Electron identification in Au+Au collisions at 1.23 GeV/u using multivariate analysis

Szymon Harabasz for the HADES Collaboration
Technische Universität Darmstadt, Schlossgartenstraße 9, 64289 Darmstadt, Germany
Jagiellonian University, ul. Reymonta 4, 30-059 Kraków, Poland
E-mail: s.harabasz@gsi.de

Abstract. Au+Au collisions at a beam kinetic energy of 1.23 GeV/u have been measured by HADES in 2012. Lepton identification in this experiment has been done using a multivariate algorithm based on an artificial neural network. In the proceedings, details of the identification method and its assessment in terms of purity of the final lepton sample are presented. The obtained purity reaches 95% and the amount of identified electrons and positrons is sufficient to perform further steps of the physics analysis with $e^+e^-$ pairs.

1. Introduction
Dilepton spectroscopy allows to investigate the properties of hadrons embedded in a strongly interacting medium. Dileptons are not subject to the strong force, which makes them able to leave a region of a nuclear reaction practically undistorted by any final-state interactions. Moreover, as leptons are radiated during all stages of nuclear collisions, they provide an insight into the hot and dense fireball as well as into first-chance interactions and the freeze-out. This is in contrast to hadrons, whose multiplicities and kinematical properties are established in the last stage, when hadronic interactions vanish. Considering that the dilepton invariant mass spectrum is also closely related to the issue of modifications of light vector meson in nuclear matter, dilepton spectroscopy offers a possibility to investigate the question of mass generation. However, production of dileptons is suppressed by a factor of the order of $\alpha^2$ ($\alpha$ - fine structure constant), thus their measurements are highly demanding from an experimental point of view.

An enhanced yield of dileptons with masses below the vector meson (i.e. $\rho^0$ and $\omega$ meson) pole mass appears to be a general feature of heavy-ion reactions, from bombarding energies as low as 1 GeV/u, studied by the former DLS experiment [1] at the Bevalac, through the range of SPS energies (40–158 GeV/u) used by the CERES [2][3] and NA60 [4] experiments at CERN, up to the highest energies (with $\sqrt{s_{NN}} = 200$ GeV) employed by the PHENIX [5] and STAR [6] experiments (the latter performed also a beam energy scan [7]) at the RHIC. This enhancement is defined as the excess of the measured pair yield over the summed-up cocktail of dileptons from long-lived sources, namely the decays of $\pi^0$, $\eta$, and $\omega$ mesons. It is hence expected to probe the early phases of the collision.

The High-Acceptance DiElectron Spectrometer (HADES) operates at the GSI Helmholtzzentrum für Schwerionenforschung in Darmstadt where it has started a systematic investigation of dilepton production in the SIS/Bevalac energy regime of 1–2 GeV/u.
One of main results is shown in Fig.1, which compares dilepton invariant mass spectra in heavy-ion and nucleon-nucleon collisions. It appears, that while the light system C+C may be treated as a superposition of individually colliding nucleons [8], this is not any longer true for heavier Ar+KCl system, where there is an additional source of dilepton radiation emerging above $M_{ee} = 150 \text{ MeV}/c^2$. Current results show the scaling of this so called ”excess” as $\propto A_{\text{part}}^{1.4}$ with the number of nucleons participating in the interaction [9].

Figure 1. Invariant mass spectrum of dielectrons measured in Ar+KCl at 1.76 GeV/u compared to a reference spectrum from elementary collisions. Opening angle and single particle momentum cuts are applied and a ”trivial” component of $\eta$ (with $T_{1/2}$ longer than hot and dense matter stage) is identified and subtracted.

To confirm this result and to further investigate dilepton production enhancement, an experiment with the heaviest system available at SIS18, Au+Au at 1.23 GeV/u was performed in April–May 2012. In this experiment, due to a large hadron multiplicity, traditional methods of electron identification have been proved to be insufficient. In this paper, results of a multivariate lepton identification in Au+Au experiment are presented.

2. The HADES Experiment

HADES is a charged-particle detector consisting of a 6-coil toroidal magnet centered on the beam axis and six identical detection sections covering polar angles between $18^\circ$ and $85^\circ$ with nearly complete azimuthal coverage. The main component serving for electron and positron selection is a hadron-blind Ring-Imaging Cherenkov detector (RICH). Further particle identification power is provided by combined effect of (i) the time-of-flight (TOF) measurement in a plastic scintillator and Resistive-Plate Chamber (RPC) walls, (ii) the electromagnetic shower characteristics observed in a PreShower detector, and (iii) the energy-loss signals from the scintillators of the TOF wall as well as from the four planes of drift chambers serving as tracking stations. A detailed description of the HADES setup can be found in [10].

