Nuclear Science and Techniques (2026) 37:124
https://doi.org/10.1007/s41365-026-01943-0

Performance of the FARICH-based particle identification in realistic environment at charm superfactories using machine learning

Marina Chadeeva[1] · Platon Rogozhin[1] · Timofey Uglov[1,2]

Received: 8 July 2025 / Revised: 10 September 2025 / Accepted: 29 September 2025 / Published online: 15 April 2026

Abstract
A detailed study of particle identification by the Focusing Aerogel Ring-Imaging Cherenkov subsystem at the future charm superfactory detector is presented. A dedicated signal ring reconstruction algorithm was implemented in the detector simulation, considering realistic operating conditions. The algorithm performance was tested using single particles generated within the Aurora framework. Two boosted decision trees-based classifiers for particle identification were developed for moderate and the most conservative assumptions about photosensor noise levels. The approach is validated with the analysis of the $D^0 \to K\mu\nu_\mu$ decays, for which the systematic uncertainty and background contribution related to the π/μ separation performance can be minimized owing to the high efficiency of the particle identification algorithm.

Keywords Charm superfactories · Particle identification · Cherenkov detector · Boosted decision trees

1 Introduction

The Standard Model (SM)—a theory describing our universe at the most fundamental level—despite its obvious successes, is far from complete. The inability to describe the nature of 19 free parameters arising in the model and the failure to incorporate dark matter and dark energy phenomena motivate searches for effects beyond the SM, that is, New Physics (NP). There are two main approaches to NP detection. Collisions at the highest available energies, such as those in LHC experiments, are expected to yield new types of particles. Another way is to precisely measure the SM processes and search for indirect manifestations of the NP in interference processes such as CP violation or rare and forbidden particle decays. The latter approach requires a huge amount of accumulated data and outstanding detector performance. The new beam collision scheme, “crab-waist”, proposed by Raimondi [1] in 2006, revolutionized electron–position collider technologies, and opened the opportunity for the creation of flavor superfactories. The future charm superfactories, next-generation symmetric e^+e^- colliders, and their detectors are designed for precision measurements of the parameters of charmed hadrons and τ leptons. There are two competing projects, the Russian Super Charm-Tau Factory (SCTF) [2] and the Chinese Super Tau-Charm Facility (STCF) [3], which share physical tasks and are based on similar experimental techniques.

The lepton flavor universality (LFU) test is one of the most challenging tasks in the charm superfactories physical program. According to the SM, weak interactions are not sensitive to the lepton flavor, that is, in similar processes, weak matrix elements for all lepton flavors must be exactly the same. The difference in the decay branching ratios originates from the phase space and hadronic form factors. For the semileptonic D meson decays the observable to test LFU is the branching fractions ratio[1]:

$$\mathcal{R}_{\mu/e} = \frac{\mathcal{B}(D \to K\mu\nu_\mu)}{\mathcal{B}(D \to Ke\nu_e)}. \quad (1)$$

The accuracy of $\mathcal{R}_{\mu/e}$ measurement is strongly limited by the detector particle identification (PID) capabilities owing

This work was partially supported by the project “Mirror Laboratories” of the HSE University.

✉ Marina Chadeeva
mchadeeva@gmail.com

1 P. N. Lebedev Physical Institute of the RAS, 53, Leninskiy prospekt, Moscow, Russia 119991

2 Higher School of Economics, 20, Mjasnitskaya str., Moscow, Russia 101000

[1] Charge conjugation is implied throughout the paper.

to inevitable backgrounds, for example, from $D^0 \to K^-\pi^+\pi^0$ decays, which mimic the signal $D^0 \to K^-\mu^+\nu_\mu$ in the case of π/μ misidentification. The combined PID system for charm superfactories is aimed at providing very efficient K/π and π/μ separation within all reachable momentum ranges. In this study, the performance of the dedicated PID subsystem proposed for the SCTF was studied using single particles and semileptonic D meson decays.

2 Particle identification subsystem at charm superfactories

2.1 Detector technologies for charm superfactories

The detector for a charm factory is a general-purpose high-hermiticity magnetic spectrometer optimized for precision studies of the production and decay properties of various particles. The tracking system consists of an inner tracker aimed at providing superior spatial resolution at the interaction point (IP) and a main tracker for precise momentum measurement. For the inner tracker micropattern gaseous detector, complementary metal–oxide–semiconductor (CMOS) maps and time projection chamber options are considered. The technology for the main tracker is a drift chamber in a 1.0 T axial magnetic field provided by a superconducting solenoid. The particle identification system is based on Cherenkov light detection using DIRC-like time-of-flight or ring-imaging Cherenkov detector (RICH) detectors. Photons are detected using an electromagnetic calorimeter, which is built using pure CsI crystals to cope with high backgrounds. Muons and K_L mesons are identified in the instrumented (resistive plate chambers or organic scintillator) flux-return yoke. A detailed description of the SCTF and STCF detector designs can be found elsewhere [3, 4]. Unless explicitly stated otherwise, the SCTF project paradigm is assumed throughout this study.

2.2 Particle identification in the SCTF detector

Particle identification at the SCTF is based on several subsystems, and their contributions to PID and performance have been previously estimated using a fast simulation approach [5]. The dE/dx measurements inside the drift chamber are efficient for separating kaons (below ~ 700 MeV/c) and protons (below ~1200 MeV/c) from pions and muons. A CsI crystal electromagnetic calorimeter (with a crystal transversal size of approximately 1 Molière radius and a longitudinal depth of 1 nuclear interaction length) was used to separate hadrons from photons and electrons. Its contribution to muon–hadron separation is negligible for momenta <2 GeV/c, where hadronic showers (if any) can hardly be distinguished from minimum ionizing particle tracks. The main task of the muon system is the efficient identification of hard muons (with momenta >700 MeV/c).

For the middle-momentum range from ~0.5 GeV/c to ~ 1.5 GeV/c, a dedicated subsystem is proposed, namely the Focusing Aerogel Ring-Imaging Cherenkov (FARICH) detector. The FARICH radiator is composed of a few layers of aerogel with slightly different refraction indices, resulting in the focusing of Cherenkov photons in the thinner ring at the photodetector plane compared to that from a homogeneous radiator. Silicon photomultiplier matrices are considered as an option for photon detection devices. A detailed description of the concept and test beam results for the prototype can be found in Ref. [6].

2.3 Simulation and digitization of the FARICH subsystem

Dedicated software packages are being developed for charm factory detector simulations and data processing. As many other modern HEP projects, both Aurora package [7] for the SCTF and OSCAR package [8] for the STCF are based on the DD4hep framework [9]. Within the SCTF Geant4 [10] detector model, the FARICH system was placed between the tracker and electromagnetic calorimeter. Hereinafter, the simulations of the FARICH barrel part are discussed, which has the following layout: It is composed of 27 equivalent trapezoid blocks starting at 800 mm from the IP with a total thickness of 40 mm, and each block contains four aerogel layers (radiator). The layer of the sensitive detector material was separated from the radiator by a 200-mm air gap. The number of aerogel layers and their optical properties in the FARICH subdetector model were identical to those in Ref. [11]. Figure 1 shows a schematic view of the FARICH subsystem of the SCTF detector on the left and a sketch of the focusing effect from the multilayer aerogel in the middle.

