particles

MDPI

Article

A Scintillation Hodoscope for Measuring the Flux of Cosmic Ray Muons at the Tien Shan High Mountain Station

Alexander Shepetov [1,*], Aliya Baktoraz [2,3], Orazaly Kalikulov [2], Svetlana Mamina [1], Yerzhan Mukhamejanov [2,3,4], Kanat Mukashev [5], Vladimir Ryabov [1], Nurzhan Saduyev [2], Turlan Sadykov [6], Saken Shinbulatov [2,3], Tairzhan Skokbayev [2], Ivan Sopko [3], Shynbolat Utey [2,3], Ludmila Vildanova [1], Nurzhan Yerezhep [2,3] and Valery Zhukov [1]

1 P. N. Lebedev Physical Institute of RAS (LPI), Leninsky Ave., 53, Moscow 119991, Russia
2 Institute of Nuclear Physics, 1 Ibragimov St., Almaty 050032, Kazakhstan
3 Faculty of Physics and Technology, Al-Farabi Kazakh National University, 71 Al-Farabi Ave., Almaty 050040, Kazakhstan
4 Joint Institute for Nuclear Research, 6 Joliot-Curie St., Dubna 141980, Russia
5 Institute of Communication and Space Engineering, G. Daukeyev Almaty University of Power Engineering and Telecommunications, 126/1 Baytursynuli St., Almaty 050013, Kazakhstan
6 Institute of Physics and Technology, Satbayev University, 22a Satbayev St., Almaty 050013, Kazakhstan
* Correspondence: ashep@tien-shan.org

Abstract

For further investigation of the properties of the muon component in the core regions of extensive air showers (EASs), a new underground hodoscopic set-up with a total sensitive area of 22 m^2 was built at the Tien Shan High Mountain Cosmic Ray Station. The hodoscope is based on a set of large-sized scintillation charged particle detectors with an output signal of analog type. The installation ensures a (5–8) GeV energy threshold of muon registration and a $\sim 10^4$ dynamic range for the measurement of the density of muon flux. A program facility was designed that uses modern machine learning techniques for automated search for the typical scintillation pulse pattern in an oscillogram of a noisy analog signal at the output of the hodoscope detector. The program provides a ∼99% detection probability of useful signals, with a relative share of false positives below 1%, and has a sufficient operation speed for real-time analysis of incoming data. Complete verification of the hardware and software tools was performed under realistic operation conditions, and the results obtained demonstrate the correctness of the proposed method and its practical applicability to the investigation of the muon flux in EASs. In the course of the installation testing, a preliminary physical result was obtained concerning the rise of the multiplicity of muon particles around an EAS core in dependence on the primary EAS energy.

Keywords: cosmic ray muons; extensive air showers; EAS; muon hodoscope; machine learning

check for updates

Academic Editors: Armen Sedrakian and Francesco Nozzoli

Received: 22 June 2025
Revised: 23 July 2025
Accepted: 29 July 2025
Published: 4 August 2025

Citation: Shepetov, A.; Baktoraz, A.; Kalikulov, O.; Mamina, S.; Mukhamejanov, Y.; Mukashev, K.; Ryabov, V.; Saduyev, N.; Sadykov, T.; Shinbulatov, S.; et al. A Scintillation Hodoscope for Measuring the Flux of Cosmic Ray Muons at the Tien Shan High Mountain Station. *Particles* **2025**, *8*, 73. https://doi.org/10.3390/particles8030073

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

By the investigation of the muon component of cosmic rays at the Tien Shan High Mountain Cosmic Ray Station (Northern Tien Shan, 3340 m a.s.l.), a phenomenon of an accelerated rise in the energy deposit left after the interaction of the muon particles moving in the core region of extensive air showers (EASs) was discovered [1]. This effect, which starts to be observable at a primary EAS energy of a few PeV and above, was revealed by registration of the signal of evaporation neutrons originating from the interaction of

cosmic ray muons with surrounding material. For this purpose, a special neutron detector was applied, which is installed in an underground room of the Tien Shan station, under a rock absorber with a thickness of 20 m of water equivalent. It was found that the rate of multiplicity growth of evaporation neutrons produced in the interaction of the muons of the EAS core suddenly starts to rise disproportionately fast with increasing EAS energy E_0, and this peculiarity reveals itself in the range of $E_0 \sim$ (2–3) PeV, i.e., around the prominent "knee" in the energy spectrum of primary cosmic rays. As is known, in recent decades, a number of other anomalous effects were revealed in the "knee" region, while any attempts at a detailed and mutually consistent explanation of these phenomena have faced significant challenges. Among such poorly understood observations, the following could be mentioned: anomalous effects found in the interaction of ultra-high-energy cosmic ray hadrons with matter [2,3]; the signs of reduced absorption of such hadrons [4–6] and peculiarities in the development of the cascades of secondary particles produced by them [7–9]; the violation of normal scaling behavior with rising energy in the interaction of the hadronic component of EASs [10]; and the difficulties with interpretation of the chemical composition of primary cosmic rays based on the properties of EASs in the energy range of (1–1000) PeV [11–14].

The anomalies in the behavior of the muonic component of EASs, which have been found up to the present time at the Tien Shan station and described in detail in [1], today follow the same trend as the so-called "muon deficit" problem actively discussed in recent years. The problem consists of a systematically lower multiplicity of muon production in simulations based on modern models of high-energy hadronic interaction as compared with the measurement data accumulated in cosmic ray experiments. Thus, the investigations of both single cosmic ray muons [15–21] and the muonic component of EASs [22–26] have steadily reported detection of an excessive muon flux in comparison with simulation predictions. As concluded in a systematic study of this problem [27], a deficit of produced muons was found in simulated EASs for each of the six post-LHC interaction models tested there, and the value of this deficit increased with the shower energy. A similar result was reported in a more contemporary study [28].

Since the high-energy muons in an EAS are immediate decay products of the pions born at the leading development stages of an EAS cascade, any peculiarity in the flux of such muons may be considered as a consequence of some unsuspected effect existing among the most energetic processes that take place at the interaction of primary cosmic ray particles and nuclei coming into the atmosphere from the Universe. This circumstance necessitates further study of the effects related to the behavior of cosmic ray muons.

The method based on detection of the evaporation neutrons originating from muon interaction, which has been applied at the Tien Shan station so far, has some unique advantages for investigation of the muonic component of EASs. As such, a high, of an order of ∼1 TeV, energy threshold of effective muon registration, the possibility of energy estimation of interacting muons, and the absence of any principal upper limit on the energy of detected muons should be mentioned [1]. A principle drawback of the method is that it cannot clearly distinguish between the following two possibilities: (1) either an accelerated rise in the average number of muons $\langle N_\mu \rangle$ occurs in the EASs, with the primary energy E_0 exceeding the threshold of several PeV, or (2) an increase in the mean muon energy $\langle E_\mu \rangle$ takes place in the above-the-threshold events. Indeed, the total multiplicity of evaporation neutrons produced by penetrative particles of an EAS core, and directly registered by the detector, depends on both factors, $\langle N_\mu \rangle$ and $\langle E_\mu \rangle$.

To resolve this confusion, it is necessary to try another type of registration device, sensitive only to the number of muon particles traveling through. This was the reason for constructing a hodoscopic set-up at the Tien Shan station, which should be applied for direct measurement of the flux of cosmic ray muons, regardless of their energy. The in-

stallation uses the modern scintillation detectors of charged particles, and an up-to-date operation technique is realized for processing the raw measurement data acquired from these scintillators. A description of the newly developed investigation method of the muonic component of EAS is the subject of the present publication.

2. Hardware Description

The muon hodoscope is based on large-sized scintillation detectors with a built-in high-voltage power converter designed for use in cosmic ray experiments. The detector consists of a $(1 \times 1 \times 0.01)\,m^3$ polystyrene scintillator plate connected to a photomultiplier tube by wavelength-shifting fibers. The design ensures the 99% registration efficiency of a single-charged relativistic particle, an inhomogeneity of scintillation light output no worse than 8%, and a $\sim$10 ns or lower decay time of scintillation flash [29].

