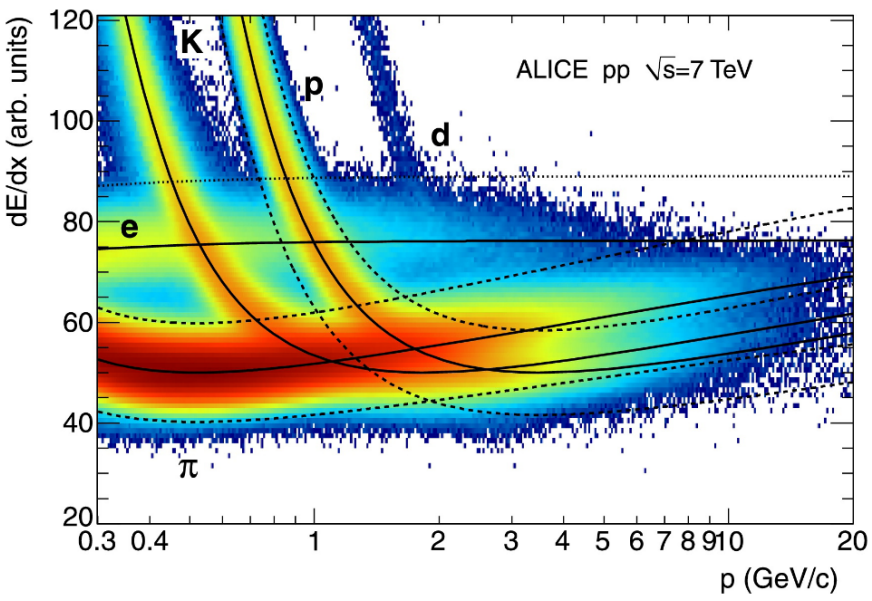


Figure 3.3.: Specific energy loss ($\mathrm{d}E/\mathrm{d}x$) in the TPC as a function of momentum p [Ali11]. The solid lines represent the Bethe-Bloch lines for various particle species. The dashed lines show exclusion bands for pions and protons, used in [Ali11], and are not important in this context. The specific energy loss is shown for electrons (e), kaons (K), protons (p), deuterons (d) and pions (π).

of electrons and pions with a momentum ≥ 1 GeV/c and serving as a fast trigger on high p_T particles [TRD01]. The TRD is able to identify electrons with an efficiency of 90% and a pion rejection above 100 [Wil09]. It consists of 18 super-modules, each one comprising 30 individual detector modules, the ReadOut Chambers [TRD01]. One super-module is subdivided in five stacks with six layers each.

TOF (Time Of Flight): The TOF detector separates protons and kaons from pions by measuring the time between the collision and the arrival of the particles at the TOF detector [TOF00]. Together with the ITS and the TPC, the TOF is able to identify pions and kaons with a momentum up to 2.5 GeV/c and protons with a momentum up to 4 GeV/c with a good π/K and K/p separation [ALI08].

The ITS, TPC, TRD, and TOF have an acceptance of $|\eta| \leq 0.9$ and the full azimuth of $\varphi = 360°$.

PHOS (PHOton Spectrometer): The PHOS detector is a calorimeter with an acceptance $|\eta| \leq 0.12$ and $\Delta\varphi \approx 100°$. It is positioned at the bottom of ALICE and

detects direct photons, π^0s, and ηs by measuring their decay products [PHO99]. By measuring high p_T π^0s and photons, PHOS and EMCAL allow the study of γ-jet correlations and jet-quenching.

EMCal (ElectroMagnetic Calorimeter): The EMCal is, like PHOS, a calorimeter to detect photons and is installed approximately opposite to PHOS. It has a bigger acceptance of $|\eta| \leq 0.7$ and $\Delta\varphi \approx 107°$, but a worse resolution [EMC08].

HMPID (High-Momentum Particle IDentification): The task of the HMPID detector is the identification of hadrons with a high momentum. It has an acceptance of $|\eta| \leq 0.6$ and $1.2° \leq \varphi \leq 58.8°$ [HMPID98].

ACORDE (ALICE COsmic Ray DEtector): The ACORDE detector is placed on the top of the L3 magnet. It studies high-energy cosmic rays and is also used for calibration and alignment.

3.2.2. The Muon Arm and the Forward Detectors

The muon spectrometer consists of a dipole magnet, absorbers, tracking chambers, and trigger chambers. It covers a pseudorapidity range of -4.0 $\leq \eta \leq$ -2.5 and the full azimuth. Its task is to measure quarkonia (J/Ψ, Ψ', Υ, Υ',Υ'') through their dimuon decays. The mass resolution has to be good enough to separate the single states. Furthermore, the production of beauty and open charm can be studied.
The forward detectors are ZDC, PMD, FMD, T0, and V0. The ZDC (Zero-Degree Calorimeter) is located 116 m away from the collision point at both sides. It measures the energy in beam direction, which is correlated to the number of spectator nucleons and therefore to the centrality of the collision [ZDC99]. The PMD (Photon Multiplicity Detector) is 5.8 m away from the collision point and covers an η range of $1.8 \leq \eta \leq 2.6$ and full azimuth (opposite to the muon spectrometer). It measures the multiplicity distribution of photons close to the beam direction [PMD99]. The FMD (Forward Multiplicity Detector) collects information about the multiplicity of charged particles in -3.4 $\leq \eta \leq$ -1.7 and 1.7 $\leq \eta \leq$ 5.0. The T0 detector is used as a trigger and creates the start signal for TOF. The V0 detector is an online trigger for minimum bias events and centrality with an acceptance of -3.3 $\leq \eta \leq$ -2.9 and 4.5 $\leq \eta \leq$ 5.0 [FWD04].

4. Theory of Artificial Neural Networks

The goal of this thesis is to distinguish between prompt photons and decay photons. For both kinds of photons, variables exist in which they differ partially. With these variables, a pattern can be formed which is characteristic of the photons. The patterns of signal (= prompt photons) and background (= decay photons) form two classes. An artificial neural network is able to distinguish between classes of patterns. In this chapter, the theory of artificial neural networks will be presented.

4.1. Introduction

An inspiring example of a neural network is the human brain. It has a lot of abilities which, nowadays, no computer can reach, for instance understanding languages or developing complicated movements. These advantages over computers are caused by the capability of learning. Otherwise, computers can perform calculations much faster than humans. An artificial neural network is the attempt to emulate the properties of a brain by using the capability of fast calculating.
Typical applications of neural networks are classification of patterns, approximation of functions, predictions, e.g. of share prices and others [Sch97]. In particle physics, neural networks are often used for identification of particles [Wil09], reconstruction of particle tracks and classification of decays [Kun95].
Some advantages of neural networks are learning aptitude, robustness, speed, and generalization aptitude [Sch97]. An artificial neural network is able to learn which output it has to produce, depending on the input (learning aptitude). It is also able to handle noisy data (robustness) and can associate the correct output to unknown input (generalization aptitude). The training of the network will take some time, but after that, the generating of the output will be fast (speed). A disadvantage is that neural networks are black boxes. The output is a non-linear function of the input and therefore, it is not possible to find out which part of an input is responsible for a certain output.

4.2. The Network Structure

A neural network is composed of 'neurons' and 'synapses'. The neurons are like small processors which are able to perform simple calculations, for instance the summation of the input. The synapses connect the neurons to each other, refer the output of one neuron to another neuron, and amplify or damp the output meanwhile. This weighting of the output of a neuron can be done by multiplying a number (weight) to the output (see figure 4.1). Changing the weight means changing the output of the network. To solve a certain problem with a neural network, a network topology, a learning algorithm, some training data, and some validation data are needed.

There are many types of neural networks, which can be divided into feedforward and feedback networks. Feedforward networks have a hierarchical structure, therefore, the neurons can be grouped in individual layers. This kind of network has one input layer, some hidden layers, and one output layer. The synapses only connect different layers, therefore no connections inside a layer exist (see figure 4.2). These networks are called perceptrons. The information only flows in one direction from the input layer, through the hidden layers, to the output layer. The synapses can only connect a layer with the next layer (first order) or additionally with other higher layers (second order).
In figure 4.2, a multilayer perceptron of first order with one hidden layer is shown. The following explanations are adapted from references [Wil09] and [Wil04]. The input neurons I_i can be written as an input vector $\vec{I} = (I_1,I_2,I_3,I_4)$, the weights to the hidden layer $w_{i,j}$ as a matrix $(w_{i,j})$

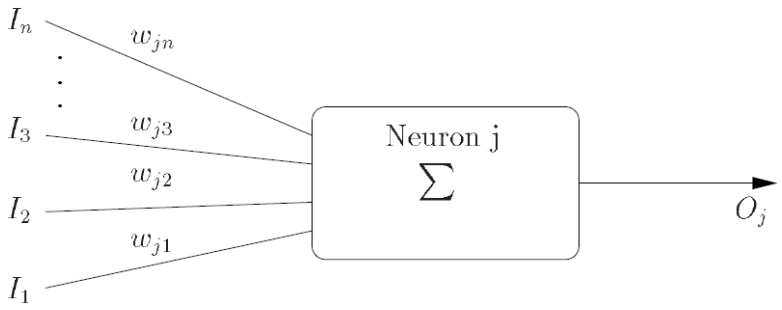


Figure 4.1.: Illustration of a neuron [Wil09]. The input signals are weighted by the synapses and summed up by the neuron.

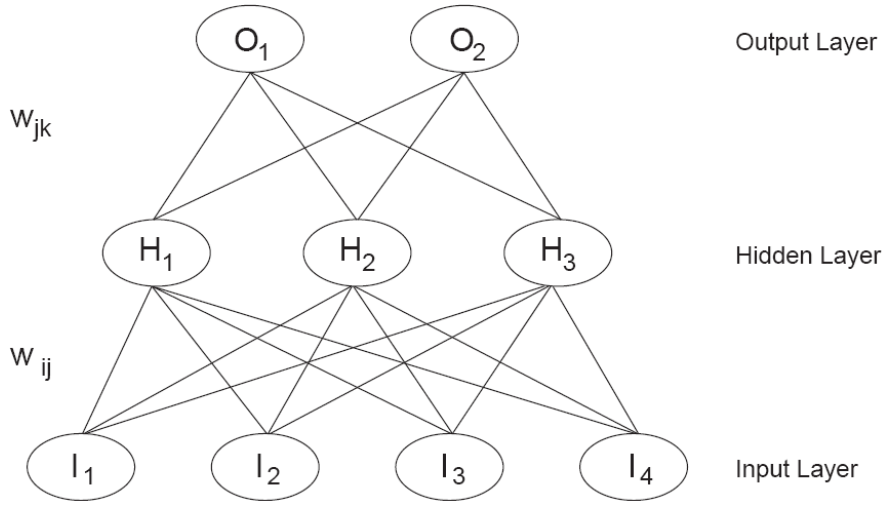


Figure 4.2.: Illustration of a neural network [Wil09]. Shown is a multilayer perceptron of first order with one hidden layer. The network has four input neurons I_i, three hidden neurons H_j, and two output neurons O_k. The weights $w_{i,j}$ and $w_{j,k}$ only connect consecutive layers.

$$(w_{i,j}) = \begin{pmatrix} w_{1,1} & w_{1,2} & w_{1,3} \\ w_{2,1} & w_{2,2} & w_{2,3} \\ w_{3,1} & w_{3,2} & w_{3,3} \\ w_{4,1} & w_{4,2} & w_{4,3} \end{pmatrix}. \tag{4.1}$$

The output of the hidden neuron H_j is calculated by using a 'propagation function' h_j and an 'activation function' f. The propagation function is typically the summation of the weighted input (another example is the multiplication of the weighted input):

$$h_j := \sum_{i=1}^{4} w_{i,j} I_i. \tag{4.2}$$

The activation function f can be the identity function, the sigmoid function, or others. In this thesis, a sigmoid function is used as activation function for hidden neurons

$$H_j = f(h_j) = f(\sum_{i=1}^{4} w_{i,j} I_i) = \frac{1}{1 + e^{-h_j}} = \frac{1}{1 + \exp(-(\sum_{i=1}^{4} w_{i,j} I_i))}. \tag{4.3}$$

These outputs are weighted again by the weights to the output layer $(w_{j,k})$

$$(w_{j,k}) = \begin{pmatrix} w_{1,1} & w_{1,2} \\ w_{2,1} & w_{2,2} \\ w_{3,1} & w_{3,2} \end{pmatrix}. \tag{4.4}$$

For output neurons, the activation function is the softmax function

$$f(x_k) = \frac{e^{x_k}}{\sum_k e^{x_k}}. \tag{4.5}$$

This ensures that the values of the output neurons are between zero and one and that their sum is one (compare section 4.6). The output of the network is given by

$$O_k = \frac{\exp\left(\sum_{j=1}^{3} w_{j,k} H_j\right)}{\sum_k \exp\left(\sum_{j=1}^{3} w_{j,k} H_j\right)}. \tag{4.6}$$

Feedback networks have no hierarchical structure. Mostly, a separation of different layers is not possible and the neurons are connected with any other neurons. This kind of networks will not be used in this thesis.

4.3. Backpropagation Algorithm

Training of the neural network means changing the weights in such a way that the output of the network is approaching the target output. In this thesis, the target output of signal is (t_1,t_2) = (1,0) and the target output of background is (t_1,t_2) = (0,1). The backpropagation algorithm is able to train feedforward networks of first and second order [Sch97].
In appendix A.4, the backpropagation algorithm is derived. The modification of the weight $w_{i,j}$ is proportional to the variation of the error ϵ_p with the weight,

$$\Delta_p w_{i,j} := -\eta \frac{\partial \epsilon_p}{\partial w_{i,j}}, \tag{4.7}$$

and can be calculated with

$$\Delta_p w_{i,j} = \eta(t_{p,j} - o_{p,j}) \cdot f'(\sum_i w_{i,j} o_{p,i}) o_{p,i} \tag{4.8}$$

for output units and with

$$\Delta_p w_{i,j} = \eta(\sum_k \delta_{p,k} w_{j,k}) \cdot f'(\sum_i w_{i,j} o_{p,i}) o_{p,i} \tag{4.9}$$

for other units. $t_{p,j}$ is the target output and $o_{p,j}$ the real output of the jth output neuron for a pattern p. The learning parameter η is a real, constant number and determines the speed of the training. In this thesis, the class `TMultiLayerPerceptron` of the analysis framework ROOT[1] is used to create the neural networks. In this class, the training of the network is done using equation 4.7, which corresponds to equation 4.8 and 4.9.

[1] More about ROOT can be found in [ROO].

4.4. Training and Validation

The training of the network starts with the input vector, which is supplied to the input neurons. The input is multiplied by the weights (which have random values) and sent to the first layer of hidden neurons. The output of the hidden neurons is calculated and sent to the next hidden layer or the output layer (feedforward). In this way, an output for the network is produced. The resulting output is compared with the target output and the error $\epsilon_p = \frac{1}{2}\sum_j (t_{p,j} - o_{p,j})^2$ of the network is calculated. The modification of each single weight is computed with equation 4.8 and 4.9. First, the modifications for the output layer are determined, then the modifications for the previous layer and so on, up to the input layer. The weights are modified from back to front (backpropagation). The error of the network is a function of the weights and therefore high-dimensional. The backpropagation algorithm is trying to find the global minimum of the error function by modifying the weights in such a way that the steepest slope of the error function is followed (see figure 4.3).

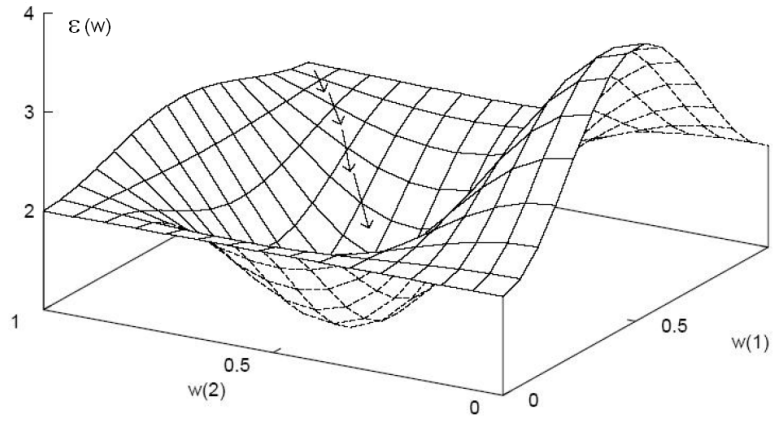


Figure 4.3.: Example of an error function $\epsilon(w)$ for two weights $w(1)$ and $w(2)$ [Sam05]. The modification of the weights follows the steepest slope of the error function.

To train the network, some training data are needed. The minimization of the error function is done for every input vector, which means for every pattern of the training data. A loop over all training patterns is called an 'epoch'. The network is trained several epochs and after each one, some test data are used to test the generalization ability[2] of the network. The patterns of the test data are propagated through the network and their error is calculated. If the error of the test data becomes larger during

[2]The ability of the network to identify unknown patterns correctly.

the training, the generalization ability of the network will decrease. This phenomenon is called 'Overfitting' and means an exclusive adaption to the training data. Therefore, the ability of the network to identify unknown patterns (= patterns which are not used for training) correctly is lost and the network can only identify the training patterns. To validate the separation performance of the neural network, a third data set is needed. This data set is propagated through the network after the training and the output distribution of signal and background is plotted. A good separation performance means well-separated output distribution of signal and background.

4.5. Input Space

Feedforward networks have an input layer with n input neurons. These input neurons are the basis of the n-dimensional input space.

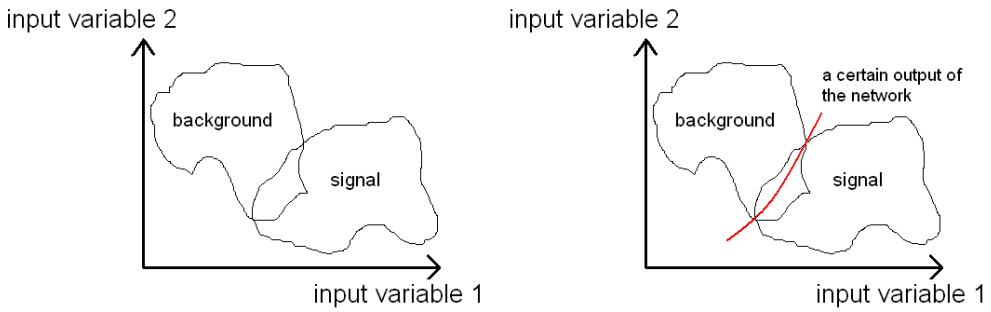


Figure 4.4.: Left: Example of an input space of a feedforward network with two input neurons. The patterns of signal and background will form two sets of numbers in the input space, which are not disjoint in most cases. Right: Input space with a separation plane, calculated by the neural network.

For two input neurons, a possible input space is shown in figure 4.4. Every pattern of signal and background is a point in the two-dimensional input space and together they will form two sets of numbers. The sets of signal and background will partly overlap in most cases. By training the network, 'separation planes' will be automatically calculated in the input space (compare figure 4.4, right). This separation plane is synonymic with a certain output of the network, e.g. 0.5. On the left (right) side of the line in figure 4.4, the output of the network will be less (greater) than 0.5 and the network will identify these patterns as background (signal). This separation plane can be a straight

line as well as a complicated curve. The set of background and signal are not disjoint, therefore, the neural network cannot completely distinguish between background and signal. Compared to the use of cuts, the advantage of a neural network is that the 'cut' on variable 2 is not fixed, but depends on the value of variable 1.
For more input variables (and higher dimensions), it works in the same way, the network will calculate separation planes in the high-dimensional input space. The main challenge is to find suitable input variables to disjoin the set of background and signal in the input space.

4.6. Interpretation of the Network Output as Likelihood

Under certain conditions, the output of a neural network can be interpreted as Bayesian a posteriori likelihood. That means that the output of the i-th output neuron of the network, for a certain pattern, is the likelihood that this pattern belongs to the class of patterns c_i. The class of patterns c_1 must have the target output (t_1,t_2) = (1,0) and the class of patterns c_2 must have the target output (t_1,t_2) = (0,1). The conditions are [LIP91, Ott96]

- use of a quadratic error function,
- one output neuron per class of patterns,
- the a priori likelihood of all distributions of the training data has to be the same as in the experiment, and
- the global minimum of the error function must be reached.

The Bayesian a posteriori likelihood that a pattern p belongs to the class of patterns c_i is given by [Ott96]

$$P(c_i|p) = \frac{P(p|c_i)P(c_i)}{\sum_j P(p|c_j)P(c_j)}. \tag{4.10}$$

P(p|c_i) is the likelihood that the class of patterns c_i creates a pattern p, P(c_i) is the a priori likelihood of the class of patterns c_i[3](i=1,2). As it can be seen in equation 4.10, the output of the network is generally weighted by the relative frequencies of c_1 and c_2 in the training data.
For the same relative frequencies of c_1 and c_2, P(c_1) = P(c_2), equation 4.10 can be simplified to

$$P(c_i|p) = \frac{P(p|c_i)}{P(p|c_1) + P(p|c_2)}. \tag{4.11}$$

[3]This is equal to the relative frequency of c_i in the training data.

In chapter 6 and 8, the neural networks are trained, tested, and validated using the same relative frequencies. In chapter 7, the neural networks are trained and tested using the same relative frequencies and validated using realistic relative frequencies. Also in the application of the network in measured data realistic relative frequencies will be used for validation, since more decay photons than prompt photons exist.

5. Monte-Carlo Simulation of Data

In this thesis, a neural network is used to distinguish between prompt photons and decay photons. To train the neural network and to test it, data with special properties are needed. The requirements are not fulfilled by the existing simulated data, therefore, new simulations are implemented. The settings of the simulations are presented in this chapter.

5.1. Principles of Data Simulation

As mentioned in chapter 3, ALICE measures the final state particles which are created in proton-proton collisions and Pb-Pb collisions at the LHC. The collisions can be simulated by using 'event generators', which are able to simulate the particles and their interactions among each other. No new physics can be discovered in simulations, but the event generators can reproduce the particles and their distributions (e.g. their p_T distribution and η distribution), which were measured in earlier experiments or predicted theoretically.
In the event generator, all produced particles can be identified by their 'PDG code'[1]. The PDG code is a numerical code to label the type of the particle (e.g. 111 = π^0, 22 = photon). Besides the PDG code, the mother-daughter relationships of every particle, all kinematic variables like the transverse momentum, global event information like the p_T-hard[2], and other information are available.
In this thesis, the event generator PYTHIA is used [PYT06]. PYTHIA simulates the particles until a short time after the collision. The program GEANT propagates the particles from PYTHIA to the detector and simulates interactions and particle decays [GEA3]. GEANT is also able to simulate the detector's layout and calculates the interaction of the particles with the detector material when they pass through. The response of the detector is stored as 'digits', which can be processed to 'raw data'. These raw data are also produced by the detectors in ALICE and measured and simulated

[1] PDG = Particle Data Group.
[2] See section 5.2.

raw data should ideally be equal.
The data, measured by ALICE or simulated, are processed using the software framework AliRoot, which is written in the object-oriented programming language C++ and based on ROOT. With AliRoot, the raw data can be used to reconstruct 'clusters', which represent a set of adjacent signals in the detector generated by the same particle. A possible track of a particle is created in such a way that the track fits well to the reconstructed clusters. This track represents the measured particle.

5.2. Pure PYTHIA Simulation

To get prompt photons, a simulation of γ-jet events is needed (see section 2.5). At the beginning of the collision, two partons scatter hard with a momentum transfer $\hat{q}$. The transverse component of $\hat{q}$ is called 'p_T-hard'. The transverse momentum of the prompt photon ($p_T^{\gamma(\text{prompt})}$) is approximately the same as the p_T-hard. In reality, the p_T-hard distribution will follow a power law, which means that it is small in most events and only rarely high. Usually in simulations, the p_T-hard distribution is divided in several 'p_T-hard bins', which means that a high p_T-hard is as frequent as a small p_T-hard and the p_T-hard follows the power law only inside a p_T-hard bin. This is done to get high statistics of events also for high p_T-hard.
As mentioned in section 4.6, it is necessary that the a priori likelihood of all distributions of the training data is the same as in the experiment. That means that the distribution of the p_T-hard (and of $p_T^{\gamma(\text{prompt})}$) for the training data has to be identical with the p_T-hard distribution measured in the experiment. Therefore, a new simulation of γ-jet events is needed.
In the simulation of γ-jet events, the p_T-hard is set to p_T-hard ≥ 10 GeV/c in a single p_T-hard bin at a center-of-mass energy of $\sqrt{s} = 7$ TeV. The magnetic field is fixed to -0.5 Tesla and jet-quenching is switched off. The prompt photons must have a pseudorapidity of $|\eta| \leq 0.7$ and an azimuthal angle φ of $79° \leq \varphi \leq 191°$, which corresponds to the position of EMCal.
The decays of π^0s and ηs are the main source of decay photons. Here, only photons from π^0s are investigated, representative for all decay photons. To get decay photons with high transverse momentum $p_T^{\gamma(\pi^0)}$, jet-jet events are used. A π^0 with high transverse momentum ($p_T \gtrsim 10$ GeV/c) will be produced only in a jet. Therefore it does not matter whether the condition that a jet is produced is set or not. As in the existing γ-jet events, the p_T-hard of the existing jet-jet-simulations is divided in several p_T-hard

bins and is not as in reality. Therefore the existing jet-jet events cannot be used. Besides jet-jet events, 'minimum bias' events exist with a realistic p_T-hard distribution. The requirement for a minimum bias event is that a collision happened. There are no further requirements, like the creation of a jet or a special p_T-hard. In minimum bias events, the p_T spectrum of the produced particles is the same as in the experiment, therefore, the probability to get a π^0 with high p_T is very small. To get a realistic p_T-hard distribution and many high p_T π^0s simultaneously, a new simulation of jet-jet events is needed.