Optical photon transport simulations are known to be extremely time-consuming. There are two simulation modes within the Aurora package: with and without the Geant4 optical photon functionality implemented for the FARICH subdetector. In this study, the signal reconstruction and particle identification performance were investigated using particle propagation from the interaction point with a 1 T magnetic field up to the FARICH subdetector and subsequent full Geant4 simulation of optical photon transport through the FARICH components. The vector of the particle momentum at the track endpoint in the tracker was obtained from the generator-level information and projected to the entrance point of the FARICH radiator.

The simulated response of the FARICH subsystem was produced by Cherenkov photons generated and propagated to sensitive detectors using Geant4. To obtain realistic hits, the quantum and geometrical efficiencies of photosensors

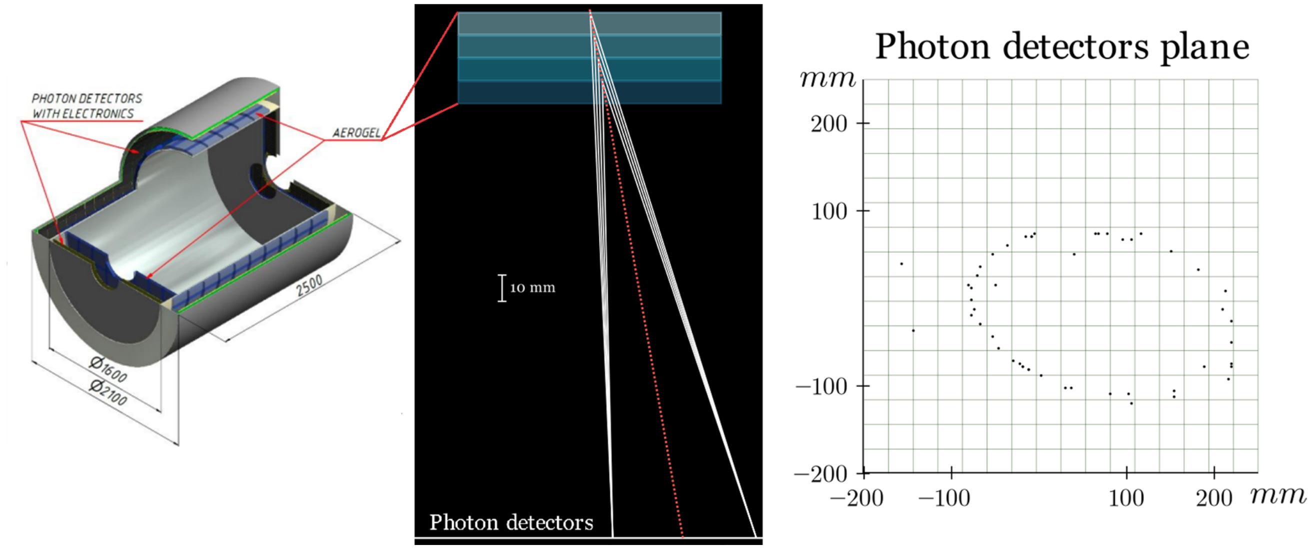

Fig. 1 (Color online) Left: a schematic view of the FARICH subdetector at the SCTF. Middle: FARICH design with four aerogel layers and Cherenkov photon path up to the photodetector plane. Right: FARICH photodetector plane with the signal from a simulated 1.5-GeV charged pions. The square grid shows the SiPM matrices, and each point corresponds to the signal from a single SiPM in the matrix. See text for details

were implemented as a digitization procedure of the photon detection: The Geant4-generated Cherenkov photon spectrum was convolved with the quantum efficiency curve and an additional factor to account for an effective geometrical efficiency. The energy dependence of the quantum efficiency was obtained from Ref. [12] for the KETEK 8×8 SiPM matrix (SiPM: silicon photomultiplier) proposed as one of the photosensor options. These photodetectors have pixel (SiPM) sizes of 3.36 mm ×3.36 mm and geometrical fill factors of ~88%. Subsequently, the photon arrival points in the detector plane were placed on a square grid with a size corresponding to the aforementioned pixel size. One or more photons registered in a single pixel forced the pixel to produce one hit. An example of a Cherenkov ring in the FARICH subdetector from a simulated pion is shown in Fig. 1 (right).

The noise hits corresponding to various assumptions about the SiPM dark count rates (DCR) were also superimposed at the digitization step. The number of noise hits for each event was sampled from a Poisson distribution with a mean corresponding to the expected noise level in a pixel area. Two spatial coordinates of each noise hit at the photodetector plane were generated uniformly, and the times of hits were uniformly distributed over the event time frame of 7 ns.

2.4 Ring reconstruction algorithm

The typical reconstructed hit image from a single particle in the FARICH is not a regular ring but a distorted elliptical figure because of the inclination of the particle trajectory with respect to the photodetector plane and the additional air gap between the radiator and photosensors. The distortion effect depends on the incident angle of the particle. Moreover, the restricted efficiency of photon detection results in stochastic fluctuations in the number of hits for the same particle velocity β.

To account for these peculiarities, a dedicated algorithm for ring reconstruction was developed and tested on simulated events in the FARICH prototype [11]. The value of β is extracted using the incident particle momentum and direction at the photodetector plane, the relative number of hits per ring around the projected particle momentum, and the photon arrival times. The intersection of the track projection from the tracker with the FARICH photosensor plane serves as a "seed" for Cherenkov ring reconstruction.

The analyzed value is the ratio $S = N(R, T)/R$, where N is the number of hits calculated in bins at the (R, T) plane, where R is the hit radius with respect to the "seed" point defined above and T is the hit time. The outputs of the algorithm are the reconstructed radius, R_{reco}, which corresponds to the maximum value of the ratio, $S_{\max}$, and the signal z-score parameter, Z_{DCR}, which is defined as follows:

$$Z_{\text{DCR}} = \frac{S_{\max} - S_{\text{DCR}}}{V_{\text{DCR}}}, \tag{2}$$

where S_{DCR} and V_{DCR} are the mean and standard deviation of the ratio S estimated from the pure noise events for a given DCR value, respectively. The parameter Z_{DCR} characterizes the signal significance over the given photosensor dark count rate and is used for identification. An alternative ring reconstruction approach based on machine learning was presented in Ref. [13] and demonstrated comparable performance.

2.5 Boosted decision trees classifier

The particle identification landscape at charm superfactory energies is expected to contain several species: electrons, muons, pions, kaons, and protons. In the tracker with a radius of 800 mm positioned inside the 1 T magnetic field, only particles with momenta above ~310 MeV/c will reach the FARICH subdetector at incident angles smaller than 45°. Tracks with larger incident angles were not considered for analysis owing to the poor performance of the ring reconstruction algorithm. To produce Cherenkov light in the radiator with given parameters (see Sect. 2.3), the particle momentum should exceed approximately 350 MeV/c for muons, 500 MeV/c for pions, and 1800 MeV/c for kaons. Electrons and protons were excluded from the analysis because electrons always produce a maximal Cherenkov ring, whereas protons do not produce any rings. Moreover, electrons are expected to be reconstructed in the electromagnetic calorimeter with high efficiency. A particle species is identified using the boosted decision tree (BDT) technique. For the current study, the $\mu/\pi/$K classifier is considered.

2.5.1 BDT training and optimization

The classifiers were constructed and trained using the XGBoost library version 2.1.2 [14] for several noise levels. Four variables serve as BDT inputs: particle momentum and incident angle at the photodetector plane, reconstructed ring radius (R_{reco}), and signal z-score parameter defined in Eq. 2 and dependent on the chosen photosensor dark count rate (see Sect. 2.4). The particle momentum and incident angle in the real detector are reconstructed in the tracker and extrapolated to the radiator plane closer to the tracker.