Since the detectors of scintillation type are sensitive only to the number of passing relativistic particles and not to their energy, the use of such detectors in the muon hodoscope solves the abovementioned problem of differentiation between the growth of muon multiplicity or average muon energy in EAS events.

The current arrangement scheme of the hodoscope detectors in the $(11 \times 9)\,m^2$ area of the underground room is shown in Figure 1, top. Currently, there are 22 scintillators located in the upper tier of the system, another four are placed below, and they form a coincidence telescope for determining the direction of the charged particles passing through. Further installation of an additional 10–15 detectors is planned as well, increasing the total sensitive area of the hodoscope up to $\sim$(30–40) m^2.

The underground room with the muon detectors is situated just under the central *carpet* of the surface system of EAS particle detectors (Figure 1, bottom); so, the center of the muon hodoscope is near the origin of the coordinate frame used for the analysis of the EAS axes position (point *C* in the scheme). Such arrangement of the detector systems allows for an effective study of the muon flux around the core region of the EASs belonging to the energy range of (1–100) PeV [30]. The rocky ground above the room, of a $\sim$2000 g/cm^2 thickness, ensures an energy threshold of muon detection about (5–8) GeV, depending on the zenith angle of the muon trajectory. Thus, the hodoscope has a significantly lower energy threshold in comparison with the technique based on the signal of evaporation neutrons, which was applied earlier; so, both methods are complementary.

The hodoscope set-up is supposed to analyze the spatial distribution of the muon density around the EAS core based on the number of the muons, which, at the moment of an EAS, have passed through the scintillating radiator of each detector. In turn, the number of muons should be estimated according to the peak amplitudes of scintillation pulses registered in the detector. To measure these amplitudes, a special multichannel amplitude-to-digital conversion (ADC) system is assembled, which consists of a VME crate equipped with the fast V1730 type digitizer modules of CAEN production [31]. Each module allows for the simultaneous digitization of 16 analog signals, with a 2 ns time resolution and a digitization precision of 14-binary bits. The last parameter determines the dynamic range for the measurement of the intensity of charged particles flux to be $2^{14} \approx 10^4$, which is quite appropriate for studying the flow of muons in the core region of an EAS.

To control the measurement process, a custom set of software tools was designed on the basis of the CAENDigitizer library provided by the CAEN company [32].

During the measurements, the digitization of scintillation pulses from the hodoscope detectors is synchronized by a trigger signal elaborated at the moment of EAS passage by the surface shower installation of the Tien Shan station and transmitted to the underground room via a coaxial cable. Since the trigger pulse reaches the hodoscope with a noticeable time delay of about (20–23) μs, the output oscillograms of hodoscope detectors are stored for

36 μs immediately preceding the trigger moment. Thus, the volume of the numerical data on an EAS event transmitted from the ADC equals $1.8 \cdot 10^4$ values per detector. Of these, 5000 amplitude values in each detector signal waveform, corresponding to the period of $(-27 \ldots -10)$ μs before the trigger, are selected to be permanently kept in the database for further off-line analysis. The time borders ensure hitting the signal of the particles of the shower front at approximately the middle of that period.

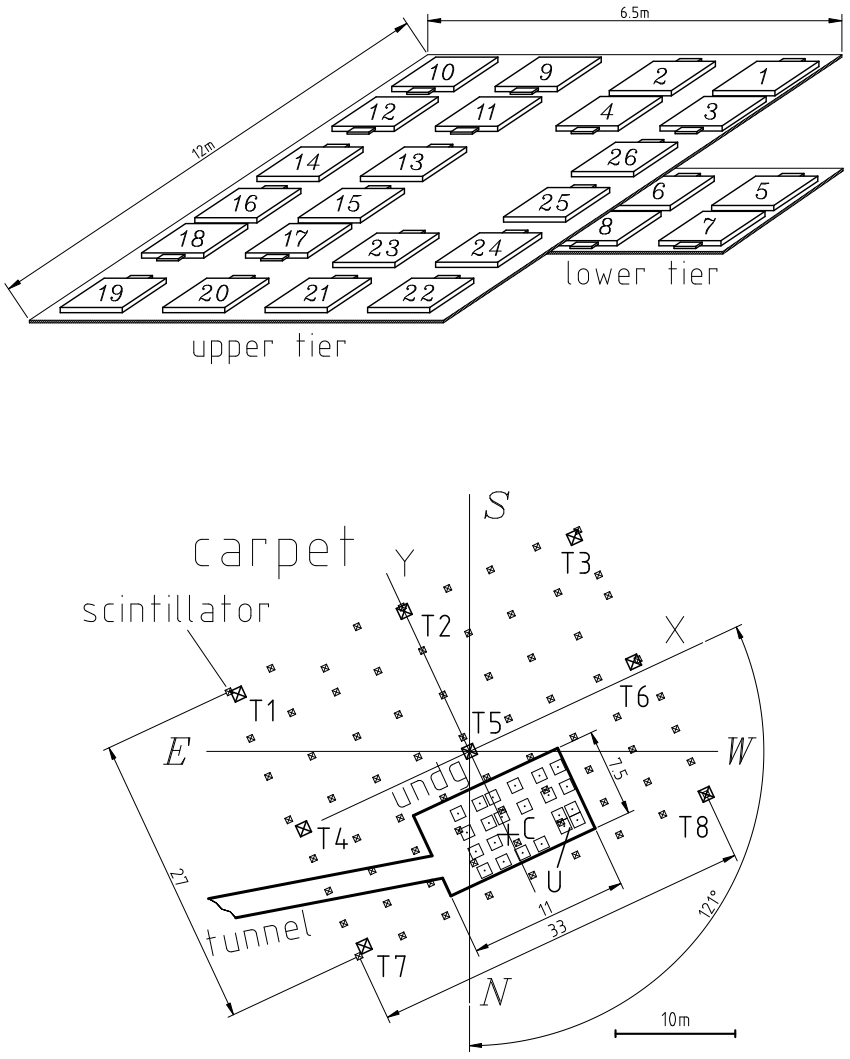


Figure 1. (**Top**): schematic of the muon hodoscope detectors installed in the underground room at the Tien Shan station. (**Bottom**): mutual disposition of the underground room (*undg*), detectors of the muon hodoscope (dotted squares *U*), and the surface *carpet* of scintillation detectors of EAS particles (crossed squares) in the central area of the Tien Shan shower installation. The *C* marker indicates the origin of the coordinate frame used for representation of the EAS axis position.

A sample of scintillation signal oscillograms synchronously obtained from the detectors of the hodoscope in an event of an EAS passage is shown in Figure 2. The two pictures in the figure are window screenshots of a program facility especially designed to control both the collection of the hodoscope data, processing them in real time, and storing the results in database tables. The program simultaneously supports two visualization modes: *Scope* and *Operation*; switching to each is possible on a corresponding tab.