In the simulation of jet-jet events, the p_T-hard is set to p_T-hard ≥ 20 GeV/c in a single p_T-hard bin at a center-of-mass energy of $\sqrt{s} = 7$ TeV. The magnetic field is fixed to -0.5 Tesla and a possible jet-quenching is switched off. The transverse energy[3] of the jet E_T has to be between 10 GeV and 1000 GeV, the transverse momentum of at least one π^0 has to be greater than 17 GeV/c, and its η and φ have to be in the EMCal acceptance.

In reference [Bal05], prompt photons with $p_T \geq 20$ GeV/c were investigated. In this thesis, photons with a similar value of $p_T \geq 17$ GeV/c are simulated. The analysis of this simulated data is a first try of using a neural network for prompt photon identification. Therefore, no GEANT detector simulation is done after the PYTHIA simulation. This simulated data will be used in section 6.1.

5.3. PYTHIA Simulation with a GEANT Detector Simulation

As mentioned in section 5.1, after the PYTHIA simulation, a GEANT detector simulation is usually done to simulate the interaction of the particles with the detector and the response of the detector. A new PYTHIA simulation of γ-jet and jet-jet events is done, together with a GEANT detector simulation, to get a more realistic distribution of the possible input variables for the neural network.

In the PYTHIA simulation of γ-jet events, the p_T-hard is set to p_T-hard ≥ 7 GeV/c in a single p_T-hard bin. All other settings are the same as in the simulation of γ-jet events of section 5.2. The complete ALICE experimental setup is simulated by GEANT, including all detectors. For the TRD, only seven supermodules are taken into account; for all other detectors, the final layout is simulated.

In the simulation of jet-jet events, the p_T-hard is set to p_T-hard ≥ 13 GeV/c in a single p_T-hard bin. The transverse momentum of at least one π^0 has to be greater

[3]The transverse energy E_T can be calculated by $E_T = \mathrm{E} \cdot \sin(\vartheta)$.

than 10 GeV/c and its pseudorapidity has to be in the acceptance of the central barrel, $|\eta| \leq 0.9$. The change of the minimal p_T of the π^0 from 17 GeV/c to 10 GeV/c is done to get better statistics of π^0s. All other settings are the same as in the simulation of jet-jet events of section 5.2. For jet-jet events, only a part of the ALICE setup is simulated by GEANT. The muon spectrometer, its absorber, FMD, PHOS, TRD, ZDC, EMCal, ACORDE, and the VZERO detector are not simulated because they are not important for this analysis.
This simulated data will be used in section 6.2.

6. Photon Separation in Monte-Carlo Simulations

6.1. Photon Separation in a Pure PYTHIA Simulation

In this section, the separation of prompt photons and decay photons created in the pure PYTHIA simulation is presented. For the settings of the pure PYTHIA simulation, see section 5.2. After a description of the analysis procedure, the results of the isolation cut and the neural network are presented.

6.1.1. Analysis Procedure

For the simulation of the high-energy proton-proton collision, PYTHIA 6.4 is used. In this context, decay photons are all photons originating from a π^0, which means that the mother of the photon has to be a π^0. Prompt photons are all photons without a mother particle. In figure 2.8 in section 2.5, it can be seen that the mother of a prompt photon should be a quark or a gluon. In this PYTHIA simulation, the mother of a prompt photon is not set, therefore the prompt photons have no mother. The decay photons and the prompt photons are called 'trigger photons'. Their transverse momentum[1] has to be greater than 17 GeV/c and they must be in a pseudorapidity interval of $|\eta| \leq 0.5$. To separate decay photons and prompt photons, an 'isolation cut' can be used. As mentioned in section 2.5, the prompt photon and the jet are back-to-back correlated in φ, therefore, no particle with high p_T should be in the immediate surrounding of the prompt photon. In contrast, a π^0 with high transverse momentum will be produced only in a jet, which results in a lot of high p_T particles in the immediate surrounding of the decay photon (compare figure 6.1). The immediate surrounding of the trigger photon is defined by a cone with a radius R_{max}.

[1] In reference [Bal05], the isolation cut was applied to prompt photons with $p_T \geq 20$ GeV/c. In this thesis, the analysis started using a similar value of 17 GeV/c as a first try.

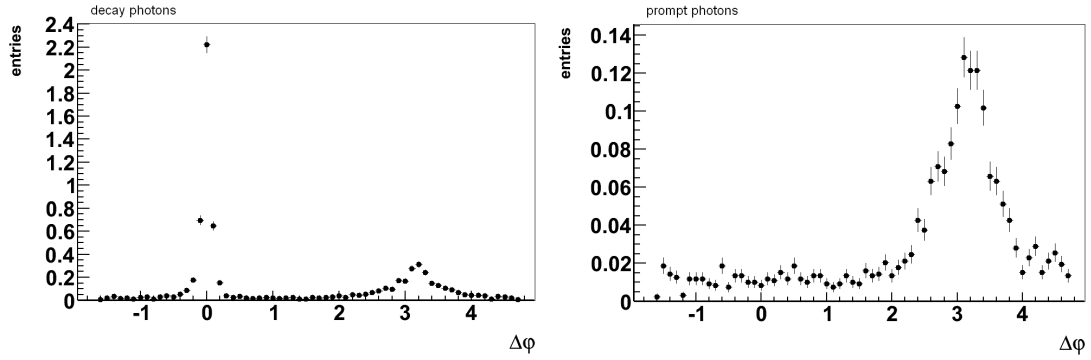


Figure 6.1.: $\Delta\varphi = \varphi_{\text{particle}}$ - $\varphi_{\text{trigger }\gamma}$ distribution of high p_T particles ($p_T \geq 2$ GeV/c) for decay photons (left) and prompt photons (right) normalized to the number of photons. Particles are all charged PYTHIA 'final-states' in $\left|\eta_{\text{particle}}\right| \leq 0.9$ and all final-states photons in the EMCal acceptance except the trigger photon and the second decay photon. It can be seen that the prompt photon and the jet are back-to-back correlated in φ and that a decay photon originating from a π^0 with high transverse momentum ($p_T \gtrsim 17$ GeV/c) is surrounded by a jet.

The distance R of a particle to the trigger photon is defined as

$$R = \sqrt{(\eta_{\text{particle}} - \eta_{\text{trigger }\gamma})^2 + (\varphi_{\text{particle}} - \varphi_{\text{trigger }\gamma})^2}. \tag{6.1}$$

The particles inside the cone, which are called 'associated particles', fulfill the following conditions:

- They are charged particles or photons.
- The distance to the trigger photon R is less than R_{max}.
- Their transverse momentum is greater than a certain threshold, $p_T^{\text{particle}} \geq p_T^{\text{threshold}}$.
- They are PYTHIA 'final-states'[2].
- The charged particles have $|\eta_{\text{ch}}| \leq 0.9$ and $0 \leq \varphi_{\text{ch}} \leq 2\pi$, the photons have $|\eta_\gamma| \leq 0.7$ and $79° \leq \varphi_\gamma \leq 191°$ (EMCal acceptance).
- The photon is not the trigger photon.

[2]PYTHIA final-states are particles which are not decayed in the PYTHIA simulation.

- The photon is not the second decay photon originating from the same π^0 as the trigger photon[3].

In this thesis, the isolation cut is done by investigating the number of particles inside the cone. In case no particle with $p_T^{\text{particle}} \geq p_T^{\text{threshold}}$ is inside the cone with radius R_{max}, the trigger photon is identified as a prompt photon, otherwise it is identified as a decay photon. The isolation cut can be done with a different cut value and with other variables, like 'sum of p_T of all particles inside the cone'. For the neural network, different variables are tested.
Due to the fact that photons are only taken into account if they are in the EMCal acceptance, the performance of the isolation cut and the neural network will depend on the azimuthal angle φ in this section. This dependence is not investigated, since using an isolation cut and a neural network to identify prompt photons is a first try in this section.
In figure 6.2, the p_T spectrum of the simulated decay photons and prompt photons, respectively, is plotted.

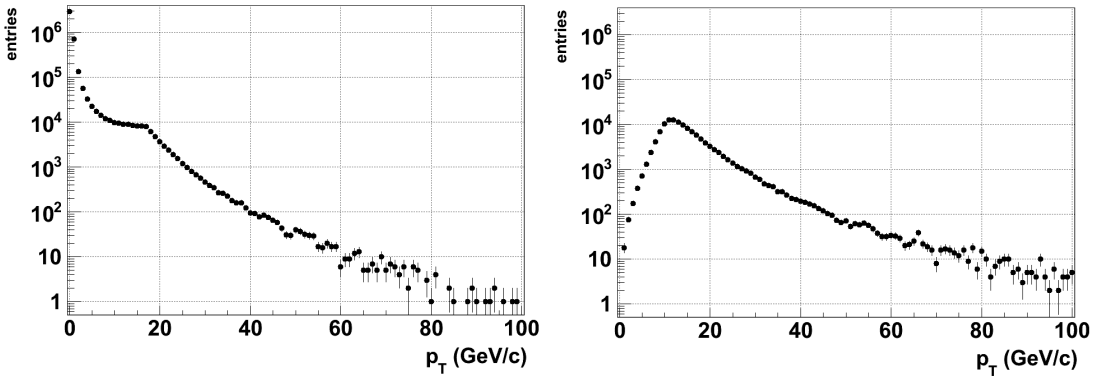


Figure 6.2.: Left: p_T spectrum of all simulated photons from π^0s in $|\eta| \leq 0.5$. Right: p_T spectrum of all simulated prompt photons in $|\eta| \leq 0.5$.

In the left plot, it can be seen that the number of decay photons is almost uniformly distributed for 10 GeV/c $\leq p_T(\gamma) \leq$ 17 GeV/c. The uniform distribution of decay photons occurs because in every event the transverse momentum of at least one π^0 has to be greater than 17 GeV/c and because the distribution of the asymmetry of the decay photon energies is a flat distribution if both photons originate from the same π^0 (see

[3]As mentioned in section 2.5, the second decay photon is difficult to separate.

[KB04]). In the right plot, it can be seen that a maximum occurs for the p_T spectrum around 10 GeV/c. The reason for this is that the p_T-hard of the collision of the γ-jet events is set to p_T-hard $\geq$ 10 GeV/c in a single p_T-hard bin. Both distributions follow a power law and are realistic for $p_T \geq$ 17 Gev/c.
Above 17 Gev/c, 38020 prompt photons and 38880 decay photons exist which can be used to train, test, and validate the neural network.

6.1.2. Results of the Isolation Cut

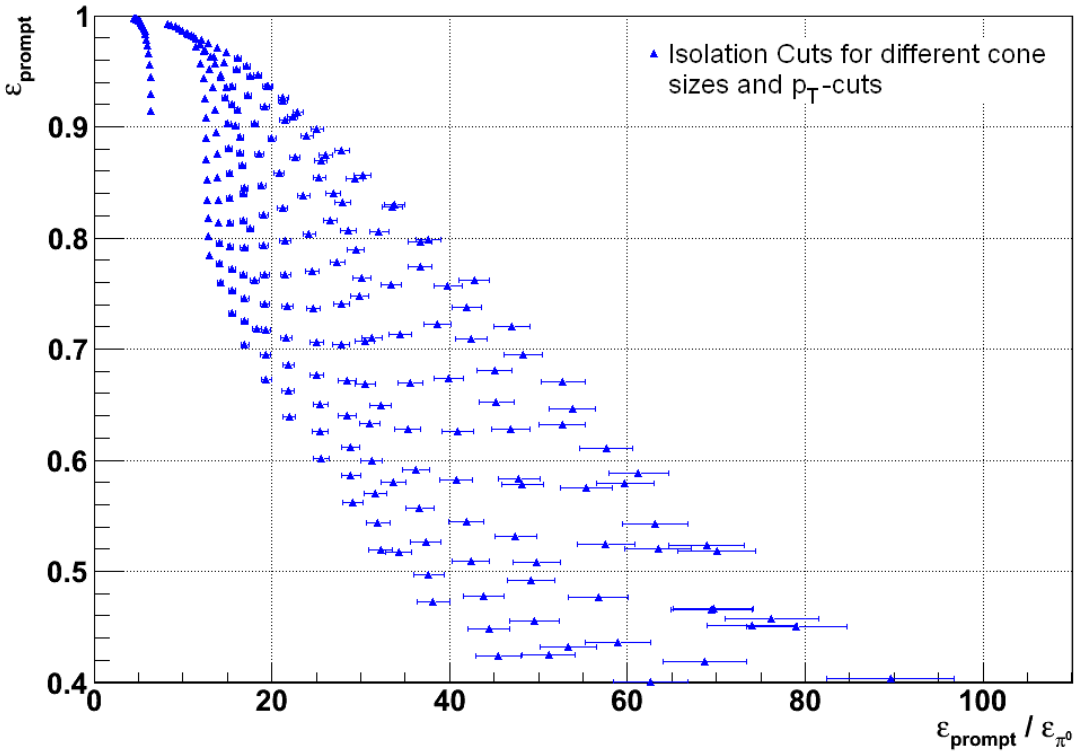


Figure 6.3.: Performance plot of the investigated isolation cuts. The signal efficiency is plotted against the improvement of the signal to background ratio. Every triangle corresponds to a certain cone size and a certain $p_T^{\text{threshold}}$.

In figure 6.3, the performance of the investigated isolation cuts is plotted. Different cone sizes and different p_T-cuts ($p_T^{\text{threshold}}$) are tested. The cone size is varied between $R_{\text{max}} = 0.1$ and $R_{\text{max}} = 0.8$ in steps of 0.1. In addition to these cones, a very large isolation region of the trigger photon is tested. For this region, the distance in φ has to be smaller than $\frac{\pi}{2}$ ($|\Delta\varphi| = \left|\varphi_{\text{particle}} - \varphi_{\text{trigger }\gamma}\right| \leq \frac{\pi}{2}$), and the distance in η has to be smaller than 1.4 ($|\Delta\eta| = \left|\eta_{\text{particle}} - \eta_{\text{trigger }\gamma}\right| \leq 1.4$). The $p_T^{\text{threshold}}$ of

the associated particles is simultaneously varied between $p_T^{\text{threshold}} = 0.0$ GeV/c and $p_T^{\text{threshold}} = 2.0$ GeV/c in steps of 0.1 GeV/c. All combinations of the cone size R_{max} and $p_T^{\text{threshold}}$ are tested for the isolation cut. Every triangle in figure 6.3 corresponds to a certain radius and a certain $p_T^{\text{threshold}}$.
The prompt photon efficiency ϵ_{prompt} is defined by

$$\epsilon_{\text{prompt}} = \frac{\text{number of prompt photons after the cut}}{\text{number of prompt photons before the cut}} \tag{6.2}$$

and describes the fraction of correctly identified prompt photons. The decay photon efficiency ϵ_{π^0} is defined by

$$\epsilon_{\pi^0} = \frac{\text{number of decay photons which are identified as prompt photons, after the cut}}{\text{number of decay photons before the cut}} \tag{6.3}$$

and describes the fraction of misidentified decay photons. Starting from $\mathrm{n}_{\text{prompt}}$ prompt photons and n_{π^0} decay photons before the cut, the ratio of signal to background before the cut v_1 is given by

$$v_1 = \frac{\mathrm{n}_{\text{prompt}}}{\mathrm{n}_{\pi^0}}. \tag{6.4}$$

The ratio of signal to background after the cut v_2 is given by

$$v_2 = \frac{\mathrm{n}_{\text{prompt}} \cdot \epsilon_{\text{prompt}}}{\mathrm{n}_{\pi^0} \cdot \epsilon_{\pi^0}}. \tag{6.5}$$

The improvement of the ratio of signal to background is given by

$$\frac{v_2}{v_1} = \frac{\epsilon_{\text{prompt}}}{\epsilon_{\pi^0}}, \tag{6.6}$$

which is plotted on the x-axis of figure 6.3. It can be seen that for a prompt photon efficiency of e.g. 80%, an improvement of the ratio of signal to background up to 40 is reached by using the isolation cut.
The uncertainties of ϵ_{prompt} and ϵ_{π^0} are determined by assuming a binomial distribution

$$\sigma_{\epsilon_{\text{prompt}}} = \sqrt{\frac{1}{\mathrm{n}_{\text{prompt}}} \epsilon_{\text{prompt}}(1 - \epsilon_{\text{prompt}})},$$

$$\sigma_{\epsilon_{\pi^0}} = \sqrt{\frac{1}{\mathrm{n}_{\pi^0}} \epsilon_{\pi^0}(1 - \epsilon_{\pi^0})}. \tag{6.7}$$

The uncertainty of the improvement of the ratio of signal to background is calculated using the Gaussian law of propagation of uncertainty

$$\frac{\epsilon_{\text{prompt}}}{\epsilon_{\pi^0}} =: G,$$

$$\Delta G = \sqrt{(\frac{\partial G}{\partial \epsilon_{\text{prompt}}} \cdot \sigma_{\epsilon_{\text{prompt}}})^2 + (\frac{\partial G}{\partial \epsilon_{\pi^0}} \cdot \sigma_{\epsilon_{\pi^0}})^2}$$
$$= \sqrt{\frac{1}{\epsilon_{\pi^0}^2}\sigma_{\epsilon_{\text{prompt}}}^2 + \frac{\epsilon_{\text{prompt}}^2}{\epsilon_{\pi^0}^4}\sigma_{\epsilon_{\pi^0}}^2}. \quad (6.8)$$

In figure 6.3, it can be seen that the uncertainty of the improvement of the ratio of signal to background, ΔG, is much larger than the uncertainty of the signal efficiency, $\sigma_{\epsilon_{\text{prompt}}}$, although $\sigma_{\epsilon_{\pi^0}}$ is of the same order of magnitude as $\sigma_{\epsilon_{\text{prompt}}}$. The reason for this is that ϵ_{π^0}, which is very small, is in the denominator in equation 6.8.

6.1.3. Results Using a Neural Network

As mentioned in section 4.5, the network will automatically calculate separation planes in the high-dimensional input space and the main challenge is to find suitable input variables to disjoin the set of background and signal in the input space. For the separation performance of the neural network, some possible input variables are tested:
The network has to distinguish between particles forming a jet and particles forming the underlying event. In this context, the underlying event is defined as everything except the particles originating from the hard scattering partons at the beginning of the collision. This hard scattering results in the production of two jets or in the production of a prompt photon and a jet. The underlying event consists of initial and final-state radiation and of remnants of the two beams (compare figure 2.7 in section 2.4 and compare section 2.5).
A jet consists of a lot of high p_T particles being close together, whereas the underlying event consists of a few low p_T particles homogeneously distributed in a large region. Therefore, e.g. the overall p_T (and the number) of the particles nearby a decay photon should be greater than the overall p_T (and the number) of the particles nearby a prompt photon. The overall distance, $\sum(R)$, of the particles should be similar for prompt photons and decay photons. On the one hand the particles are close to the decay photon but on the other hand a lot of particles exist, resulting in a medium overall distance. The distance of the associated particles to a prompt photon should be large, but only a few particles exist, also resulting in a medium overall distance. Therefore the variable $\sum(R)$ of the particles should be similar for both photon species. In contrast, the variable mean distance (= $\sum(R)$ of the particles / Number of particles) should be larger for prompt photons than for decay photons. With these considerations in mind, the following variables are tested:

For particles with $p_T \geq p_T^{\text{threshold}}$ inside the cone:

- $\sum(p_T)$ of the particles
- Number of particles
- Mean distance of the particles
- $\sum(p_T/R)$ of the particles
  p_T of a particles should be large, whereas R of a particles should be small for decay photons. Therefore $\sum(p_T/R)$ should be large for decay photons and small for prompt photons.
- $\sum(p_T \times R)$ of the particles
  This variable should be similar for decay photons and prompt photons. Nevertheless, it is tested for completeness.
- $\sum(E_T)$ of photons
  It is possible that photons carry a larger (smaller) energy fraction of a jet than of the underlying event.
- $\sum(E_T)$ of the particles
- Distance to the particle with maximum p_T
  The particle with maximum p_T should be closer to a decay photon than to a prompt photon. In case no particle with $p_T \geq p_T^{\text{threshold}}$ is inside the cone, the variable is set to 4.
- p_T of the particle with maximum p_T
  The p_T of the particle with maximum p_T should be larger for decay photons than for prompt photons. In case no particle with $p_T \geq p_T^{\text{threshold}}$ is inside the cone, the variable is set to 0.

And in addition:

- Distance R to the closest particle with $p_T \geq p_T^{\text{threshold}}$ in $|\Delta\varphi| \leq \frac{\pi}{2}$ and $|\Delta\eta| \leq 1.4$
  The closest particle to the trigger photon should be closer for decay photons than for prompt photons. In case no particle with $p_T \geq p_T^{\text{threshold}}$ is inside the surrounding, the variable is set to 4.
- p_T of the closest particle $p_T \geq p_T^{\text{threshold}}$ in $|\Delta\varphi| \leq \frac{\pi}{2}$ and $|\Delta\eta| \leq 1.4$
  The p_T of the closest particle should be larger for decay photons than for prompt photons.

- p_T of the trigger photon

  It is possible that correlations exist between p_T of the trigger photon and the other variables. For example, $\sum(p_T)$ of the particles should be large for large p_T of the trigger photon, since the p_T of the original parton is large.

All these variables are tested for different cone sizes and different p_T-cuts. In the same way as for the isolation cut in section 6.1.2, the cone size is varied between $R_{\max} = 0.1$ and $R_{\max} = 0.8$ in steps of 0.1 and in addition to these cones, a very large isolation region of the trigger photon ($|\Delta\varphi| \leq \frac{\pi}{2}$, $|\Delta\eta| \leq 1.4$) is tested. Also the $p_T^{\text{threshold}}$ of the associated particles is simultaneously varied between $p_T^{\text{threshold}} = 0.0$ GeV/c and $p_T^{\text{threshold}} = 2.0$ GeV/c in steps of 0.1 GeV/c. The cone size and the $p_T^{\text{threshold}}$ can be independently changed and the neural network can have input variables which are calculated for different cone sizes and different $p_T^{\text{threshold}}$.
Due to the nine different cone sizes and twenty-one p_T-cuts, almost two thousand different input variables are possible for the neural network. The more input neurons, the more weights between the neurons exist, which have to be calculated with equation 4.8 and equation 4.9, respectively for every pattern. This results in a long duration of the training; therefore, the best input variables have to be identified.
To determine which input variables result in a good performance of the neural network, a network with one input variable is trained at first. All possible variables are tested as input and their performance is calculated. Only the best variable, var_1, is used further as input variable one. The best variable is identified by adding up the improvements of the ratio of signal to background, $G(\epsilon_{\text{prompt}})$, for every signal efficiency ϵ_{prompt}. This is done for signal efficiencies between 1.00 and 0.70 in steps of 0.01 and for signal efficiencies between 1.00 and 0.40 in steps of 0.01.

$$G_A = \sum_{\epsilon_{\text{prompt}}=0.70}^{1.00} G(\epsilon_{\text{prompt}}) = G(0.70) + G(0.71) + ... + G(0.99) + G(1.00). \quad (6.9)$$

$$G_B = \sum_{\epsilon_{\text{prompt}}=0.40}^{1.00} G(\epsilon_{\text{prompt}}) = G(0.40) + G(0.41) + ... + G(0.99) + G(1.00). \quad (6.10)$$

G_A or G_B has to be larger for var_1 than for any other variable to call var_1 'best variable'. The choice whether G_A or G_B is used as well as the upper and lower limits (0.40, 0.70, and 1.00) is not determined. In this thesis, G_A or G_B are chosen in such a way that $G(\epsilon_{\text{prompt}})$ is large between 1.00 and 0.70 as well as between 0.70 and 0.40. After identifying var_1, a network with two input variables is trained. Besides var_1, all other possible variables are tested and only the best pair of variables, (var_1, var_2),

is used in the following. In this way the best input variables are determined. The performance of the neural network should increase for more input neurons and saturate at a certain number. In this section, four input variables are used for the network, since its performance is not significantly improving for more input neurons. The four best input variables are:

- var_1: $\sum(p_T/R)$ of the particles in a cone with radius $R_{\text{max}} = 0.4$, $p_T^{\text{particle}} \geq 1.0$ GeV/c
- var_2: $\sum(p_T/R)$, $R_{\text{max}} = 0.6$, $p_T^{\text{particle}} \geq 0.0$ GeV/c
- var_3: Mean distance, in $|\Delta\varphi| \leq \frac{\pi}{2}$ and $|\Delta\eta| \leq 1.4$, $p_T^{\text{particle}} \geq 0.4$ GeV/c
- var_4: p_T of the trigger photon

In figure 6.4, the structure of the used neural network is plotted. Besides four input neurons (left side) for the four input variables, two hidden layers and one output layer (right side) can be seen.

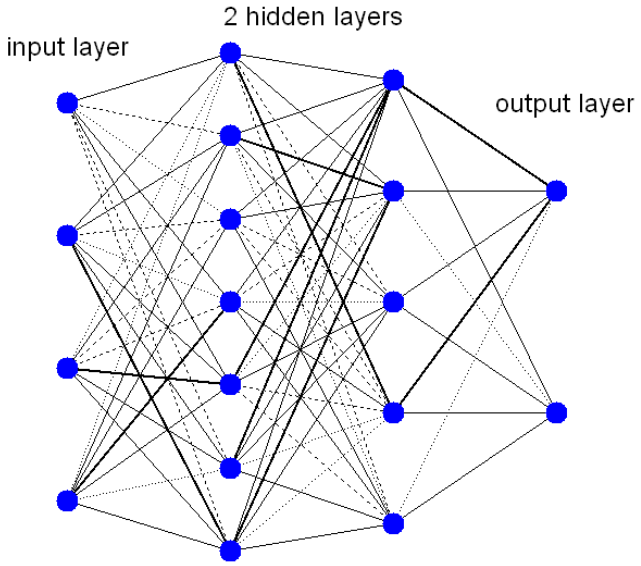


Figure 6.4.: Structure of the used neural network. The blue dots represent the neurons, the lines between the neurons represent the synapses. A thick line means a high absolute value of the weight. The network consists of one input layer (left side) with four input neurons, two hidden layers with seven and five neurons, respectively, and one output layer with two output neurons.