The reconstruction of the particle velocity is an intermediate step in BDT-based classification. Figure 2 shows the results of the preliminary stage of the algorithm tuning, that is, the accuracy of velocity reconstruction by the FARICH subsystem for single pions simulated in the SCTF detector for different DCR values of photosensors. The BDT-based approach implemented in the full-scale detector model provides uncertainties of velocity reconstruction by a factor of ~ 1.5 smaller than the previous results from Ref. [11] obtained without BDT for a standalone FARICH prototype.

The reconstruction accuracy achieved by the BDT-based algorithm is better than 0.001 for $\beta < 0.999$ at incident angles up to 45° and photodetector dark count rates of to 10^5 Hz/mm^2. The performance slightly worsens at higher dark count rates, which is expected due to the damage of SiPMs by irradiation under SCTF conditions. The "nσ" separation between the pion and muon mass distributions is shown in Fig. 2 in terms of the requirements for β reconstruction accuracy, as it gives the dominant contribution compared to the momentum reconstruction accuracy in the drift chamber[2]. Assuming the same accuracy for pions and muons, the difference between means of pion and muon reconstructed mass distributions is larger than 5 standard deviations (5σ separation) up to $\beta = 0.985$ even for the worst case of DCR = 1 MHz/mm^2. For DCR ≤ 100 kHz/mm^2, the 5σ FARICH-based separation is available up to $\beta = 0.995$.

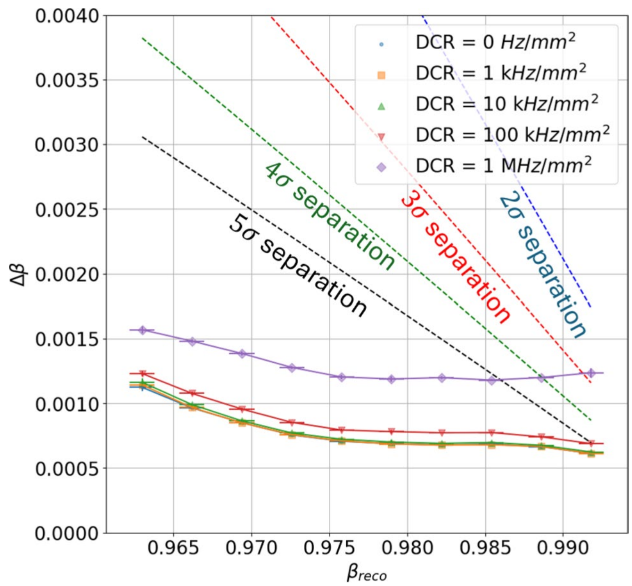

Fig. 2 Uncertainty of β_{reco} from the simulations of single pion response in the FARICH subdetector of the SCTF at different DCR levels. See text for details about $n\sigma$ separation

As shown in Fig. 2, the reconstruction algorithm shows very similar performance for dark count rates from 100 kHz/mm^2 and below. Therefore, two separate classifiers were trained for noise levels of 100 kHz/mm^2 and 1 MHz/mm^2 and used for further analysis. For training, 450,000 events were generated, 150,000 for each muon, charged pion, and charged kaon sample. The particle momentum for each particle type was uniformly distributed from 310 MeV/c (the minimum momentum detected in the FARICH subsystem) to 2500 MeV/c. Figure 3 shows the loss function behavior during training. For both noise levels, approximately 250 boosting rounds were sufficient to achieve good performance without overtraining. A study of feature importance using

[2] The value of required accuracy $\Delta\beta$ for $n\sigma$ separation between mass distributions is calculated as $\Delta\beta = \Delta m_{\pi\mu} \cdot \beta \cdot (1 - \beta^2)/m_\pi/n$, where $\Delta m_{\pi\mu}$ is the difference between pion and muon masses, m_π is the pion mass.

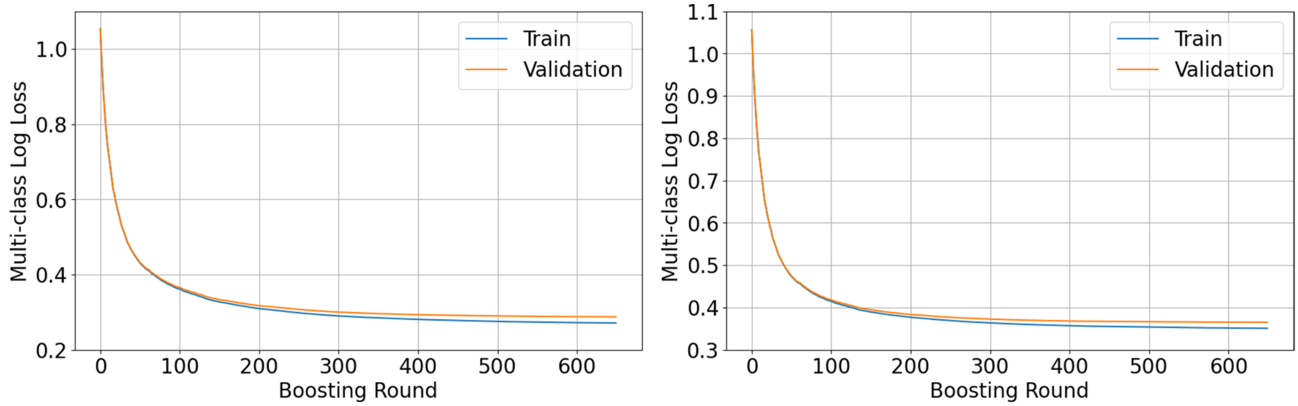

Fig. 3 Multiclass logistic loss function dependence on the number of boosting rounds for training (blue) and validation (orange) samples for DCR = 100 kHz/mm^2 (left) and DCR = 1 MHz/mm^2 (right)

the standard tool from the XGBOOST package showed that the importance scores were quite similar for all four input variables; the difference between the highest (for ring radius) and lowest (for incident angle) scores did not exceed 15%.

The classifier hyperparameters were optimized on the grid by maximizing the performance characteristic ROC-AUC using the tool provided in the XGBoost library. An additional 50,000 events per particle species (150,000 events in total) were generated for the test samples, which were used to demonstrate the classifier quality. Both the training and test samples of the single particles had the same fractions of the three particle species.

2.5.2 BDT performance for single particles

The single BDT classifier was trained for the entire momentum range. However, the PID performance differs significantly from the momentum. Therefore, the confusion matrices for the test sample are presented in six momentum ranges in Figs. 4 and 5 for the two DCR values. At low momenta, the reconstruction efficiency suffers from the low number of detected Cherenkov photons, which is further reduced by the implementation of realistic photon detection and geometrical efficiencies in digitization. As a result, the first momentum bin with $300 \leq p \leq 600$ MeV/c is the one where kaon misidentification has a noticeable probability, the pion–kaon separation being more affected than muon–kaon because pions have higher Cherenkov threshold than muons.

The assumed SiPM geometrical resolution makes it impossible to distinguish between muons and pions with momenta above ~1800 MeV/c, which leads to an inevitable drop in classification quality in this range with FARICH as the only PID subsystem. Therefore, the last bin covers the momentum range above 1800 MeV/c, although the information on the number of signal hits in each event, which is incorporated into the z-score value, allows a certain level of identification above this threshold.

