Contents lists available at ScienceDirect

Nuclear Inst. and Methods in Physics Research, A

Full Length Article

# The online data filter for the KM3NeT neutrino telescopes

O. Adriani [2,1], S. Aiello [3], A. Albert [4,56], A.R. Alhebsi [5], M. Alshamsi [6], S. Alves Garre [7],
A. Ambrosone [9,8], F. Ameli [10], M. Andre [11], L. Aphecetche [12], M. Ardid [13], S. Ardid [13],
J. Aublin [14], F. Badaracco [16,15], L. Bailly-Salins [17], Z. Bardačová [19,18], B. Baret [14],
A. Bariego-Quintana [7], Y. Becherini [14], M. Bendahman [8], F. Benfenati Gualandi [21,20],
M. Benhassi [22,8], M. Bennani [17], D.M. Benoit [23], E. Berbee [24], E. Berti [2], V. Bertin [6], P. Betti [2],
S. Biagi [25], M. Boettcher [26], D. Bonanno [25], S. Bottai [2], A.B. Bouasla [57], J. Boumaaza [27],
M. Bouta [6], M. Bouwhuis [24], C. Bozza [28,8], R.M. Bozza [9,8], H. Brânzaş [29], F. Bretaudeau [12],
M. Breuhaus [6], R. Bruijn [30,24], J. Brunner [6], R. Bruno [3], E. Buis [31,24], R. Buompane [22,8],
J. Busto [6], B. Caiffi [16], D. Calvo [7], A. Capone [10,32], F. Carenini [21,20], V. Carretero [30,24], T. Cartraud [14],
P. Castaldi [33,20], V. Cecchini [7], S. Celli [10,32], L. Cerisy [6], M. Chabab [34], A. Chen [35],
S. Cherubini [36,25], T. Chiarusi [20], M. Circella [37], R. Clark [38], R. Cocimano [25], J.A.B. Coelho [14],
A. Coleiro [14], A. Condorelli [14], R. Coniglione [25], P. Coyle [6], A. Creusot [14], G. Cuttone [25],
R. Dallier [12], A. De Benedittis [22,8], G. De Wasseige [38], V. Decoene [12], P. Deguire [6], I. Del
Rosso [21,20], L.S. Di Mauro [25], I. Di Palma [10,32], A.F. Díaz [39], D. Diego-Tortosa [25], C. Distefano [25],
A. Domi [40], C. Donzaud [14], D. Dornic [6], E. Drakopoulou [41], D. Drouhin [4,56], J.-G. Ducoin [6],
P. Duverne [14], R. Dvornický [19], T. Eberl [40], E. Eckerová [19,18], A. Eddymaoui [27], T. van Eeden [24],
M. Eff [14], D. van Eijk [24], I. El Bojaddaini [42], S. El Hedri [14], S. El Mentawi [6], A. Enzenhöfer [6],
G. Ferrara [36,25], M.D. Filipović [43], F. Filippini [20], D. Franciotti [25], L.A. Fusco [28,8], T. Gal [40],
J. García Méndez [13], A. Garcia Soto [7], C. Gatius Oliver [24], N. Geißelbrecht [40], E. Genton [38],
H. Ghaddari [42], L. Gialanella [22,8], B.K. Gibson [23], E. Giorgio [25], I. Goos [14], P. Goswami [14],
S.R. Gozzini [7], R. Gracia [40], B. Guillon [17], C. Haack [40], H. van Haren [44], A. Heijboer [24],
L. Hennig [40], J.J. Hernández-Rey [7], A. Idrissi [25], W. Idrissi Ibnsalih [8], G. Illuminati [20],
R. Jaimes [7], O. Janik [40], D. Joly [6], M. de Jong [45,24],*, P. de Jong [30,24], B.J. Jung [24],
P. Kalaczyński [58,46], J. Keegans [23], V. Kikvadze [47], G. Kistauri [48,47], C. Kopper [40],
A. Kouchner [49,14], Y.Y. Kovalev [50], L. Krupa [18], V. Kueviakoe [24], V. Kulikovskiy [16],
R. Kvatadze [48], M. Labalme [17], R. Lahmann [40], M. Lamoureux [38], G. Larosa [25], C. Lastoria [17],
J. Lazar [38], A. Lazo [7], S. Le Stum [6], G. Lehaut [17], V. Lemaître [38], E. Leonora [3], N. Lessing [7],
G. Levi [21,20], M. Lindsey Clark [14], F. Longhitano [3], F. Magnani [6], J. Majumdar [24], L. Malerba [16,15],
F. Mamedov [18], A. Manfreda [8], A. Manousakis [51], M. Marconi [15,16], A. Margiotta [21,20],
A. Marinelli [9,8], C. Markou [41], L. Martin [12], M. Mastrodicasa [32,10], S. Mastroianni [8], J. Mauro [38],
K.C.K. Mehta [46], G. Miele [9,8], P. Migliozzi [8], E. Migneco [25], M.L. Mitsou [22,8], C.M. Mollo [8],
L. Morales-Gallegos [22,8], N. Mori [2], A. Moussa [42], I. Mozun Mateo [17], R. Muller [20],
M.R. Musone [22,8], M. Musumeci [25], S. Navas [52], A. Nayerhoda [37], C.A. Nicolau [10], B. Nkosi [35],
B. Ó Fearraigh [16], V. Oliviero [9,8], A. Orlando [25], E. Oukacha [14], L. Pacini [2], D. Paesani [25],

* Corresponding author at: Nikhef, National Institute for Subatomic Physics, PO Box 41882, Amsterdam, 1009 DB, Netherlands.
  E-mail addresses: km3net-pc@km3net.de, mjg@nikhef.nl (M.d. Jong).
  a also at Faculty of Mathematics and Physics, Charles University in Prague, Prague, Czech Republic.
  b Deceased.

J. Palacios González [7], G. Papalashvili [37,47], P. Papini [2], V. Parisi [15,16], A. Parmar [17], E.J. Pastor Gomez [7], C. Pastore [37], A.M. Păun [29], G.E. Păvălaş [29], S. Peña Martínez [14], M. Perrin-Terrin [6], V. Pestel [17], M. Petropavlova [18,a], P. Piattelli [25], A. Plavin [50,59], C. Poirè [28,8], V. Popa [29,b], T. Pradier [4], J. Prado [7], S. Pulvirenti [25], C.A. Quiroz-Rangel [13], N. Randazzo [3], A. Ratnani [53], S. Razzaque [54], I.C. Rea [8], D. Real [7], G. Riccobene [25], J. Robinson [26], A. Romanov [15,16,17], E. Ros [50], A. Šaina [7], F. Salesa Greus [7], D.F.E. Samtleben [45,24], A. Sánchez Losa [7], S. Sanfilippo [25], M. Sanguineti [15,16], D. Santonocito [25], P. Sapienza [25], M. Scaringella [2], M. Scarnera [38,14], J. Schnabel [40], J. Schumann [40], J. Seneca [24], N. Sennan [42], P.A. Sevle Myhr [38], I. Sgura [37], R. Shanidze [47], Chengyu Shao [60,6], A. Sharma [14], Y. Shitov [18], F. Šimkovic [19], A. Simonelli [8], A. Sinopoulou [3], B. Spisso [8], M. Spurio [21,20], O. Starodubtsev [2], D. Stavropoulos [41], I. Štekl [18], D. Stocco [12], G. Takadze [47], Y. Tayalati [27,53], H. Thiersen [26], S. Thoudam [5], I. Tosta e Melo [3,36], B. Trocmé [14], V. Tsourapis [41], E. Tzamariudaki [41], A. Ukleja [46], A. Vacheret [17], V. Valsecchi [25], V. Van Elewyck [49,14], G. Vannoye [6,16,15], E. Vannuccini [2], G. Vasileiadis [55], F. Vazquez de Sola [24], A. Veutro [10,32], S. Viola [25], D. Vivolo [22,8], A. van Vliet [5], E. de Wolf [30,24], I. Lhenry-Yvon [14], S. Zavatarelli [16], D. Zito [25], J.D. Zornoza [7], J. Zúñiga [7]