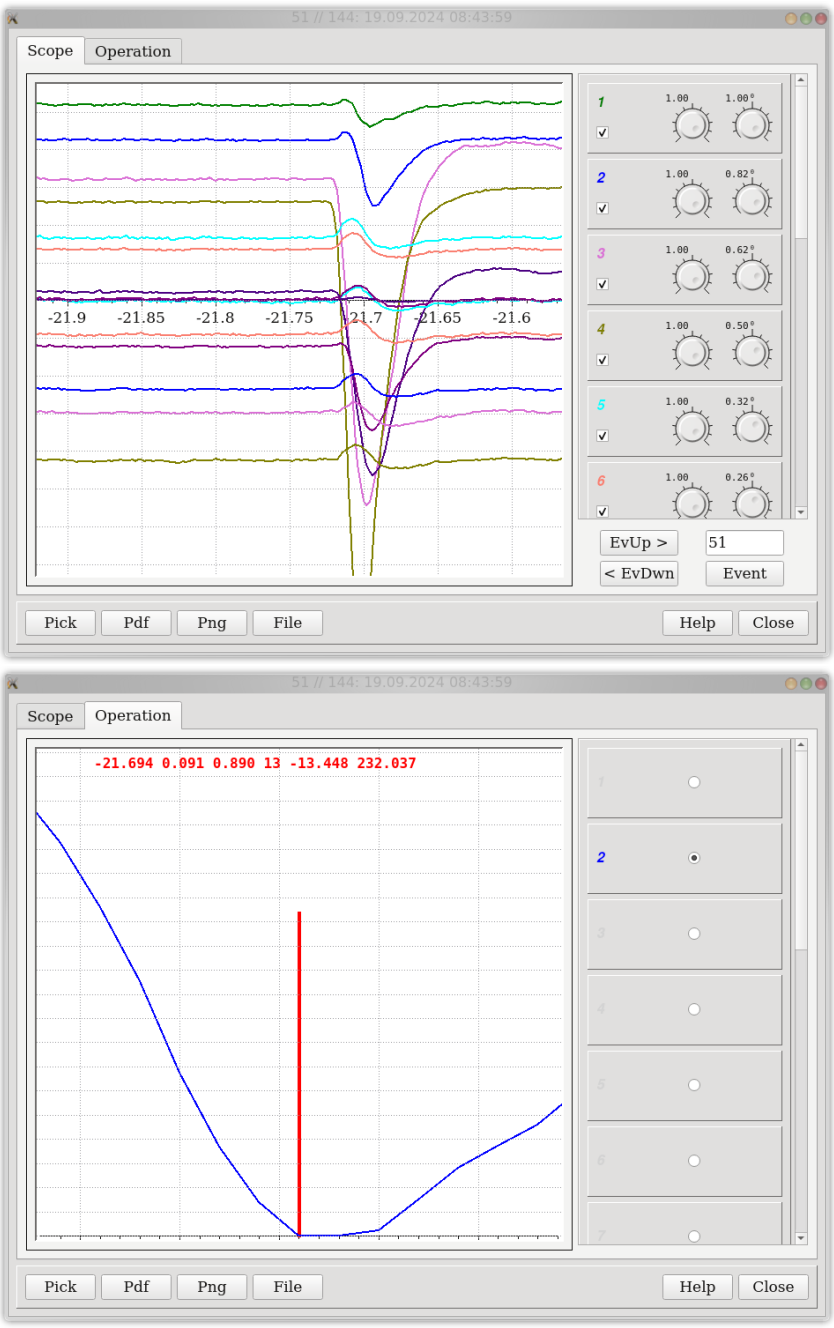


Figure 2. The window of the program designed to control the measurements at the underground hodoscope (see text). (**Top**)—working in the *Scope* mode: a set of the oscillograms of scintillation pulse signal from different detectors fixed by registration of the muon flux near an EAS core. The numbers at horizontal axis mean the time, in microseconds, preceding the arrival of the trigger from the surface installation of shower particles detectors; the units along the vertical axis are arbitrary. (**Bottom**)—the *Operation* mode: a normalized 1 μs long fragment of the oscillogram of one of the data channels (corresponding to the detector #2), automatically cut and pre-processed for application of the machine learning procedure searching for scintillation pulses. The vertical red line is a marker of the global minimum found in the waveform curve, and the inscription indicates the numerical parameters of the considered oscillogram to be used as features in the searching procedure.

The top frame of Figure 2 illustrates the outlook of the program window with the *Scope* mode activated, in which case the oscillograms of all registered signals are indicated in its graphic area. Nearly simultaneous deep dips in all oscillograms correspond to the short pulses (of negative polarity) generated by the hodoscope scintillators at the moment of

passage of the EAS core muons through the detector system. As follows from this picture, there is a significant difference between the oscillograms of various channels, both in the amplitude of the negative pulse and in the fine details of the waveform shape (e.g., relative value of the positive bump preceding the pulse). Such variability can be caused by both the features specific for a concrete EAS event (e.g., a unique distribution of the muon flux density over the hodoscope detectors) and constant peculiarities of the characteristics of each detector (such as the properties of the scintillator, parameters of the electronic signal transmission circuit, matching quality of the signal cable line, the intensity of crosstalk and power main interference on the cable, etc.). In order to automatically extract useful information from the data obtained from the detectors with such scattering of randomly changed or badly controlled individual parameters, it would be appropriate to resort to the help of modern methods of machine learning. An alternative might be an attempt to strictly tune the operation and signal transfer conditions individually for each detector, which is much more complicated, time consuming, and does not guarantee proper functioning stability over the long period of the hodoscope set-up exploitation; so, the second decision should be preferably rejected.

Another argument in favor of the application of a machine learning technique is the rather intensive flux of the raw original data in the measurements on the underground hodoscope: with a typical rate of EAS triggers from the shower installation of (250–300) h^{-1}, the hourly amount of registered oscillograms for a $\sim$20 detectors installation is 4000–6000. At the same time, the oscillograms with the presence of a scintillation pulse signal make up only (5–7)% of the statistics. Automated selection of such rare useful signals against an overwhelming flux of background cases is a relevant problem, again, for the algorithms of machine learning.

3. Processing the Hodoscope Data

An initial stage in processing a digitized oscillogram of an analog signal registered at the output of hodoscope detector is the identification of short pulses exhibiting a typical exponential shape. Generally, such pulses correspond to scintillation flashes from muon transits, but they may be obscured by noise oscillations originating from various electromagnetic interferences. To address this problem, a specialized software procedure employing a machine learning-based approach is proposed, which enables automated identification of useful pulse-like signatures in oscillograms and facilitates automatic tuning of the program's internal parameters according to the individual characteristics of each particular signal. During program development, several automatic classification algorithms were evaluated: Stochastic Gradient Descent (SGD), Support Vector Machines (SVM), and Random Forest (RF) [33].

Internally, the SGD classification model minimizes a cost function (defined as the mean squared error between known values in a labeled dataset and model predictions based on element features) recurrently during training. The tuned model parameters are subsequently used to classify new data presented to the model.

The SVM model establishes a boundary hypersurface in the multi-dimensional feature space that maximally separates the classes of input data. The position and shape of this boundary are defined by training the model on a labeled dataset, such that the boundary tries to maximize its distance from the closest elements of the training dataset (named *support vectors*).

Finally, Random Forest is an ensemble model utilizing a collection of Decision Tree (DT) type classifiers. In turn, each DT model seeks to split its input data into a few groups based on the feature values of every element in the dataset. The numerical thresholds used for splitting constitute a set of the model's internal parameters, tuned in the process of its

training. The training aims to minimize the inhomogeneity (*impurity*) of the content of each resulting group.

The practical implementation of all considered classifiers utilized the corresponding tools from the Scikit-Learn program library [34]. All classifiers operated using their default hyperparameter sets defined in the library.

When analyzing muon content in registered EAS events, the developed program treats the signals, which have come from different hodoscope detectors, fully independently of each other. Processing of a registered signal waveform $I(t)$ begins with normalizing the time series of signal data: $i(t) = (I(t) - I_{min})/(I_{max} - I_{min})$, where $I(t)$ is the measured signal amplitude at time t, and I_{min} and I_{max} are its minimum and maximum values over the entire oscillogram. The normalized distribution $i(t)$ thus varies in the limits $[0, 1]$, inclusively.

Next, the time t_{min} is identified, at which $i(t)$ reaches its global minimum (because of normalization, $i(t_{min}) = 0$), and a segment of the oscillogram within the limits $[t_{min} - t_g, t_{min} + t_g]$ is selected, where t_g, the *time gate*, is a fixed parameter of the pulse searching algorithm.

The correctness of data pre-processing can be verified in real time using the above-mentioned control facility switched to *Operation* mode. The bottom frame of Figure 2 illustrates this verification on a ± 0.5 μs fragment automatically cut from the normalized signal oscillogram of the detector #2. All other signals registered in that event were similarly pre-processed before application of the classification algorithms.