The muon identification efficiency and misidentification probability of pions as muons are shown for the test sample in Fig. 6 for two DCR levels in several momentum bins. For a DCR level of 1 MHz/mm^2, the drop in identification performance can be observed in all momentum ranges. The worsening of the PID quality in the momentum range below 600 MeV/c is related to the significant confusion of both pions and muons with kaons.

3 Projection of PID performance to physics analysis

The analysis of the measurement precision of the branching fraction of $D^0 \to K\mu\nu$ decay was chosen to demonstrate the FARICH performance for particle identification at the SCTF. It should be noted that the full reconstruction of the SCTF detector model is not yet available, and the estimates are obtained under different assumptions about the tracker and calorimeter performances.

The decays of $\psi(3770)$ resonance, which is an admixture of $2S$ and $1D$ $c\bar{c}$ states, can be considered as a natural source of D meson pairs in e^+e^- collider experiments. The state $\psi(3770)$ decays into $D^0\bar{D}^0$ pair with the probability of 52^{+4}_{-2}% without any extra final-state particles [15]. With the design peak instantaneous luminosity about 10^{35} cm^{-2}s^{-1}, the annual integrated SCTF luminosity at $\psi(3770)$ peak reaches ~150 fb^{-1}, which corresponds to $\mathcal{O}(10^9)$ of $\psi(3770)$ produced [2]. Taking advantage of this enormous data sample, the experiment at the SCTF can test the LFU with unprecedented precision. As mentioned above, promising channels for this task include $D^0 \to K\mu\nu$ and $D^0 \to Ke\nu$. While the efficiency of the latter channel relies mainly on the electromagnetic calorimeter, the reconstruction of the

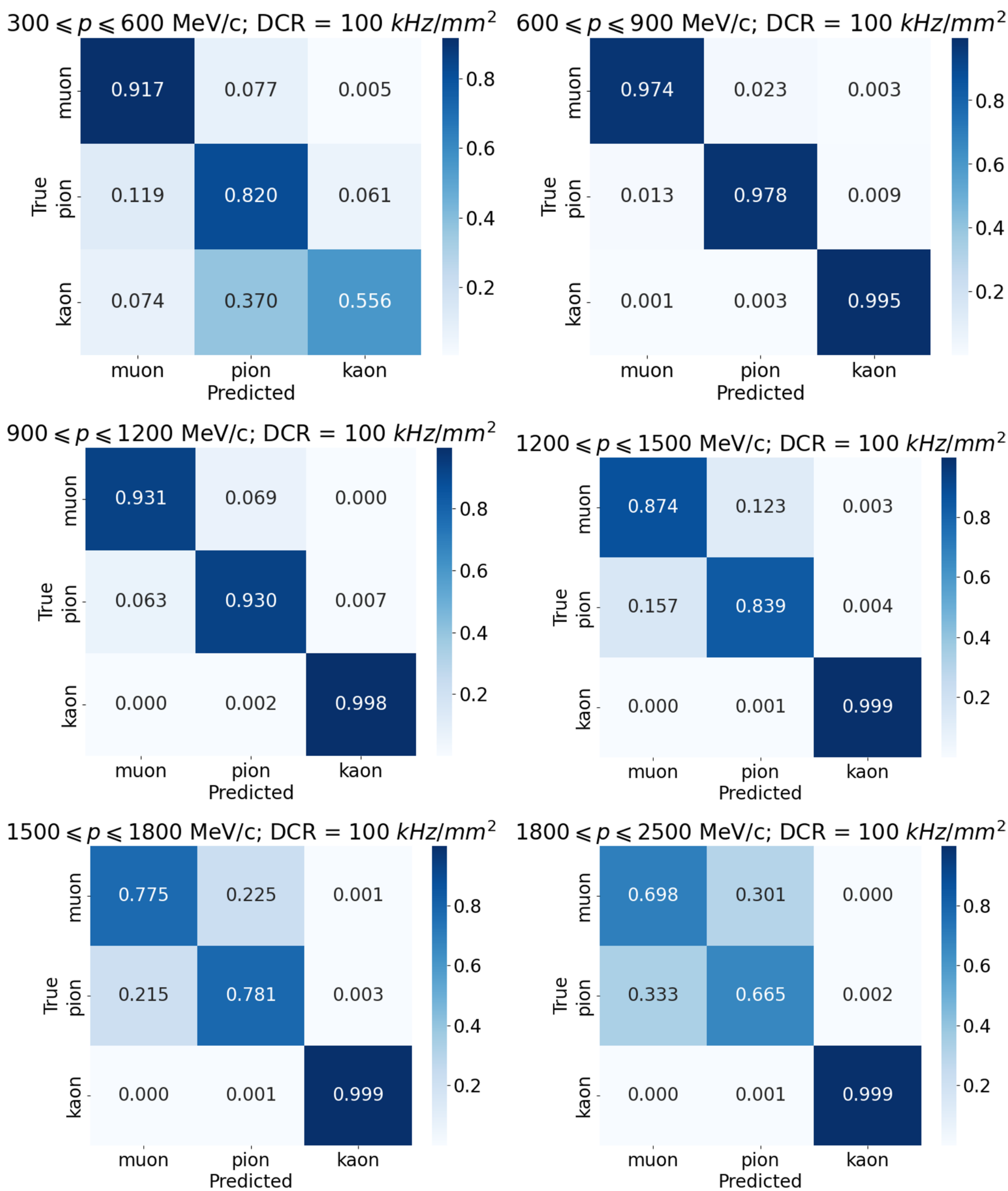

Fig. 4 (Color online) Confusion matrices (fraction of predicted type for a given true type) for the test sample of simulated single particles in the SCTF for DCR = 100 kHz/mm^2 in six momentum ranges

channel $D^0 \to K\mu\nu$ significantly depends on the particle identification (PID) performance.

The estimates made in this study follow the strategy developed by the BESIII collaboration [16], which implies event tagging by reconstructing one of the D mesons, while only the kaon and muon are reconstructed from the other. As a conservative estimate, we use the efficiency of the decay chain reconstruction averaged over three tag-side channels: $\bar{D}^0_{\text{tag}} \to K^+\pi^-$, $K^+\pi^-\pi^-\pi^+$, $K^+\pi^-\pi^0$. The