The output layer consists of two output neurons, since signal (= prompt photons) and background (= decay photons) form two classes of patterns. The network has two hidden layers with seven and five neurons, respectively. The lines represent the synapses and their thickness the absolute value of the weight. In this thesis all neural networks consist of one input layer, two hidden layers and one output layer.
In figure 6.5, the output distribution of the first output neuron of the neural network is plotted. As mentioned in section 4.3, the target output of signal is $(t_1,t_2) = (1,0)$ and the target output of background is $(t_1,t_2) = (0,1)$. Therefore signal patterns should ideally create a 1 and background patterns a 0 in this plot.

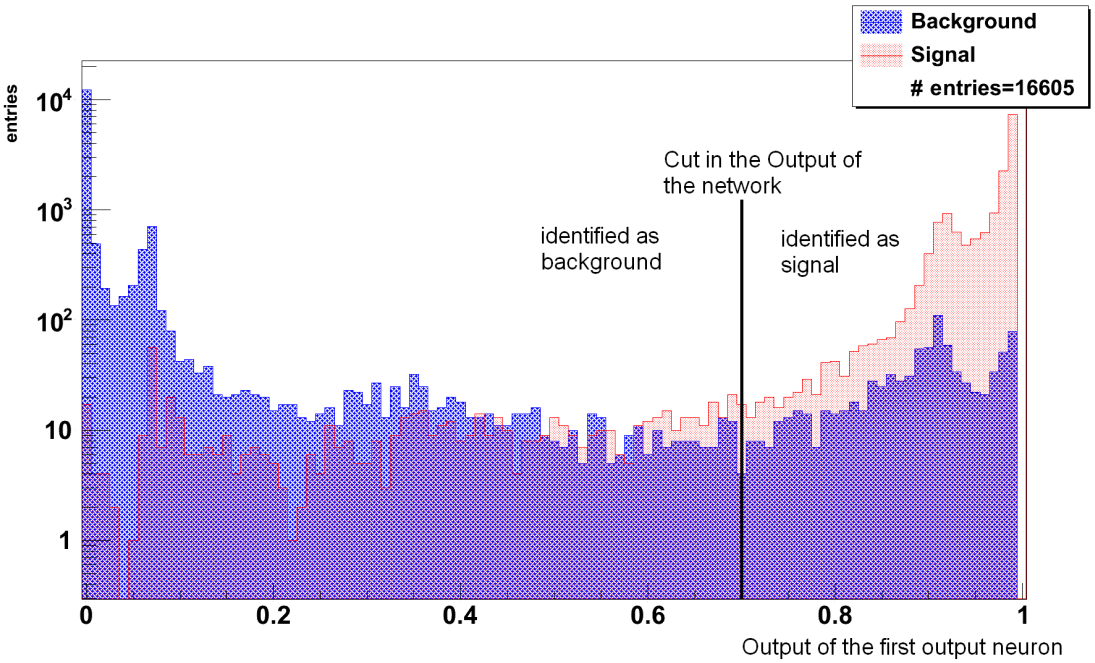


Figure 6.5.: Output distribution of the neural network. Since the outputs of the output neurons of the network have to sum up to 1, only the distribution of the first output neuron is plotted.

To calculate a certain prompt photon efficiency and a certain decay photon efficiency, it is necessary to distinguish between patterns identified as background and patterns identified as signal. This can be done by defining a cut in the output of the network. In figure 6.5, this cut is set to 0.7 in the output distribution. All patterns with an output greater than 0.7 are identified as signal, all patterns with an output less than 0.7 are identified as background. The efficiency of signal and background, and as a consequence the improvement of the ratio of signal to background, can be changed by changing the

cut in the output of the network.

In figure 6.6, the performance of the neural network together with the investigated isolation cuts are plotted. The results of the neural network are shown as red stars. Every star corresponds to a certain cut in the output of the neural network. It can be seen that the improvement of the signal to background ratio for the same prompt photon efficiency is higher for the neural network than for any isolation cut.

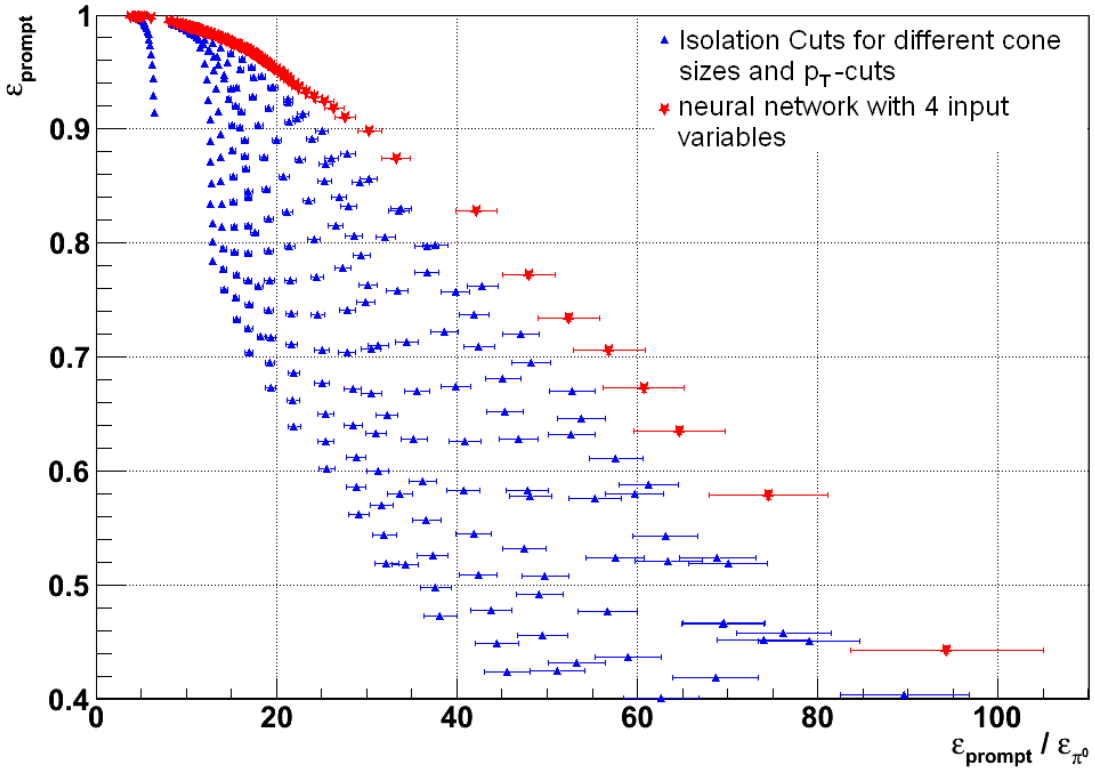


Figure 6.6.: Performance plot of the investigated isolation cuts and the neural network. The signal efficiency is plotted against the improvement of the ratio of signal to background.

6.2. Photon Separation in a PYTHIA and GEANT Simulation

In this section, the separation of prompt photons and decay photons is presented. The data used are created by a PYTHIA simulation with a GEANT detector simulation. For the settings of the simulation, see section 5.3.

The identification of prompt photons and decay photons is done in the same way as described in section 6.1.1. Prompt photons are all photons without a mother particle,

decay photons are all photons originating from a π^0. In the experiment, the trigger photon can be measured with EMCal/PHOS or via conversion[4]. If the photon from the π^0 is measured via conversion, the second decay photon will probably not be measured, since the probability of conversion is very small ($\approx 8\%$) and only the transverse momentum of the photon is known, but not the p_T of the π^0. If the photon from the π^0 is measured by PHOS or EMCal, both decay photons will be measured. The single signals are difficult to separate and only the p_T of the π^0 is known, but not the transverse momenta of the single photons. In principle, it is possible that the first decay photon is measured by PHOS or EMCal and the second decay photon converts before it reaches EMCal/PHOS. In this case, only the transverse momentum of the photon is known. In reality, this case cannot be distinguished from the case that both photons are measured with EMCal/PHOS. Therefore, this case is not taken into account separately.
One neural network is trained using the p_T of the trigger photon (for photons identified via conversion) and a second neural network is trained using the p_T of the π^0 for background and the p_T of the prompt photon for signal (for photons identified using EMCal/PHOS). A third neural network is trained using neither the p_T of the photon nor the p_T of the π^0. This neural network is used to test the possibility of using 'random cones' instead of simulated prompt photons to get the patterns for signal (compare chapter 8).
For the first and third network, the transverse momenta of the photons have to be greater than 10 GeV/c. For the second network, the p_T of the π^0 and the p_T of the prompt photon have to be greater than 10 GeV/c. The trigger particles must have $|\eta| \leq 0.5$ and for the prompt photons, the azimuthal angle φ has to be in the EMCal acceptance.
The associated particles are all reconstructed tracks, except the daughter particles of the trigger photon. The trigger photons themselves and the second decay photon are excluded, since photons carry no electric charge and are therefore not part of the reconstructed tracks.
Independently from the measurement of the trigger photon, the second decay photon can be measured with EMCal/PHOS, it can convert, or it is not measured. Only in the case that the second decay photon converts, it will have daughters (electron and positron), which contribute to the associated particles.
The isolation cut is done in the same way as in section 6.1.1 and 6.1.2. Only the

[4] A photon can create an electron-positron pair by pair production. The charged particles can be measured with the TPC and reconstructed to a photon. See section 7 for details.

number of particles inside the cone is investigated for the isolation cut for different cone sizes ($R_{\max} = 0.1, 0.2, ..., 0.8$ and $|\Delta\varphi| \leq \frac{\pi}{2}$, $|\Delta\eta| \leq 1.4$) and different p_T-cuts ($p_T^{\text{threshold}} = 0.0$ GeV/c, 0.1 GeV/c, ..., 2.0 GeV/c) in all combinations. In case no particle with $p_T^{\text{particle}} \geq p_T^{\text{threshold}}$ is inside the cone with radius $R_{\max}$, the trigger photon is identified as a prompt photon, otherwise it is identified as a decay photon.

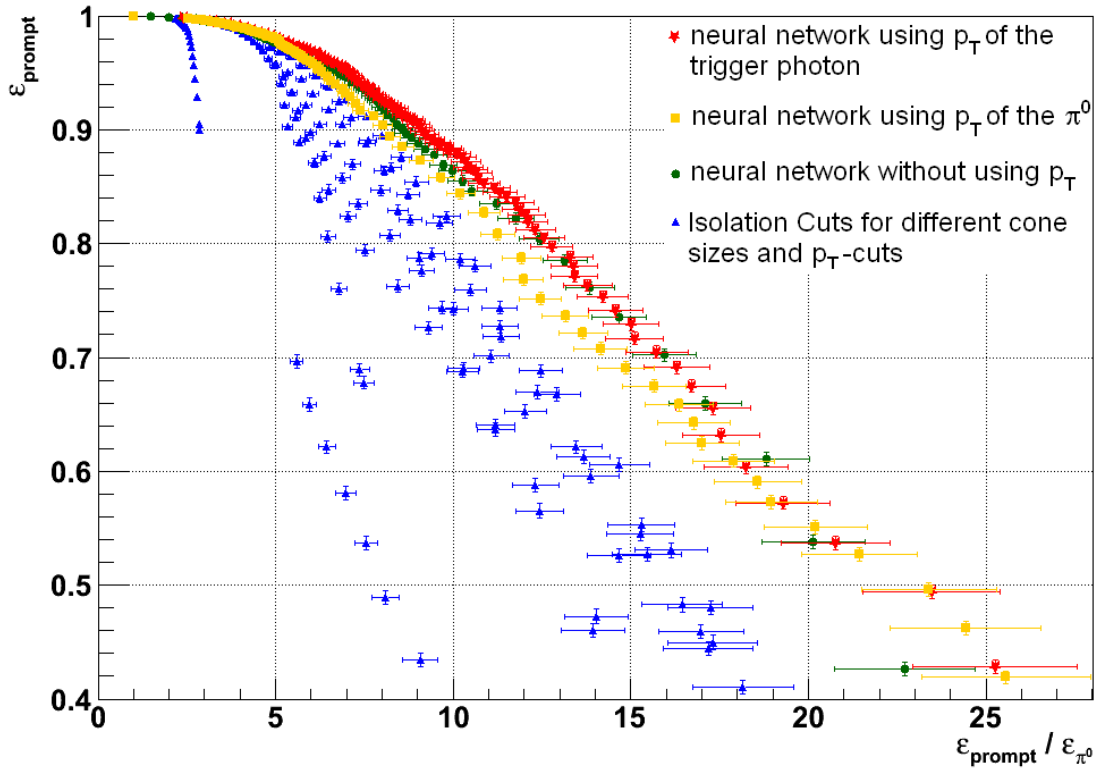


Figure 6.7.: Performance plot of the investigated isolation cuts and three neural networks. The signal efficiency is plotted against the improvement of the ratio of signal to background.

The separation using a neural network is done in a similar way as described in section 6.1.3. One difference in this analysis is that the input variable 'sum of E_T of photons' cannot be used, since photons are not part of the associated particles. In figure 6.7, the performance of the three neural networks together with the investigated isolation cuts is plotted. The results of the neural networks are shown as red stars (conversion), yellow squares (EMCal/PHOS), and green circles (without p_T), respectively. It can be seen that the improvement of the signal to background ratio for the same prompt photon efficiency is higher for the neural networks than for any isolation cut.
The five best input variables determined by the optimization algorithm, which is ex-

plained in section 6.1.3, are:
For the neural network using the p_T of the trigger photon:

- $\sum(p_T/R)$, $R_{\text{max}} = 0.5$, $p_T^{\text{particle}} \geq 0.0$ GeV/c
- p_T of the trigger photon
- $\sum p_T$, $R_{\text{max}} = 0.6$, $p_T^{\text{particle}} \geq 0.0$ GeV/c
- Distance to the particle with maximum p_T, $R_{\text{max}} = 0.6$, $p_T^{\text{particle}} \geq 0.3$ GeV/c
- Number of particles in $|\Delta\varphi| \leq \frac{\pi}{2}$ and $|\Delta\eta| \leq 1.4$, $p_T^{\text{particle}} \geq 0.7$ GeV/c

For the neural network using the p_T of the π^0 for background and the p_T of the prompt photon for signal:

- $\sum(p_T/R)$, $R_{\text{max}} = 0.7$, $p_T^{\text{particle}} \geq 0.1$ GeV/c
- p_T of the π^0 and p_T of the prompt photon, respectively
- Distance to the closest particle in $|\Delta\varphi| \leq \frac{\pi}{2}$ and $|\Delta\eta| \leq 1.4$, $p_T^{\text{particle}} \geq 1.2$ GeV/c
- p_T of the particle with maximum p_T, $R_{\text{max}} = 0.3$, $p_T^{\text{particle}} \geq 0.3$ GeV/c
- Mean distance of the particles in $|\Delta\varphi| \leq \frac{\pi}{2}$ and $|\Delta\eta| \leq 1.4$, $p_T^{\text{particle}} \geq 0.8$ GeV/c

For the neural network without using p_T:

- $\sum(p_T/R)$, $R_{\text{max}} = 0.5$, $p_T^{\text{particle}} \geq 0.0$ GeV/c
- Mean distance of the particles, $R_{\text{max}} = 0.5$, $p_T^{\text{particle}} \geq 1.1$ GeV/c
- Number of particles, $R_{\text{max}} = 0.7$, $p_T^{\text{particle}} \geq 0.1$ GeV/c
- Number of particles in $|\Delta\varphi| \leq \frac{\pi}{2}$ and $|\Delta\eta| \leq 1.4$, $p_T^{\text{particle}} \geq 0.6$ GeV/c
- Number of particles in $|\Delta\varphi| \leq \frac{\pi}{2}$ and $|\Delta\eta| \leq 1.4$, $p_T^{\text{particle}} \geq 0.0$ GeV/c

In figure 6.7, the improvement of the signal to background ratio is lower compared to section 6.1.3. The reasons for this is that in this section photons are not part of the associated particles, detector effects are taken into account, and the minimum p_T of the trigger photon is changed from 17 GeV/c to 10 GeV/c. Changing the minimum p_T of the trigger photon to a lower value is probably the main reason for the lower performance.

The variables $\sum(p_T/R)$, p_T of the trigger photon, mean distance, and number of particles are used more than once in this section and in section 6.1.3. As described in section 6.1.3, it was expected that these variables are different for decay photons and prompt photons. However, the choice of the variables, the cone sizes, and $p_T^{\text{threshold}}$ is not determined and depends on the choice of G_A or G_B and the upper and lower limits in the optimization algorithm.
In figure 6.8, the performance of the neural network without using p_T is shown for different number of input neurons. The input neurons are determined using the optimization algorithm, which is explained in section 6.1.3.

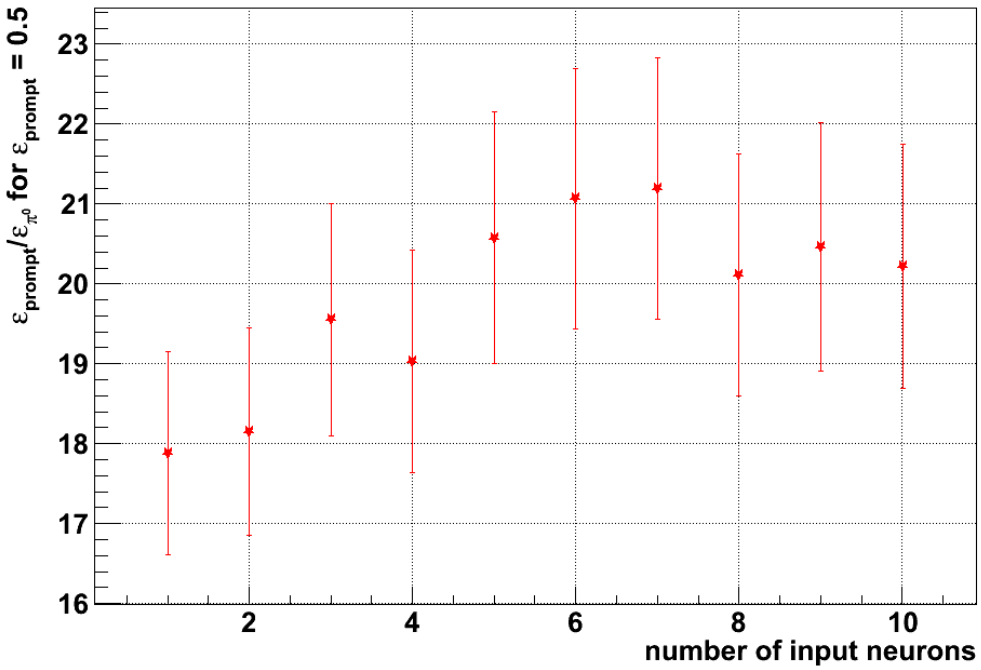


Figure 6.8.: Improvement of the signal to background ratio for $\epsilon_{\text{prompt}} = 0.5$ versus different number of input neurons for the neural network without using p_T. It can be seen that the performance of the networks converges after an increasing.

It can be seen that the improvement of the signal to background ratio increases for a larger number of input neurons, since more information is available for the network. For five or six input neurons, the performance saturates and a larger number of input neurons will not lead to a larger improvement of the signal to background ratio.

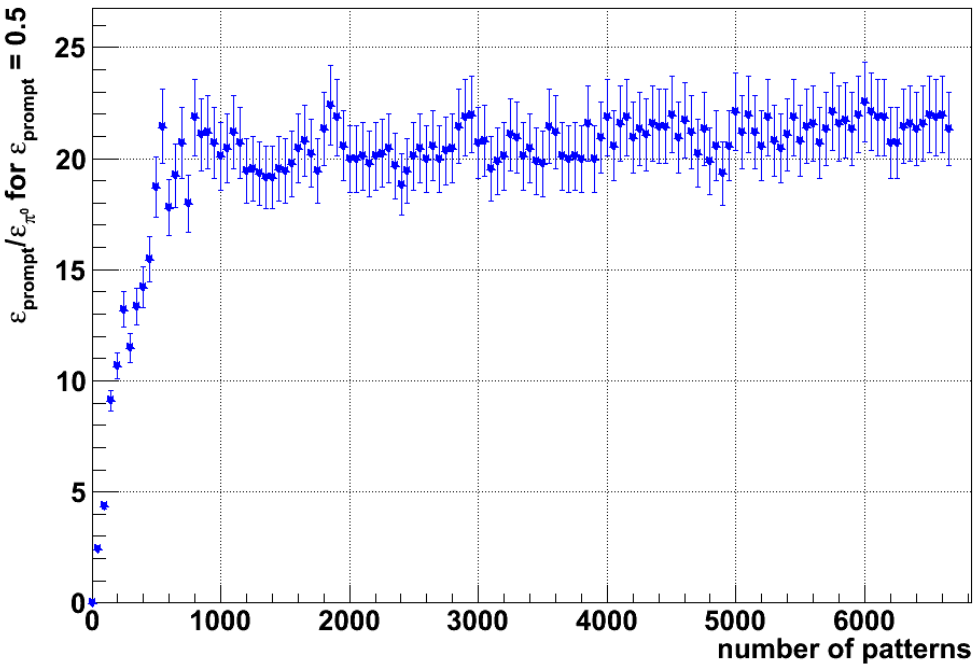


Figure 6.9.: Improvement of the signal to background ratio for $\epsilon_{\text{prompt}} = 0.5$ versus number of pattern used for training and testing. The performance of the neural network without using p_T is investigated. It can be seen that the performance increases slowly for more than 800 patterns.

In figure 6.9, the performance of the neural network without using p_T is shown for different number of patterns. Half of the patterns are used to train and half are used to test the network. It is validated with around 7000 patterns, independent from the number of patterns for training and testing. It can be seen that the performance increases fast for a small number of patterns. For more than 800 patterns the performance increases only slowly and will perhaps saturate at a certain number.

7. Identification of Photons from π^0s

In chapter 6, the separation of prompt photons and decay photons in Monte-Carlo simulations was presented. In the analyses of these simulations, the prompt photons and the decay photons were identified by using the PDG code of the particles. In the experiment, this information is not available and a different possibility to identify the photons is needed. One possibility to identify decay photons will be presented in this chapter. After identifying the decay photons, their input variables for the neural network can be calculated not only in simulated data, but also in measured data.

7.1. Secondary Vertex Algorithm - V0s

Using the ITS and the TPC, charged particles can be measured and identified, their tracks can be reconstructed and propagated back to the vertex. The direct measurement of neutral particles can be done by the two calorimeters EMCal and PHOS, but these detectors are not used in this analysis. Some sorts of neutral particles (mainly K^0_S, Λ, and $\overline{\Lambda}$) decay into charged particles and can be detected by the ITS and the TPC by measuring the decay products. Photons, which are also neutral particles, can create an electron-positron pair in matter by pair production. This pair production is not a decay of a particle, it is called 'conversion' of the photon. Due to conversion, photons can be measured indirectly by the ITS and the TPC.
The reconstruction of the neutral particle by measuring its charged daughters is done by the 'secondary vertex algorithm'. The outputs of this algorithm are 'V0s', which represent the neutral particle candidates. For the reconstruction of V0s, an 'On-the-fly-Finder' and an 'Offline-Finder' exist. The Offline-Finder starts to reconstruct V0s after the reconstruction of all tracks, whereas the On-the-fly-Finder starts during the reconstruction of the tracks. Both finders will partly find the same V0s, therefore, only one finder can be used in an analysis. In this analysis, only the On-the-fly-Finder is used. In the following, the Offline-Finder will be shortly explained, since this finder is easier to understand [ALI06]. The On-the-fly-Finder works in a similar way, but can use more track-parameters, because it starts during the reconstruction of the tracks.

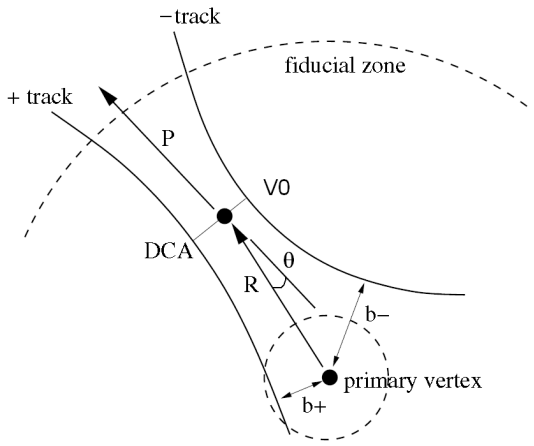


Figure 7.1.: Geometrical selections for secondary vertex reconstruction [ALI06]. The impact parameter b, the DCA, the position, and the pointing angle must fulfill certain conditions to accept the V0.

