Methodology Used in Borexino for the Identification of Cosmogenic Long-Time Decay Background

M. Agostini,[1,2] K. Altenmüller,[2] S. Appel,[3] V. Atroshchenko,[3] Z. Bagdasarian,[4,5] D. Basilico,[6] G. Bellini,[6] J. Benziger,[7] R. Biondi,[8] D. Bravo,[6,9] B. Caccianiga,[6] F. Calaprice,[10] A. Caminata,[11] P. Cavalcante,[12,8] A. Chepurnov,[13] D. D'Angelo,[6] S. Davini,[11] A. Derbin,[14] A. Di Giacinto,[8] V. Di Marcello,[8] X.F. Ding,[10] A. Di Ludovico,[10] L. Di Noto,[11] I. Drachnev,[14] A. Formozov,[6,15] D. Franco,[16] C. Galbiati,[10,17] C. Ghiano,[8] M. Giammarchi,[6] A. Goretti,[10,8] A.S. Göttel,[4,18] M. Gromov,[13,15] D. Guffanti,[19] Aldo Ianni,[8] Andrea Ianni,[10] A. Jany,[20] D. Jeschke,[2] V. Kobychev,[21] G. Korga,[22,23] S. Kumaran,[4,18] M. Laubenstein,[3,24] E. Litvinovich,[3,24] P. Lombardi,[6] I. Lomskaya,[14] L. Ludhova,[4,18] G. Lukyanchenko,[3] L. Lukyanchenko,[3] I. Machulin,[3,24] J. Martyn,[19] E. Meroni,[6] M. Meyer,[25] L. Miramonti,[6] M. Misiaszek,[20] V. Muratova,[14] B. Neumair,[2] M. Nieslony,[19] R. Nugmanov,[3,24] L. Oberauer,[2] V. Orekhov,[19] F. Ortica,[26] M. Pallavicini,[11] L. Papp,[2] L. Pelicci,[4,18] Ö. Penek,[4] L. Pietrofaccia,[10] N. Pilipenko,[14] A. Pocar,[27] A. Porcelli,[19,28,a)] G. Raikov,[3] M.T. Ranalli,[8] G. Ranucci,[6] A. Razeto,[8] A. Re,[6] M. Redchuk,[4,18,29] A. Romani,[26] N. Rossi,[8] S. Schönert,[2] D. Semenov,[14] G. Settanta,[4] M. Skorokhvatov,[3,24] A. Singhal,[4,18] O. Smirnov,[15] A. Sotnikov,[15] Y. Suvorov,[3,8,30] R. Tartaglia,[11] G. Testera,[11] J. Thurn,[25] E. Unzhakov,[14] A. Vishneva,[15] R.B. Vogelaar,[12] F. von Feilitzsch,[2] M. Wojcik,[20] M. Wurm,[19] S. Zavatarelli,[11] K. Zuber,[25] and G. Zuzel[20]

[1)] *Department of Physics and Astronomy, University College London, London, UK*
[2)] *Physik-Department and Excellence Cluster Universe, Technische Universität München, 85748 Garching, Germany*
[3)] *National Research Centre Kurchatov Institute, 123182 Moscow, Russia*
[4)] *Institute for Nuclear Physics IKP-2, Forschungszentrum Jülich, 52428 Jülich, Germany*
[5)] *Present Address: Department of Physics, University of California, Berkeley, Berkeley, CA 94720, USA*
[6)] *Dipartimento di Fisica, Università degli Studi e INFN, 20133 Milano, Italy*
[7)] *Chemical Engineering Department, Princeton University, Princeton, NJ 08544, USA*
[8)] *INFN Laboratori Nazionali del Gran Sasso, 67010 Assergi (AQ), Italy*
[9)] *Present Address: Universidad Autónoma de Madrid, Ciudad Universitaria de Cantoblanco, 28049 Madrid, Spain*
[10)] *Physics Department, Princeton University, Princeton, NJ 08544, USA*
[11)] *Dipartimento di Fisica, Università degli Studi e INFN, 16146 Genova, Italy*
[12)] *Physics Department, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, USA*
[13)] *Lomonosov Moscow State University Skobeltsyn Institute of Nuclear Physics, 119234 Moscow, Russia*
[14)] *St. Petersburg Nuclear Physics Institute NRC Kurchatov Institute, 188350 Gatchina, Russia*
[15)] *Joint Institute for Nuclear Research, 141980 Dubna, Russia*
[16)] *AstroParticule et Cosmologie, Université Paris Diderot, CNRS/IN2P3, CEA/IRFU, Observatoire de Paris, Sorbonne Paris Cité, 75205 Paris Cedex 13, France*
[17)] *Gran Sasso Science Institute (INFN), 67100 L'Aquila, Italy*
[18)] *Physics Institute IIIB, RWTH Aachen University, 52062 Aachen, Germany*
[19)] *Institute of Physics and Excellence Cluster PRISMA$^+$, Johannes Gutenberg-Universität Mainz, 55099 Mainz, Germany*
[20)] *M. Smoluchowski Institute of Physics, Jagiellonian University, 30059 Krakow, Poland*