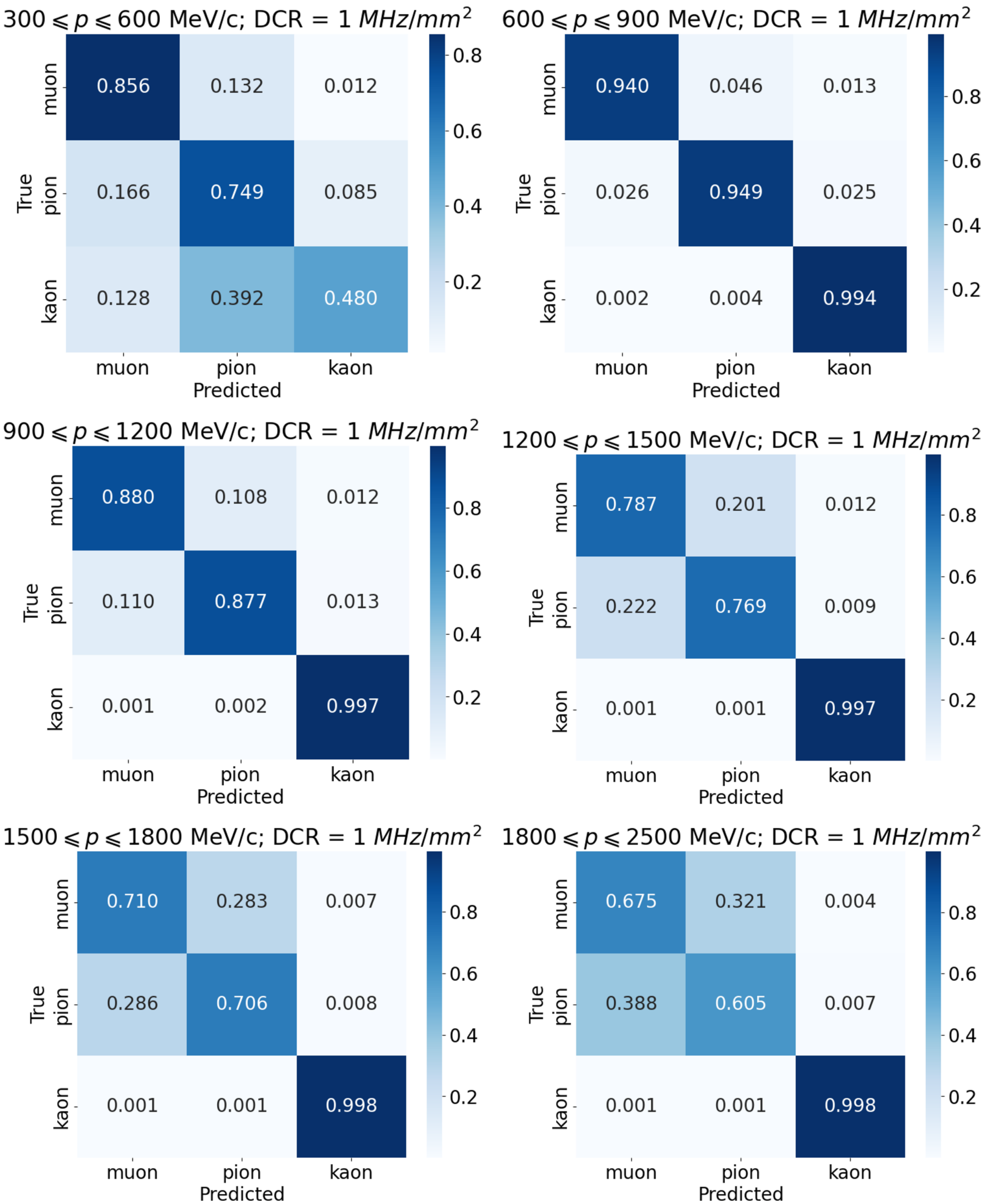

Fig. 5 (Color online) Confusion matrices (fraction of predicted type for a given true type) for the test sample of simulated single particles in the SCTF for DCR = 1 MHz/mm^2 in six momentum ranges

efficiency and total number of tagged events taken from Ref. [16] are $\varepsilon = 58.93\%$ and $N_{\text{tag}}^{\text{BESIII}} = 2341408$, respectively. The number of the tags is then scaled with expected integrated luminosity by factor of $\frac{\mathcal{L}^{\text{SCTF}}}{\mathcal{L}^{\text{BES III}}} = \frac{150\ \text{fb}^{-1}}{2.93\ \text{fb}^{-1}}$ yielding $N_{\text{tag}}^{\text{SCTF}} = 799 \times 10^8$ tagged D^0 mesons.

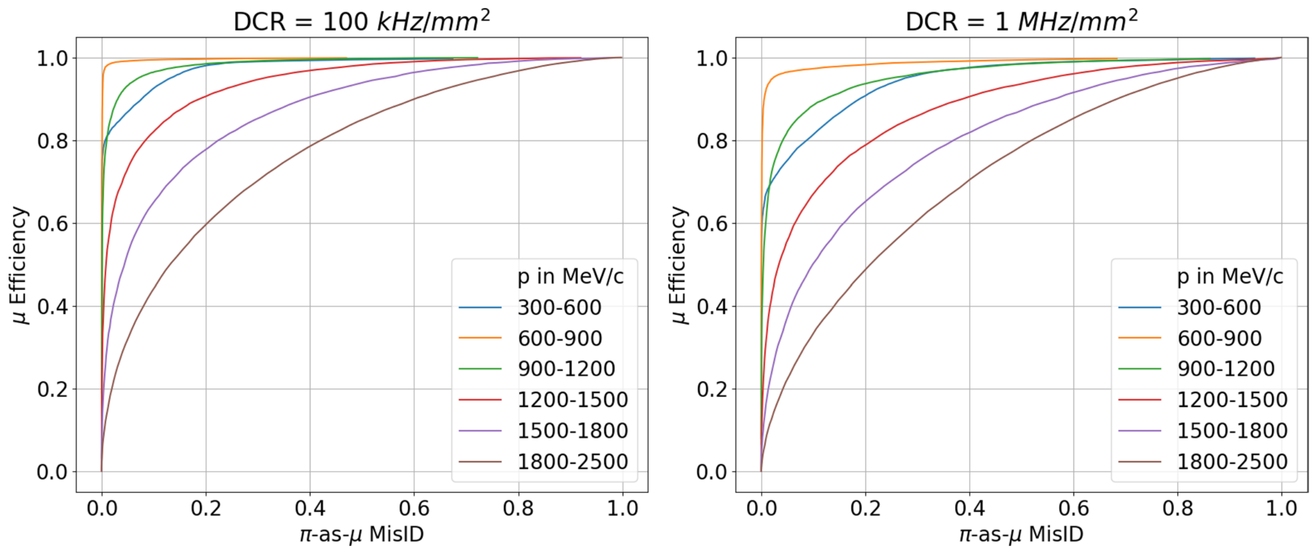

Fig. 6 (Color online) Muon identification efficiency versus pion-as-muon misidentification probability (π-as-μ MisID) for the test sample of simulated single particles in the SCTF for DCR = 100 kHz/mm^2 (left) and DCR = 1 MHz/mm^2 (right) in six momentum ranges

There are some backgrounds that could mimic the signal. Those associated with the tag impurity (no D^0 mesons in the decay) can be described by a smooth threshold function in both the tag and signal distributions. The peaking background from the $D^0 \to K^-\pi^+$ decays in the case of π^+/μ^+ misidentification is eliminated by the requirement that the invariant mass of the $K^-\mu^+$ combination is lower than 1.56 GeV. The other possible background sources either have a tiny probability (requiring double misidentification) or shift the peak in the signal distribution by at least the mass of the lost particles. The sole remaining peaking background comes from the $D^0 \to K^-\pi^+\pi^0$ decay in the case of lost π^0 and π^+/μ^+ misidentification, which is discussed in detail below.

3.1 Signal and background simulation and event selection

To estimate the performance of the developed PID algorithm for the physics analysis discussed above, 550,000 events with decays $\psi(3770) \to D^0\bar{D}^0$ at the center-of-mass energy of e^+e^- beams $\sqrt{s} = 3770$ MeV have been generated and propagated through the tracker and FARICH subsystem of the SCTF detector. The produced $\bar{D}^0$ was simulated to decay into invisible particles and was ignored. The D^0 meson is forced to decay as either $D^0 \to K^-\mu^+\nu_\mu$ (signal) or $D^0 \to K^-\pi^+\pi^0$ (background). The propagation of particles in a drift chamber volume in a 1 T magnetic field, as well as the generation of Cherenkov photons in the FARICH subsystem and hits detection, were simulated within the Aurora framework.