[1] Università di Firenze, Dipartimento di Fisica e Astronomia, via Sansone 1, Sesto Fiorentino, 50019, Italy
[2] INFN, Sezione di Firenze, via Sansone 1, Sesto Fiorentino, 50019, Italy
[3] INFN, Sezione di Catania, (INFN-CT), Via Santa Sofia 64, Catania, 95123, Italy
[4] Université de Strasbourg, CNRS, IPHC UMR 7178, F-67000 Strasbourg, France
[5] Khalifa University of Science and Technology, Department of Physics, PO Box 127788, Abu Dhabi, United Arab Emirates
[6] Aix Marseille Univ, CNRS/IN2P3, CPPM, Marseille, France
[7] IFIC - Instituto de Fí,sica Corpuscular (CSIC - Universitat de València), c/Catedrático José Beltrán, 2, 46980 Paterna, Valencia, Spain
[8] INFN, Sezione di Napoli, Complesso Universitario di Monte S. Angelo, Via Cintia ed. G, Napoli, 80126, Italy
[9] Università di Napoli "Federico II", Dip. Scienze Fisiche "E. Pancini", Complesso Universitario di Monte S. Angelo, Via Cintia ed. G, Napoli, 80126, Italy
[10] INFN, Sezione di Roma, Piazzale Aldo Moro, 2 - c/o Dipartimento di Fisica, Edificio, G.Marconi, Roma, 00185, Italy
[11] Universitat Politècnica de Catalunya, Laboratori d'Aplicacions Bioacústiques, Centre Tecnològic de Vilanova i la Geltrú, Avda. Rambla Exposició, s/n, Vilanova i la Geltrú, 08800, Spain
[12] Subatech, IMT Atlantique, IN2P3-CNRS, Nantes Université, 4 rue Alfred Kastler - La Chantrerie, Nantes, BP 20722 44307, France
[13] Universitat Politècnica de València, Instituto de Investigación para la Gestión Integrada de las Zonas Costeras, C/Paranimf, 1, Gandia, 46730, Spain
[14] Université Paris Cité, CNRS, Astroparticule et Cosmologie, F-75013 Paris, France
[15] Università di Genova, Via Dodecaneso 33, Genova, 16146, Italy
[16] INFN, Sezione di Genova, Via Dodecaneso 33, Genova, 16146, Italy
[17] LPC CAEN, Normandie Univ, ENSICAEN, UNICAEN, CNRS/IN2P3, 6 boulevard Maréchal Juin, Caen, 14050, France
[18] Czech Technical University in Prague, Institute of Experimental and Applied Physics, Husova 240/5, Prague, 110 00, Czech Republic
[19] Comenius University in Bratislava, Department of Nuclear Physics and Biophysics, Mlynska dolina F1, Bratislava, 842 48, Slovak Republic
[20] INFN, Sezione di Bologna, v.le C. Berti-Pichat, 6/2, Bologna, 40127, Italy
[21] Università di Bologna, Dipartimento di Fisica e Astronomia, v.le C. Berti-Pichat, 6/2, Bologna, 40127, Italy
[22] Università degli Studi della Campania "Luigi Vanvitelli", Dipartimento di Matematica e Fisica, viale Lincoln 5, Caserta, 81100, Italy
[23] E. A. Milne Centre for Astrophysics, University of Hull, Hull, HU6 7RX, United Kingdom
[24] Nikhef, National Institute for Subatomic Physics, PO Box 41882, Amsterdam, 1009 DB, Netherlands
[25] INFN, Laboratori Nazionali del Sud, (LNS) Via S. Sofia 62, Catania, 95123, Italy
[26] North-West University, Centre for Space Research, Private Bag X6001, Potchefstroom, 2520, South Africa
[27] University Mohammed V in Rabat, Faculty of Sciences, 4 av. Ibn Battouta, B.P. 1014, R.P. 10000 Rabat, Morocco
[28] Università di Salerno e INFN Gruppo Collegato di Salerno, Dipartimento di Fisica, Via Giovanni Paolo II 132, Fisciano, 84084, Italy
[29] Institute of Space Science - INFLPR Subsidiary, 409 Atomistilor Street, Magurele, Ilfov, 077125, Romania
[30] University of Amsterdam, Institute of Physics/IHEF, PO Box 94216, Amsterdam, 1090 GE, Netherlands
[31] TNO, Technical Sciences, PO Box 155, Delft, 2600 AD, Netherlands
[32] Università La Sapienza, Dipartimento di Fisica, Piazzale Aldo Moro 2, Roma, 00185, Italy
[33] Università di Bologna, Dipartimento di Ingegneria dell'Energia Elettrica e dell'Informazione "Guglielmo Marconi", Via dell'Università 50, Cesena, 47521, Italy
[34] Cadi Ayyad University, Physics Department, Faculty of Science Semlalia, Av. My Abdellah, P.O.B. 2390, Marrakech, 40000, Morocco
[35] University of the Witwatersrand, School of Physics, Private Bag 3, Johannesburg, Wits 2050, South Africa
[36] Università di Catania, Dipartimento di Fisica e Astronomia "Ettore Majorana", (INFN-CT), Via Santa Sofia 64, Catania, 95123, Italy
[37] INFN, Sezione di Bari, via Orabona, 4, Bari, 70125, Italy
[38] UCLouvain, Centre for Cosmology, Particle Physics and Phenomenology, Chemin du Cyclotron, 2, Louvain-la-Neuve, 1348, Belgium
[39] University of Granada, Department of Computer Engineering, Automation and Robotics/CITIC, 18071 Granada, Spain
[40] Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU), Erlangen Centre for Astroparticle Physics, Nikolaus-Fiebiger-Straße 2, 91058 Erlangen, Germany
[41] NCSR Demokritos, Institute of Nuclear and Particle Physics, Ag. Paraskevi Attikis, Athens, 15310, Greece
[42] University Mohammed I, Faculty of Sciences, BV Mohammed VI, B.P. 717, R.P. 60000 Oujda, Morocco
[43] Western Sydney University, School of Computing, Engineering and Mathematics, Locked Bag 1797, Penrith, NSW 2751, Australia
[44] NIOZ (Royal Netherlands Institute for Sea Research), PO Box 59, Den Burg, Texel, 1790 AB, The Netherlands
[45] Leiden University, Leiden Institute of Physics, PO Box 9504, Leiden, 2300 RA, Netherlands
[46] AGH University of Krakow, Al. Mickiewicza 30, 30-059 Krakow, Poland
[47] Tbilisi State University, Department of Physics, 3, Chavchavadze Ave., Tbilisi, 0179, Georgia
[48] The University of Georgia, Institute of Physics, Kostava str. 77, Tbilisi, 0171, Georgia
[49] Institut Universitaire de France, 1 rue Descartes, Paris, 75005, France
[50] Max–Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn, Germany
[51] University of Sharjah, Sharjah Academy for Astronomy, Space Sciences, and Technology, University Campus - POB 27272, Sharjah, United Arab Emirates
[52] University of Granada, Dpto. de Fí,sica Teórica y del Cosmos & C.A.F.P.E., 18071 Granada, Spain
[53] School of Applied and Engineering Physics, Mohammed VI Polytechnic University, Ben Guerir, 43150, Morocco