In the experiment discussed here, a gold beam with a kinetic energy of 1.23 GeV/u was used to bombard a 15-fold segmented gold target (for details see Tab.1).

3. Electron identification using multivariate analysis

In the data analysis, $e^+$ and $e^-$ are identified by applying appropriate selection cuts to the RICH ring observables, time-of-flight, PreShower and energy loss signals. Particle momenta were obtained by tracking the charged particles through the HADES magnetic field.

In a first attempt one could impose one-/two-dimensional conditions (so called hard cuts) on various observables in order to suppress hadrons and to keep as many electrons as possible. However, to find the optimal set of cuts is very difficult. Applying or modifying a requirement on one quantity changes what would be regarded as best condition on all the others. Another difficulty is that a one- or two-dimensional cut ignores all other dimensions (other variables). This makes the selection less flexible. In addition, it is possible to accept a particle, which just
Table 1. Basic facts about April-May 2012 Au+Au beam-time.

| Property                        | Value                          |
|---------------------------------|--------------------------------|
| Beam kinetic energy            | 1.23 GeV/u                     |
| Total beam-on-target time      | 557 hours                      |
| Beam intensity                 | $(1.2 - 1.5) \cdot 10^6$ ions/s|
| Trigger rate                   | 8 kHz                          |
| Data rate                      | 200 MB/s                       |
| Total events collected         | $7.3 \cdot 10^9$               |
| Total data collected           | 140 TB                         |

slightly passes all tests and on the other hand reject a particle which meets well all conditions but one.

![Figure 2](image1.png)

**Figure 2.** Neural network output (denoted as MLP for multilayer perceptron) versus particle’s momentum for tracks falling into TOF scintillator wall.

![Figure 3](image2.png)

**Figure 3.** Particle’s velocity versus momentum for (a) all particles, (b) particles, for which the neural network output function has values larger than 0.6. On the left side there are particles measured in the RPC detector, on the right side – in the TOF scintillator.

All these difficulties may be addressed by classification algorithms of multivariate analysis, which allow to create a multidimensional decision boundary optimizing it in all dimensions (variables) at once. In the current work, the implementation available as a part of the ROOT data analysis framework (TMVA [11]) was used. From the number of algorithms offered, the Multilayer Perceptron (MLP) model of artificial neural network was chosen.

The algorithm has to be trained with well defined samples of signal and background particles, to determine characteristic properties of the two and be able to classify tracks of the whole data sample later on. There are two challenges in training the neural network to give the correct response: one is the selection of pure signal and background training samples. The other one is the choice of variables with the highest separation power between hadron and lepton tracks. The
signal and background samples are taken from experimental data (not from simulation), using a strong cut on the matching of track to RICH ring, which has sufficiently large separation power to generate a pure electron sample. The matching can be quantified by track-ring matching quality:

\[ C = \sqrt{(\Delta \theta_{\text{track-ring}})^2 + (\Delta \phi_{\text{track-ring}} \cdot \sin \theta)^2}. \]

Electron tracks are well matched to rings and in their case \( C \) distribution is peaked around 0 (cf. Fig.4 (b)), while for any other particle species it has a broad shape. For the training of the neural network tracks with \( C < 0.5 \)◦ are treated as signal, while those with \( C > 7\)◦ are regarded as background.

**Figure 4.** Spatial correlation between MDC tracks and RICH rings: (a) experimental data (green), rotated RICH (background - orange), difference of the two (signal - dark blue hatched); UrQMD simulation without (b) and with (c) rotated RICH, depicting different particle types components.

The neural network output is shown in Fig.2 as a function of particle momentum. The closer to value 1, the more certainly the particle is a lepton; events closer to 0 are related to background. Finding a boundary which separates between signal and background is a matter of compromise between purity and efficiency of the identification. In the further analysis, tracks with MLP values smaller than 0.6 have been discarded. A comparison of velocity-momentum plots before and after applying this MLP cut is presented in Fig.3. It can be seen that most of hadrons have been removed and a significant statistics of electrons has been kept.