The charged particles will create tracks in the ITS and in the TPC. These tracks are propagated back to the primary vertex and the impact parameter b is calculated. The impact parameter is the closest distance of the track to the primary vertex (compare figure 7.1). Secondary tracks, that means tracks which do not originate from the primary vertex but e.g. from a particle decay, will have larger impact parameters than primary tracks. Only tracks with an impact parameter greater than a certain value will be taken into account for the creation of V0s. All selected tracks with opposite charges[1] are combined and the distance of closest approach (DCA) of the two tracks is calculated. Tracks originating from the same neutral particle will have a small DCA. Therefore, only V0s with a DCA smaller than a certain value are accepted.
The pointing angle θ is the angle between the reconstructed momentum of the neutral particle and the vector to the position of the DCA. For small θ, the momentum of the neutral particle points well to the primary vertex. Therefore, the pointing angle has to be smaller than a certain value to accept the combined tracks as a V0.
Beside the impact parameters, DCA and pointing angle, the momentum, the invariant mass, η, φ, and other parameters of the V0 can be calculated.

[1]The charge of a track can be detected by the orientation of the curvature of the track. The curve is created due to the Lorentz force on a moving, charged particle in a magnetic field.

7.2. Identification of Photons from V0s

As described in section 7.1, converted photons can be measured by using the secondary vertex algorithm. Besides converted photons, the decay of K^0_S, Λ, and $\overline{\Lambda}$ can result in a reconstruction of a V0. Therefore, a V0 can represent one of these particles and a method is needed to identify the photons from V0s. Moreover, it is possible that two tracks are combined to a V0 which do not originate from the same particle. This kind of V0 is called 'fake-V0'. For the identification of photons, V0s from converted photons are defined as signal, whereas fake-V0s and V0s from K^0_S, Λ, and $\overline{\Lambda}$ are defined as background in this part of the analysis.

For signal and background, some variables exist in which they differ partially. The investigated variables are

- cosine of the pointing angle of the V0,
- position, p_T, and η of the V0,
- Ψ_{Pair} of the V0,
- invariant mass of the V0,
- χ^2/NDF of the V0,
- Armenteros variables α and q_T of the V0,
- DCA of the two combined tracks,
- TPC refit of the two combined tracks,
- number of clusters in the TPC of the two combined tracks,
- χ^2 per TPC cluster of the two combined tracks,
- found number of TPC clusters divided by the findable number of TPC clusters of the two combined tracks,
- kinks, charge, $\mathrm{d}E/\mathrm{d}x$, p_T, and η of the two combined tracks,
- number of contributors of the primary vertex of the event,
- the particle identification (PID) response of the TPC.

Ψ_{Pair} of a V0 is the angle between the plane perpendicular to the magnetic field and the plane which is defined by the V0-daughter momenta.
The invariant mass m_{inv} of a V0 can be calculated with ($c = 1$)

$$m_{inv} = m_1^2 + m_2^2 + 2E_1E_2 - 2\vec{p}_1\vec{p}_2, \tag{7.1}$$

where m_i, E_i, and $\vec{p}_i$ are the mass, the energy, and the momentum of the V0-daughter i (i=1,2). The mass and the energy of a particle are not measured. Therefore, it is necessary to assume the species of the V0-daughter particles. The neutral particle decays, which can create a V0 are

- $K_S^0 \rightarrow \pi^+\pi^-$,
- $\Lambda \rightarrow p\pi^-$,
- $\overline{\Lambda} \rightarrow \overline{p}\pi^+$,
- $\gamma \rightarrow e^-e^+$.

For the different neutral particle decays, the known masses of π^+, π^-, p, $\overline{p}$, e^-, and e^+, respectively, are used to calculate the invariant mass of the V0.
For the reconstruction of the decayed particle, a Kalman Filter can be used to fit the decayed particles [CBM07]. The quality of the fit is described by χ^2 and by the number of degrees of freedom (NDF). For $\chi^2/\text{NDF} \approx 1$, the fit matches the data well.
The Armenteros variables α and q_T of the V0 are defined by

$$\alpha \;=\; \frac{|\vec{p}_{L+}| - |\vec{p}_{L-}|}{|\vec{p}_{L+}| + |\vec{p}_{L-}|}, \tag{7.2}$$

where $\vec{p}_{L+}$ ($\vec{p}_{L-}$) is the longitudinal momentum of the positive (negative) V0-daughter and

$$q_T \;=\; |\vec{p}_+| \cdot \sin(\vartheta^+), \tag{7.3}$$

where $\vartheta^+ = \arccos(\vec{p}_+ \cdot \vec{p}/\,|\vec{p}_+| \cdot |\vec{p}|)$ is the angle between the momentum of the positive V0-daughter, $\vec{p}_+$, and the momentum of the V0, $\vec{p}$.
The track of a particle is reconstructed for each detector. Afterwards, the track is refitted to the next (previous) detector. The variable 'TPC refit' is the output of this algorithm and describes whether the refit of the track matches a track in a different detector.
If a particle crosses a sensitive element of a detector, it will create a signal. A cluster is a set of adjacent signals generated by the same particle. For a high number of clusters,

the track of a particle can be reconstructed more easily.
χ^2 per TPC cluster is the χ^2 of the track reconstruction divided by the number of clusters in the TPC, which are created by the particle.
The findable number of TPC clusters is equal to the number of crossings of the track with the pad rows of the TPC.
In figure 7.2, the origin of kinks of a track is shown. A charged particle decays into a charged particle and a neutral particle (e.g. $\pi^- \rightarrow \mu^- + \overline{\nu}_\mu$). Only the charged particles will be measured and they will look like a single track with a kink.

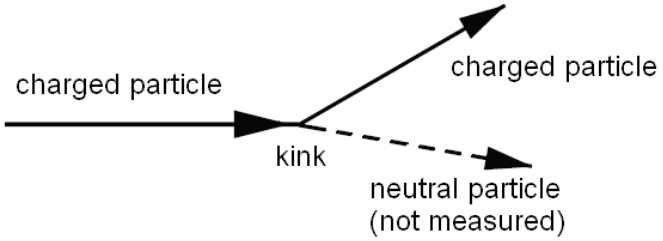


Figure 7.2.: Schematic view of the origin of kinks. A charged particle decays into a charged particle and a neutral particle. The neutral particle will not be measured, therefore the decay looks like a single track with a kink.

For the identification of photons from V0s, two existing methods and two neural networks are tested. In the ALICE analysis framework, the class `AliESDv0KineCuts` can be used to identify γs, K^0_Ss, Λs, and $\overline{\Lambda}$s from V0s and the function `NextV0` from the class `AliV0Reader` can be used to identify only γs from V0s. The class `AliESDv0KineCuts` uses only kinematic information of the V0 and no detector information about the identity of the V0-daughters, whereas the function `NextV0` is not subjected to these restrictions. For this analysis, minimum bias Monte-Carlo events are used at a center-of-mass energy of $\sqrt{s} = 7$ TeV and a magnetic field of -0.5 Tesla (production number LHC10f6a). The investigated V0s have to be found by the On-the-fly-Finder and their pseudorapidity must be in the acceptance of the central barrel, $|\eta| \leq 0.9$. There is no p_T-cut on the V0s and also no p_T-cut on the reconstructed tracks, which means that the p_T spectrum of the V0s is as it is generated in minimum bias Monte-Carlo events (compare figure 7.3).
The performance of the method `AliESDv0KineCuts` and the method `NextV0` are compared with the performance of a neural network, whereas the input variables of the neural network are the same variables of the method `AliESDv0KineCuts` and `NextV0`, respectively. The seventeen input variables of the neural network, which will be com-

pared with the method `AliESDv0KineCuts`, are listed in appendix A.7. Cuts[2] on the same variables are done in the method `AliESDv0KineCuts` to identify γs.
The twenty-six input variables of the neural network, which will be compared with the method `NextV0`, are also listed in appendix A.7. Cuts on the same variables are done in the method `NextV0`[3]. Additionally, for the method `NextV0` and the corresponding neural network, some other variables exist. Cuts are done on these variables for both `NextV0` and the corresponding neural network, since the value of the variables is almost always the same. Using these variables as input for the neural network would result in a longer duration of the training without significant improvement of the performance of the network. The variables are the kink-index of the positive and negative track, the charge of the first track times the charge of the second track, and the TPC refit of the positive and negative track.

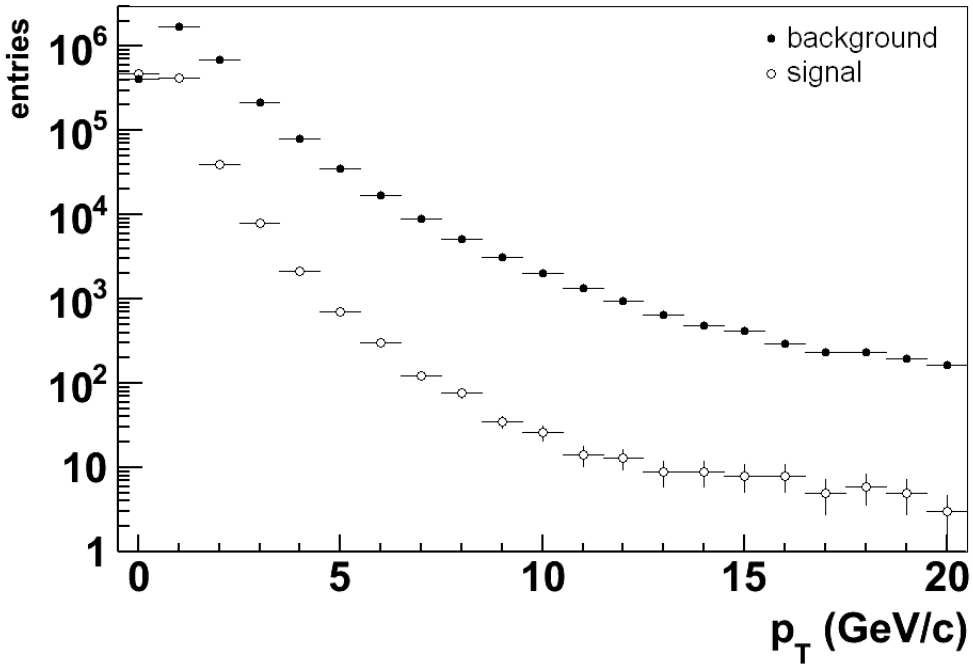


Figure 7.3.: p_T spectrum of all converted photons (signal) and p_T spectrum of all other V0s (background) created in a minimum bias Monte-Carlo simulation. It can be seen that more background V0s than photons exist, especially for high p_T. 8,787,519 events are used, production number LHC10f6a.

[2]The standard cuts of `AliESDv0KineCuts` are used here. See class `AliESDv0KineCuts` of the AliRoot version v5-01-Rev-08.
[3]Cut-Number 900366208010033211360000000090.

Variables which are used in the method `NextV0` but not in the class `AliESDv0KineCuts` are some detector information (TPC response and TPC $\mathrm{d}E/\mathrm{d}x$), event information (number of contributors of the primary vertex), and some global information of the V0 and the tracks (z-position, momenta, and η value). Variables which are used in the method `AliESDv0KineCuts` but not in the method `NextV0` are some kinematic information (cosine of the pointing angle, DCA, Ψ_{Pair}, Armenteros variable α, and invariant mass of the V0 for e^-/e^+) and some information about the quality of the tracks (found number of TPC clusters divided by the findable number of TPC clusters and χ^2 per TPC cluster). That means that both methods do not use all available information and both methods can be further improved.
In figure 7.4, the performance of the method `AliESDv0KineCuts` and the method `NextV0` compared with the performance of the associated neural network for identifying γs from V0s is plotted. The contamination of the signal, depending on the signal efficiencies, is shown for different p_T of the V0. The contamination of the signal is defined as

$$\text{contamination} = 1 - \frac{\text{number of real photons from V0s after the cut}}{\text{number of accepted V0s after the cut}}. \tag{7.4}$$

The accepted V0s after the cut are all V0s which are identified as photons.
The efficiency of the signal is defined as

$$\text{efficiency} = \frac{\text{number of real photons from V0s after the cut}}{\text{number of all real photons from V0s before the cut}}. \tag{7.5}$$

In figure 7.4 it can be seen that the contamination of the signal is much lower for the neural network with input from `AliESDv0KineCuts` (red stars) than for `AliESDv0KineCuts` (red triangle) itself. This is true for the same signal efficiency and also for higher efficiencies, for $p_T(\mathrm{V0}) \leq 5.5$ GeV/c (see appendix A.8 for the performance plots of all investigated p_T bins). Except for 1.5 GeV/c $\leq p_T(\mathrm{V0}) \leq$ 2.5 GeV/c, the neural network with input from `NextV0` (blue stars) also has a better performance than the corresponding method `NextV0` (blue triangle). For $p_T(\mathrm{V0}) \geq 5.5$ GeV/c, the statistical uncertainties are too big to compare the different methods.
In the lower plot of figure 7.4, some steps in the contamination can be seen. Only a small number of high p_T V0s are used for this plot and therefore the number of background V0s for an efficiency of e.g. 0.1 is very small. The contamination for this efficiency only consists of a few background V0s and a step in the contamination occurs every time one more background V0 belongs to the contamination.
The minimum contamination, which is reached with the neural networks, is plotted in

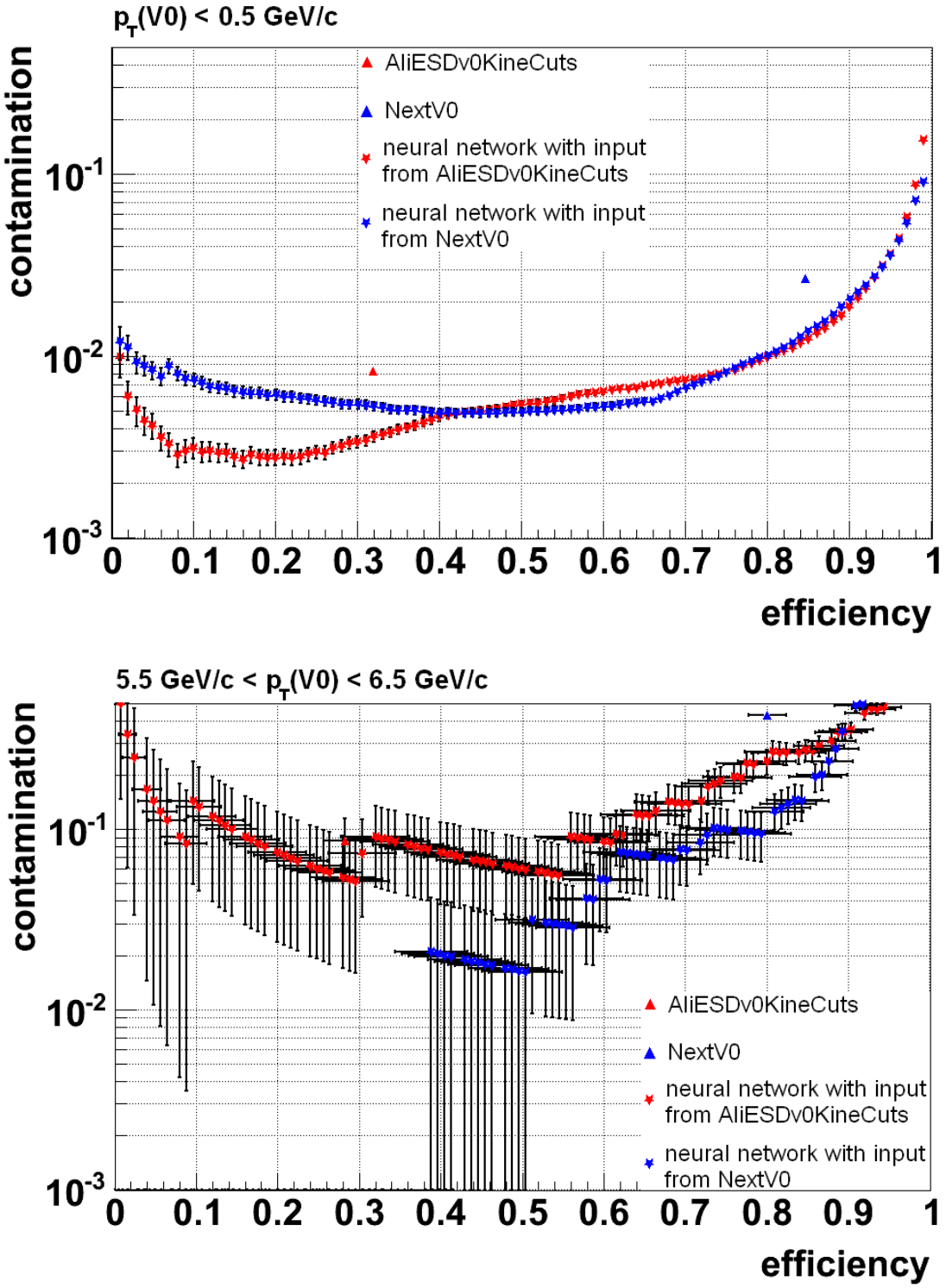


Figure 7.4.: Performance of the method `AliESDv0KineCuts` and the method `NextV0` compared with the performance of the associated neural network for identifying γs from V0s. The contamination of the signal is plotted against the signal efficiency for different p_T of the V0. Only statistical uncertainties are shown. See appendix A.8 for the performance plots of all investigated p_T bins.

figure 7.5 for different p_T of the V0. The minimum contamination is defined as the lowest reached contamination, independent from the efficiency. For $p_T(\mathrm{V0}) \leq 9.5$ GeV/c, the minimum contamination is on a low level ($\lesssim 10\%$), but increases exponentially up to roughly 60% for 12.5 GeV/c $\leq p_T(\mathrm{V0}) \leq$ 13.5 GeV/c. The statistical uncertainties are the same as the uncertainties of the lowest contamination of figure 7.4.
In this analysis, the two neural networks are trained with minimum bias data. In this data, the photons have a p_T spectrum as shown in figure 7.3. Therefore the networks are trained with V0s which predominantly have a small p_T. Using different networks for different p_T bins could result in a lower contamination for high p_T photons.

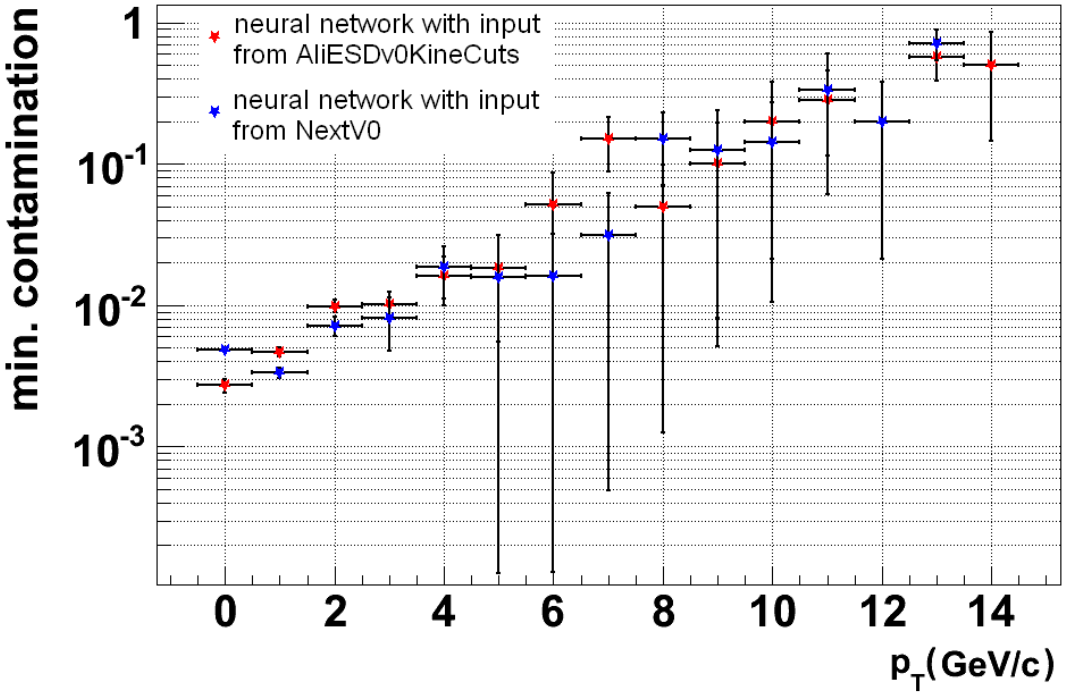


Figure 7.5.: Minimum contamination of signal for the neural network with input from `AliESDv0KineCuts` (red stars) and `NextV0` (blue stars) versus p_T of the V0. Only statistical uncertainties are shown. It can be seen that the contamination is low for small p_T of the V0 and increases exponentially for higher p_T.

In figure 7.6, the TPC dE/dx for measured data is plotted. The TPC dE/dx is shown for the negative daughters of all V0s (upper plot) and for the negative daughters of the V0s which are identified as photons using the neural network with input from `AliESDv0KineCuts` (left plot) and with input from `NextV0` (right plot). In addition,

solid lines are shown representing the Bethe-Bloch lines for electrons (e), protons (p), and pions (π) to guide the readers eye.

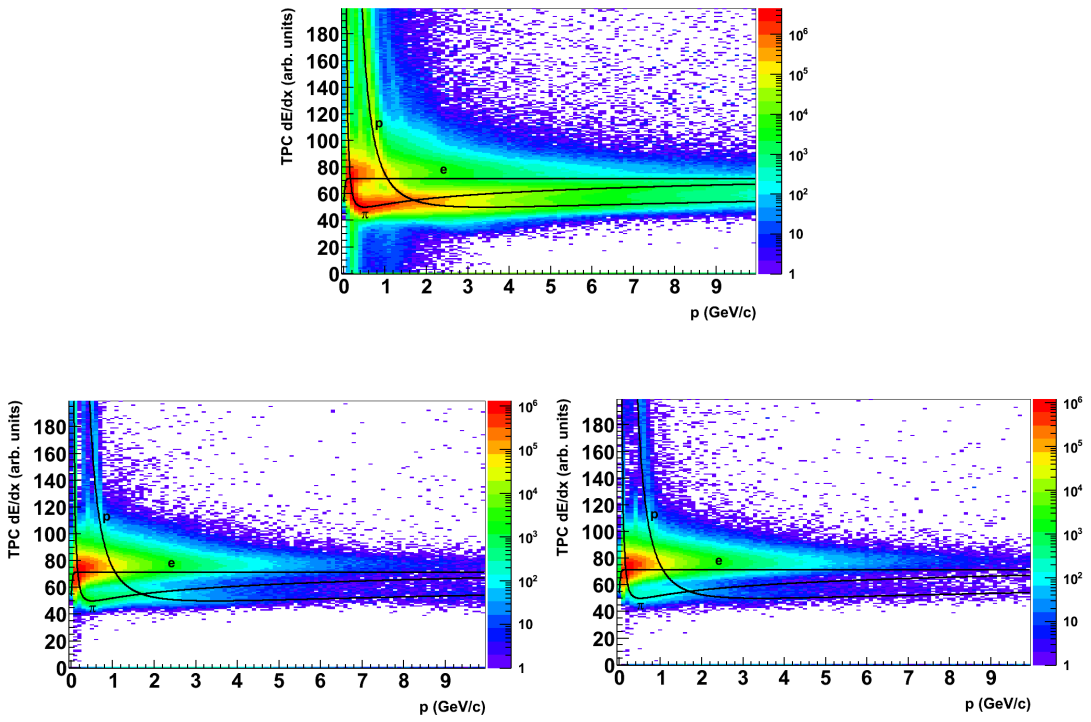


Figure 7.6.: Measured TPC dE/dx for negative V0-daughters (data: LHC10b.pass2, LHC10c.pass2, LHC10d.pass2, LHC10e.pass2). Upper plot: Negative daughters of all V0s. Left (right) plot: Negative daughters of all V0s which are identified as photons using the neural network with input from `AliESDv0KineCuts` (`NextV0`). The Bethe-Bloch lines are shown for electrons (e), protons (p), and pions (π) to guide the readers eye. Most of the V0-daughters are pions (dE/dx around 40-60) and electrons (dE/dx around 70-80). See also figure 3.3.

In the upper plot, it can be seen that most of the V0-daughters are pions (dE/dx around 40-60) and electrons (dE/dx around 70-80) (see also figure 3.3). It can be concluded that the two neural networks are also able to identify photons from V0s in measured data with a low contamination, since almost only electrons remain in the two lower plots. See appendix A.5 for a direct comparison of Monte-Carlo simulation and measured data.

7.3. Identification of π^0s and Photons from π^0s

For the identification of decay photons, it is necessary to know whether a photon is originating from a π^0 or not, since photons from π^0s are representative for all decay photons in this thesis. For this, an invariant mass analysis can be used: The photons are identified by using the PDG code of the particles and by using the neural network with input from `AliESDv0KineCuts`, and `NextV0`, respectively. See appendix A.6 for a comparison of Monte-Carlo simulation and measured data. Two photons are combined and their invariant mass m_{inv} is calculated ($c = 1$)

$$m_{inv} = m_1^2 + m_2^2 + 2E_1E_2 - 2\vec{p}_1\vec{p}_2, \tag{7.6}$$

where m_i, E_i, and $\vec{p}_i$ are the mass, the energy, and the momentum of the photon i (i=1,2). The mass of a photon should be equal to 0 and the energy should be equal to the absolute value of the momentum times speed of light. To ensure this, the mass is constrained to 0 in the reconstruction of the photons. For two photons originating from the same π^0, their invariant mass should be the rest mass of the π^0, $m_{\pi^0} = (134.9766 \pm 0.0006)$ MeV/c^2 [PDG11]. Two photons originating from different particles will not have a determined invariant mass. These photon pairs form the 'combinatorial background'. In figure 7.7, the invariant mass of all photon pairs is shown. These spectra consist of three contributions: a peak at $m_{\gamma\gamma} \approx 0.135$ GeV/c^2, which corresponds to two photons originating from the same π^0; a peak at $m_{\gamma\gamma} \approx 0.550$ GeV/c^2, which corresponds to two photons originating from the same η[4]; and a smooth curve, which corresponds to the combinatorial background (see also appendix A.6). Photon pairs with an invariant mass of around 0.135 GeV/c^2 can originate from the same π^0, but it is also possible that two photons originating from different particles have an invariant mass of around 0.135 GeV/c^2. To identify decay photons with a high purity, the combinatorial background has to be reduced.