[54] *University of Johannesburg, Department Physics, PO Box 524, Auckland Park, 2006, South Africa*

[55] *Laboratoire Univers et Particules de Montpellier, Place Eugène Bataillon - CC 72, Montpellier Cédex 05, 34095, France*

[56] *Université de Haute Alsace, rue des Frères Lumière, 68093 Mulhouse Cedex, France*

[57] *Université Badji Mokhtar, Département de Physique, Faculté des Sciences, Laboratoire de Physique des Rayonnements, B. P. 12, Annaba, 23000, Algeria*

[58] *AstroCeNT, Nicolaus Copernicus Astronomical Center, Polish Academy of Sciences, Rektorska 4, Warsaw, 00-614, Poland*

[59] *Harvard University, Black Hole Initiative, 20 Garden Street, Cambridge, MA 02138, USA*

[60] *School of Physics and Astronomy, Sun Yat-sen University, Zhuhai, China*

ARTICLE INFO

ABSTRACT

The KM3NeT research infrastructure comprises two neutrino telescopes located in the deep waters of the Mediterranean Sea, namely ORCA and ARCA. KM3NeT/ORCA is designed for the measurement of neutrino properties and KM3NeT/ARCA for the detection of high-energy neutrinos from the cosmos. Neutrinos are indirectly detected using three-dimensional arrays of photo-sensors which detect the Cherenkov light that is produced when relativistic charged particles emerge from a neutrino interaction. The analogue pulses from the photo-sensors are digitised offshore and all digital data are sent to a station on shore where they are processed in real time using a farm of commodity servers and custom software. In this paper, the design and performance of the software that is used to filter the data are presented. The performance of the data filter is evaluated in terms of its efficiency, purity and capacity. The efficiency is measured by the effective volumes of the sensor arrays as a function of the energy of the neutrino. The purity is measured by a comparison of the event rate caused by muons produced by cosmic ray interactions in the Earth's atmosphere with the event rate caused by the background from decays of radioactive elements in the sea water and bioluminescence. The capacity is measured by the minimal number of servers that is needed to sustain the rate of incoming data. The results of these evaluations comply with all specifications. The count rates of all photo-sensors are measured with a sampling frequency of 10 Hz. These data are input to the simulations of the detector response and will also be made available for interdisciplinary research.

1. Introduction

The KM3NeT research infrastructure comprises two neutrino telescopes located in the deep waters of the Mediterranean Sea, namely ORCA and ARCA [1]. They are located off the shores of Toulon, France and Portopalo di Capo Passero (Sicily), Italy, respectively.

The configuration of ORCA is optimised for the study of atmospheric neutrinos with energies between $2-100$ GeV and the determination of the neutrino mass ordering. The configuration of ARCA is optimised for the detection of neutrinos from the cosmos with energies in excess of 1 TeV.

Neutrinos are indirectly detected via the Cherenkov light that is produced when relativistic charged particles emerge from a neutrino interaction. The chosen photo-sensor is a 3-inch photo-multiplier tube (PMT) which combines accurate timing, high quantum efficiency, low dark count and large area [2]. To operate the PMTs in the deep sea, 31 of them are housed in a pressure-resistant glass sphere, referred to as an optical module [3]. Each optical module also houses the necessary electronics to process the signals from the PMTs and to send the data to shore. To deploy the optical modules, 18 of them are mounted on a narrow flexible structure, referred to as a string [4]. The strings are anchored to the seabed and kept vertical by the buoyancy of the optical modules and a 1.35 kN buoy on top. The spacing between the optical modules in the ARCA and ORCA neutrino telescopes is optimised for the energies of interest. The vertical spacing between the optical modules is about 36 m and 9 m and the horizontal spacing between the strings about 90 m and 20 m, respectively. A set of strings is operated as a standalone detector. A total of 115 strings is referred to as a building block. The geometrical volume of a building block is defined by the smallest cylinder encompassing all PMTs in the detector. For ARCA and ORCA, it is about 5.6×10^{-1} km^3 and 7.6×10^{-3} km^3, respectively. The ARCA and ORCA neutrino telescopes are under construction. At completion, they will consist of two and one building blocks, respectively. The telescopes are operated during construction.

The interaction of a neutrino, the passage of a muon or other signals of interest are referred to as events. The primary task of the online data filter is to detect the wanted events and to reject the background due to decays of radioactive elements in the sea water and bioluminescence.

An overview of the data processing system is presented in Section 2, covering the offshore data acquisition, the data transmission, the on shore data processing and the data recording. The definitions of the various types of data that are recorded are introduced in this section. The top-level specifications are summarised in Section 3. The event signatures and the main algorithm are presented in Section 4. The performance of the online data filter is assessed in Section 5 and the conclusions are summarised in Section 6.

2. Data processing

The data processing system of KM3NeT is based on the all-data-to-shore concept. The analogue pulses from the PMTs are digitised offshore and all digital data are sent to shore where they are processed in real time using a farm of commodity servers and custom software (see Fig. 1). This concept has been successfully pioneered in the ANTARES project [5].

2.1. Offshore data acquisition

The analogue output of a PMT is processed via an amplifier with pulse shaping and a time-over-threshold discriminator. The typical threshold of the discriminator corresponds to 0.25 photo-electron equivalent (p.e.). Each analogue pulse that passes this threshold is digitised. The digital data contain a PMT identifier (1 byte), the time stamp of the leading edge (4 bytes) and the length of the pulse (1 byte), jointly referred to as a hit. The least significant bits of the time stamp and the pulse length correspond to 1 ns. The dynamic ranges of the time stamp and pulse length are thus 2^{32} ns and 2^8 ns, respectively. The digitisation of the signals from the different PMTs is implemented inside a field programmable gate array (FPGA) that is located inside the optical module [6]. The FPGA also hosts a microprocessor which is connected to a station on shore via a fibre-optic network. The time stamp is obtained from a local clock that runs inside the FPGA which is synchronised to a master clock on shore using the White Rabbit system [7]. The time stamp from the local clock is reset every 100 ms. The master clock is locked to GPS. The actual time is maintained in UTC and included in the data throughout the processing. The count rate of a PMT typically amounts to about 7 kHz [8]. This rate is primarily due to decays of radioactive elements in the sea water but can increase

for certain periods due to varying contributions from bioluminescence. The dark count rate *in situ* is a few hundred Hertz. The total data rate amounts to about 25 Gb/s per building block.