4. **Background estimations**

To assess the quality of electron identification, one has to estimate the amount of background still present in the sample. This can be done by adding 60° to the azimuthal position of each RICH ring (thus moving it to the neighboring sector of the spectrometer) and then perform the usual ring-track matching procedure and subsequent analysis. True matches, corresponding to leptons, should then be destroyed, but false i.e. random matches should be statistically reproduced in the spectra. The overlay of a distribution of not rotated events with the background prepared in such a way is shown in Fig.4 (a). In Fig.4 (b) and (c) one can compare contributions of different particle species from simulation (UrQMD [12][13], Au+Au at 1.25 GeV/u, \( b_{\text{max}} = 9 \) fm) analyzed in the same manner as the experimental data. The agreement of pions, protons and muons spectra between data and data with and without rotated rings confirms that the RICH rotating procedure allows to reproduce perfectly all components of background.

The difference between the original data and the reproduced background can be regarded as a signal. The purity of the electron sample may be then calculated as \( P = \frac{\text{signal}}{\text{signal+background}} \).
is shown in Fig. 5 for different momentum bins, compared to the best hard cut result obtained before the application of the multivariate analysis method. Global values are 84.8% for hard cut and 94.8% for multivariate analysis. A way to calculate the identification efficiency is currently under investigation, but comparing total number of identified leptons $2.45 \times 10^8$ with $1.71 \times 10^8$ from hard-cut analysis supports the superiority of the multivariate approach.

![Figure 5. Lepton sample purity in momentum bins, calculated as described in text, for multi-variate analysis using neural network approach compared to the best result obtained with hard cut analysis.](image)

5. Summary
The multilayer perceptron, which is one of the algorithms of multivariate analysis of the TMVA framework provides a unique way to obtain a pure and efficient sample of electrons in a strongly hadron dominated environment of the Au+Au at 1.23 GeV/u experiment. Such a sample is necessary to obtain reliable spectra of dileptons with properly treated calculation of the combinatorial background which has to be subtracted from the pair yield. Multivariate analysis is superior to the traditional hard cut approach and will be applied in further steps of dielectron analysis.

Acknowledgments
The collaboration gratefully acknowledges the following funding: INFN-LNS Catania (Italy); LIP Coimbra (Portugal): PTDC/FIS/113339/2009; SIP JUC Cracow (Poland): NN20198639; GSI Darmstadt (Germany): Helmholtz Alliance HA216/EMMI; TU Darmstadt (Germany): VH-NG-823, Helmholtz Alliance HA216/EMMI; HZDR, Dresden (Germany): 283286, 05P12CRGHE; Goethe-University, Frankfurt (Germany): Helmholtz Alliance HA216/EMMI, HIC for FAIR (LOEWE), GSI F&E, BMBF 06FY91001; TU München, Garching (Germany): BMBF 06MT7180; JLU Giessen (Germany): BMBF:05P12RGGHM; University Cyprus, Nicosia (Cyprus): UCY/3411-23100; IPN Orsay, Orsay Cedex (France): CNRS/IN2P3; NPI AS CR, Rez, (Czech Republic): MSMT LG 12007, GACR 13-06759S.

References
[1] Porter R J et al. (DLS Collaboration) 1997 Phys. Rev. Lett. 79 1229
[2] Agakichiev G et al. (CERES Collaboration) 2005 Eur. Phys. J. C 41 475
[3] Adamova D et al. (CERES Collaboration) 2003 Phys. Rev. Lett. 91 042301
[4] Arnaldi R et al. (NA60 Collaboration) 2006 Phys. Rev. Lett. 96 162302
[5] Adare A et al. (PHENIX Collaboration) 2010 Phys. Rev. C 81 034911
[6] Huang B et al. (STAR Collaboration) 2013 Nucl. Phys. A 904-905
[7] Guertz F et al. (STAR Collaboration) 2012 EPJ Web of Conferences 36
[8] Agakishiev G et al. (HADES Collaboration) 2010 Phys. Lett. B 690 118
[9] Agakishiev G et al. (HADES Collaboration) 2011 Phys. Rev. C 84 014902
[10] Agakishiev G et al. (HADES Collaboration) 2009 Eur. Phys. J. A 41 243-277
[11] Hoecker A et al. 2007 PoS ACAT 040 (Preprint physics/0709309)
[12] Bass S et al. 1998 Prog. Part. Nucl. Phys. 41 255 – 369
[13] Bleicher M et al. 1999 J. Phys. G 25 1859