In this analysis, only photons originating from a high p_T π^0 ($p_T(\pi^0) \geq 10$ GeV/c) are investigated. Therefore, the invariant mass spectrum has to be only considered for photon pairs with an overall p_T of more than 10 GeV/c,

$$|\vec{p}_T(\gamma_1) + \vec{p}_T(\gamma_2)| \geq 10 \text{ GeV/c}, \tag{7.7}$$

where $\vec{p}_T(\gamma_i)$ is the transverse momentum of photon i (i=1,2).

[4] $m_\eta = (547.853 \pm 0.024)$ MeV/c^2 [PDG11].

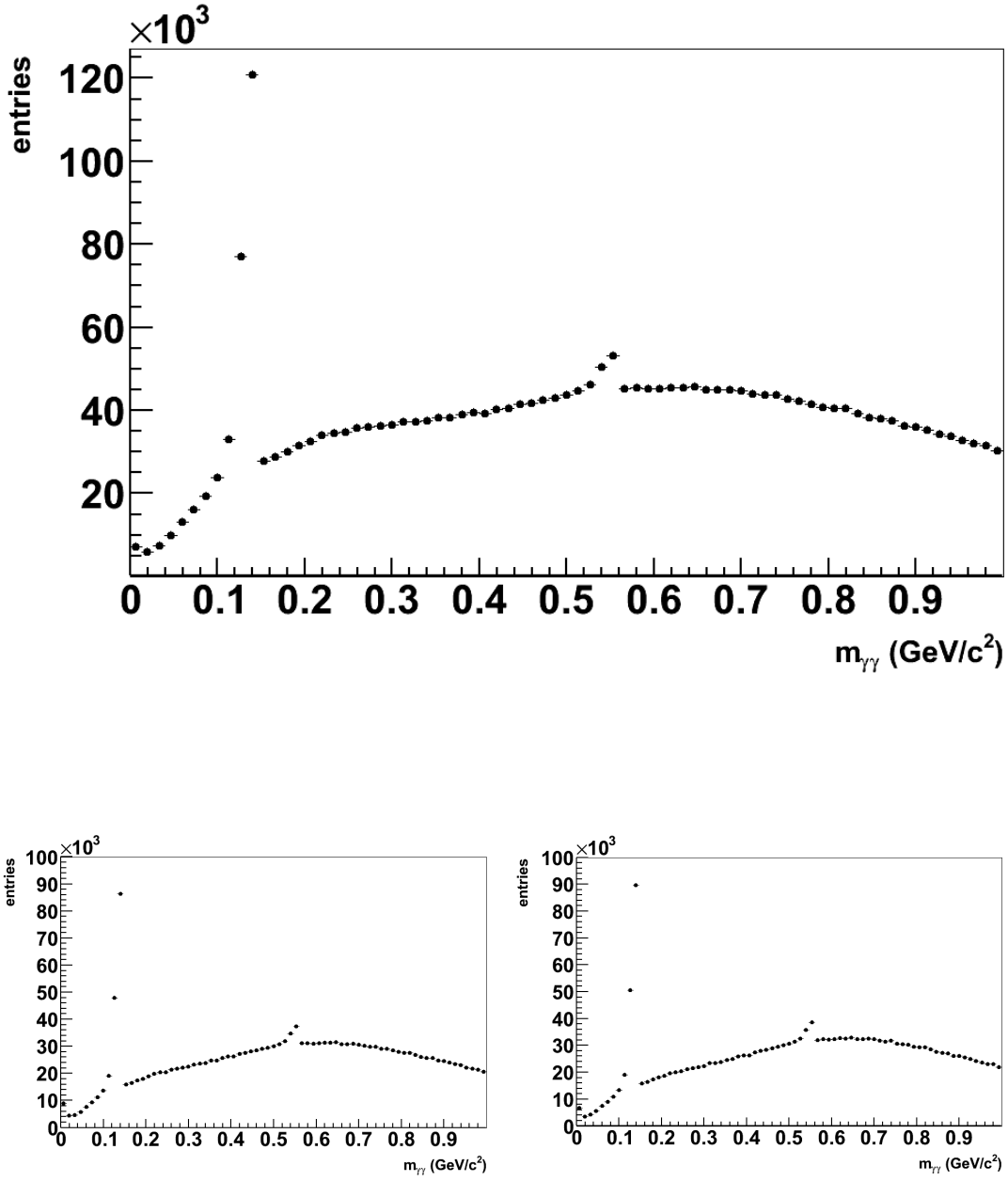


Figure 7.7.: Invariant mass distribution of all photon pairs. The photons are identified by using the PDG code of the particles (upper plot), the neural network with input from `AliESDv0KineCuts` (left plot), and `NextV0` (right plot), respectively. The pseudorapidity of both photons has to be in the acceptance of the central barrel, $|\eta| \leq 0.9$, and the V0 has to be found by the On-the-fly-Finder. 4.24×10^8 minimum bias Monte-Carlo events are used (production numbers LHC10d1, LHC10d2, LHC10d4, LHC10d4a, LHC10f6, LHC10f6a, LHC10e20, LHC10e21).

In the 4.24×10^8 minimum bias Monte-Carlo events which are used for figure 7.7, only 68 π^0 exist where

- both photons are converted and found by the On-the-fly-Finder,
- $|\vec{p}_T(\gamma_1) + \vec{p}_T(\gamma_2)|$ is greater than 10 GeV/c, and
- $|\eta(\gamma_{1,2})|$ is smaller than 0.9.

To train a neural network with such photons from π^0s as trigger photons, roughly 10^3-10^4 photons are needed, which corresponds to $10^4 \cdot 4.24 \times 10^8/68 \approx 6 \times 10^{10}$ minimum bias events. This large amount of simulated events does not exist in the ALICE repository. The neural network can also be trained with measured data[5]. In this case, a high p_T electron trigger can be used to select events with high p_T electrons and therefore, probably with converted high p_T photons. This trigger will be elaborated by Uwe Westerhoff in his Ph.D. thesis.
In figure 7.8, the invariant mass of all photon pairs with an overall p_T of more than 10 GeV/c and a distance smaller than $R = 0.05$ is plotted. The photons are identified by using the neural network with input from `AliESDv0KineCuts` (left plot), and `NextV0` (right plot), respectively.

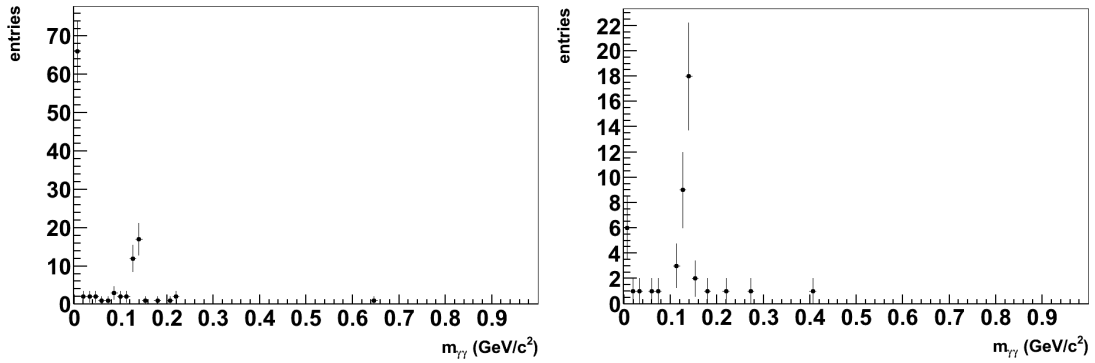


Figure 7.8.: Invariant mass distribution of all photon pairs with $|\vec{p}_T(\gamma_1) + \vec{p}_T(\gamma_2)|$ is greater than 10 GeV/c, the distance of the two photons in (η,φ) is smaller than $R = 0.05$, and $|\eta(\gamma_{1,2})|$ is smaller than 0.9. The V0 has to be found by the On-the-fly-Finder. Left: Using the neural network with input from `AliESDv0KineCuts`. Right: Using the neural network with input from `NextV0`. 4.24×10^8 minimum bias Monte-Carlo events are used. See appendix A.6 for the same plots in measured data.

[5]See chapter 8 for a method to calculate prompt photon patterns.

Besides the combinatorial background of two photons, another source of contamination for decay photons exists. Due to the fact that a certain contamination of neutral particles other than photons and fake-V0s exists for the two neural networks, a photon can be combined with a background V0 or two background V0s can be combined. These combinations also will not a have a determined invariant mass and will contribute to the combinatorial background. At $m_{\gamma\gamma} \approx 0$ GeV/c^2 another peak occurs in the left plot of figure 7.8. This peak originates from photon pairs combining photons with pairwise permuted electrons/positrons or from photon pairs combining identical photons.
Using the neural network with input from `AliESDv0KineCuts` (`NextV0`), 29 (24) true photon pairs originating from the same π^0 exist in the π^0 peak[6]. In addition, 1 (2) true photon pairs originating from the different particles and 0 (4) pairs with at least one background V0 exist in the π^0 peak. This results in a contamination of $\frac{1}{1+29} \approx 3\%$ ($\frac{2+4}{2+4+24} = 20\%$) for identifiying π^0s. The π^0s can be used to train a neural network for photons identified using EMCal/PHOS (using the p_T of the π^0 for background and the p_T of the prompt photon for signal).
To train a neural network for photons identified via conversion, it is necessary to know how many of the V0s are photons originating from a π^0. Here, only the first V0s of the two combined V0s are investigated, since only these V0s are used as trigger photons. To identify π^0s, the overall p_T of the two V0s has to be greater than 10 GeV/c. For photons identified via conversion, the transverse momentum of the first V0 has to be greater than 10 GeV, whereas the p_T of the second V0 is not subjected to any restrictions. With this condition, the invariant mass of the accepted V0 pairs is calculated. Using the neural network with input from `AliESDv0KineCuts` (`NextV0`), 5 of 5 (2 of 5) first V0s in the π^0 peak are photons originating from a π^0. For the network with input from `NextV0`, 2 of the first V0s are fake-V0s and 1 V0 is a Λ particle. This results in a contamination of 0% and 60%, respectively. However, due to the low statistics, the uncertainties of the contaminations are very large.

7.4. Contamination Studies

In section 7.3, it was concluded that decay photons can be identified with a certain contamination due to combinatorial background of two photons, fake-V0s, and neutral particles other than photons. To investigate the effect of contamination on the

[6]The invariant mass of the photon pairs in the π^0 peak has to be in [0.115 GeV/c^2, 0.145 GeV/c^2], 0.115 GeV/c^2 $\leq m_{\gamma\gamma} \leq$ 0.145 GeV/c^2.

separation performance, different neural networks are trained with different contaminations for background. The contamination is investigated for the neural network without using p_T (see section 6.2), since this neural network is used to test the possibility of using 'random cones' instead of simulated prompt photons to get the patterns for signal (compare chapter 8). The five input variables of the networks are:

- $\sum(p_T/R)$, $R_{\max} = 0.5$, $p_T^{\text{particle}} \geq 0.0$ GeV/c
- Mean distance of the particles, $R_{\max} = 0.5$, $p_T^{\text{particle}} \geq 1.1$ GeV/c
- Number of particles, $R_{\max} = 0.7$, $p_T^{\text{particle}} \geq 0.1$ GeV/c
- Number of particles in $|\Delta\varphi| \leq \frac{\pi}{2}$ and $|\Delta\eta| \leq 1.4$, $p_T^{\text{particle}} \geq 0.6$ GeV/c
- Number of particles in $|\Delta\varphi| \leq \frac{\pi}{2}$ and $|\Delta\eta| \leq 1.4$, $p_T^{\text{particle}} \geq 0.0$ GeV/c

The networks are trained and tested using $(1-x_1-x_2) \times 6700$ decay photons, $x_1 \times 6700$ V0s, and $x_2 \times 6700$ prompt photons for background and 6700 prompt photons for signal. The $x_1 \times 6700$ V0s are fake-V0s and neutral particles other than photons. The contamination due to V0s, x_1, is varied between 0.01 and 0.5, whereas the contamination due to prompt photons, x_2, is varied between 0.00 and 0.06, since V0s will contribute more to the contamination than prompt photons and only a small number of prompt photons exist.
The transverse momentum of the V0s/photons must be greater than 10 GeV/c, they must be found by the On-the-fly-Finder and they must be in the acceptance of the central barrel ($|\eta| \leq 0.9$). All networks are trained and validated using prompt photons simulated in γ-jet events (see section 5.3) for signal as in section 6.2. For background, the networks are validated using decay photons as trigger photons (see section 5.3).
In figure 7.9, the improvement of the signal to background ratio for $\epsilon_{\text{prompt}} = 0.5$ is plotted versus different contaminations due to V0s, x_1, for two different contaminations due to prompt photons, $x_2 = 0.00$ and $x_2 = 0.06$, respectively. It can be seen that the performance of the networks is almost constant for contaminations due to V0s, x_1, smaller than 0.16, decreases for larger contaminations, and does not depend on the contaminations due to prompt photons, x_2, in the tested range.
In figure 7.10, the performance of different neural networks is shown. The first network is trained using converted decay photons for background (red stars, 'neural network using decay photons for training'). In section 6.2, the neural networks are trained and tested using 6728 simulated decay photons and 6728 simulated prompt photons. In

the 4.24×10^8 minimum bias Monte-Carlo events only 1813 converted decay photons with $p_T \geq 10$ GeV/c exist. These 1813 photons are used to train, test, and validate the network. Therefore, the first network is trained and tested with 1188 patterns for signal and 1188 patterns for background instead of 6728 patterns. The network is validated using the 625 remaining decay photons for background and 625 simulated prompt photons for signal.

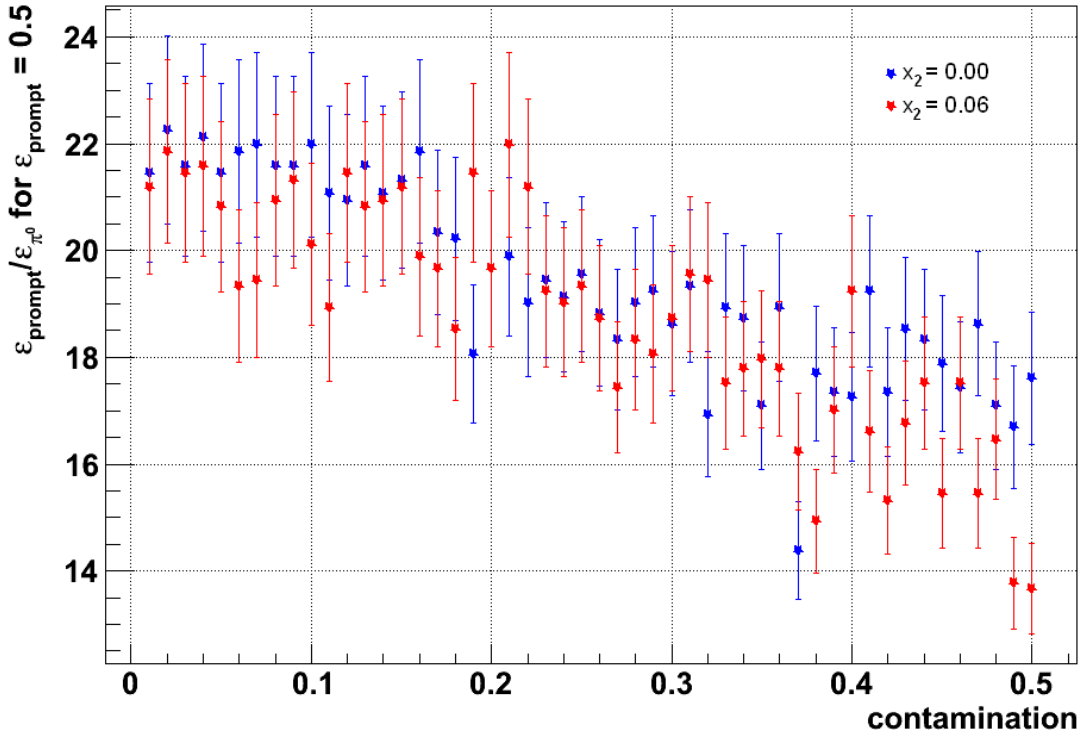


Figure 7.9.: Improvement of the signal to background ratio for $\epsilon_{\text{prompt}} = 0.5$ versus different contaminations due to V0s x_1 for two different contaminations due to prompt photons $x_2 = 0.00$ (blue) and $x_2 = 0.06$ (red), respectively.

To check whether the decay photons have to be identified using an invariant mass analysis, two neural networks are trained using the accepted V0s for background. The V0s have to be accepted by the neural networks with input from `AliESDv0KineCuts` (yellow squares, 'neural network using NN KineCuts for training') and `NextV0` (green circles, 'neural network using NN NextV0 for training'), respectively. The cut in the output the networks is set to 0.99 and 0.96, respectively. The two networks are trained and tested using 1188 accepted V0s for background and 1188 simulated prompt photons for signal. They are validated using 625 decay photons and 625 prompt photons. In figure

7.10, it can be seen that independent of training with all accepted V0s for background or training without contamination the ability of separating decay photons and prompt photons is equal within the uncertainties. That means that the decay photons do not have to be identified using an invariant mass analysis, but all accepted V0s can be used as background to train the network.

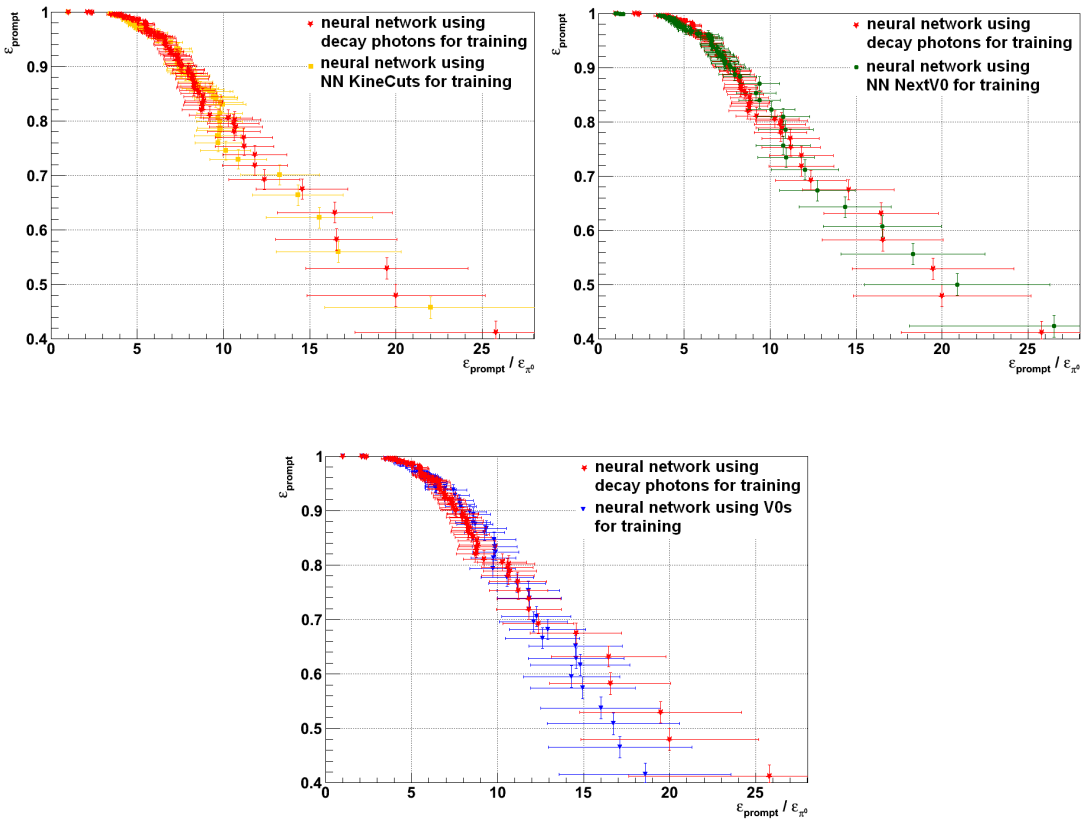


Figure 7.10.: Performance plot of different neural networks to study the effect of contamination. The first neural network is trained without background contamination (red stars). Two neural networks are trained using NN KineCuts (NN NextV0) for background (yellow squares and green circles, respectively). A fourth network is trained using all V0s for background (blue triangles).

To check whether the photons have to be identified at all, a neural network is trained and tested using 1188 V0s for background (blue triangles, 'neural network using V0s for training') and 1188 simulated prompt photons for signal. The network is validated using 625 decay photons and 625 prompt photons. In figure 7.10, it can be seen that the performance of this network is similar to the performance of the neural network without

contamination. That means that the photons from V0s do not have to be identified to use them as background for the neural network. However, due to the low statistics the calculated performances are only rough estimations and further studies are necessary to verify these results.

8. Prompt Photon Patterns in Minimum Bias Events

In chapter 7, the identification of decay photons using an invariant mass analysis was presented. After the identification of the trigger photon, the associated particles in its immediate surrounding can be used to calculate the input variables for the neural network. These input variables will form the characteristic patterns for background (= decay photons). Since prompt photons cannot be identified, a different possibility to calculate the characteristic patterns for signal is needed. One possibility to estimate these patterns will be presented in this chapter.

As mentioned in section 2.5, the prompt photon and the jet are back-to-back correlated in φ (compare also figure 6.1 in section 6.1.1). Therefore the immediate surrounding of a prompt photon should only consist of particles from the underlying event and no particles from a jet should appear. To get particles from the underlying event and to calculate the characteristic patterns for signal, minimum bias events can be used. The creation of a jet is not a requirement for a minimum bias event, therefore no jet will be produced in a lot of these events, but only the underlying event. The 'trigger direction', defined by the prompt photon up to now, can be imitated by using 'random cones'. For the η and the φ-value of the cone axis, uniformly distributed pseudo random numbers in the interval η =]-0.5, 0.5] and φ =]0, 2π] are used. The patterns inside the random cone in minimum bias events are formed by the particles inside the cone. These patterns have to be equal to the patterns in a cone around a prompt photon to estimate the patterns for prompt photons with random cones in minimum bias events. In case it is possible to calculate prompt photon patterns in minimum bias events in simulated data, it should also be possible in measured data, if the random cones are chosen without too strict conditions.
In figure 8.1, the number of reconstructed tracks in simulated γ-jet events is shown (red). See section 5.3 for the settings of the γ-jet simulation. It can be seen that in most of the events, the number of reconstructed tracks in an event is roughly between 10

and 100 and that the distribution looks similar to a Landau distribution. The number of reconstructed tracks in minimum bias events is shown in yellow in figure 8.1.

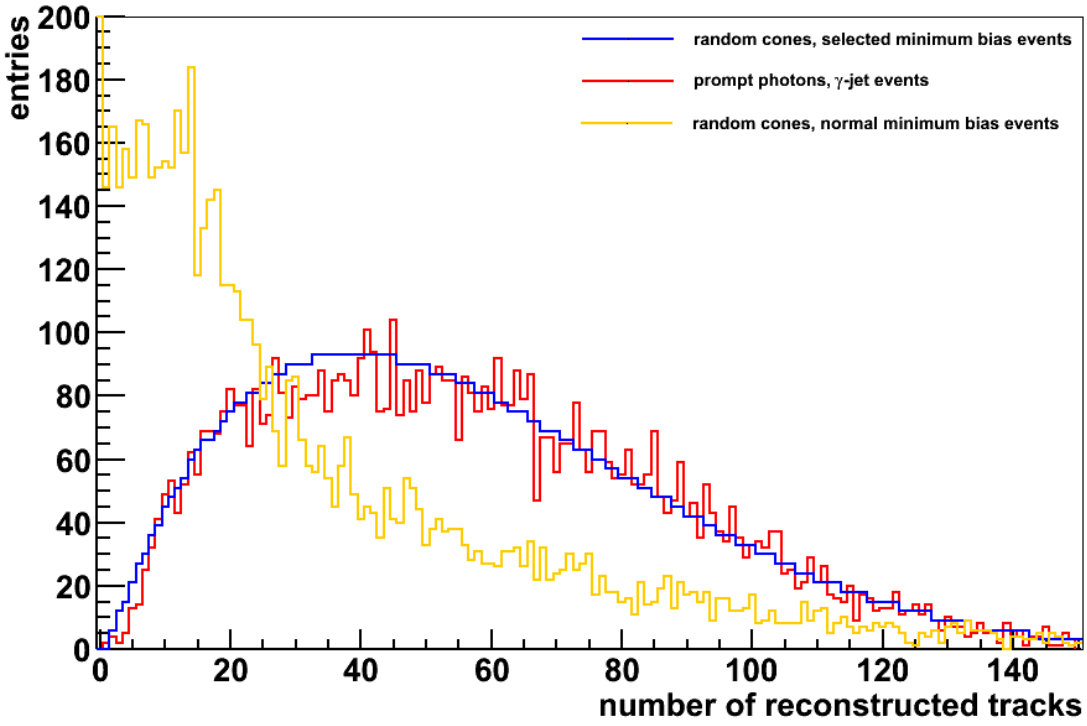


Figure 8.1.: Red: Number of reconstructed tracks in simulated γ-jet events. See section 5.3 for the settings of the simulation. Yellow: Number of reconstructed tracks in minimum bias events. Blue: Number of reconstructed tracks in selected minimum bias events. The events are chosen in such a way that the number of reconstructed tracks distribution is nearly the same as in γ-jet events.