2.2. Data transmission

For the distribution of the data on shore, the continuous data streams from the optical modules are sliced in time by the FPGA. The duration of the time slice is set to 100 ms which matches the reset of the time stamp. The data are organised into jumbo frames of 9000 bytes using a FIFO and sent to shore using UDP. Each slice in time yields about 15 jumbo frames. Counters are included in the header of each jumbo frame to reassemble the UDP frames on shore and to monitor possible losses of UDP packets. To limit the probability of packet losses somewhere in the network due to anomalous high count rates in some of the PMTs, the maximal number of hits per PMT that can be recorded during a time slice is set to 2000. The remaining data are sent to shore anyway and are used for monitoring. This limit corresponds to 20 kHz and is referred to as the high-rate veto. A temporary increase of the background due to a burst of bioluminescence is thereby filtered. As a result, the probability of a packet loss is normally less than 10^{-6}. Seasonal variations of the bioluminescence have been observed at the site of ORCA which lead to a general increase of the background. This is included in the specifications of the data filter. Due to the high-rate veto, the total data rate is limited any time to about 75 Gb/s per building block.

2.3. On shore data processing

Three software applications have been developed in the C++ programming language to process the data on shore, namely "data queue", "data filter" and "data writer". Multiple instances of the data queue and data filter applications run in parallel on the servers. The neutrino telescopes as well as the applications to process the data are configured and operated via the control unit [9]. All data corresponding to the same time slice are sent to the same data filter. For each time slice, the UDP frames that originated from the same optical module are first assembled into a single frame by one of the data queues and then distributed to the data filters using TCP/IP. The distribution is based on the UTC time of the data and sustained using round-robin scheduling of the data filters. As a prerequisite, the time to process a slice of data by a data filter must be shorter than the duration of the time slice multiplied by the number of data filters. The farm of servers should thus provide the capacity needed to process the incoming data in real time. All data produced by the data filters are sent via a dispatcher to one central data writer using TCP/IP. Normally, the neutrino telescopes are operated around the clock. Apart from sea operations during the construction of the telescope, maintenance of the shore stations and taking of calibration data, there is no significant dead-time. During operation, data are continuously recorded using consecutive runs of typically 3 h, each of which is identified by a unique number. For each run, a separate file is written by the data writer which is archived in multiple computer centres. Each of these files represents a single set of data with a well-defined and reproducible configuration of the whole system.

An event is triggered by the identification of a cluster of hits which complies with predefined criteria. Following a trigger, all telescope data within a time window around the time of the event are recorded. The reduction of the data is thus given by the product of duration of this time window and the trigger rate. Different trigger algorithms are operated simultaneously. In addition, level-0 (L0), level-1 (L1), level-2 (L2), Supernova (SN), and summary data can be recorded according to predefined samplings. The L0 data contain all hits from all PMTs in the detector. The L1 data are a subset of the L0 data and contain only local coincidences between two (or more) L0 hits in one optical module within a predefined time window. The L2 and SN data are a subset of
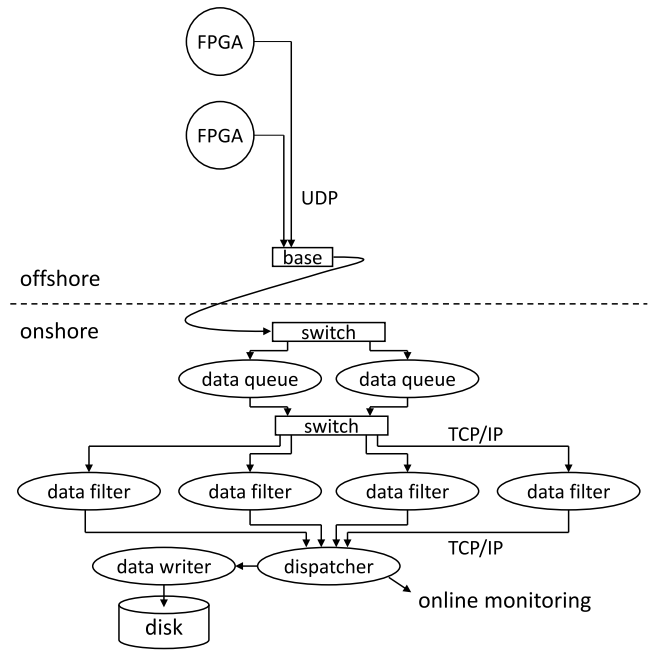

Fig. 1. Schematic view of the data flow (not to scale). The part above the dashed line corresponds to offshore and the part below to on shore. The circles correspond to the optical modules which are deployed in the deep sea and house the PMTs as well as the FPGA. The ellipses and rectangular boxes correspond to processes and switch fabric, respectively. The base provides for the connection to the sea-floor network and is mounted on the anchor of the string.

the L1 data subject to additional constraints. These constraints apply to the time difference between hits, the space angle between PMT axes and the number of hits. The L0, L1 and L2 data correspond to the first processing steps of the data filter. They are used for monitoring of the data quality and for the calibration of the detector. The SN data are used to study low-energy neutrinos (MeV) from e.g. Supernovae. The criteria for L0, L1, L2 and SN data are similar for ORCA and ARCA. The summary data contain the rate and status of the offshore data acquisition of all PMTs in the detector per slice in time. They are used to monitor the operation of the detector and the conditions in the deep sea. To limit the volume of the summary data, the measured rate of each PMT is compressed to a single byte using the logarithmic value of the counts per unit time. The dynamic range is from 2 kHz (1) to 2 MHz (256). A zero value (0) corresponds to any rate that is less than the lower limit. The high-rate veto and FIFO status from the FPGA are recorded for each PMT inside a given optical module using two values of 32 bits. The UDP frame counters are included in the summary data using two values of 16 bits.

2.4. Data recording

Normally, all triggered events, all summary data and all SN data are recorded. In addition, a small fraction of the L1 data (typically 1%) are recorded for the calibration of the PMTs. Optionally, each data filter can also maintain a circular buffer containing the L0, L1, L2 or SN data from a predefined number of the latest time slices. Following the reception of an external alert (e.g. from other telescopes), these data are written to the local disks of the servers and collected afterwards for offline analyses. The maximal time that can be recovered depends on the available random access memory (RAM) in the CPUs of the servers and could reach several minutes of history.

The neutrino telescope data are filtered online using calibrations that have been produced beforehand. These calibrations are archived in

the central database and communicated to the data filters by the control unit. The calibrations include the position, orientation and time offset of each PMT. Since the filtered data contain an unmodified copy of the original raw data, the recorded data can be analysed offline using calibrations that are produced afterwards. The attainable resolution of the neutrino telescope is thereby not affected by the calibrations used during operation. The summary data are also input to the simulations of the detector response. In these simulations, the livetime of the detector as well as the status of the offshore data acquisition and count rate of each PMT are taken into account at any given time. This reproduces the operation of the detector and the conditions in the deep sea according to the sampling of once per time slice which corresponds to a frequency of 10 Hz.