For the performance analysis, events were selected using generator-level information about the particle momenta. The selection efficiencies listed below affect the signal-to-background ratio and do not depend on the PID performance.

- The ratio of generated numbers of signal and background events is taken from the corresponding branching fractions from PDG [15]: $\mathcal{B}_{\text{signal}} = \mathcal{B}(D^0 \to K^-\mu^+\nu_\mu) = 0.034$ and $\mathcal{B}_{\text{bgd}} = \mathcal{B}(D^0 \to K^-\pi^+\pi^0) = 0.144$.
- From the background events, only those with lost π^0 are retained. The probability of this case, $P_{\text{lost}\pi^0}$, is combined from two efficiencies, $\epsilon_{2\gamma}$ and ϵ_{π^0} ($P_{\text{lost}\pi^0} = 1 - \epsilon_{2\gamma} \cdot \epsilon_{\pi^0}$). The efficiency $\epsilon_{2\gamma}$ is the probability of detecting both γs from π^0, that is, that both γs reach a calorimeter (barrel or endcap) and have energies $E_\gamma > E_{\gamma\text{cut}}$. The geometrical acceptance estimated for the STCF calorimeter covers the polar angles $|\theta| < 70.8°$ [3] and is fixed for this study, amounting to 0.95 for a single γ. The efficiency ϵ_{π^0} is the probability of reconstructing π^0 from two γs and depends on the calorimeter performance, which is not yet known. Therefore, several combinations of ϵ_{π^0} and $E_{\gamma\text{cut}}$ values were considered for the analysis.
- Both charged particles (tracks) must go through the barrel part of the SCTF detector with the FARICH subsystem. For the fixed detector geometry and very similar kinematics of signal and background, this efficiency is $\epsilon_{\text{tr}}^{\text{signal}} \approx \epsilon_{\text{tr}}^{\text{bgd}} \approx 0.58$ for two tracks.
- The momenta of both charged particles (tracks) must exceed 310 MeV/c, i.e., the limit sensitivity of the FARICH subsystem. The estimated efficiencies were $\epsilon_{\text{F}}^{\text{signal}} \approx 0.84$ and $\epsilon_{\text{F}}^{\text{bgd}} \approx 0.78$.

The numbers of signal and background events, $N_{\mathrm{sel}}(K\mu\nu)$ and $N_{\mathrm{sel}}(K\pi\pi^0)$, selected before identification, can be calculated from the efficiencies listed above as

$$N^{\mathrm{signal}}_{\mathrm{sel}} = N^{\mathrm{signal}}_{\mathrm{gen}} \cdot \epsilon^{\mathrm{signal}}_{\mathrm{tr}} \cdot \epsilon^{\mathrm{signal}}_{\mathrm{F}}, \tag{3}$$

$$N^{\mathrm{bgd}}_{\mathrm{sel}} = N^{\mathrm{bgd}}_{\mathrm{gen}} \cdot \epsilon^{\mathrm{bgd}}_{\mathrm{tr}} \cdot \epsilon^{\mathrm{bgd}}_{\mathrm{F}} \cdot P_{\mathrm{lost}\pi^0}, \tag{4}$$

where $N^{\mathrm{signal}}_{\mathrm{gen}}$ is the number of generated signal events (about 100,000) and $N^{\mathrm{bgd}}_{\mathrm{gen}} = N^{\mathrm{signal}}_{\mathrm{gen}} \cdot \mathcal{B}_{\mathrm{bgd}}/\mathcal{B}_{\mathrm{signal}}$. As can be seen from Eqs. (3) to (4), in addition to particle identification characteristics, the efficiency of π^0 reconstruction also affects the background admixture to the signal sample under study.

Two constraints on the minimal photon energy detected in the calorimeter were investigated, representing the baseline and the most conservative assumptions about calorimeter performance, 50 MeV and 200 MeV for $E_{\gamma\mathrm{cut}}$ values, respectively. The first value is widely used in Belle analyses[3] [17]. Although the SCTF calorimeter is designed to detect photons starting from 20 MeV, the expected number of background clusters increases drastically with a decrease in the cluster energy cutoff, which, in combination with increased luminosity, results in a large combinatorial background and degradation of the reconstruction efficiency [18]. And the fact that the current analysis suffers from the signal impurity rather than from the lack of statistics justifies the choice of such a tight energy cut for the most conservative estimate. The efficiencies of signal and background event reconstruction and numbers of selected events for the four probabilities of lost π^0 are summarized in Table 1.

The normalized momentum distributions of the final-state particles for both the signal and background channels are shown in Fig. 7. For the background, the two most different options are shown: the baseline and the most conservative one. The sharp low-momentum edge of the after-cut distributions is due to the sensitivity limit of the FARICH subsystem, as stated in the last item of the aforementioned list of constraints.

3.2 Estimate of PID performance

The ring reconstruction algorithm described above and the BDT inference to identify a particle species are applied to selected events, and the number of identified signals and misidentified background events are calculated. The efficiency of kaon identification, ϵ_{K}, is considered: Negatively charged tracks are required to be identified as kaons. The numbers of selected and reconstructed signal events identified as signals can be represented as follows:

$$N^{\mathrm{signal}}_{\mathrm{reco}}(K\mu) = N^{\mathrm{signal}}_{\mathrm{sel}} \cdot \epsilon_{\mathrm{PID}}; \quad \epsilon_{\mathrm{PID}} = \epsilon_{\mathrm{K}} \cdot \epsilon_{\mu}, \tag{5}$$

where ϵ_{μ} is the efficiency of muon identification in the FARICH subsystem. The number of background events misidentified as signals can be calculated as follows:

$$N^{\mathrm{bgd}}_{\mathrm{reco}}(K\mu) = N^{\mathrm{bgd}}_{\mathrm{sel}} \cdot \tau_{\mathrm{PID}}; \quad \tau_{\mathrm{PID}} = \epsilon_{\mathrm{K}} \cdot \tau_{\pi/\mu}, \tag{6}$$

where $\tau_{\pi/\mu}$ is the misidentification probability of pion as muon. From Eqs. (5) to (6), the pairs of ϵ_{PID} and τ_{PID} values can be obtained for various BDT working points defined by the classifier threshold at different assumptions about the probabilities of lost π^0. The efficiency of signal identification ϵ_{PID} versus the probability of background misidentification τ_{PID} is shown in Fig. 8 for the two noise levels. The belts shown represent the maximal variation between the results obtained with different assumptions about $\epsilon_{2\gamma}$ and ϵ_{π^0} from Table 1. As shown in Fig. 8, significant differences between π^0 reconstruction efficiencies result in a similar π/μ separation performance.

The unprecedented expected luminosity of charm superfactories, together with the achieved performance of particle identification, leads to a negligible statistical uncertainty for the $D^0 \to K\mu\nu$ branching ratio measurements at a level well below 0.1%. Figure 9 shows the percentage of background admixture in the reconstructed signal sample, f_{bgd}, for various assumptions regarding the π^0 reconstruction efficiencies and for two noise levels. With the possibility of moving the background contribution below 0.5%, the systematic uncertainty can also be reduced, opening the opportunity to achieve a new level of LFU measurement precision at charm superfactories.