It can be seen that the number of reconstructed tracks will very often be smaller than in γ-jet events. To estimate the characteristic patterns for signal with random cones in minimum bias events, the events are chosen in such a way that the distribution of the number of reconstructed tracks is nearly the same as in γ-jet events. The exact distribution is shown in blue in figure 8.1. Using these minimum bias events and the cone axis of the random cones, the five input variables from section 6.2 are calculated:

- $\sum(p_T/R)$, $R_{\text{max}} = 0.5$, $p_T^{\text{particle}} \geq 0.0$ GeV/c
- Mean distance of the particles, $R_{\text{max}} = 0.5$, $p_T^{\text{particle}} \geq 1.1$ GeV/c
- Number of particles, $R_{\text{max}} = 0.7$, $p_T^{\text{particle}} \geq 0.1$ GeV/c

- Number of particles in $|\Delta\varphi| \leq \frac{\pi}{2}$ and $|\Delta\eta| \leq 1.4$, $p_T^{\text{particle}} \geq 0.6$ GeV/c
- Number of particles in $|\Delta\varphi| \leq \frac{\pi}{2}$ and $|\Delta\eta| \leq 1.4$, $p_T^{\text{particle}} \geq 0.0$ GeV/c

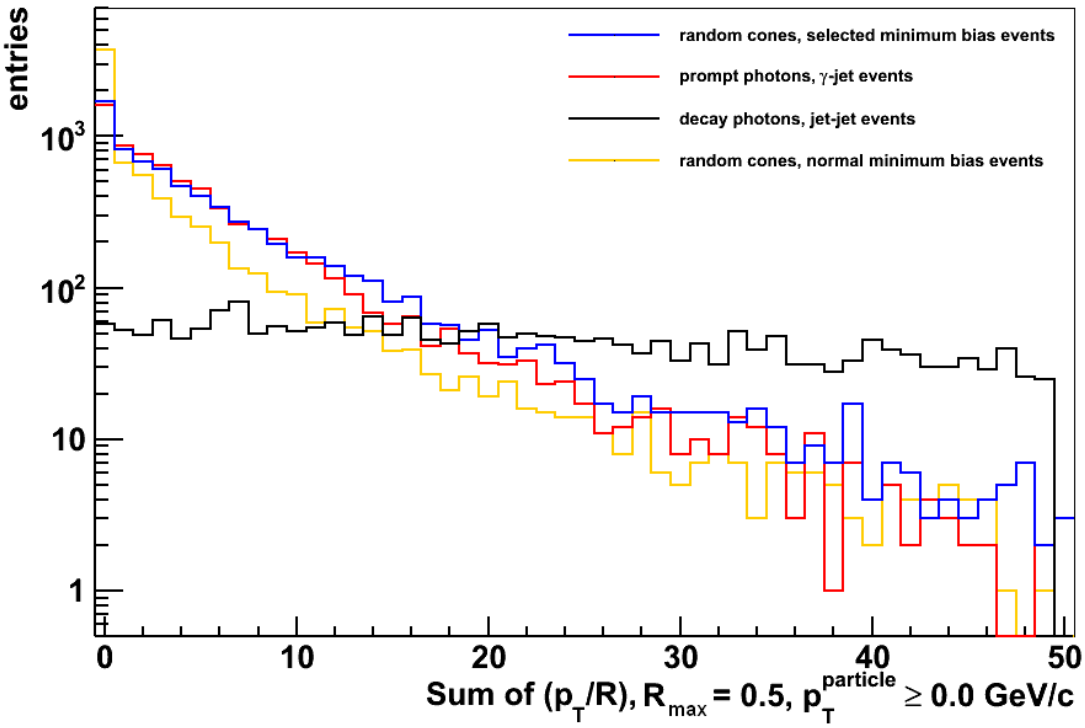


Figure 8.2.: Distribution of the first input variable from section 6.2. The input variable is calculated in γ-jet events (red) and in selected minimum bias events (blue). The events are chosen in such a way that the distribution of number of reconstructed tracks is nearly the same as in γ-jet events. Yellow: Distribution of the first input variable from section 6.2 in minimum bias events which are not selected in a special way. Black: Distribution of the first input variable for background. See appendix A.9 for the distribution of the other input variables.

The transverse momentum of the trigger photon is not used, because no trigger photon exists. To calculate this input variable, the p_T spectrum of prompt photons has to be fitted and the generated values must be used as input in consideration of all correlation to the other input variables. This is not done and therefore, the transverse momentum of the trigger photon is not used for the following studies.
In figure 8.2, the calculation of the first input variable is shown. The distribution of the input variable calculated in γ-jet events is shown in red. It can be seen that the

distribution is nearly the same as in selected minimum bias events (blue), whereas the distribution of the input variable calculated in minimum bias events which are not selected in a special way (yellow) differs from the distribution of the input variable calculated in γ-jet events. However, both distributions differ from the distribution of the input variable for background (black). Therefore, the neural network should be able to distinguish between signal and background.

To calculate the patterns of prompt photons, random cones in selected minimum bias events can be used. The minimum bias events are chosen in such a way that the distribution of number of reconstructed tracks is nearly the same as in γ-jet events. The number of reconstructed tracks of γ-jet events is not known in measured data. Therefore the selection of minimum bias events in this way is a too strict condition to the random cones and cannot be used in measured data.

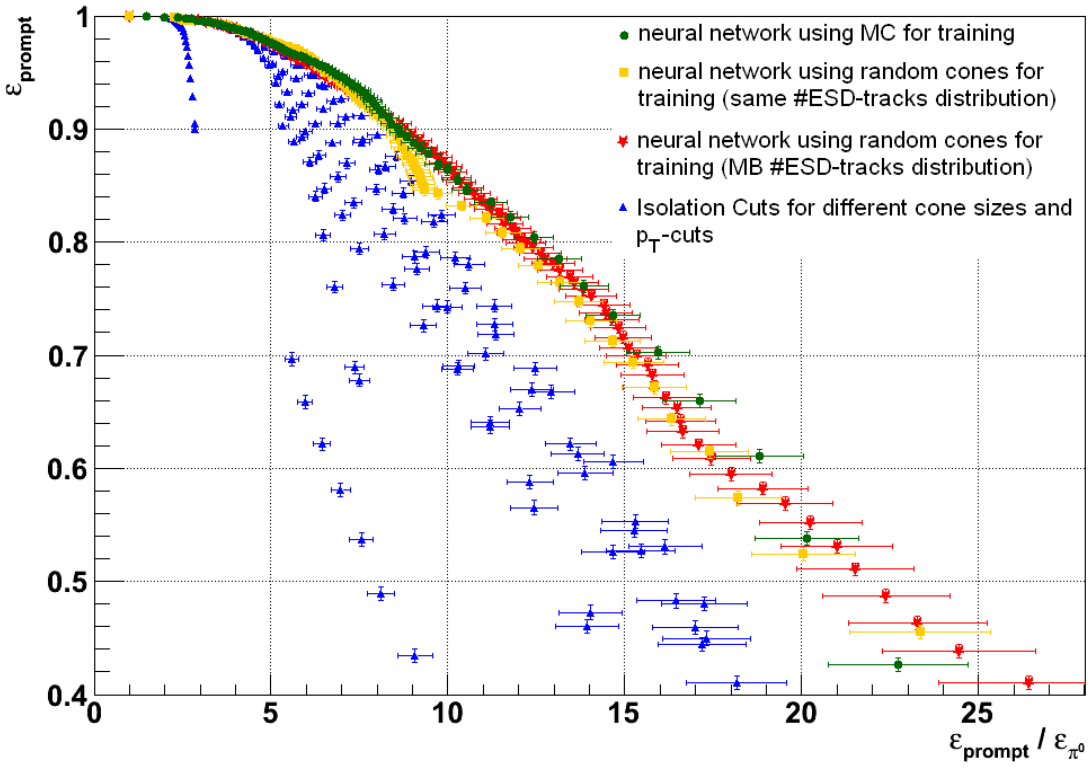


Figure 8.3.: Performance plot of the three neural networks.

To check whether only random cones in the selected minimum bias events or also random cones in normal minimum bias events can be used as a substitute for prompt photon patterns, three neural networks are trained. The first network is trained and validated with input variables for signal, which are calculated in γ-jet events ('using

MC for training') and the second (third) network is trained with input variables for signal which are calculated in selected (normal) minimum bias events ('using random cones for training') and validated with input variables for signal which are calculated in γ-jet events. The input variables for background are calculated in jet-jet events, which are described in section 5.3, for all three networks.
In figure 8.3, the performance of the three neural networks is plotted. Independent of training with input variables from γ-jet events or training with input variables from (selected or normal) minimum bias events, the ability of separating decay photons and prompt photons is almost equal.
It should be explicitly mentioned that the neural network which is trained with input variables calculated in normal minimum bias events has the same ability of separating decay photons and prompt photons as the other two networks, although the training patterns differ from the validating patterns. This clearly shows the robustness and generalization aptitude of the neural network and is a natural advantage compared to using cuts.

9. Summary and Outlook

In this thesis, the separation of decay photons and prompt photons in simulated proton-proton collisions at $\sqrt{s} = 7$ TeV in the ALICE experiment was presented.
The ALICE experiment studies the quark-gluon plasma (QGP), a state of matter which is generated in Pb-Pb collisions at the LHC and presumably existed until 10^{-5} s after the Big Bang. The quarks and gluons of the QGP cannot be directly observed, since only colorless particles can be measured. Therefore, some probes are needed to investigate the QGP, for example prompt photons. Using prompt photons, the fragmentation function of hadrons can be calculated in γ-jet events and also jet-quenching in Pb-Pb collisions can be investigated. The aim of this thesis was to separate prompt photons from the large background of decay photons on an event-by-event basis using a neural network.
Contrary to the conventional method (isolation cut), a neural network is able to take correlations of different variables into account. For the training of the neural network, some training patterns for signal and background are needed.
To get training patterns, the prompt photons and decay photons were identified in Monte-Carlo simulations using the PDG code of the particles. The training patterns are formed by the particles in the immediate surrounding of the trigger photon. In section 6.2, the neural networks were trained with these patterns and it was concluded that the separation performance is higher for the networks than for the investigated isolation cuts.
In chapter 7, the identification of decay photons using an invariant mass analysis was presented. Two neural networks were trained to identify photons from V0s and compared to two existing methods. It was concluded that photons can be identified with a higher efficiency and a lower purity using these networks instead of the conventional methods. The invariant mass of photon pairs was calculated and photons from π^0s were identified. Training patterns of decay photons can be calculated not only in simulated data, but also in measured data, using the invariant mass analysis or only the two neural networks to identify photons from V0s.

To train the neural network using measured data, the training patterns of prompt photons have to be calculated. In chapter 8, a method was presented to estimate these patterns in minimum bias events. Despite the patterns calculated in minimum bias events differs from the prompt photon patterns, the neural network is able to identify prompt photons without loss of performance. This robustness and generalization aptitude of the network is a natural advantage compared to using cuts.
Using the results of chapters 7 and 8 a neural network can be trained using measured data. A high p_T electron trigger can be used to select events with converted high p_T photons to get enough high p_T converted photons.
To study jet-quenching and other phenomenons, this analysis can also be done in Pb-Pb collisions. Some possible problems are the identification of photons from V0s with a small contamination of fake-V0s, the jet-quenching itself, and the much larger underlying event. Further studies are needed to reduce and estimate the influence of these problems to the separation performance.

A. Appendix

A.1. Kinematic Variables

In special relativity, the 'four-momentum' P of a particle is defined by[1]

$$P = (E, \vec{p}) = (E, p_x, p_y, p_z),$$

where E is the energy of the particle and $\vec{p}$ its 'three-momentum'. The square of P can be calculated using the Minkowski metric $\eta_{\mu\nu} = \text{diag}(+1,-1,-1,-1)$:

$$P^2 = \eta_{\mu\nu} P^\mu P^\nu = E^2 - p^2. \tag{A.1}$$

The relation between the energy and the three-momentum of a particle with rest mass m_0 is given by

$$E^2 = p^2 + m_0^2.$$

In a $2 \to 2$ scattering process, it is useful to define the 'Mandelstam variables' s,t, and u,

$$s = (P_1 + P_2)^2 = (P_3 + P_4)^2,$$

$$t = (P_1 - P_3)^2 = (P_2 - P_4)^2,$$

$$u = (P_1 - P_4)^2 = (P_2 - P_3)^2,$$

where P_1 and P_2 are the four-momenta of the incoming particles and P_3 and P_4 are the four-momenta of the outgoing particles. $\sqrt{s}$ is the total energy of a reaction in the center-of-mass system and $\sqrt{t}$ the momentum transfer.
The particle's momentum $\vec{p}$ can be split into a longitudinal component $\vec{p}_L$, which points in the direction of the beam axis, and a transverse component $\vec{p}_T$, which is in the plane perpendicular to the beam axis. Using the polar angle ϑ of the ALICE coordinate system (see appendix A.2), p_L and p_T can be written as

$$p_L = |\vec{p}_L| = p \cdot \cos(\vartheta) = p_z,$$

[1] In all equations is set: $\hbar = 1 = c$.

$$p_T = |\vec{p}_T| = p \cdot \sin(\vartheta) = \sqrt{p_x^2 + p_y^2},$$

where p is

$$p = |\vec{p}| = \sqrt{p_L^2 + p_T^2}.$$

The transverse momentum is invariant under Lorentz transformation, the longitudinal momentum is not. Therefore, the rapidity y, defined as

$$y = \frac{1}{2} \ln(\frac{E + p_L}{E - p_L}),$$

is often used. The rapidity y is additive under Lorentz transformation, that means

$$y = y^{'} + y_{S'},$$

where y is the rapidity in system S, $y^{'}$ is the rapidity in system $S^{'}$, and $y_{S'}$ is the rapidity of $S^{'}$ as measured in S.
For E $\gg m_0$, the rapidity is approximately equal to the pseudorapidity η,

$$y \approx \frac{1}{2} \ln(\frac{p + p_L}{p - p_L}) = -\ln[\tan(\frac{\vartheta}{2})] := \eta.$$

Contrary to y, η only depends on the polar angle ϑ and can be measured for every particle without identifying it.
In figure A.1, the pseudorapidity η can be seen in the ALICE coordinate system.

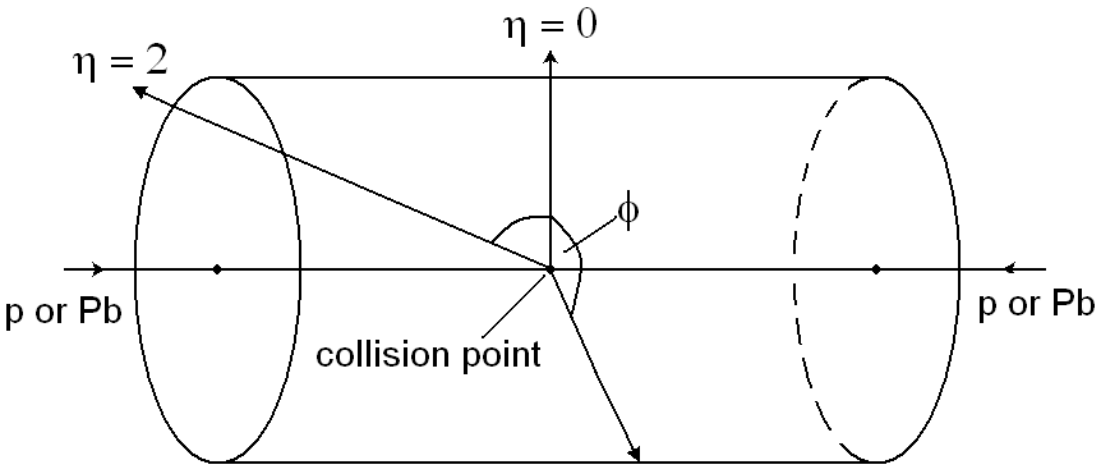


Figure A.1.: Schematic view of the ALICE coordinate system angles

A.2. The ALICE Coordinate System

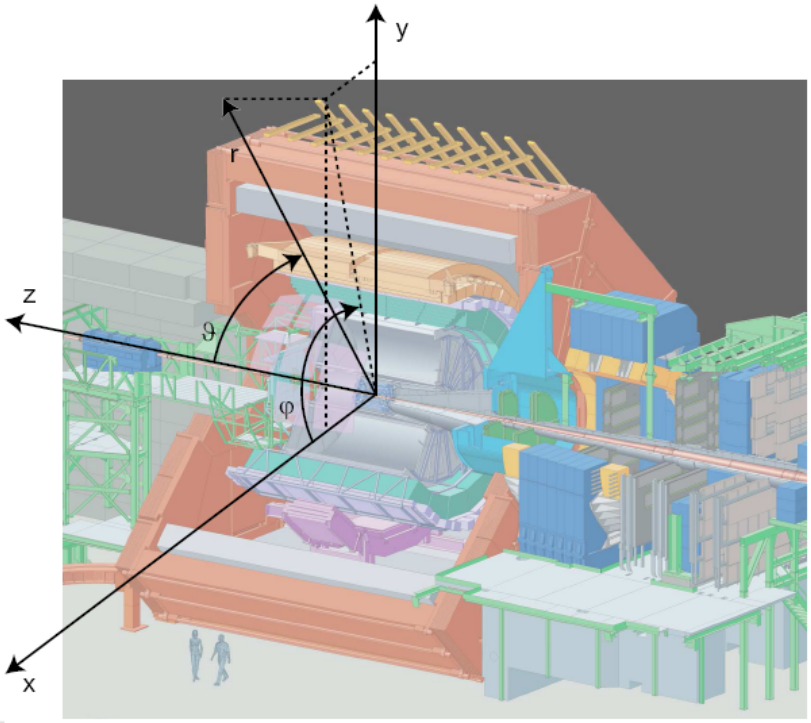


Figure A.2.: The ALICE Coordinate System [GO09].

The point of origin of the ALICE coordinate system is the beam interaction point [ALI03], it is defined as x,y,z = 0. The coordinate system is a right-handed orthogonal Cartesian system. The individual axes and angles are defined as follows:

- x-axis: The x-axis is perpendicular to the beam direction and parallel to the local horizon. Positive x is from the point of origin towards the accelerator center.
- y-axis: The y-axis is perpendicular to the beam direction and perpendicular to the x axis. Positive y is from the point of origin upward.
- z-axis: The z-axis is parallel to the beam direction. Negative z is from the point of origin towards the Muon Arm.
- azimuthal angle φ: φ is the angle between the positive x-axis and the positive y-axis. $\varphi = 0$ means y = 0 and $\varphi = \pi/2$ means x = 0.
- polar angle ϑ: ϑ increases from the positive z-axis ($\vartheta = 0$) to the x-y-plane ($\vartheta = \pi/2$) to the negative z-axis ($\vartheta = \pi$).

A.3. Vacuum Polarization

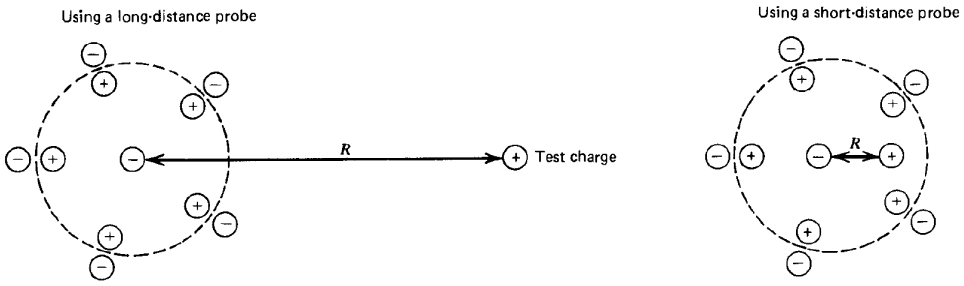


Figure A.3.: Measuring the electric charge [Hal84].

In quantum field theory, an electron can emit a photon that converts to an electron-positron pair. Therefore, every electron is surrounded by some of these electron-positron pairs and due to the attraction of opposite charges, the positrons will be closer to the electron (compare figure A.3). This results in a screening of the charge of the electron by a polarization of the vacuum. Measuring the charge of the electron by using a test charge, the result will depend on the distance R of the test charge to the original electron. For large distances, the screened electric charge will be measured, for small distances, the non-screened electric charge.

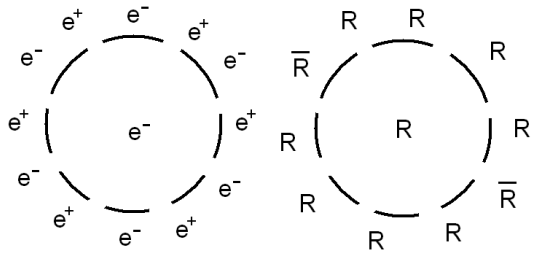


Figure A.4.: Schematic view of the screening of an electron and the anti-screening of a quark.

For color-charged particles, the situation is different. A red-colored quark will emit gluons, which can turn into pairs of gluons because of the SU(3) invariance of the QCD Lagrangian. Due to the fact that gluons carry color, the quark is not only surrounded by anti-red-colored anti-quarks, but also by red-colored gluons. This results in an 'anti-screening' of the color charge of the original quark and the measured color will be stronger for large distances.

In optics, the angular resolution of an object depends on the wavelength of the incoming light. For small wavelengths, the resolution is high and small details of an object can be resolved. In scattering experiments, like in Rutherford scattering, high energy of an incoming particle means small distance to the target particle and a good resolution of

the target. The coupling constant of the strong force α_s is proportional to the square of the color charge g, $\alpha_s \propto g^2$. Therefore, the coupling constant measured in scattering experiments will be higher for small energies (or momentum transfers) and smaller for high energies.

A.4. Backpropagation Algorithm

The following derivation is adapted from reference [Rum85]. The error ϵ_p of the network for a pattern p is given by

$$\epsilon_p := \frac{1}{2}\sum_j (t_{p,j} - o_{p,j})^2. \tag{A.2}$$

$t_{p,j}$ is the target output and $o_{p,j}$ the real output of the j-th output neuron for a pattern p. The weight from the i-th to the j-th neuron is called $w_{i,j}$ and $\Delta_p w_{i,j}$ is the modification of this weight due to pattern p. The new weight $w_{i,j}^{new}$ is given by

$$w_{i,j}^{new} := w_{i,j} + \Delta_p w_{i,j}. \tag{A.3}$$

The modification of the weight $w_{i,j}$ is proportional to the variation of the error with the weight,

$$\Delta_p w_{i,j} := -\eta \frac{\partial \epsilon_p}{\partial w_{i,j}}. \tag{A.4}$$

The learning parameter η is a real, constant number and determines the speed of the training. Applying the chain rule, the derivation of ϵ_p can be written as

$$\frac{\partial \epsilon_p}{\partial w_{i,j}} = \frac{\partial \epsilon_p}{\partial o_{p,j}} \frac{\partial o_{p,j}}{\partial w_{i,j}}. \tag{A.5}$$

With equation A.2 follows

$$\frac{\partial \epsilon_p}{\partial o_{p,j}} = o_{p,j} - t_{p,j} =: -\delta_{p,j}. \tag{A.6}$$

In case the activation function is the identity function, $o_{p,j}$ is given by

$$o_{p,j} = \sum_i w_{i,j} i_{p,i}. \tag{A.7}$$

$i_{p,i}$ is the input of the i-th input neuron for pattern p. Then, for the derivation of $o_{p,j}$, it follows

$$\frac{\partial o_{p,j}}{\partial w_{i,j}} = i_{p,i} \tag{A.8}$$

and for the derivation of ϵ_p

$$\frac{\partial \epsilon_p}{\partial w_{i,j}} = -\delta_{p,j} i_{p,i}. \tag{A.9}$$

For activation functions f, which are not the identity, $o_{p,j}$ is given by

$$o_{p,j} = f(\sum_i w_{i,j} o_{p,i}), \tag{A.10}$$

where f is differentiable and $o_{p,i}$ equal to $i_{p,i}$ if i is an input unit. The derivation of ϵ_p is then

$$\frac{\partial \epsilon_p}{\partial w_{i,j}} = \frac{\partial \epsilon_p}{\partial(\sum_i w_{i,j} o_{p,i})} \frac{\partial(\sum_i w_{i,j} o_{p,i})}{\partial w_{i,j}}. \tag{A.11}$$

The second factor can be directly calculated,

$$\frac{\partial(\sum_i w_{i,j} o_{p,i})}{\partial w_{i,j}} = o_{p,i}. \tag{A.12}$$

Consistent with equation A.6, $\delta_{p,j}$ is defined as

$$\delta_{p,j} := -\frac{\partial \epsilon_p}{\partial(\sum_i w_{i,j} o_{p,i})} \tag{A.13}$$

and can be written as

$$\delta_{p,j} := -\frac{\partial \epsilon_p}{\partial(\sum_i w_{i,j} o_{p,i})} = -\frac{\partial \epsilon_p}{\partial o_{p,j}} \frac{\partial o_{p,j}}{\partial(\sum_i w_{i,j} o_{p,i})}. \tag{A.14}$$

Then the modification of the weight is given by

$$\Delta_p w_{i,j} = \eta \delta_{p,j} o_{p,i}. \tag{A.15}$$

Using equation A.10, it can be seen that

$$\frac{\partial o_{p,j}}{\partial(\sum_i w_{i,j} o_{p,i})} = f'(\sum_i w_{i,j} o_{p,i}), \tag{A.16}$$

which is the derivative of the function f. For output units, the first factor of equation A.14 is given by

$$\frac{\partial \epsilon_p}{\partial o_{p,j}} = -(t_{p,j} - o_{p,j}). \tag{A.17}$$

For other units, the chain rule can be applied to calculate the derivative of ϵ_p

$$\frac{\partial \epsilon_p}{\partial o_{p,j}} = \sum_k \frac{\partial \epsilon_p}{\partial(\sum_i w_{i,k} o_{p,i})} \frac{\partial(\sum_i w_{i,k} o_{p,i})}{\partial o_{p,j}} = \sum_k \frac{\partial \epsilon_p}{\partial(\sum_i w_{i,k} o_{p,i})} w_{j,k} = \sum_k \delta_{p,k} w_{j,k}. \tag{A.18}$$

Using equation A.15 - A.18, the modification of the weight can be calculated with

$$\Delta_p w_{i,j} = \eta(t_{p,j} - o_{p,j}) \cdot f'(\sum_i w_{i,j} o_{p,i}) o_{p,i} \tag{A.19}$$

for output units and with

$$\Delta_p w_{i,j} = \eta(\sum_k \delta_{p,k} w_{j,k}) \cdot f'(\sum_i w_{i,j} o_{p,i}) o_{p,i} \tag{A.20}$$

for other units.