The selected group of $i(t)$ values, centered around the t_{min} point and bounded by the $t_{min} \pm t_g$ limits, forms the input feature set for the machine learning classifiers implemented by the program. Evidently, the length of the feature set depends on the accepted duration of the time gate. Additionally, this set is augmented with a few general parameters: the time position of the absolute amplitude minimum t_{min}, as expressed in microseconds, the width Δt of the oscillogram dip around the moment t_{min} measured at a half of its depth, in microseconds also, and the overall skew and kurtosis coefficients of the oscillogram as a whole. The latter two parameters characterize, correspondingly, the symmetry of the general shape of the original signal curve and the degree of its sharpness. To prevent losing useful information, they are calculated over the initial $I(t)$ distribution, before the application of any transformations.

The newly developed program for the automatic identification of the narrow scintillation pulses from muon passage was tested over a set of $\sim 2.8 \cdot 10^5$ real EAS events registered at the Tien Shan installation.

The initial training of the machine learning models used by the classifiers was performed over a limited dataset of about $3.8 \cdot 10^4$ manually labeled oscillograms, which makes up about 0.8% of the total statistics of signal waveforms. The labels assigned to the training oscillograms were the simple logical values: *Yes* or *No*, i.e., whether a scintillation pulse signal exists in an oscillogram at the moment t_{min} or not. Accordingly, the scintillation pulse searching program implements the binary version of all mentioned classifiers, also returning the *Yes/No* answer.

A sample of the signal waveforms manually labeled for use in training the classifier models is shown in Figure 3. The rather obvious difference of the *Yes*-class signals (left column) is seen both in the presence of a characteristically narrow asymmetric scintillation pulse on the oscillogram and in the relative depth of the global minimum in the normalized time distribution.

For evaluation of the automated pulse searching program, the labeled dataset was split into two parts, one of which, containing about 80% of labeled data, was used to train the classifiers, and the rest was used to test the results they returned. Before splitting,

the timestamps of labeled events were randomly shuffled over the whole period of data accumulation in order to avoid the influence on the training process of any slowly changing time-dependent bias, which could be present in the input data. For shuffling and splitting the dataset, as well as training the classifier models and verification of the results of their operation, the standard utilities of the Scikit-Learn program library were used.

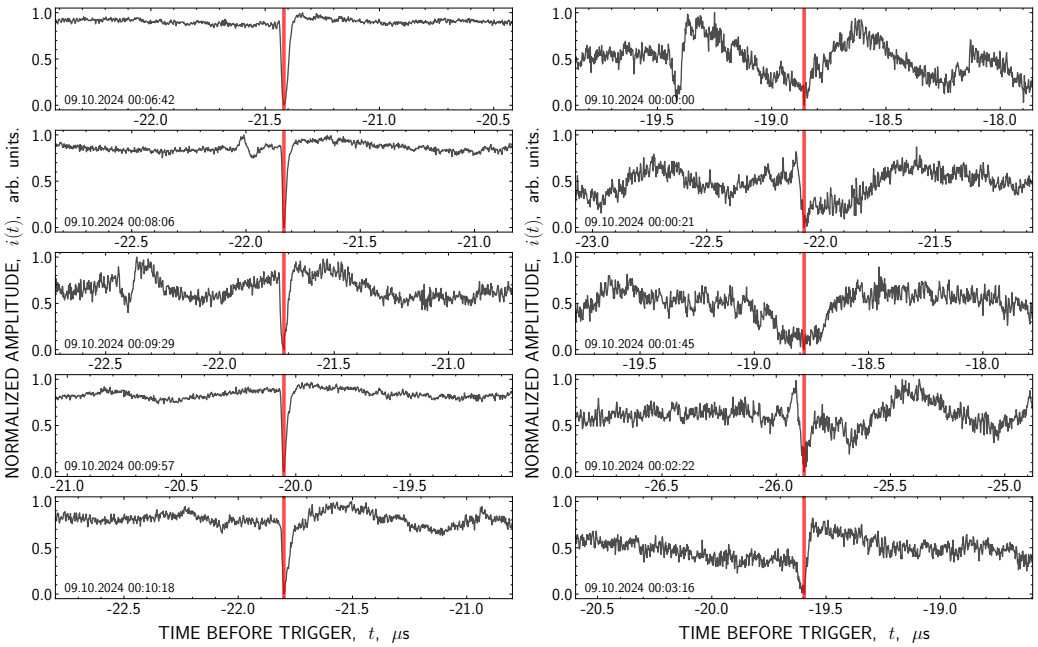


Figure 3. An example of labeled elements from the training dataset. The waveforms shown in the **left** column were manually marked as having a scintillation pulse signal and those in the **right** as not. Each graph presents a 1000-point long fragment of the normalized $i(t)$ distribution centered between the $t_{min} \pm t_g$ time limits ($t_g = 1$ μs); the red straight lines in the middle of every plot indicate the position of the global minimum in the complete distribution $I(t)$ at the moment t_{min} (see text).

The evaluation procedure was applied to all three classifiers mentioned above: SGM, SVM, and RF. Additionally, a Voting Classifier (VTG) based on the successive application of SVM and RF to the measured waveforms was also examined. In each case, four variants of input data preparation were tested, with the time gate duration parameter t_g set to 0.1, 0.5, 1.0, and 2 μs. The preliminary choice of the latter values was made experimentally, as a tradeoff between the rising amount of useful information at larger t_g (since the number of $i(t)$ counts accessible to classifier multiplies with the expansion of the time gate), which is good, and the simultaneous increase in the influence of random noise and parasitic oscillations on stability of classifier operation, which is bad.

The result of these calculations is illustrated by a set of receiver operating characteristics (ROC) plotted in Figure 4. Each ROC curve shows the percentage of true positive vs. false positive predictions made by a binary classifier, and ideally it should have a rectangular shape with the upper left corner coinciding with the point $(0, 1)$. In reality, the classifier performs the better, the closer its ROC approaches to this point. Judging from the plot in Figure 4, for the task of finding narrow pulse patterns among realistic scintillation detector signal waveforms, the VTG and SVM classifiers appear to be superior to the other two.

Detailed verification metrics for the binary classifiers are listed in Table 1. For each combination of "classifier type/time gate", the cells of this table present the values of precision (the percentage of correct estimates among all positive answers given by the classifier), recall (the percentage of true positive answers relative to the known amount of scintillation pulses present in the labeled test dataset), and the area under the ROC curve (the closer to one unit, the better). Accordingly to Table 1, the optimal option that

maximizes the performance appears to be the VTG classifier with the time gate t_g equal to 0.1 or 1 μs (the corresponding cells in the table are in bold). Further investigation showed that with $t_g = 1$ μs, the pulse search procedure better generalizes most real signals obtained from different scintillators, that were not a part of the training set for the machine learning algorithms. Apparently, this is due to the larger volume of useful information available to the classifier in the case with $t_g = 1$ μs; so, this option was chosen for further practical use. According to the corresponding cell in Table 1, in the case of the finally selected combination, VTG with $t_g = 1$ μs, the identification probability of a useful signal slightly exceeds 99%, and the admixture of falsely recognized pulses is about 0.6%.

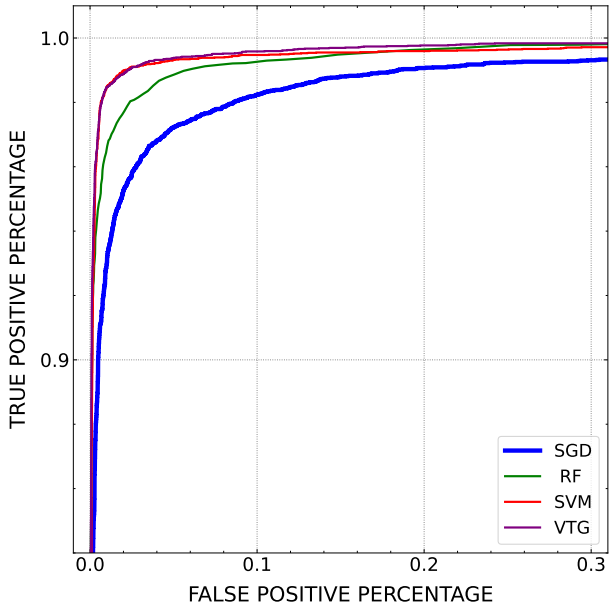


Figure 4. Classification quality score functions (ROC curves) for the program of automated narrow scintillation pulse searching trained on a set of 1 μs-long oscillograms using various algorithms of machine learning: stochastic gradient descent (SGD), support vector machines (SVM), and random forest (RF); as well as an ensemble of the latter two combined by voting classifier (VTG).