4 Conclusion

The performance of the FARICH-based particle identification for the future charm superfactory was investigated using simulated samples of single particles and simulations of D meson decay processes. The algorithms for Cherenkov ring reconstruction and BDT-based classification were developed and tested for different event reconstruction conditions. The algorithm robustness under various SiPM dark count rates was proved: It was shown that π/μ separation at the level of 5 standard deviations is available up to ~700 MeV/c for the most conservative DCR = 1 MHz/mm^2 and up to ~ 1.3 GeV/c for the baseline DCR = 100 kHz/mm^2.

The physics analysis of the $D^0 \to K\mu\nu$ decay, the FARICH performance, is further explored. Considering the most conservative assumptions about the subsystem

[3] The CsI-based calorimeter of the Belle detector is similar to those proposed for the SCTF.

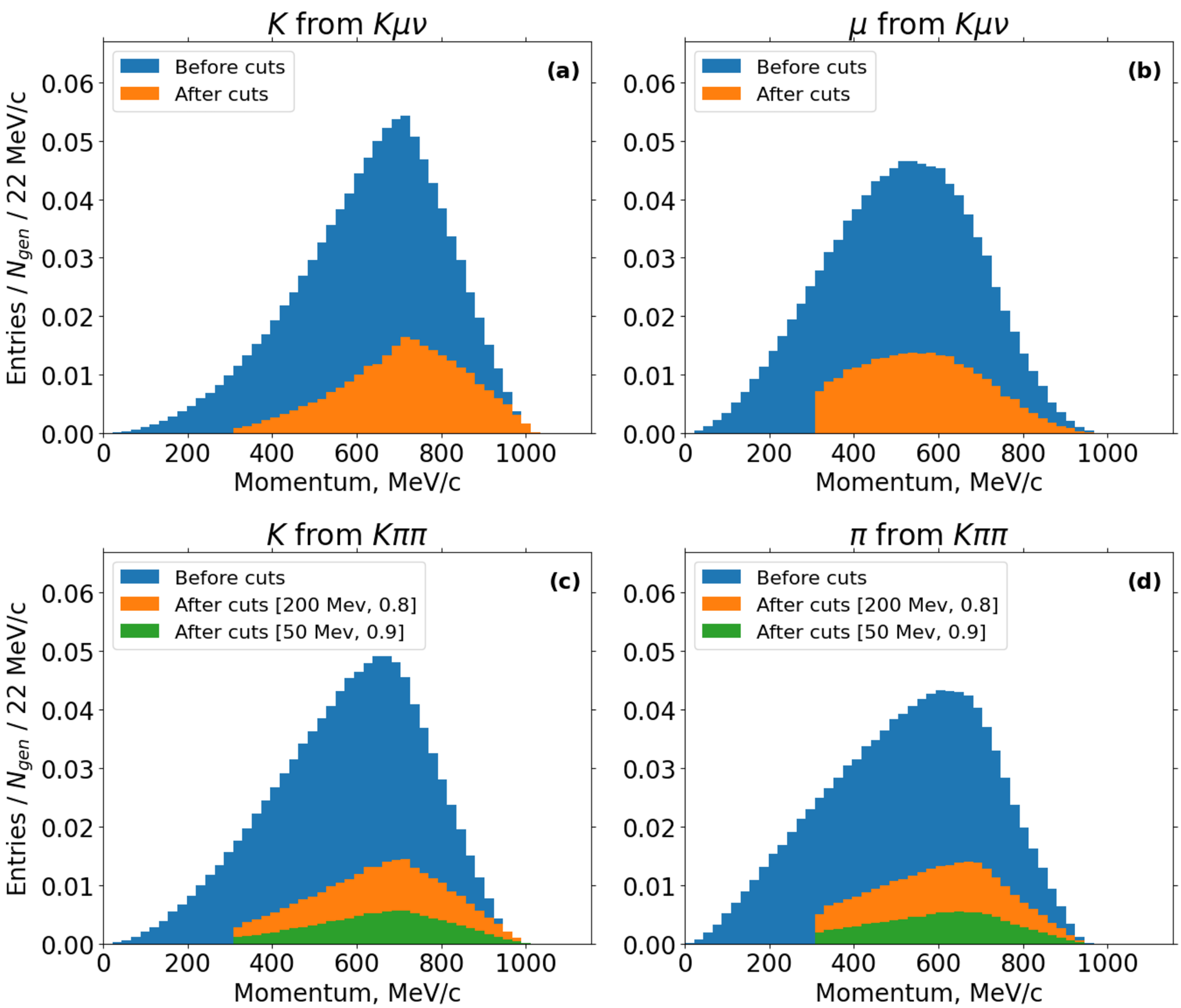

Fig. 7 (Color online) Distributions of particle momentum for signal channel kaons (**a**) and muons (**b**) and background channel kaons (**c**) and charged pions (**d**): before selections (blue), after signal selections (orange at **a**, **b**) and after background selections (**c**, **d**) with conservative (orange) and baseline (green) assumptions about π^0 reconstruction efficiency. Histograms are normalized to generated number of signal (**a**, **b**) or background (**c**, **d**) events

Table 1 Branching fractions, reconstruction efficiencies, and numbers of selected signal and background events for the BDT performance estimates at various conditions of π^0 reconstruction at the SCTF

$\mathcal{B}_{\text{signal}}$	$\epsilon_{\text{tr}}^{\text{signal}}$	$\epsilon_{\text{F}}^{\text{signal}}$	$N_{\text{sel}}^{\text{signal}}$	$\mathcal{B}_{\text{bgd}}$	$\epsilon_{\text{tr}}^{\text{bgd}}$	$\epsilon_{\text{F}}^{\text{bgd}}$	$P_{\text{lost}\pi^0}$ $[\epsilon_{\pi^0}, E_{\gamma\text{cut}}]$	$N_{\text{sel}}^{\text{bgd}}$
							0.32 [0.9, 50 MeV]	65,293
0.034	0.58	0.84	47,415	0.144	0.58	0.78	0.40 [0.8, 50 MeV]	80,352
							0.81 [0.9, 200 MeV]	164,219
							0.83 [0.8, 200 MeV]	168,285

efficiencies and background admixture due to misidentification, as well as the planned luminosity of the charm superfactory, the role of the particle identification efficiency provided by a single subsystem of the SCTF was demonstrated. For this physics case, the dominant contribution to background rejection comes from particle identification with the FARICH subsystem, even for conservative expectations of photosensor dark count rates. As follows from the analysis, the performance of the developed algorithms allows for significant mitigation of the background contribution and, therefore, improvement of the measurement precision.