A.5. TPC dE/dx Plots

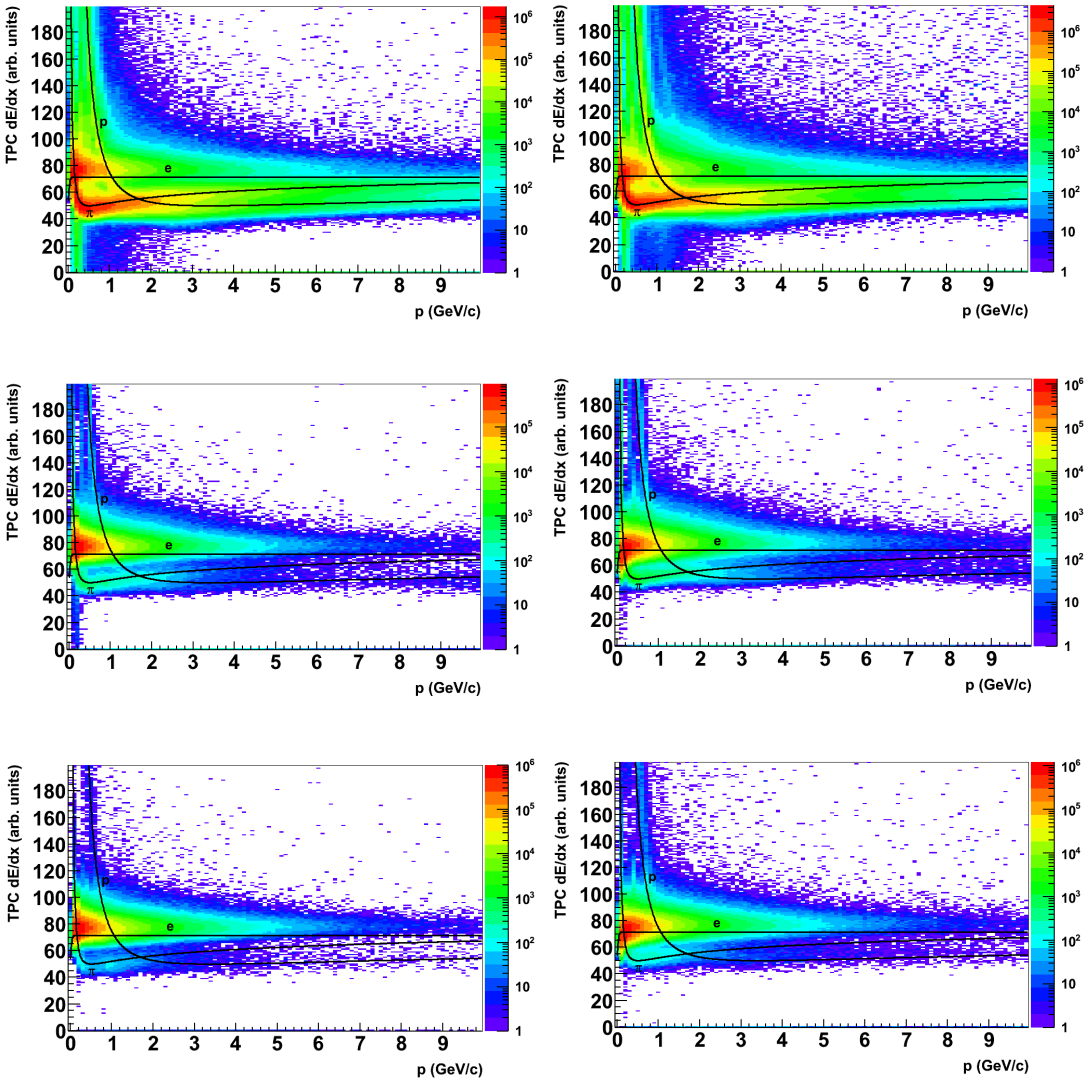


Figure A.5.: TPC dE/dx for negative V0-daughters. Left: For a Monte-Carlo simulation. Right: For measured data. From top to bottom: negative V0-daughters of all V0s, negative daughters of all V0s which are identified as photons using the neural network with input from `AliESDv0KineCuts` and `NextV0`, respectively. The Bethe-Bloch lines are shown for electrons (e), protons (p), and pions (π) to guide the readers eye. It can be seen that simulation and measured data are very similar and that the identification of photons from V0s is a little bit better in the simulation than in the measured data, since less pions remain in the left plots.

A.6. Invariant Mass Plots

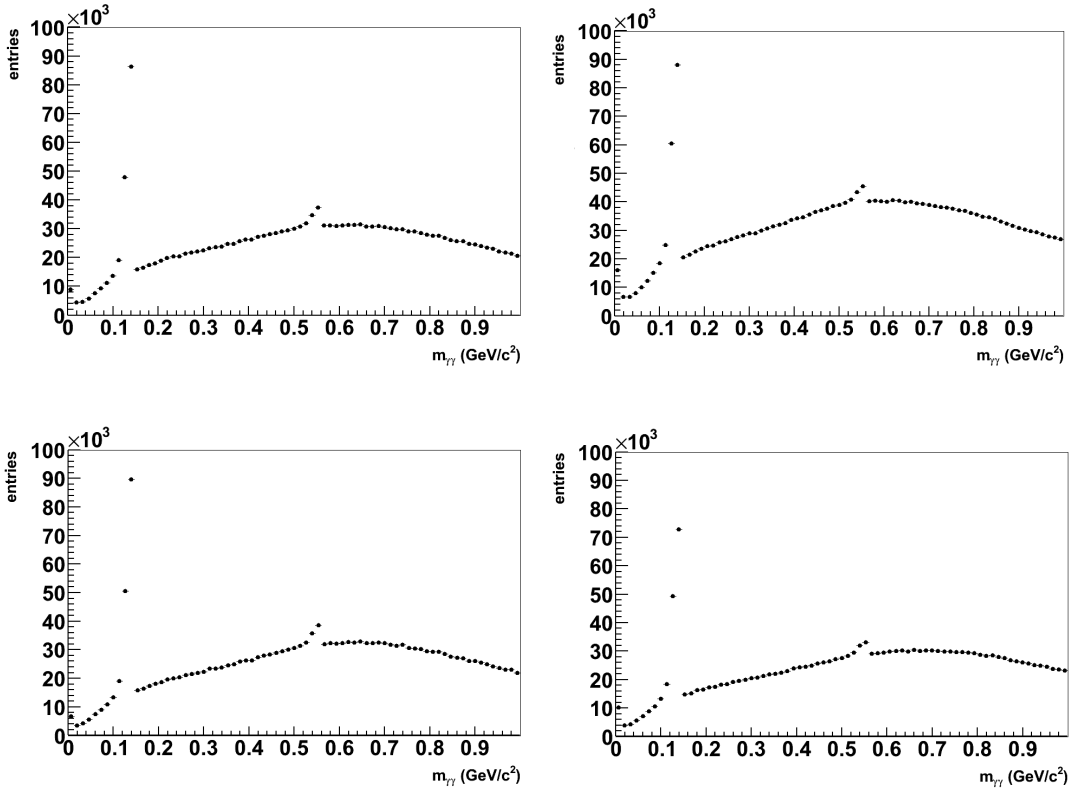


Figure A.6.: Invariant mass distribution of all photon pairs. The photons are identified by using the neural network with input from `AliESDv0KineCuts` (upper plots) and `NextV0` (lower plots). The pseudorapidity of both photons has to be in the acceptance of the central barrel, $|\eta| \leq 0.9$, and the V0 has to be found by the On-the-fly-Finder. 4.24×10^8 minimum bias Monte-Carlo events (left plots) and 4.43×10^8 measured minimum bias events (right plots) are used, respectively. It can be seen that π^0s can be identified using the neural network for both simulated and measured data.

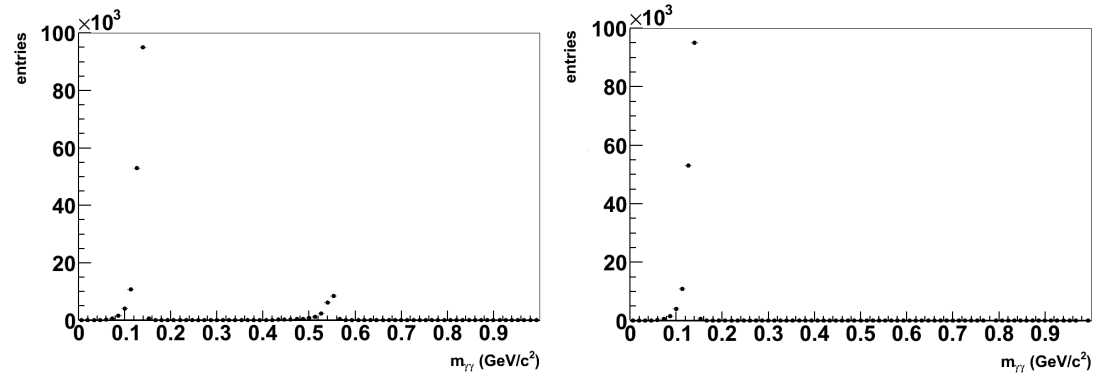


Figure A.7.: Invariant mass distribution of all photon pairs with same mother (left) and same π^0 as mother (right). The photons are identified by using the PDG code of the particles. The pseudorapidity of both photons has to be in the acceptance of the central barrel, $|\eta| \leq 0.9$, and the V0 has to be found by the On-the-fly-Finder. 4.24×10^8 minimum bias Monte-Carlo events are used. It can be seen that for photon pairs with the same mother, no combinatorial background occurs and that photon pairs with the same π^0 as mother have an invariant mass of $m_{\gamma\gamma} \approx 0.135$ GeV/c^2.

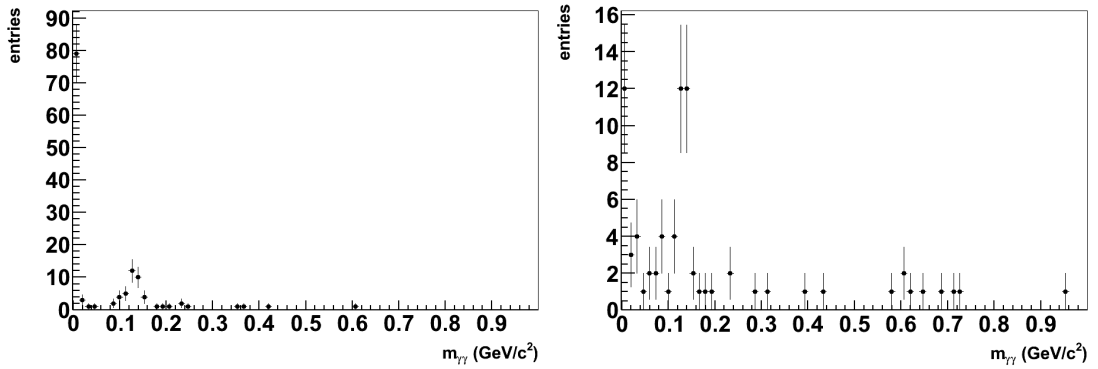


Figure A.8.: Invariant mass distribution of all photon pairs with $|\vec{p}_T(\gamma_1) + \vec{p}_T(\gamma_2)|$ is greater than 10 GeV/c, the distance of the two photons in (η,φ) is smaller than $R = 0.05$, and $|\eta(\gamma_{1,2})|$ is smaller than 0.9. The V0 has to be found by the On-the-fly-Finder. Left: Using the neural network with input from `AliESDv0KineCuts`. Right: Using the neural network with input from `NextV0`. 4.43×10^8 measured minimum bias events are used (LHC10b.pass2, LHC10c.pass2, LHC10d.pass2, LHC10e.pass2). It can be seen that background is larger compared to simulation data (figure 7.8), but decay photons can still be identified.

A.7. Input Variables for the two Neural Networks to Identify Photons from V0s

The seventeen input variables of the neural network, which will be compared with the method `AliESDv0KineCuts`, are

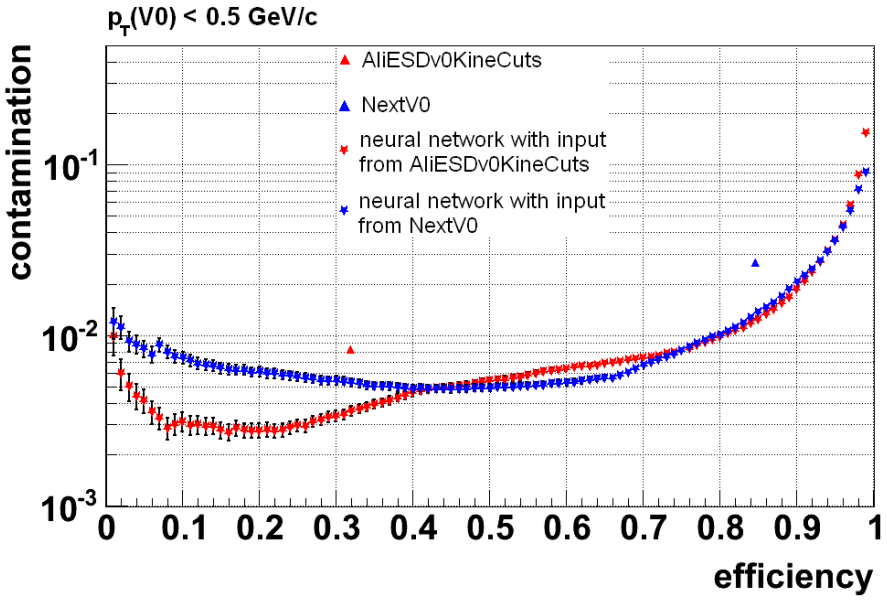
- the invariant mass of the V0 for e^-/e^+ as V0-daughter,
- the cosine of the pointing angle of the V0,
- the DCA of the two combined tracks,
- the distance of the V0 to the primary vertex for z=0,
- Ψ_{Pair} of the V0,
- χ^2/NDF of the V0,
- the number of clusters in the TPC of the positive track,
- the number of clusters in the TPC of the negative track,
- χ^2 per TPC cluster of the positive track,
- χ^2 per TPC cluster of the negative track,
- found number of TPC clusters divided by the findable number of TPC clusters of the positive track,
- found number of TPC clusters divided by the findable number of TPC clusters of the negative track,
- the kink-index of the positive track,
- the kink-index of the negative track,
- the Armenteros variables α and q_T of the V0,
- the charge of the positive track times the charge of the negative track.

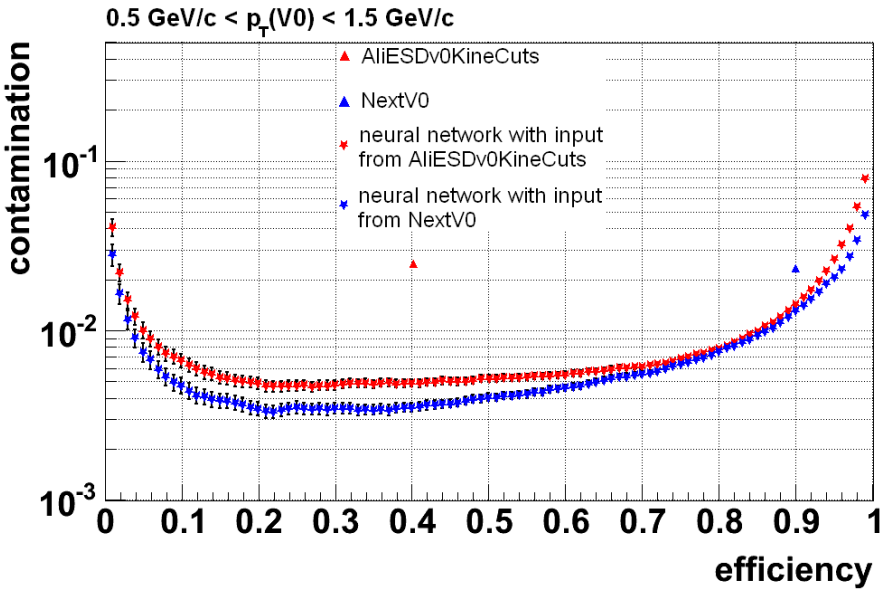
The twenty-six input variables of the neural network, which will be compared with the method `NextV0`, are

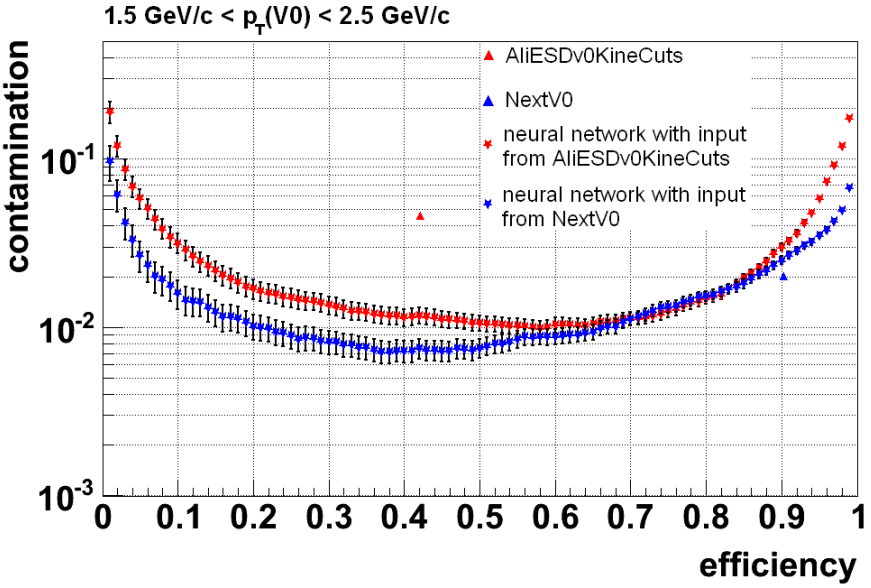
- the radial distance of the V0 to the primary vertex for z=0,

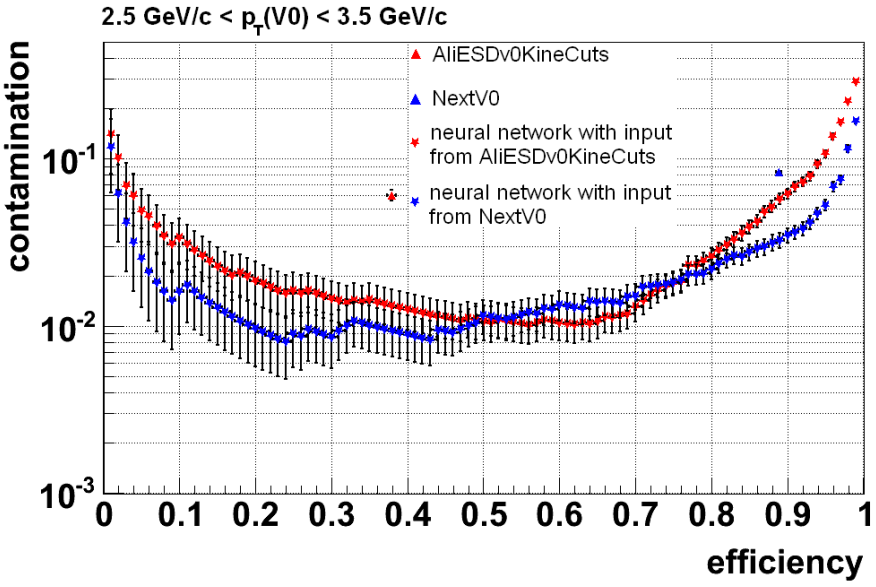
- the absolute value of the z position of the V0,
- ($|z|$·tan(2 arctan($e^{-0.9}$)))-7 (some regions of the detector are not used for reconstruction in the method `NextV0`)
- the number of sigmas distance of the dE/dx of the positive/negative V0-daughter to the $e^{-/+}$ dE/dx for the given momentum of the particle,
- the number of sigmas distance of the dE/dx of the positive/negative V0-daughter to the $\pi^{-/+}$ dE/dx for the given momentum of the particle,
- the number of sigmas distance of the dE/dx of the positive/negative V0-daughter to the $K^{-/+}$ dE/dx for the given momentum of the particle,
- the number of sigmas distance of the dE/dx of the positive/negative V0-daughter to the proton dE/dx for the given momentum of the particle,
- the momentum of the positive V0-daughter,
- the momentum of the negative V0-daughter,
- the Armenteros variable q_T of the V0,
- the p_T of the V0,
- the number of contributors of the primary vertex of the event,
- the probability that the negative particle is an electron (using the TPC),
- the probability that the positive particle is a positron (using the TPC),
- the number of clusters in the TPC of the positive track,
- the number of clusters in the TPC of the negative track,
- the p_T of the positive V0-daughter,
- the p_T of the negative V0-daughter,
- χ^2/NDF of the V0,
- η of the V0, the positive and the negative V0-daughter.

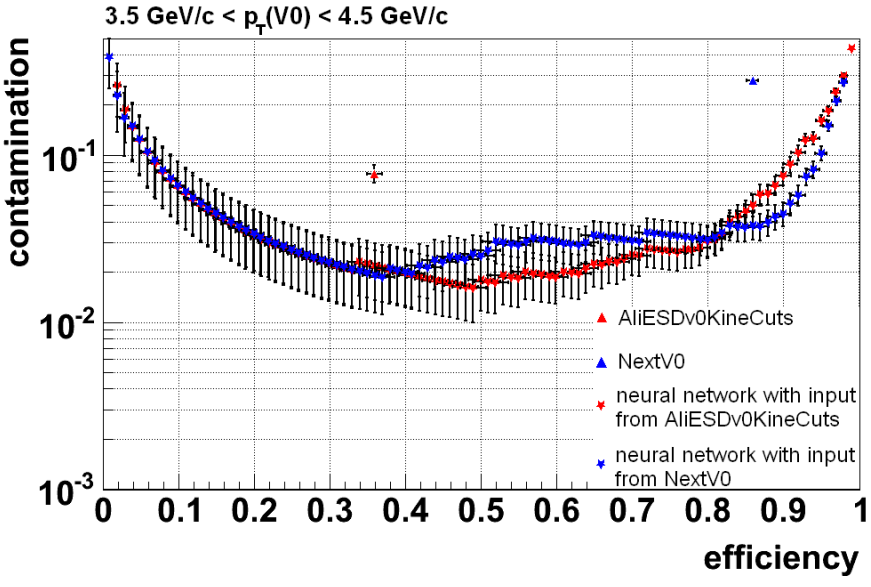
A.8. Performance Plots of the two Neural Networks to Identify Photons from V0s