The program implementation of Random Forest Classifier in the Scikit-Learn library provides a useful ability to compose, during training, an importance score distribution of the features attributed to the items of input data from the viewpoint of the classification task. Figure 5 shows two such distributions calculated during testing the algorithms of the pulse search program. Here, the relative importance of each feature assigned to the oscillogram (the coefficient characterizing how much taking this feature into account reduces the impurity of the two classes into which the input dataset should be divided) is presented as a function of the feature's ordinal number in the full list of features: the position of the absolute minimum t_{min} (feature #0), the width of the negative depression of the oscillogram Δt (#1), the skew and kurtosis of the recorded oscillogram curve (#2 and #3), and the array $y(t)$ of normalized amplitude values centered around the point t_{min}. (Remember, that since the number of $y(t)$ samples depends on the length of the time gate t_g, the larger the time gate t_g is in a given data processing option, the more features are taken into account by the machine learning algorithm).

As expected, for the given classification problem, the most useful information is contained in the amplitude data describing the middle of the considered part of the oscillogram, where the front of a narrow scintillation pulse, if any, should be located. In addition, several leading characteristics describing the general shape of the curve are of great importance, i.e., t_{min}, Δt, etc., and in the case of the time gate extended up to 1 μs—also the tail of selected $i(t)$ distribution fragment, which allows excluding the influence of interfering fluctuations

that are more likely to affect the dataset perceived by the classifier when using a wider time gate.

Table 1. Performance metrics (precision, recall, and the area under ROC curve) for various binary classifiers trained on the labeled oscillograms of the hodoscope detector signals.

Method \ Time Gate	$t_g = 0.1$ μs	$t_g = 0.5$ μs	$t_g = 1.0$ μs	$t_g = 2.0$ μs
Stochastic Gradient Descent	prec.: 0.989 rec.: 0.988 area: 0.991	prec.: 0.988 rec.: 0.992 area: 0.993	prec.: 0.948 rec.: 0.988 area: 0.982	prec.: 0.976 rec.: 0.991 area: 0.990
Support Vector Machine	prec.: 0.991 rec.: 0.996 area: 0.996	prec.: 0.989 rec.: 0.995 area: 0.995	prec.: 0.987 rec.: 0.995 area: 0.994	pre.: 0.993 rec.: 0.992 area: 0.994
Random Forest	prec.: 0.993 rec.: 0.991 area: 0.994	prec.: 0.994 rec.: 0.992 area: 0.995	prec.: 0.993 rec.: 0.989 area: 0.993	pre.: 0.994 rec.: 0.988 area: 0.992
Voting Classifier	**prec.: 0.991** **rec.: 0.994** **area: 0.995**	prec.: 0.990 rec.: 0.994 area: 0.994	**prec.:0.991** **rec.: 0.994** **area: 0.995**	pre.: 0.993 rec.: 0.992 area: 0.995

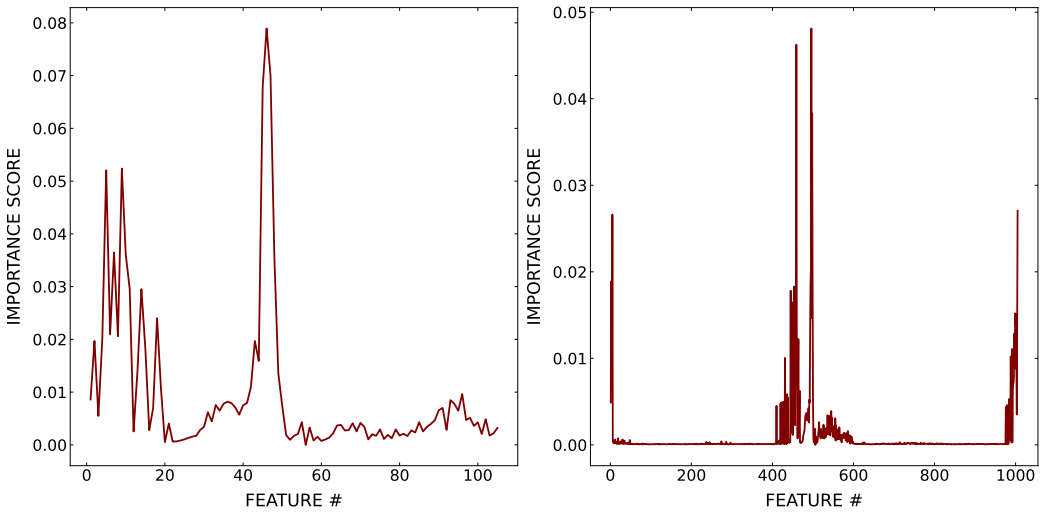


Figure 5. The score of relative features' importance as defined when training the random forest classifier over the 0.1 μs- and 1 μs-long oscillogram fragments.

Lastly, it should be noted that an analysis of the entire set of hodoscope data using a trained classifier of the considered type takes only a few tens of milliseconds on a fairly typical modern computer configuration (an Intel Core i7–12700 processor with a 1400 MHz clock frequency, Intel, Santa Clara, CA, USA). Since, in practice, the EASs registered by the Tien Shan installation follow to each other with a typical rate of 3–5 events per minute, the said operation speed allows us to implement processing of the EAS events in real time, with immediate visualization of the measurement results. This may be beneficial in some situations, especially when fine-tuning the installation's hardware.

4. Verification of the Muon Hodoscope Data Processing Procedure

4.1. Calibration of Hodoscope Detectors

The capabilities of the proposed method for studying the muon component of cosmic rays with the help of the underground hodoscope were tested by applying the procedure

described above to the experimental data obtained during the joint operation of the hodoscope and the Tien Shan shower installation, when $2.8 \cdot 10^5$ EAS events were registered. Since only the first 16 scintillators were operating in the underground hodoscope at that time, the amount of EASs corresponds to $4.6 \cdot 10^6$ analog signal oscillograms recorded at the output of hodoscope detectors. Using the considered procedure, $3.2 \cdot 10^5$ scintillation pulses were identified among these waveforms.

The estimate of the intensity of muon flux in each specific scintillator of the underground hodoscope is based on the depth of the negative pulse in the time distribution function $I(t)$ describing its waveform. The pulse amplitude should be determined at the global minimum point t_{min} and within the limits $t_{min} \pm t_g$. Indeed, if the pulse search program has recognized the presence of a fast scintillation pattern on the oscillogram, the front of the negative polarity pulse corresponding to the scintillation flash must be located at the point t_{min}, and the relative amplitude A, defined as the difference between the mean background level of the time series and the minimal amplitude at the scintillation peak,

$$A = \langle I(t) \rangle - I(t_{min}), t \in [t_{min} - t_g, t_{min} + t_g], \quad (1)$$

may be used as a measure of scintillation intensity. Here, the mean background level $\langle I(t) \rangle$ is calculated as the average value of the time series $I(t)$ within the limits $t_{min} \pm t_g$, and constraining the considered fragment to a double duration of the time gate allows avoiding the influence of random interferences on the pulse amplitude estimated in this way. Since the values $I(t)$ are initially expressed as arbitrary units (momentary codes of ADC conversion), the use of the dimensionless parameter A is quite adequate for the task.

The algorithm of the signal amplitude calculation is illustrated by the left graph in Figure 6. In this fairly typical case, the standard deviation of random fluctuations σ equals 10.5, while the useful pulse has an amplitude of $A = 147$; so, the signal-to-noise ratio is about 14.