3. Specifications

The data filter should sustain the data rates from the telescopes. The maximal rate of data being written to disk is set to 25 Mb/s. This corresponds to a reduction of the data by a factor of about 1×10^3. The main event rate is due to muons produced by interactions of cosmic rays in the atmosphere above the detector but there is a finite probability of an accidental trigger due to the background produced by decays of radioactive elements in the sea water and bioluminescence. For optimal use of the data recording capacity, the rate of accidental triggers should be kept well below the expected event rate due to muons. The number of CPU cores in the servers required to process the data in real time is determined by the time to process a time slice. To limit the costs and the size of the network, it is foreseen to operate each building block with less than 1000 CPU cores with 15 HS06 per core.

4. Data filter

The data from the neutrino telescope are processed by a set of data filters that run in parallel on the servers. The data are collected first in the RAM of the CPUs of the servers. The collected data corresponding to a given time slice are processed as soon as the list of optical modules is complete. When the total amount of collected data or the time difference between the earliest and latest data exceed predefined limits, the data from the earliest time slice are processed anyway. Possible intermittent delays in the network can thus be covered for as long as these can be accommodated in the available RAM. The filtering of the data proceeds in several consecutive steps, commonly referred to as levels.

4.1. Level-0

By construction, the data from a PMT are sorted in time. Inside the FPGA, the data from the different PMTs inside the same optical module are multiplexed before they are sent to shore. They are de-multiplexed at the start of the data processing on shore. While the data are de-multiplexed, the time stamp of each hit is corrected for the time calibration of the PMT. For each optical module, 31 arrays of time sorted hits are obtained (one per PMT). These hits are referred to as level-0 (L0) hits.

4.2. Level-1

In the level-1 (L1) step of the data processing, coincidences between two (or more) L0 hits from different PMTs inside the same optical module are selected. The 31 arrays with L0 hits are first multiplexed using a custom implementation of the merge sort algorithm. An L1 hit is obtained when the time difference between two (or more) consecutive hits is less than a predefined time window. A suitable value for the time window is 20 ns. The time of the L1 hit is determined by the time of the earliest L0 hit. Optionally, the time-over-threshold is included. The time-over-threshold of the L1 hit is determined by the first leading

edge and the last trailing edge of the L0 hits. Multiple L1 hits within the same time window are combined to suppress double counting of hits. The time of the first L1 hit is then maintained and the number of L0 hits incremented. The typical rate of L1 hits per optical module amounts to about 1 kHz. About half of this rate is due to genuine coincidences from decays of radioactive elements in the sea water and a small fraction to the passage of muons [8]. The remainder can be attributed to random coincidences.

4.3. Level-2

In the level-2 (L2) step of the data processing, the L1 data are subject to additional constraints. The list of constraints includes a time window, a maximal space angle between the PMT axes and a minimal number of L0 hits. To improve the signal-to-noise ratio, the time window is set to 10 ns, the maximal space angle to 90 deg and the minimal number of L0 hits to 2. As a result, the rate of hits per optical module is reduced to about 700 Hz.

4.4. Supernova

Another selection of the L1 data is made for the detection of low-energy neutrinos (MeV) from Supernovae that is parallel to the L2 step. The time window and the maximal space angle between PMT axes are usually set to the same values as for the L2 selection but the minimal number of L0 hits is set to 5 instead. As a result, the rate of hits per optical module is reduced to few Hz. A sizeable fraction of this rate is then due to muons produced in the atmosphere above the neutrino telescope [8].

4.5. Summary

For each time slice, a summary of the L0 data is made. The rate of each PMT is computed by the number of hits divided by the duration of the time slice. In case of a high-rate veto, the time between the first and last hit is used instead.

4.6. Trigger

The trigger constitutes the final step in the data filtering. Normally, L2 hits are input to this step. An event is triggered when a cluster of L2 hits is found which complies with predefined criteria. For each event, all telescope data within a time window around the trigger are recorded. This is referred to as the "snapshot". The snapshot ranges from some extra time before the first hit to some extra time after the last hit in the cluster. To include all data that could be causally related to the event, the extra time ΔT covers the largest diameter of the detector D.

$$\Delta T = \frac{n}{c} D \tag{1}$$

where c corresponds to the velocity of light in vacuum and n to the ratio thereof and the group velocity of light. The latter depends on the wavelength of light and is around 1.35. For ORCA and ARCA, a suitable value for the extra time is 2.5 μs and 10 μs, respectively. The durations of the event and the snapshot are small compared to the duration of a time slice. As a result, the impact of edge effects due the distribution of data is limited to few times 10^{-4}.

The detection of an event proceeds in four main steps, namely:

1. sort hits in time;
2. for each primary hit, collect following hits within a predefined time window;
3. select hits that are causally related to the primary hit;
4. find largest subset of selected hits that are all causally related.

The predefined time window in step 2 is the same as in Eq. (1). The general causality relation for a pair of hits can be formulated as follows:

$$\left| t_i - t_j \right| \leq \frac{n}{c} \left| \bar{x}_i - \bar{x}_j \right| \tag{2}$$

where $t_{i(j)}$ corresponds to the time of hit $i(j)$ and $\bar{x}_{i(j)}$ to the position of the optical module. The causality relation is not transitive. As a consequence, the guaranteed solution for finding the largest subset of hits that are all causally related from a set of n hits requires $\mathcal{O}(n!)$ operations. To limit the number of operations in step 4 to an acceptable level, a custom implementation of the clique algorithm is used [10]. This algorithm requires $\mathcal{O}(n^2)$ operations.

A condition is applied to the number of hits and the number of optical modules which recorded the hits after steps 2–4. When one of these numbers is less than the required minimum, the processing of data continues with the next primary hit in step 2. A suitable value for the minimal number of hits and optical modules to trigger an event is 5. When the number of hits after step 2 is very large, the event is immediately triggered thereby avoiding waste of CPU time in steps 3–4. This number is referred to as factory limit. A suitable value is 100.

4.7. Event signatures

The rate of accidental triggers is determined by the number of optical modules involved and the probability that two hits accidentally comply with the causality relation. When all optical modules in the detector are used, the combinatorial background is too large to reduce the data. Therefore, two specific signatures are considered, namely a muon traversing the detector and a neutrino interacting inside the detector producing a shower of secondary particles without a muon. These signatures are used in designated implementations of the trigger processing step. The Cherenkov light from a muon can be described by a cone which propagates in a straight line with the speed of light in vacuum. The Cherenkov light from a shower can be described by a sphere which expands with the group velocity of light in water. The peak intensity of the light reduces between linearly (light cone) and quadratically (spherical) with the covered distance. The light is further attenuated by absorption and scattering. The range of Cherenkov light can therefore be limited without significant loss of the signal. Because the range of a muon can be larger than that of light, a longer distance between optical modules should be considered. This distance can be decomposed into a part related to the muon and a part related to light [11]. The corresponding velocity is then c and c/n, respectively. This reduces the time window in Eq. (2). The resulting volumes for a muon and shower correspond to a cylinder and a sphere, respectively. For ARCA and ORCA, suitable values for the radius of the cylinder are 120 m and 45 m and for the diameter of the sphere 250 m and 45 m, respectively. For a given primary hit, the combinatorial background is then limited to the number of optical modules inside these volumes.