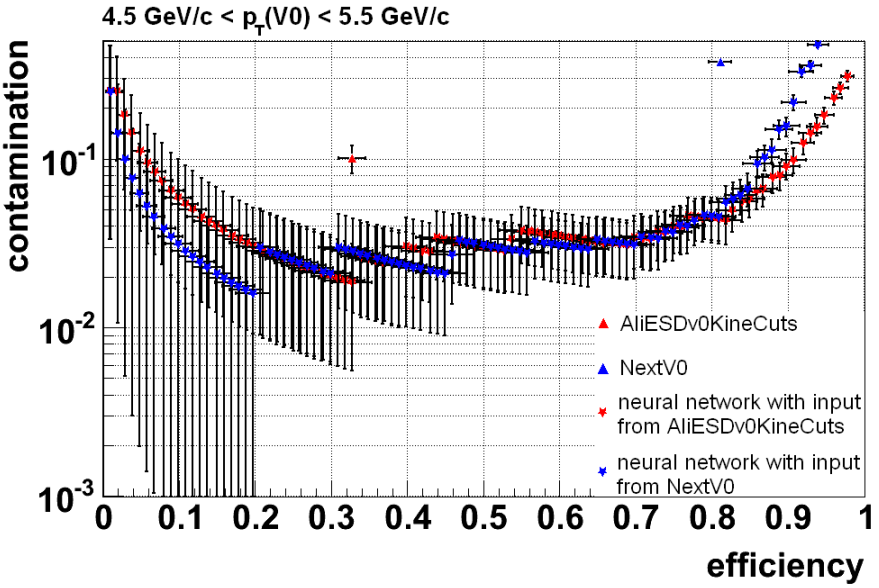


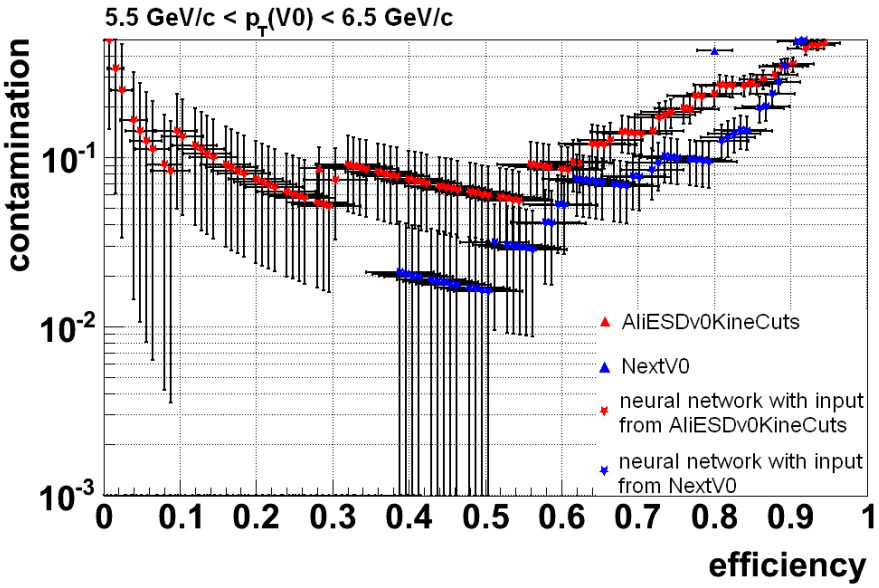


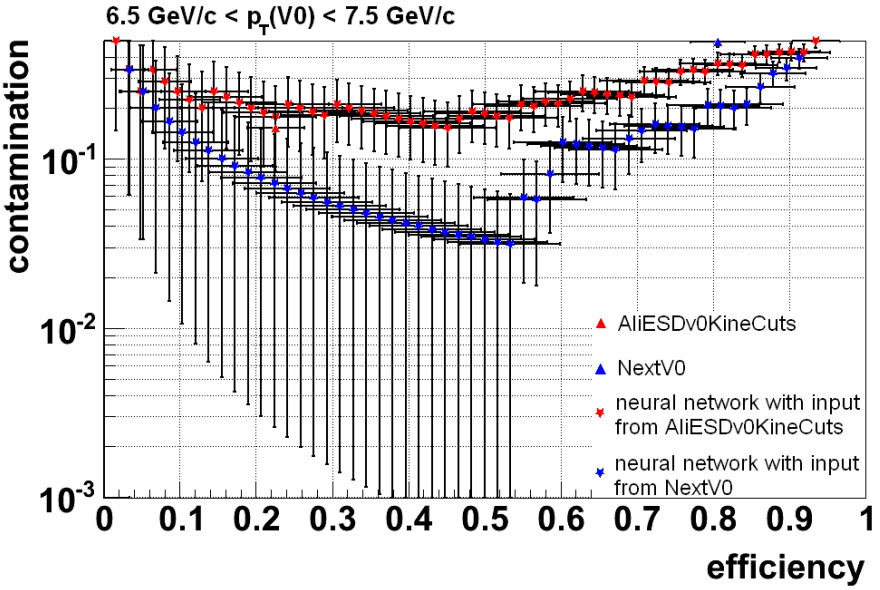

1.5 GeV/c < p_T(V0) < 2.5 GeV/c
AliESDv0KineCuts
NextV0
neural network with input from AliESDv0KineCuts
neural network with input from NextV0
contamination
efficiency


2.5 GeV/c < p_T(V0) < 3.5 GeV/c
AliESDv0KineCuts
NextV0
neural network with input from AliESDv0KineCuts
neural network with input from NextV0
contamination
efficiency


3.5 GeV/c < p_T(V0) < 4.5 GeV/c
AliESDv0KineCuts
NextV0
neural network with input from AliESDv0KineCuts
neural network with input from NextV0
contamination
efficiency


4.5 GeV/c < p_T(V0) < 5.5 GeV/c
contamination
efficiency
AliESDv0KineCuts
NextV0
neural network with input from AliESDv0KineCuts
neural network with input from NextV0


5.5 GeV/c < p_T(V0) < 6.5 GeV/c
contamination
efficiency
AliESDv0KineCuts
NextV0
neural network with input from AliESDv0KineCuts
neural network with input from NextV0


6.5 GeV/c < p_T(V0) < 7.5 GeV/c
contamination
efficiency
AliESDv0KineCuts
NextV0
neural network with input from AliESDv0KineCuts
neural network with input from NextV0

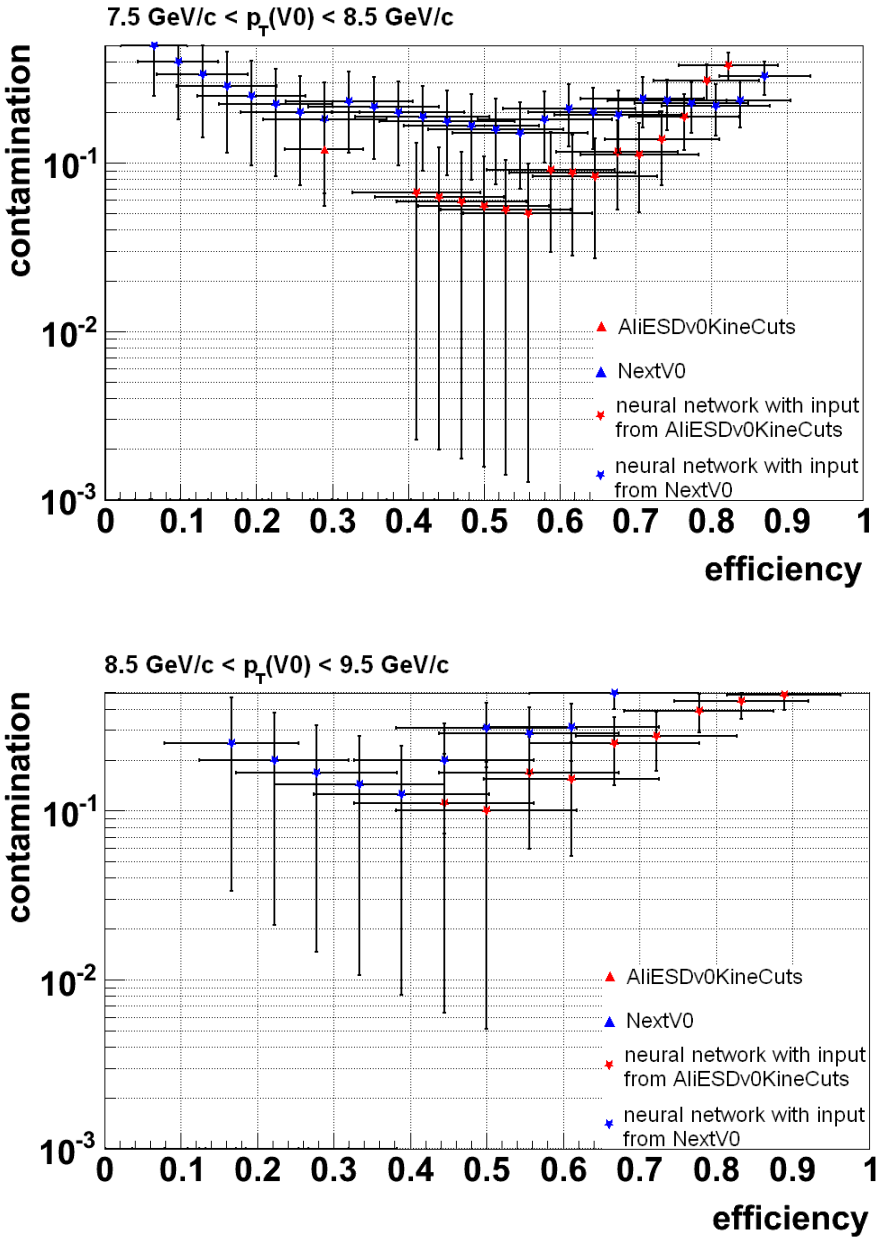


Figure A.9.: Performance of the method `AliESDv0KineCuts` and the method `NextV0` compared with the performance of the associated neural network for identifying γs from V0s. The contamination of the signal is plotted against the signal efficiency for different p_T of the V0. Only statistical uncertainties are shown.

A.9. Distribution of the Input Variables

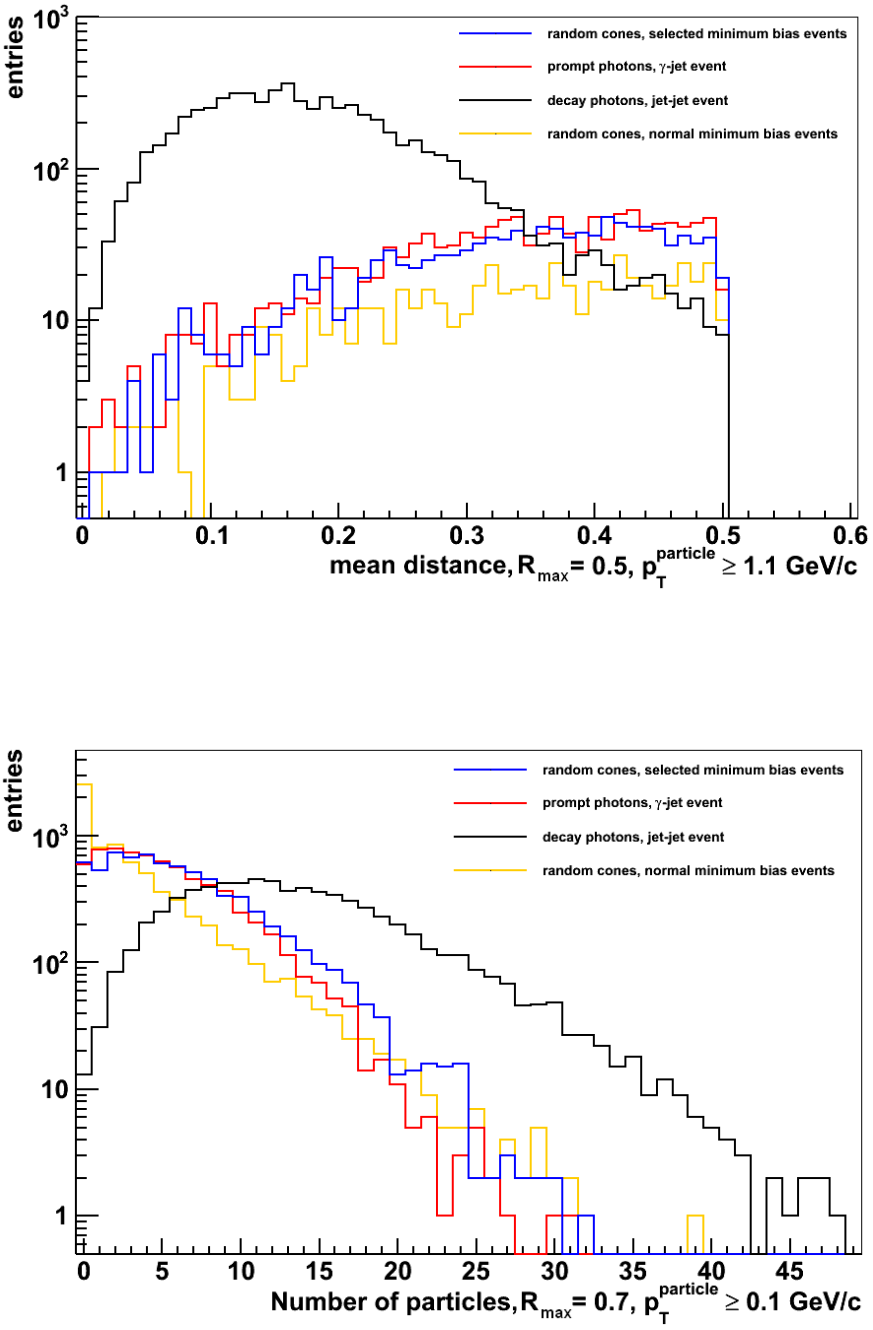


Figure A.10.: Distribution of two input variables from section 6.2 for γ-jet events (red), selected minimum bias events (blue), normal minimum bias events (yellow), and jet-jet events (black).

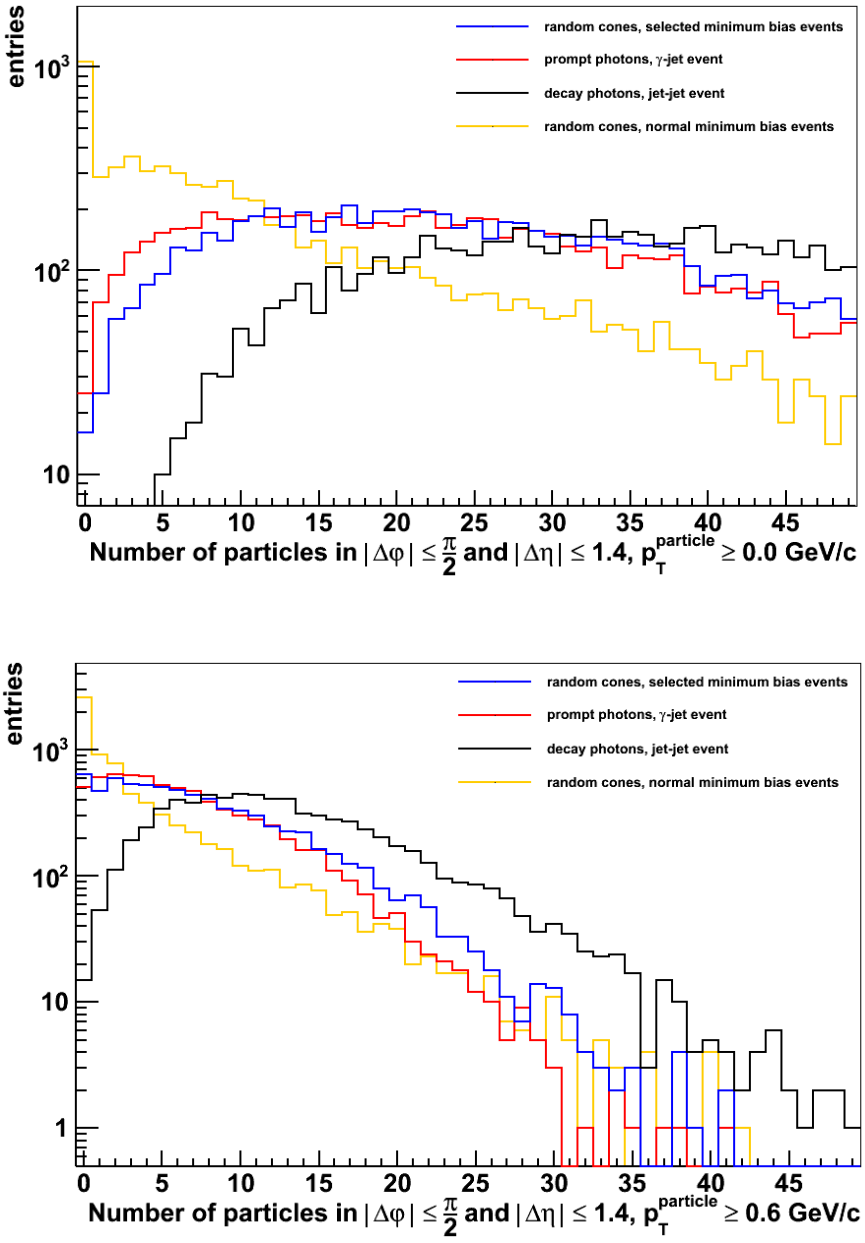


Figure A.11.: Distribution of two input variables from section 6.2 for γ-jet events (red), selected minimum bias events (blue), normal minimum bias events (yellow), and jet-jet events (black).

Bibliography

[ALI95] ALICE collaboration. *ALICE: Technical Proposal for A Large Ion Collider Experiment at the CERN LHC*. CERN/LHCC-95-71, 1995.

[ALI03] ALICE collaboration. *Definition of the ALICE Coordinate System and Basic Rules for Sub-detector Components Numbering*. ALICE-INT-2003-038, 2003.

[ALI04] ALICE collaboration. *Physics Performance Report Volume I*. J. Phys. G: Nucl. Part. Phys. 30 No 11 1517-1763, 2004.

[ALI06] ALICE collaboration. *Physics Performance Report Volume II*. J. Phys. G: Nucl. Part. Phys. 32 1517-1763, 2006.

[ALI08] ALICE collaboration. *The ALICE experiment at the CERN LHC*. JINST, 3:S08002, 2008.

[ALI10] ALICE collaboration. *Suppression of Charged Particle Production at Large Transverse Momentum in Central Pb-Pb Collisions at* $\sqrt{s_{NN}}$ *= 2.76 TeV*. arxiv:1012.1004.

[Ali11] ALICE collaboration. *Rapidity and transverse momentum dependence of inclusive J/*Ψ *production in pp collisions at* $\sqrt{s}$ *= 7 TeV*. doi:10.1016/j.physletb.2011.09.054.

[ATL99] ATLAS collaboration. *ATLAS Detector and Physics Performance Technical Design Report, Volume II*. CERN/LHCC-99-015, 1999.

[Bal05] Gustavo Conesa Balbastre. *Identification of particles and hard processes with the spectrometer PHOS of the ALICE experiment*. Ph.D. thesis, University of Valencia, 2005.

[Bet09] S. Bethke. *The 2009 World Average of* α_s. arXiv:0908.1135 [hep-ph].

[Bjo83] J. D. Bjorken. *Highly relativistic nucleus-nucleus collisions: The central rapidity region*. Phys. Rev. D Volume 27, 1983.

[BM07] P. Braun-Munzinger and J. Stachel. *The quest for the quark-gluon plasma.* Nature Vol. 448, 2007.

[Boe02] Volker Börchers. *Pioncaré-Kovariante Parton Kaskade.* Dissertation, Institut für Theoretische Physik, Universität Bremen, 2002.

[Cab75] N. Cabibbo and G. Parisi. *Exponential Hadronic Spectrum and Quark Liberation.* Phys. Lett. B59 (1975) 67.

[CBM07] S. Gorbunov and I. Kisel *Reconstruction of decayed particles based on the Kalman Filter.* CBM-SOFT-note-2007-003, 2007.

[CER11] *The four main LHC experiments.* CERN-AC-9906026, 2011.

[Cha32] J. Chadwick. *Possible Existence of a Neutron.* Nature, 1932.

[CMS06] CMS collaboration. *CMS Physics Technical Design Report, Volume II: Physics Performance.* CERN/LHCC-2006-021, 2006.

[Dal08] John Dalton. *A New System of Chemical Philosophy Part I.* 1808.

[Die06] Thomas Dietel. *Azimuthal Correlations of High-pT Photons and Hadrons in Au+Au Collisions at STAR.* Ph.D. thesis, Institut für Kernphysik, Frankfurt am Main, 2006.

[EMC08] ALICE collaboration. *ALICE Electromagnetic Calorimeter Technical Design Report.* CERN/LHCC-2008-014.

[FWD04] ALICE collaboration. *Technical Design Report of the ALICE Forward Detectors: FMD, T0 and V0.* CERN/LHCC-2004-025, 2004.

[GEA3] *GEANT - Detector Description and Simulation Tool.* http://wwwasd.web.cern.ch/wwwasd/geant/

[Gel64] M. Gell-Mann. *A schematic model of baryons and mesons.* Physical Letter, 1964.

[GO09] J.F. Große-Oetringhaus. *Measurement of the Charged Particle Multiplicity in Proton-Proton Collisions with the ALICE Detector.* Ph.D. thesis, Institiut für Kernphysik, Münster, 2009.

[Hal84] F. Halzen, A. D. Martin. *Quarks and Leptons: An Introduction Course in Modern Particle Physics.* John Wiley and Sons, 1984.

[HMPID98] ALICE collaboration. *ALICE Technical Design Report of the High Momentum Particle Identification Detector*. CERN/LHCC-98-19.

[ITS99] ALICE collaboration. *Technical Design Report of the ALICE Inner Tracking System*. CERN/LHCC-99-012, 1999.

[Kar02] F. Karsch. *Lattice QCD at high temperature and density*. Lect. Notes Phys., 583:209, 2002.

[KB04] C. Klein-Bösing. *Production of Neutral Pions and Direct Photons in Ultra-Relativistic Au + Au Collisions*. Ph.D. thesis, Institut für Kernphysik, Münster, 2004.

[Kun95] M.Kunze. *Über die Anwendung Neuronaler Systeme in der Teilchenphysik*. Habilitationsschrift, Universität Bochum, 1995.

[LHCb05] LHCb collaboration. *The LHCb experiment: status and expected physics performance*. CERN/LHCB-2005-068, 2005.

[LHCf05] ATLAS collaboration. *Technical Proposal for the CERN LHCf Experiment: Measurement of Photons and Neutral Pions in the Very Forward Region of LHC*. CERN/LHCC-2005-032, 2005.

[LHC11] *CERN faq - LHC, the guide*. CERN-Brochure-2008-001-Eng, 2011.

[LIP91] M.D. Richard, R.P. Lippmann. *Neural Network Classifiers Estimate Bayesian a posteriori Probabilities*. Neural Computation, Vol. 3, No. 4, Pages 461-483, 1991.

[Ott96] R. Ottenhues. *Teilchenklassifizierung im Bleiglaskalorimeter mit künstlichen neuronalen Netzen*. Diplomarbeit, Institut für Kernphysik, Münster, 1996.

[PDG11] Particle Data Group. *http://pdg.lbl.gov/*. 2011.

[PHO99] ALICE collaboration. *Technical Design Report of the ALICE Photon Spectrometer*. CERN/LHCC-99-004, 1999.

[Pov09] Povh, Rith, Scholz, Zetsche. *Teilchen und Kerne*. Springer-Verlag Berlin Heidelberg, 2009.

[PMD99] ALICE collaboration. *Technical Design Report of the Photon Multiplicity Detector (PMD)*. CERN/LHCC-99-032, 1999.

[PYT06] Torbjorn Sjostrand (Lund University), Stephen Mrenna (Fermilab), Peter Skands (Fermilab). *PYTHIA 6.4 Physics and Manual*. arXiv:hep-ph/0603175v2

[ROO] *ROOT: An Object-Oriented Data Analysis Framework*. http://root.cern.ch.

[Rum85] D.E. Rumelhart, G.E. Hinton, R.J. Williams. *Learning Internal Representations by Error Propagation*. Published in D. E. Rumelhart and J. L. McCelland, Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Volume 1: Foundations Cambridge, MA: MIT Press, 1985.

[Rut11] E. Rutherford. *The Scattering of α and β Particles by Matter and the Structure of the Atom*. Philosophical Magazine, 1911.

[Sah10] B. Sahlmüller *Probing Hot and Dense Matter: Measurement of Neutral Mesons and Direct Photons in Ultrarelativistic Au+Au Collisions*. Ph.D. thesis, Institut für Kernphysik, Münster, 2010.

[Sam05] L. Sammer. *Merkmalbasierte Zeichenerkennung mittels neuronaler Netze*. Diplomarbeit, Mathematisches Institut, Universität Bayreuth, 2005.

[Sch97] A. Scherer. *Neuronale Netze: Grundlagen und Anwendungen*. Vieweg, Braunschweig, 1997.

[TOF00] ALICE collaboration. *Technical Design Report of the Time-Of-Flight System*. CERN/LHCC-2000-012, 2000.

[TOT04] ALICE collaboration. *TOTEM Technical Design Report*. CERN/LHCC-2004-002, 2004.

[TPC00] ALICE collaboration. *Technical Design Report of the ALICE Time Projection Chamber*. CERN/LHCC-2000-01, 2000.

[TRD01] ALICE collaboration. *Technical Design Report of the ALICE Transition Radiation Detector*. CERN/LHCC-2001-021, 2001.

[Wil04] A. Wilk. *Elektronen-Pionen-Separation im ALICE TRD*. Diplomarbeit, Institut für Kernphysik, Münster, 2004.

[Wil09] A. Wilk. *Particle Identification Using Artificial Neural Networks with the ALICE Transition Radiation Detector*. Ph.D. thesis, Institut für Kernphysik, Münster, 2009.

[ZDC99] ALICE collaboration. *Technical Design Report of the Zero Degree Calorimeter (ZDC).* CERN/LHCC-99-05, 1999.

[Zwe64] G.Zweig. *An SU(3) model for strong interaction symmetry and its breaking.* CERN-TH-401.

Danksagung

Abschließend möchte ich mich bei allen Menschen bedanken, die zum Gelingen dieser Arbeit beigetragen haben.

Als erstes danke ich Herrn Prof. Dr. J. P. Wessels für das interessante Thema meiner Diplomarbeit und für spannende Fahrten zur GSI und zur ALICE Physiks Week nach Finnland.
Mein großer Dank gilt Dr. Thomas Dietel für die exzellente Betreuung meiner Arbeit und das geduldige Beantworten meiner Fragen. Weiterhin möchte ich mich bei Dr. Alexander Wilk, Uwe Westerhoff, Markus Heide, Linus Feldkamp, Martin Wilde, Bastian Bathen und Dr. Christian Klein-Bösing bedanken, die immer ein offens Ohr für Probleme aller Art hatten.
Für das Korrekturlesen meiner Arbeit danke ich Uwe Westerhoff, Markus Heide, Martin Wilde, Bastian Bathen, Jennifer Bersch, Dr. Thomas Dietel und Felicitas Eilting.
Dr. Alexander Wilk, Annika Passfeld, Martin Kohn, Markus Tegeder, Uwe Westerhoff, Markus Heide, Linus Feldkamp und Martin Wilde, danke ich für die gute Stimmung im Büro während meiner gesamten Arbeitszeit.
Ich danke der gesamten Arbeitsgruppe für die tolle Arbeitsatmosphäre und die interessanten Gespräche auch außerhalb der Arbeit.
Zuletzt möchte ich mich sehr bei meinen Eltern bedanken, die mir mit ihrem Zuspruch und ihrer finanziellen Unterstützung dieses Studium ermöglicht haben.

Eigenständigkeitserklärung

Hiermit versichere ich, diese Arbeit selbstständig verfasst und keine weiteren als die angegebenen Hilfsmittel und Quellen verwendet zu haben.

Münster, 27.01.2012

Thomas Keuter