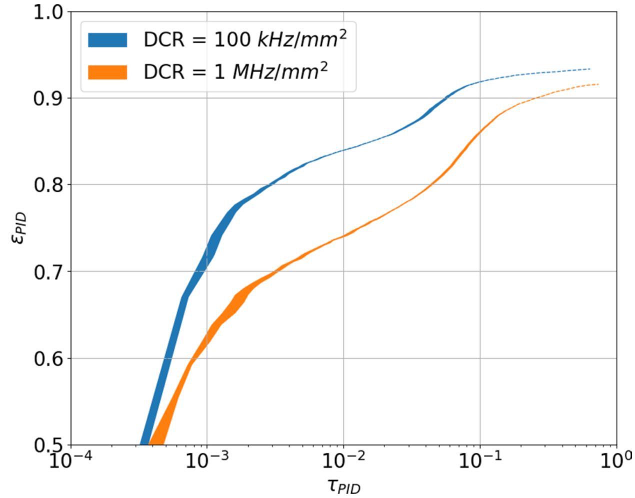

Fig. 8 (Color online) Signal identification efficiency versus background misidentification probability for the expected mixture of signal ($D^0 \to K\mu\nu$) and background ($D^0 \to K\pi\pi^0$) events for DCR = 100 kHz/mm^2 (blue) and DCR = 1 MHz/mm^2 (orange). The width of the belt corresponds to the maximal differences between the background admixture options from Table 1. See text for details

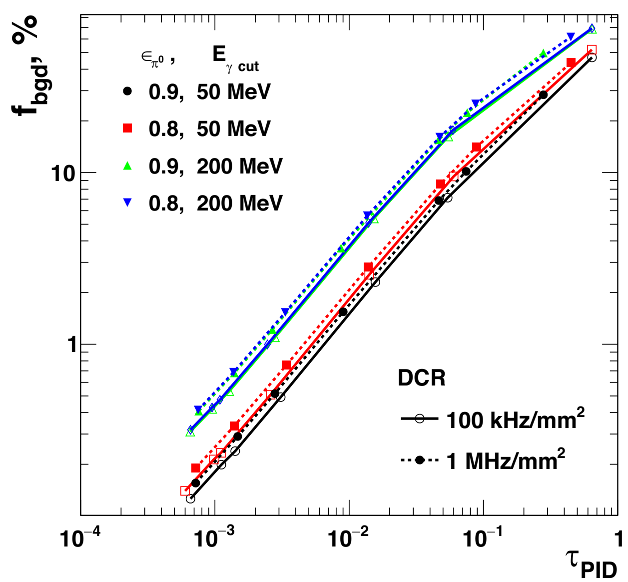

Fig. 9 Expected background fraction in the reconstructed signal sample versus the background misidentification probability for different efficiencies of π^0 reconstruction at DCR = 100 kHz/mm^2 (solid lines, open markers) and DCR = 1 MHz/mm^2 (dashed lines, filled markers). The lines are shown to guide the eye

Although the investigated FARICH subsystem plays a dominant role in the physics case studied, the implementation of a global PID technique involving other subdetectors will further improve the PID performance. The results obtained with the full simulation of the FARICH subsystem can provide a good basis for the development of a fast simulation tool for further analyses. Such a tool is important for detailed studies of a global PID technique at charm superfactories, which require a combination of fast simulation of the FARICH subsystem with full simulations of all other subdetectors.

Author contributions The study conception has been developed by T. Uglov and M. Chadeeva. Data generation and processing were performed by P. Rogozhin. All authors contributed to the data analysis. The first draft of the manuscript was written by T. Uglov and M. Chadeeva, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. P. Raimondi, M. Zobov, D. Shatilov, Beam-beam simulations for particle factories with crabbed waist. 2007 IEEE Particle Accelerator Conference (PAC), Albuquerque, NM, USA, 2007, pp. 1469-1471. https://doi.org/10.1109/PAC.2007.4440792
2. M.N. Achasov, V.E. Blinov, A.V. Bobrov et al., Experiments at the Super Charm-Tau factory. Phys. Usp. **67**, 55 (2024). https://doi.org/10.3367/UFNe.2023.10.039583
3. M. Achasov, X.C. Ai, L.P. An et al., STCF conceptual design report (Volume 1): Physics and detector. Front. Phys. **19**, 14701 (2023). https://doi.org/10.1007/s11467-023-1333-z
4. V.V. Anashin, A.V. Anisenkov, V.M. Aulchenko et al., Super Charm Tau Factory conceptual design report Part One (physics program, detector). (BINP wiki). Accessed 05 (2025)
5. A. Barnyakov, M. Belozyorova, V. Bobrovnikov et al., Fast simulation for the Super Charm-Tau Factory detector. Comput. Softw. Big Sci. **8**, 1 (2024). https://doi.org/10.1007/s41781-023-00108-7
6. A.Y. Barnyakov, V.S. Bobrovnikov, A.R. Buzykaev et al., R&D status of FARICH option for PID. Int. J. Mod. Phys. A **39**, 2442012 (2024). https://doi.org/10.1142/S0217751X24420120
7. M. Belozyorova, D. Maksimov, G. Razuvaev et al., Software framework for the Super Charm-Tau factory detector project. EPJ Web Conf. **251**, 03017 (2021). https://doi.org/10.1051/epjconf/202125103017
8. W.H. Huang, H. Li, H. Zhou et al., Design and development of the core software for STCF offline data processing. J. Instrum. **18**, P03004 (2023). https://doi.org/10.1088/1748-0221/18/03/p03004
9. M. Frank, F. Gaede, C. Grefe et al., DD4hep: A detector description toolkit for high energy physics experiments. J. Phys. Conf. Ser. **513**, 022010 (2014). https://doi.org/10.1088/1742-6596/513/2/022010
10. S. Agostinelli, J. Allison, K. Amako et al., Geant4–a simulation toolkit. Nucl. Instrum. Meth. Phys. Res. Sect. A **506**, 250–303 (2003). https://doi.org/10.1016/S0168-9002(03)01368-8
11. A.Y. Barnyakov, V.S. Bobrovnikov, S.A. Kononov et al., Calibration and reconstruction algorithm for FARICH system of the detector at super Charm-Tau factory. Bull. Lebedev Phys. Inst. **50**, 534 (2023). https://doi.org/10.3103/S1068335623120047
12. W. Hartinger, F. Wiest, P. Iskra et al., SiPM solutions realized by KETEK. (Presentation at EPS-TIG – Nov 11-12, 2013, Ravenna, Italy). Accessed 05 July 2025 https://indico.cern.ch/event/284070/sessions/53907/attachments/524514/723388/2013.11_-_KETEK_-_EPS-TIG_-_Ravenna_Italy_WHA_-_public.pdf

13. F. Shipilov, A. Barnyakov, V. Bobrovnikov et al., What machine learning can do for focusing aerogel detectors. Phys. Atom. Nucl. **86**, 864 (2023). https://doi.org/10.1051/epjconf/202429509043
14. XGBoost: Extreme Gradient Boosting. (Toolkit documentation site). Accessed 05 July 2025 https://xgboost.readthedocs.io/en/stable/
15. S. Navas, C. Amsler, T. Gutsche et al., Review of particle physics. Phys. Rev. D **110**, 030001 (2024). https://doi.org/10.1103/PhysRevD.110.030001
16. M. Ablikim, M.N. Achasov, S. Ahmed et al., Study of the $D^0 \to K^- \mu^+ \nu_\mu$ dynamics and test of lepton flavor universality with $D^0 \to K^- \ell^+ \nu_\ell$ decays. Phys. Rev. Lett. **122**, 011804 (2019). https://doi.org/10.1103/PhysRevLett.122.011804
17. G. Caria, P. Urquijo, I. Adachi et al., Measurement of $R(D)$ and $R(D^*)$ with a semileptonic tagging method. Phys. Rev. Lett. **124**, 161803 (2020). https://doi.org/10.1103/PhysRevLett.124.161803
18. T. Abe, I. Adachi, K. Adamczyk et al., Belle II technical design report. arXiv:1011.0352 [physics.ins-det]