By taking into account the propagation time of the muon, the time window is determined by the transverse distance between optical modules with respect to the muon direction. The nearest-neighbour principle can thus be applied in two dimensions to optical modules that are far apart in the other dimension. For a shower, the time window in Eq. (2) increases up to the point that the distance between optical modules is equal to the assumed range of light. Beyond that point, the time window decreases because the signal should originate from somewhere in between the optical modules. This counteracts the volumetric increase of the number of optical modules with the distance. The reduced number of optical modules and time window lower the probability of an accidental coincidence by a factor of about 100. As a result, the rate of accidental triggers is reduced by a factor of 100^{m-1} for a minimal number of hits to trigger an event equal to m.

The time windows for ARCA as a function of the distance between optical modules are shown in Fig. 2 for the different causality criteria. For ORCA, the same features as in Fig. 2 apply but the distances are

Table 1

Main parameters of the data filter.

Level	Parameter	Value	
		ORCA	ARCA
L1	Time window	20 ns	20 ns
L2	Time window	10 ns	10 ns
L2	Angle between PMT axes	90 deg	90 deg
L2	Number of L0 hits	2	2
Supernova	Time window	10 ns	10 ns
Supernova	Angle between PMT axes	90 deg	90 deg
Supernova	Number of L0 hits	5	5
Trigger	Extra time for snapshot	2.5 μs	10 μs
Trigger	Number of L2 hits	5	5
muon	Radius of cylinder	45 m	120 m
muon	Spacing of directions	10 deg	10 deg
shower	Diameter of sphere	45 m	250 m
Mixed	Number of L2 hits	1	n.a.
Mixed	Radius for vertices	35 m	n.a.
Mixed	Spacing of vertices	2 m	n.a.

shorter and time windows smaller. Finally, a small amount of extra time is added to accommodate various uncertainties. A suitable value for this extra time is 25 ns.

As a result of the application of a cylinder to the selection of optical modules and the restriction of the causality relation, the neutrino telescope is pointed to a position on the sky with a field-of-view of about 0.1 sr. To increase the field-of-view, a grid of approximately equidistant directions is used. With a spacing of 10 deg, the full sky is covered using about 200 directions.

To lower the energy threshold for the detection of neutrino interactions in ORCA, an additional trigger is used which is based on the shower signature. It requires a single L2 hit and a set of causally related L0 hits and is therefore referred to as mixed trigger. To limit the rate of accidental triggers, an additional processing step is introduced after step 4. In this step, a set of assumed vertex positions of the neutrino interactions is used. These are distributed according a spherical grid around the L2 hit. A suitable value for the radius of the sphere is 35 m and for the spacing between the vertices 2 m. For a given vertex position, the propagation time of the Cherenkov light from the vertex to the optical module can be taken into account. The remaining distance in the causality relation is then limited to the spacing between the vertex positions.

The different implementations of the trigger processing step are applied to the same data. A unique integer value is used to identify the trigger in use. This value is included in the event data via a corresponding bit in a trigger mask. When two (or more) events overlap in time, they are merged. The lists of hits in the cluster and in the snapshot from the different events are then combined, avoiding duplications. The resulting trigger mask is set to the logical OR of the underlying trigger masks. The number of times that an event has been merged is also included in the event data.

The main parameters of the data filter are summarised in Table 1. The quoted values are also suitable for the detectors that are currently in operation and can still be tuned.

5. Performance

To evaluate the performance of the data filter, simulations of the detector response to interactions of neutrinos and the passage of muons are made [12]. With these simulations, Monte Carlo data are generated that conform with the telescope data. They are subsequently processed with the same trigger algorithms as the data from the neutrino telescopes.
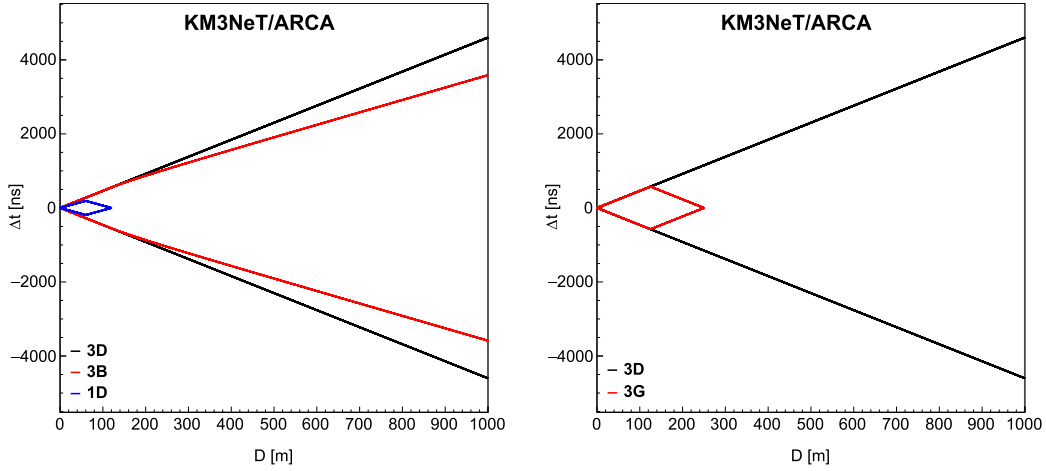
Fig. 2. The time windows for the KM3NeT/ARCA neutrino telescope as a function of the distance between optical modules for the different causality relations (see text). The left (right) figure applies to the signal from a muon (shower). The maximal distance roughly corresponds to the size of the detector. The label 3D corresponds to the general causality relation, 3B to the causality relation for which the distance covered by the signal is decomposed and 1D to the causality relation for which the propagation time of the muon is taken into account. For the 1D case, the x-axis corresponds to the transverse distance with respect to the direction of the muon. The label 3G corresponds to the causality relation for which the distance covered by the signal is limited to the radius of a sphere.
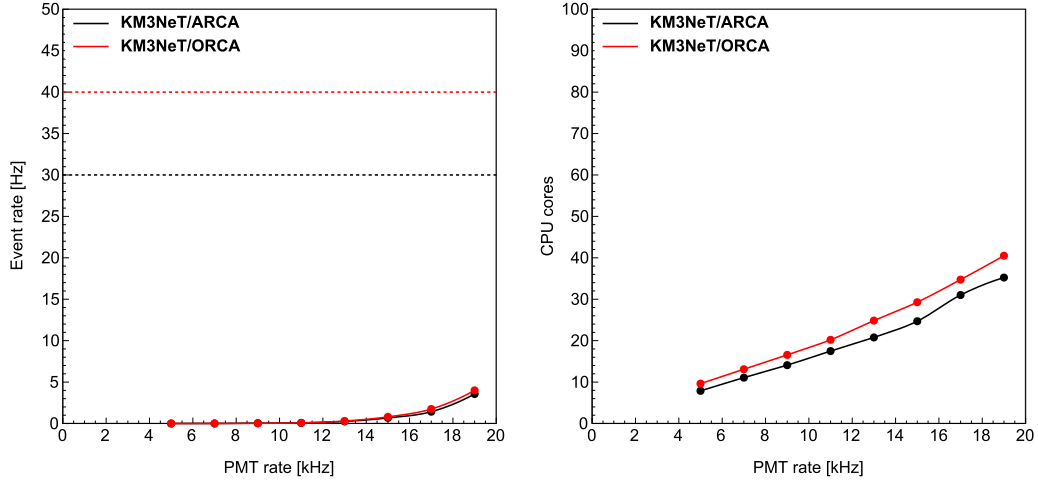
Fig. 3. The rate of events (left) and the required number of CPU cores (right) as a function of the singles rate of the PMTs for one building block of the KM3NeT/ARCA and KM3NeT/ORCA neutrino telescopes. The points correspond to the results obtained from a simulation of the random background. The dashed lines correspond to the expected rates of events due to muons produced by interactions of cosmic rays in the atmosphere above the detectors. A single CPU core of the `Intel(R) Xeon(R) E E-2478` processor is used.