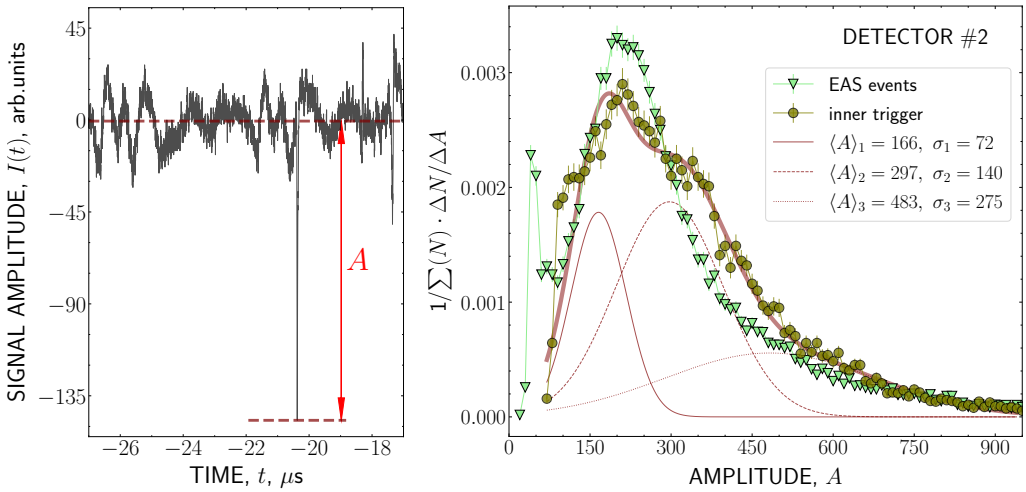


Figure 6. Calibration of the hodoscope scintillation detectors. (**Left**): definition of the pulse amplitude parameter A. (**Right**): a sample of the amplitude distribution of scintillation pulses registered in synchronization with an external EAS trigger (triangles) and with inner self-triggering of the ADC system (circles). The bold continuous curve marks an approximation of the last distribution with a combination (2) of three Gaussian functions (also shown separately with thin lines).

On the other hand, the intensity of scintillation light depends on the number of relativistic charged particles passing through detector nearly simultaneously, i.e., during a time significantly shorter than the decay time of scintillation flash. Thus, it should be expected that the amplitude A of pulse signals registered at the output of the underground scintilla-

tion detectors will be related to the density of muon flux, and the statistical distribution of the form $\Delta N/\Delta A$ will reflect the relative number of cases when 1, 2, etc. particles at once were passing through the detector.

To check the operation quality of the underground hodoscope detectors and adequacy of the proposed method for processing their data, the distributions $\Delta N/\Delta A$ were plotted of scintillation pulses found in the oscillograms of each detector. For this purpose, two types of datasets were used: the time series of detector signal synchronized with the shower trigger, which were recorded during abovementioned period of EAS registration, and the oscillograms obtained with the inner trigger of the ADC system, i.e., by starting the digitization process with the same scintillation pulse that was currently being digitized.

A typical sample of the $\Delta N/\Delta A$ distribution obtained in this way from a properly tuned and normally operating detector is shown in Figure 6, right plot. It is evident that the self-triggering of ADC system results in a rather complex spectrum of A amplitudes (shown with circles in the plot), which, nevertheless, can be approximated by a combination of three normal distributions,

$$P(A) = n_1 \cdot G_1(A, \langle A\rangle_1, \sigma_1) + n_2 \cdot G_2(A, \langle A\rangle_2, \sigma_2) + n_3 \cdot G_3(A, \langle A\rangle_3, \sigma_3), \quad (2)$$

where G_i are Gaussian functions, $\langle A\rangle_i$ and σ_i—correspondingly, their mean values and standard deviations, and n_i are normalization coefficients. In Figure 6, this approximation is represented by a bold continuous curve.

It is noteworthy, that after fitting the sum distribution $P(A)$ under experimental points, the resulting mean values of the Gaussian components, $\langle A\rangle_1 = 166$, $\langle A\rangle_2 = 297$, and $\langle A\rangle_3 = 483$, are related to each other as roughly 1:2:3. It is natural to assume that the leading component, with the mean value $\langle A\rangle_1$, corresponds to the cases of a single muon passing through the detector, and the components with $\langle A\rangle_2$ and $\langle A\rangle_3$ correspond to double and triple passages, respectively. Thus, in general, the obtained $\Delta N/\Delta A$ spectrum indicates a fairly plausible result of the detector operation and effectiveness of the suggested procedure for scintillation pulse searching.

Further on, systematic plotting of amplitude distributions of the form $\Delta N/\Delta A$, measured for each detector with the ADC system triggered by its output pulses, may be used as a standard procedure for regular diagnostics, operational check, and individual fine-tuning of the detectors that make up the underground hodoscope.

Additionally, the position of the first peak in Gaussian approximation, $\langle A\rangle_1$, can be considered as a calibration coefficient to estimate from the measured amplitude A of the scintillation pulse the absolute value of the local muon density ρ_μ at the detector point:

$$\rho_\mu = A/\langle A\rangle_1/(\epsilon \cdot S_d). \quad (3)$$

Here, $\epsilon = 0.99$ and $S_d = 1\,\mathrm{m}^2$ correspond to the registration efficiency and sensitive area of the detector.

As for the $\Delta N/\Delta A$ spectrum obtained in the cases of EAS passages accompanied by the trigger from the Tien Shan shower installation, the distribution shown with triangles in the plot of Figure 6 indicates that in majority of EAS events with non-zero muon signal, the average amplitude of the detected scintillations falls into an interval between $\langle A\rangle_1$ and $\langle A\rangle_2$, which corresponds to a density of about 1–2 muon particles per $1\,\mathrm{m}^2$ of the sensitive detector area. This is, again, a rather reasonable conclusion: according to the lateral distribution functions of the 5 GeV muons density measured in earlier experiments at the Tien Shan station and presented in [1] (see Figure 9 there), the muon flux falls below one particle per square meter already at a distance of about 10 m from the shower axis even in the $N_e \approx 10^6$ sized EASs, while the bulk of the showers considered in plotting the graph

in Figure 6 are almost an order of magnitude smaller in size, and no special selection of the events with the axis location close to the underground hodoscope was done.

4.2. Comparison with EAS Data

As follows from the experimental measurements of the EAS muon component discussed in [1], the total number of muons in an EAS depends on its size N_e (i.e., on the primary energy E_0), and the density of muon flux rapidly diminishes with the increasing distance from the shower axis. Thus, the correctness of the muon hodoscope operation and the validity of the proposed procedure of automatic search for scintillation pulses may be verified by comparing the data of the underground hodoscope obtained in the EAS triggered events with the parameters of the corresponding showers.

For further analysis, the total number of charged particles in a shower (shower size) N_e and two coordinates of EAS axis in the plane of the shower installation (x, y) were determined for all EASs according to the method described in [30]. The coordinates were used for the calculation of the distance R between the EAS axis and the installation's center point, with the underground room and the muon hodoscope situated beneath (see Figure 1).

After estimation of the shower parameters, the whole set of registered EAS events was split into several groups, depending on the mean shower size N_e and the mean distance R to the shower axis from the middle of the muon hodoscope. Then, the mean amplitude $\overline{A}$ of the underground hodoscope signal was estimated as an average of individual amplitudes A calculated in accordance with the above discussed algorithm for the 16 hodoscope detectors in all events belonging to each particular combination of (N_e, R). By definition, the value $\overline{A}$ may be considered as a single dimensionless parameter characterizing the mean relative intensity of the flux of the muons, which have come through the total area of all sensitive elements of the hodoscope in the shower events with the close values of the parameters N_e and R.

Both stages of data processing: calculating the shower characteristics for the registered EAS events and searching for scintillation pulses from muon passage in the output oscillograms of the underground hodoscope detectors were accomplished independently of each other.

The result of this operation is presented in Figure 7 as the distribution of the average amplitude of scintillation pulses in the detectors of the underground hodoscope $\overline{A}$, depending on the mean size N_e and the axis distance R of the corresponding EAS. The vertical error bars in the plot indicate the standard deviation of the amplitude estimates $\overline{A}$ obtained in the EAS events associated with every (N_e, R) cell, and the horizontal bars denote the width of the intervals over the distance parameter R.

As can be seen in the plot, the intensity of the hodoscope response to the passage of EAS-related muons obviously correlates with both the size of a shower and the distance to its core, so that the mean scintillation amplitude $\overline{A}$ systematically grows with rising N_e and diminishes further from the EAS center. Such behavior is quite natural for the muon flux in EAS; so, the correlation plot in Figure 7 once again confirms both the efficiency of the considered hardware and adequateness of the developed program toolchain for deducing the intensity of cosmic ray muons from the original data of the underground hodoscope.

It is also interesting to note that Figure 7 so far confirms the conclusion about the accelerated rise of the muon flux in the core region ($R < 10$ m) of a large-sized EAS with $N_e > 10^6$, which was previously formulated in [1] on the basis of the data on evaporation neutrons from muonic interaction, though further investigation with qualitatively larger statistics of registered EASs is necessary. Judging by the present plot, it should be primarily the growth of muon multiplicity, but not their average energy, that takes place in the central region of EAS above the 3 PeV "knee" of the primary cosmic ray spectrum.

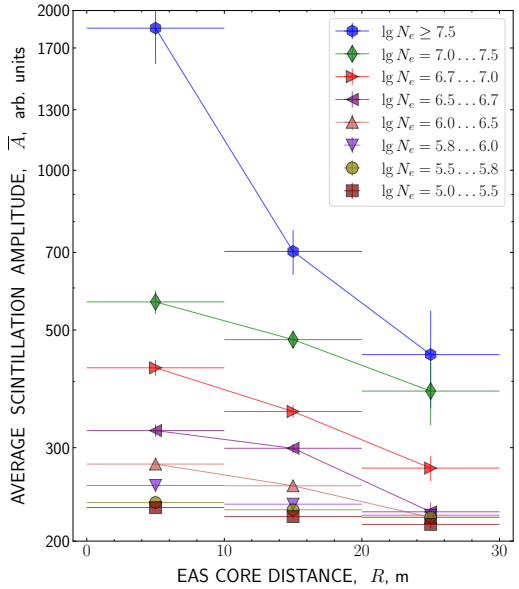


Figure 7. The correlation plot between the average amplitude of scintillation pulses $\overline{A}$ registered by the detectors of the underground muon hodoscope and the parameters (total number of electrons N_e and the distance to the shower axis R) of the corresponding EAS.

5. Conclusions

For further study of the muon component of cosmic rays at the Tien Shan High Mountain Station, a new hodoscope set-up was designed, based on the modern large-sized scintillation charged particles detectors and application of machine learning techniques to analyze the signal acquired at detector output. Installed in the underground room of the station, the hodoscope ensures registration of muons with an energy threshold of (5–8) GeV, while the dynamic range of the density measurement of muon flux, $\sim 10^4$, is quite appropriate for investigating the muon component of EAS in the region of the shower core.

The newly developed technique for automated searching for narrow pulse patterns in an output oscillogram of the scintillation detector uses modern machine learning algorithms and ensures a ∼99% detection probability of a typical scintillation signal, with a false positive rate below 1%.

Verification of the designed hardware in combination with the software tools of data processing on real EAS data demonstrated the correctness of the proposed method and its applicability to investigation of the muon flux in an EAS. In addition, a practical procedure was designed for testing, adequate tuning, and calibration of the particle detectors in the underground muon hodoscope.

Preliminary confirmation was obtained for the earlier conclusion on the change in the properties of the muon component in the EASs above the 3 PeV "knee" of the primary cosmic ray spectrum, which occurs mainly due to the increase in the multiplicity of muon particles and not the average muon energy.

Author Contributions: Conceptualization, A.S.,K.M., V.R., N.S., T.S. (Turlan Sadykov) and V.Z.; data curation, A.S.; formal analysis, A.S., A.B., O.K., Y.M., S.U. and N.Y.; funding acquisition, N.S. and T.S. (Turlan Sadykov); investigation, A.S., S.M., Y.M., N.S., S.S., T.S. (Tairzhan Skokbayev), I.S., S.U., L.V. and N.Y.; methodology, A.S., V.R. and V.Z.; project administration, K.M., V.R., N.S., T.S. (Turlan Sadykov) and V.Z.; software, A.S., I.S., S.U. and N.Y.; writing—original draft, A.S., S.U. and N.Y.; writing—review and editing, A.S., I.S., S.U. and N.Y. All authors have read and agreed to the published version of the manuscript.

Funding: 1. This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Grant No. IRN BR21881941. 2. This research was funded by the Committee of Science of the Ministry of Science and Higher Education of the Republic of Kazakhstan, Grant No. IRN AP23488396.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shepetov, A.L.; Shaulov, S.B.; Likiy, O.I.; Ryabov, V.A.; Sadykov, T.K.; Saduev, N.O.; Zhukov, V.V. The rise of muon content in extensive air showers after the 3 PeV knee of the primary cosmic ray spectrum according to data of the Tien Shan mountain installation. *Astropart. Phys.* **2021**, *133*, 102642. [CrossRef]
2. Lattes, C.M.G.; Fujimoto, Y.; Hasegawa, S. Hadronic interactions of high energy cosmic-ray observed by emulsion chambers. *Phys. Rep.* **1980**, *65*, 151. [CrossRef]
3. PAMIR Collab. Observation of a high-energy cosmic-ray family caused by a Centauro-type nuclear interaction in the joint emulsion chamber experiment at the Pamirs. *Phys. Lett. B* **1987**, *190*, 226–233. [CrossRef]
4. Yakovlev, V.I. Long flying component produced by protons at energies above 12 TeV in the ionization calorimeter. *Nucl. Phys. B Proc. Suppl.* **2003**, *122*, 417–421. [CrossRef]
5. Chacaltaya and Pamir Collaboration; Arisawa, T.; Fujimoto, Y.; Hasegawa, S.; Honda, K.; Ito, H.; Kopenkin, V.V.; Semba, H.; Tamada, M.; Yokoi, K.; et al. Observation of attenuation behavior of hadrons in extremely high energy cosmic ray interactions: New hadronic state? *Nucl. Phys. B* **1994**, *424*, 241–287. [CrossRef]
6. Borisov, A.S.; Denisova, V.G.; Galkin, V.I.; Guseva, Z.M.; Kanevskaya, E.A.; Kogan, M.G.; Mukhamedshin, R.A.; Nazarov, S.I.; Puchkov, V.S.; Sadykov, T.K. Investigation of abnormal absorption of cosmic-ray hadrons in lead calorimeters. *EPJ Web Conf.* **2019**, *208*, 06001. [CrossRef]
7. Borisov, A.S.; Guseva, Z.M.; Karpova, S.A.; Kempa, J.; Krys, A.; Maximenko, V.M.; Puchkov, V.S.; Slavatinsky, S.A. High energy gamma-families with halo and mass composition of primary cosmic rays in energy region above 10 PeV. *Nucl. Phys. B—Proc. Suppl.* **2001**, *97*, 113–115. [CrossRef]
8. Puchkov, V.S.; Borisov, A.S.; Guseva, Z.M.; Denisova, V.G.; Kanevskaya, E.A.; Maximenko, V.M.; Pyatovsky, S.E.; Slavatinsky, S.A. Mass composition of primary cosmic rays at energies of 1–1000 PeV according to data of experiment "Pamir". *Nucl. Phys. B—Proc. Suppl.* **2006**, *151*, 236–239. [CrossRef]
9. Borisov, A.; Denisova, V.; Guseva, Z.; Kempa, J.; Kanevskaya, E.; Maximenko, V.; Puchkov, V.; Pyatovsky, S.; Slavatinsky, S. Recent results and modern status of the PAMIR experiment. *Nucl. Phys.—Proc. Suppl.* **2008**, *175–176*, 143–148. . [CrossRef]
10. Shaulov, S.B.; Vil'danova, L.I.; Kupriyanova, E.A.; Ryabov, V.A.; Shepetov, A.L. Scaling violation in interaction of cosmic ray hadrons and the nature of the 3 PeV knee in the spectrum of primary cosmic rays. *J. Phys. G Nucl. Part. Phys.* **2021**, *48*, 125202. [CrossRef]
11. Antoni, T.; Apel, W.D.; Badea, A.F.; Bekk, K.; Bercuci, A.; Blümer, J.; Bozdog, H.; Brancus, I.M.; Chilingarian, A.; Daumiller, K.; et al. KASCADE measurements of energy spectra for elemental groups of cosmic rays: Results and open problems. *Astropart. Phys.* **2005**, *24*, 1–25. [CrossRef]
12. Garyaka, A.P.; Martirosov, R.M.; Ter-Antonyan, S.V.; Nikolskaya, N.M.; Gallant, Y.A.; Jones, L.W.; Procureur, J. Rigidity-dependent cosmic ray energy spectra in the knee region obtained with the GAMMA experiment. *Astropart. Phys.* **2007**, *28*, 169–181. [CrossRef]
13. Haungs, A. Cosmic Rays from the Knee to the Ankle. *Phys. Procedia* **2015**, *61*, 425–434. [CrossRef]
14. Thoudam, S.; Rachen, J.P.; van Vliet, A.; Achterberg, A.; Buitink, S.; Falcke, H.; Hörandel, J.R. Cosmic-ray energy spectrum and composition up to the ankle: The case for a second Galactic component. *Astron. Astrophys.* **2016**, *595*, A33. [CrossRef]
15. Abu-Zayyad, T.; Belov, K.; Bird, D.J.; Boyer, J.; Cao, Z.; Catanese, M.; Chen, G.F.; Clay, R.W.; Covault, C.E.; Cronin, J.W.; et al. Evidence for changing of cosmic ray composition between 10^{17} and 10^{18} eV from multicomponent measurements. *Phys. Rev. Lett.* **2000**, *84*, 4276–4279. [CrossRef]
16. Avati, V.; Dick, L.; Eggert, K.; Ström, J.; Wachsmuth, H.; Schmeling, S.; Ziegler, T.; Brühl, A.; Grupen, C. Cosmic multi-muon events observed in the underground CERN-LEP tunnel with the ALEPH experiment. *Astropart. Phys.* **2003**, *19*, 513–523. [CrossRef]
17. L3 Collaboration; Achard, P.; Adriani, O.; Aguilar-Benitez, M.; van den Akker, M.; Alcaraz, J.; Alemanni, G.; Allaby, J.; Aloisio, A.; Alviggi, M.G.; et al. Measurement of the atmospheric muon spectrum from 20 to 3000 GeV. *Phys. Lett. B* **2004**, *598*, 15–32. [CrossRef]