5.1. Purity

The impurity of the data filter is defined as the ratio of the event rates caused by background and signal. Here, decays of radioactive elements in the sea water and bioluminescence are considered as the background and muons produced by cosmic ray interactions in the atmosphere above the detector as the signal. The signal is simulated by generating muons at the depth of the detector according to parametrisations of the observed spectra [13]. The expected event rate for one building block of ARCA and ORCA is found to be 30 ± 5 Hz and 40 ± 5 Hz, respectively. The quoted uncertainty covers the uncertainties about the incident flux, the PMT characteristics and the optical properties of the sea water. The background is simulated by generating hits with random times. The times of consecutive hits follow an exponential probability distribution according to a given rate. The two-fold coincidence rate due to genuine coincidences from decays of radioactive elements in the sea water is set to 500 Hz and the singles rate is varied from 5 kHz to 20 kHz. The upper limit corresponds to the high-rate veto that is implemented in the FPGA. The simulations of the background are also used to determine the minimal number of CPU cores needed to sustain the rate of incoming data. The rate of events and the required number of CPU cores are shown in Fig. 3 as a function of the singles rate of the PMTs for one building block of ARCA and ORCA.

As can be seen from Fig. 3, the rate of events due to background is very small at the nominal rate of the PMTs. The impurity of the data filter is then also very small (<1%). At the maximal data taking rate of a PMT, the event rate is a few Hz. The corresponding impurity is then about 10%.

5.2. Capacity

The relative contributions to the CPU time of the processing steps in the data filter are summarised in Table 2. The largest contribution is from the L1 processing step. The contributions from the Calibration, L1 and L2 processing steps are common to the different trigger algorithms and depend linearly on the singles rate of the PMTs. The trigger processing step involves the largest computational complexity but contributes only 20% to the total CPU time. The overhead and the contribution of
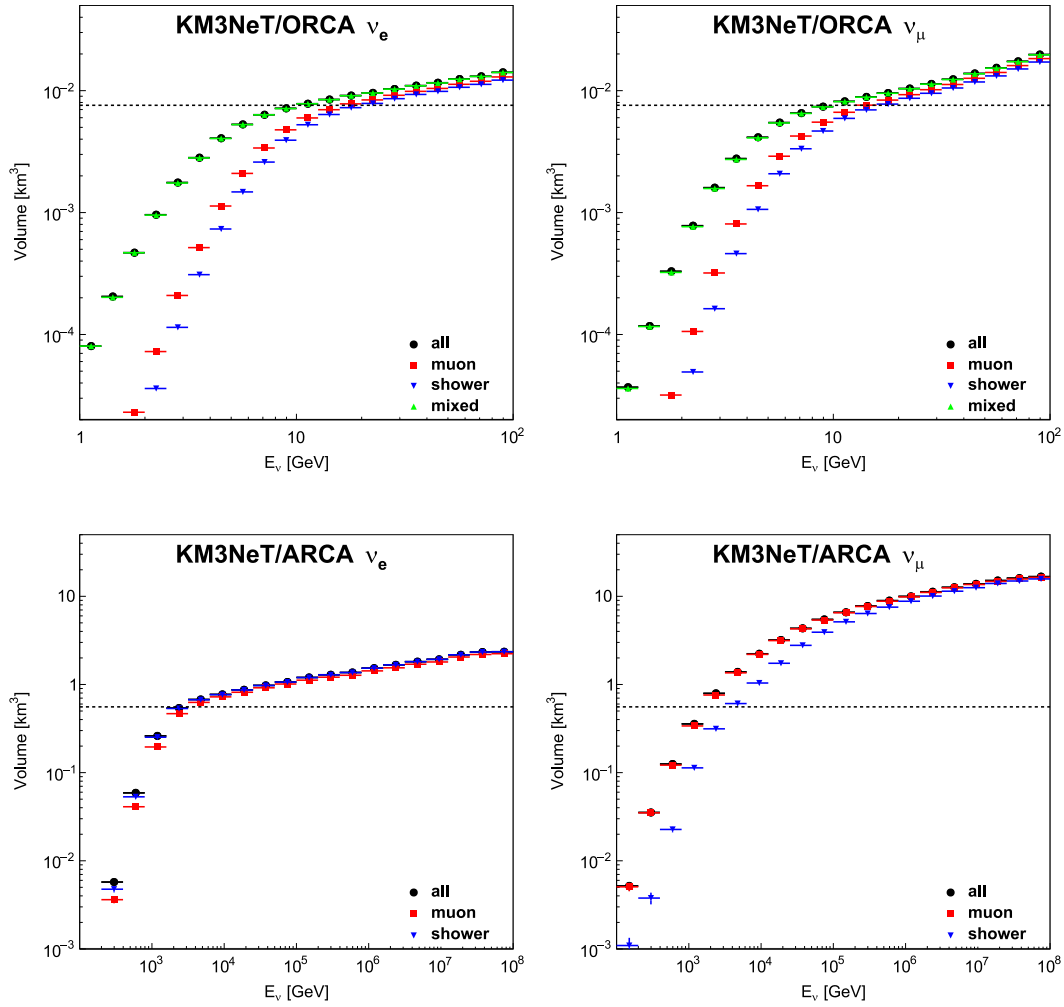
Fig. 4. The effective volume of one building block of the KM3NeT/ORCA (top) and KM3NeT/ARCA (bottom) neutrino telescopes as a function of the neutrino energy and averaged over all directions for charged-current interactions of electron (left) and muon (right) neutrinos (note the scale). The labels correspond to the signature that is used in the data filter. The dashed lines correspond to the geometrical volumes of the detectors.

Table 2

The relative contributions to the CPU time of the different processing steps in the data filter.