18. Abdallah, J.; Abreu, P.; Adam, W.; Adzic, P.; Albrecht, T.; Alemany-Fernandez, R.; Allmendinger, T.; Allport, P.P.; Amaldi, U.; Amapane, N.; et al. Study of multi-muon bundles in cosmic ray showers with the DELPHI detector at LEP. *Astropart. Phys.* **2007**, *28*, 273–286. [CrossRef]
19. Bogdanov, A.G.; Gromushkin, D.M.; Kokoulin, R.P.; Mannocchi, G.; Petrukhin, A.A.; Saavedra, O.; Trinchero, G.; Chernov, D.V.; Shutenko, V.V.; Yashin, I.I. Investigation of the properties of the flux and interaction of ultrahigh-energy cosmic rays by the method of local-muon-density spectra. *Phys. At. Nucl.* **2010**, *73*, 1852–1869. [CrossRef]
20. The ALICE Collaboration. Study of cosmic ray events with high muon multiplicity using the ALICE detector at the CERN Large Hadron Collider. *JCAP* **2016**, *2016*, 32. [CrossRef]
21. Yurina, E.A.; Barbashina, N.S.; Bogdanov, A.G.; Kindin, V.V.; Kokoulin, R.P.; Kompaniets, K.G.; Mannocchi, G.; Petrukhin, A.A.; Trinchero, G.; Khokhlov, S.S.; et al. Status of the NEVOD–DECOR Experiment on the Study of Muon Bundles Energy Deposit. *Bull. Russ. Acad. Sci. Phys.* **2021**, *85*, 455–457. [CrossRef]
22. Bakatanov, V.N.; Novosel'tsev, Y.F.; Novosel'tseva, R.V. Detection of muon groups with multiplicity $n_\mu > 1500$ at the Bacsan underground scintillation telescope. *Astropart. Phys.* **1997**, *8*, 59–66. [CrossRef]
23. Apel, W.D.; Arteaga-Velázquez, J.C.; Bekk, K.; Bertaina, M.; Blümer, J.; Bozdog, H.; Brancus, I.M.; Cantoni, E.; Chiavassa, A.; Cossavella, F.; et al. Probing the evolution of the EAS muon content in the atmosphere with KASCADE-Grande. *Astropart. Phys.* **2017**, *95*, 25–43. [CrossRef]
24. Amenomori, M.; Bi, X.J.; Chen, D.; Chen, T.L.; Chen, W.Y.; Cui, S.W.; Danzengluobu; Ding, L.K.; Feng, C.F.; Feng, Z.; et al. The cosmic ray energy spectrum measured with the new Tibet hybrid experiment. *EPJ Web Conf.* **2019**, *208*, 03001. [CrossRef]
25. Abbasi, R.; Ackermann, M.; Adams, J.; Aguilar, J.A.; Ahlers, M.; Ahrens, M.; Alameddine, J.M.; Alves, A.A.; Amin, N.M.; Andeen, K.; et al. Density of GeV muons in air showers measured with IceTop. *Phys. Rev. D* **2022**, *106*, 032010. [CrossRef]
26. Aab, A.; Abreu, P.; Aglietta, M.; Albury, J.M.; Allekotte, I.; Almela, A.; Alvarez Castillo, J.; Alvarez-Muniz, J.; Anastasi, G.A.; Anchordoqui, L.; et al. Direct measurement of the muonic content of extensive air showers between 2×10^{17} and 2×10^{18} eV at the Pierre Auger Observatory. *Eur. Phys. J.* **2020**, *C 80*, 751. [CrossRef]
27. Dembinski, H.P.; Arteaga-Velázquez, J.C.; Cazon, L.; Gonzalez, J.; Itow, Y.; Ivanov, D.; Kalmykov, N.N.; Karpikov, I.; Müller, S.; Pierog, T.; et al. Report on tests and measurements of hadronic interaction properties with air showers. *EPJ Web Conf.* **2019**, *210*, 02004. [CrossRef]
28. Ostapchenko, S.; Sigl, G. On the model uncertainties for the predicted muon content of extensive air showers. *Astropart. Phys.* **2024**, *163*, 103004. [CrossRef]
29. Britvich, G.I.; Chernichenko, S.K.; Chubenko, A.P.; Gilitsky, Y.V.; Kushnirenko, A.E.; Mamidzhanyan, E.A.; Pavlyuchenko, V.P.; Shein, I.V.; Soldatov, A.P.; Shepetov, A.L.; et al. The large scintillation charged particles detector of the Tien-Shan complex ATHLET. *Nucl. Instrum. Methods A* **2006**, *564*, 225–234. [CrossRef]
30. Chubenko, A.; Shepetov, A.; Antonova, V.; Borisov, A.; Dalkarov, O.; Kryakunova, O.; Mukashev, K.; Mukhamedshin, R.; Nam, R.; Nikolaevsky, N.; et al. New complex EAS installation of the Tien Shan mountain cosmic ray station. *Nucl. Instrum. Methods A* **2016**, *832*, 158–178. [CrossRef]
31. CAEN. UM2792—V1730/VX1730 & V1725/VX1725 User Manual Rev. 8. 2024. Available online: https://www.caen.it/products/v1730/ (accessed on 15 May 2025).
32. CAEN. User Manual UM1935 CAENDigitizer Library Rev. 21. 2021. Available online: https://www.caen.it/products/caendigitizer-library/ (accessed on 15 May 2025).
33. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed.; O'Reilly Media: Sebastopol, CA, USA, 2019.
34. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Muller, A.; Nothman, J.; Louppe, G.; et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.