Step	CPU
Calibration	5%
L1	65%
L2	5%
Trigger	20%
Overhead	5%

signals to the CPU time are small. As a result, the required number of CPU cores is less than 50. This complies with the specifications.

The total rate of data being recorded is primarily determined by the event and summary data. The recording of other data is normally configured in such a way that their contributions are negligible. The data rate due to the summary data is proportional to the number of optical modules in the detector because the sampling frequency and the data volume per optical module are fixed. For one building block, it amounts to about 10 Mb/s. The data rate due to the event data is determined by the product of the rate and data volume of the events. The latter is determined by the product of the duration of the snapshot, the total count rate of all PMTs in the detector and the size of a single hit. The maximal data rate amounts to about 10 Mb/s and 2.5 Mb/s for ARCA and ORCA, respectively. These values correspond

to a worst-case-scenario. The total data rate complies nonetheless with the specifications.

5.3. Efficiency

The efficiency of the data filter can be quantified in terms of an effective volume. The effective volume is the volume in which the interaction of a neutrino will be triggered. Each type of neutrino interaction yields a corresponding effective volume. As an example, the effective volumes of one building block of the ORCA and ARCA neutrino telescopes are shown in Fig. 4 as a function of the neutrino energy for charged-current interactions of electron and muon neutrinos. These volumes apply to a flux of neutrinos at the detector and have been averaged over all directions. As a rule of thumb, the charged-current interactions of anti-neutrinos yield a larger volume because on average more energy is transferred from the incident neutrino to the outgoing charged lepton. The neutral-current interactions of (anti-)neutrinos yield a smaller volume because on average less energy is transferred from the incident neutrino to particles producing Cherenkov light. The charged-current interactions of tau (anti-)neutrinos yield a volume that depends on the decay of the tau. The corresponding effective volumes are similar to those of the muon or electron neutrino counterparts but generally smaller due to the invisible energy of the neutrinos produced in the decay of the tau.

As can be seen from Fig. 4, the effective volume reaches the geometrical volume at 10 GeV and 1 TeV for ORCA and ARCA, respectively. The threshold can be attributed to the limited visible energy in the event. The increase of the effective volumes beyond this point can be attributed to interactions of neutrinos in the vicinity of the detector. The energy threshold of the ORCA detector is effectively lowered using the mixed trigger.

The figure-of-merit for the science objectives not only depends on the effective volume but also involves an offline analysis of the recorded data. Because the criteria to detect events are tailored to the envisaged signals in the detector, the efficiency of the data filter comes with a positive bias in the selection of events in the offline analysis.

6. Conclusions

A software data filter has been developed for the KM3NeT neutrino telescopes which has been in use since the beginning of their operation, including the detection of KM3-230213A [14]. The interaction of a neutrino, the passage of a muon and other possible signals are detected when a cluster of hits is found in the data which complies with predefined criteria. The background is suppressed by optimisation of the causality relations. All telescope data within a time window around the detected events are recorded and archived for offline analyses. In addition, summary, Supernova and calibration data are recorded. Following an external alert, unfiltered or partially filtered data can be recorded which can recover up to several minutes of history. The performance of the data filter has been evaluated in terms of the efficiency, purity and capacity. The results of these evaluations comply with all specifications. The design of the online data filter allows for tuning of the parameters and future implementations of alternative event triggers and data streams. The all-data-to-shore concept makes it possible to maintain and upgrade the computing infrastructure on shore during construction and operation of the KM3NeT neutrino telescopes. The bandwidth of a standard fibre-optic network is sufficient to transfer the data. The computing power of a modest farm of servers is sufficient to process the data. As a result, a flexible solution exists to operate neutrino telescopes in the deep sea at low cost.

CRediT authorship contribution statement

M. de Jong: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] S. Adrián-Martínez, et al., Letter of intent for KM3NeT 2.0, J. Phys. G: Nucl. Part. Phys. 43 (8) (2016) http://dx.doi.org/10.1088/0954-3899/43/8/084001.
[2] S. Aiello, et al., Characterisation of the Hamamatsu photomultipliers for the KM3NeT Neutrino Telescope, 13 (05) (2018) P05035 http://dx.doi.org/10.1088/1748-0221/13/05/P05035.
[3] S. Aiello, et al., The KM3NeT multi-PMT optical module, 17 (07) (2022) P07038 http://dx.doi.org/10.1088/1748-0221/17/07/P07038.
[4] S. Aiello, et al., Deep-sea deployment of the KM3NeT neutrino telescope detection units by self-unrolling, 15 (2020) P11027 http://dx.doi.org/10.1088/1748-0221/15/11/P11027.
[5] J. Aguilar, et al., The data acquisition system for the ANTARES neutrino telescope, Nucl. Instrum. Methods Phys. Res. Sect. A: Accel. Spectrometers Detect. Assoc. Equip. 570 (1) (2007) 107–116, http://dx.doi.org/10.1016/j.nima.2006.09.098, URL https://www.sciencedirect.com/science/article/pii/S0168900206017608.

[6] S. Aiello, et al., Architecture and performance of the KM3NeT front-end firmware, J. Astron. Telesc. Instrum. Syst. 7 (1) (2021) http://dx.doi.org/10.1117/1.JATIS.7.1.016001.

[7] P. Moreira, J. Serrano, T. Wlostowski, P. Loschmidt, G. Gaderer, White rabbit: Sub-nanosecond timing distribution over ethernet, in: 2009 International Symposium on Precision Clock Synchronization for Measurement, Control and Communication, 2009, pp. 1–5, http://dx.doi.org/10.1109/ISPCS.2009.5340196.

[8] J. Aguilar, et al., The prototype detection unit of the KM3NeT detector, Eur. Phys. J. C 76 (54) (2016) http://dx.doi.org/10.1140/epjc/s10052-015-3868-9.

[9] S. Aiello, et al., The control unit of the KM3NeT data acquisition system, Comput. Phys. Comm. 256 (2020) 107433, http://dx.doi.org/10.1016/j.cpc.2020.107433, URL https://www.sciencedirect.com/science/article/pii/S0010465520301909.

[10] R. Duncan Luce, D. Albert Perry, A method of matrix analysis of group structure, Psychometrika 14 (2) (1949) 95116.

[11] B. Bakker, Trigger studies for the Antares and KM3NeT detector (MSc Thesis), University of Amsterdam, 2011.

[12] S. Aiello, et al., GSeaGen: The KM3net GENIE-based code for neutrino telescopes, Comput. Phys. Comm. 256 (2020) 107477, http://dx.doi.org/10.1016/j.cpc.2020.107477, URL https://www.sciencedirect.com/science/article/pii/S0010465520302241.

[13] G. Carminati, et al., Atmospheric MUons from parametric formulas: a fast generator for neutrino telescopes (MUPAGE), Comput. Phys. Comm. 179.12 (2008) 915–923, http://dx.doi.org/10.1016/j.cpc.2008.07.014.

[14] S. Aiello, et al., Observation of an ultra-high-energy cosmic neutrino with KM3NeT, Nature 638 (2025) 376–382, http://dx.doi.org/10.1038/s41586-024-08543-1.