[21] *Kiev Institute for Nuclear Research, 03680 Kiev, Ukraine*
[22] *Department of Physics, Royal Holloway, University of London, Department of Physics, School of Engineering, Physical and Mathematical Sciences, Egham, Surrey, TW20 OEX, UK*
[23] *Institute of Nuclear Research (Atomki), Debrecen, Hungary*
[24] *National Research Nuclear University MEPhI (Moscow Engineering Physics Institute), 115409 Moscow, Russia*
[25] *Department of Physics, Technische Universität Dresden, 01062 Dresden, Germany*
[26] *Dipartimento di Chimica, Biologia e Biotecnologie, Università degli Studi e INFN, 06123 Perugia, Italy*
[27] *Amherst Center for Fundamental Interactions and Physics Department, University of Massachusetts, Amherst, MA 01003, USA*
[28] *Present Address: INFN Laboratori Nazionali di Frascati, 00044 Frascati (RM), Italy*
[29] *Present Address: Dipartimento di Fisica e Astronomia dell'Università di Padova and INFN Sezione di Padova, Padova, Italy*
[30] *Present Address: Dipartimento di Fisica, Università degli Studi Federico II e INFN, 80126 Napoli, Italy*

[a] *Corresponding author: alessio.porcelli@lnf.infn.it*

Abstract Borexino was a liquid scintillator detector situated underground in the Laboratori Nazionali del Gran Sasso in Italy, officially decommissioned in October 2021. Its successful and renowned physics program covered the study of solar neutrinos program and spans also across geo-neutrinos and neutrino physics. Within its solar program, Borexino successfully measured neutrinos from the fusion processes in the *pp* chain and CNO cycle. For the detection of *pep* and CNO neutrinos, an especially important background is formed by the cosmogenic radio-isotope ^{11}C that is produced by muon spallation of ^{12}C nuclei in the scintillator. Given the relatively long lifetime (30 mins) and high rate (30 cpd per 100 ton), specific signal identification is not possible. Borexino developed dedicated veto strategies in the data analysis phase to allow the detection of *pep* and CNO neutrinos. The results presented so far by Borexino relied upon a Three-Fold Coincidence (TFC) technique that exploits the time and space correlation of muons, spallation neutrons, and radioactive ^{11}C decays. This method has conservative assumptions during critical data-taking periods, such as during a board saturation case or between runs, which causes a loss of data exposure. Therefore, a new algorithm is devised to relax these TFC assumptions and deal with the critical periods by searching for space-time correlated bursts of ^{11}C events produced in cascade by the spallation. In this work, we present the state of the art of the TFC, the new algorithm working, and highlight the performance of their combination to deal with the ^{11}C background. Moreover, this method finds a general application in low radioactivity Borexino-like underground experiments when dealing with any background having a decay time too long to be identified by the triggers.

INTRODUCTION

Cosmic rays, particularly muons, can produce activated isotopes as they pass through matter, a process called cosmogenic activation. These nuclei might decay in the signal energy region, creating a background. If their average decay time is too long compared to the speed of the signal acquisition, it cannot be discriminated from signals of an unknown source. In experiments detecting neutrinos, these isotope signals might be confused with neutrino signatures. In Borexino [1], we optimized methodologies to treat this kind of background instead of trying to identify it and remove it event by event.

Borexino is situated beneath the Gran Sasso mountain in Italy, at the Laboratorio Nazionale del Gran Sasso. The about $\sim$ 1400 m of rock shielding allows the underground laboratory to have a very significant reduction of cosmic rays ($\sim$ 1 muon per hour per m^2). The experimental apparatus is schematized in Figure 1. It presents as a steel dome of 18 m in diameter and 16.9 m in height filled with 2.1 kt of ultra-pure water. This dome is the Outer Detector (OD) and is instrumented with 208 photomultiplier tubes (PMTs). It allows for an extremely efficient detection and tracking of cosmic muons via the Cherenkov light emitted during their passage through the water [2]. Inside there is a Stainless Steel Sphere (SSS) of 13.7 m diameter holding 2212 8" PMTs, referred to as the Inner Detector (ID).. The PMTs are inward-facing and detect the scintillation light caused by particle interactions in the central region. The SSS contains the active neutrino target: 278 t of organic scintillator composed of the solvent PC (1,2,4-trimethylbenzene) doped with the wavelength shifter PPO (2,5-diphenyloxazole) at a concentration of 1.5 g/l. The scintillator mixture is contained in a spherical and transparent nylon Inner Vessel (IV) with a diameter of 8.5 m and a thickness of 125 µm. To shield this central target from external γ-ray backgrounds and to absorb emanating radon, the IV is surrounded by two layers of buffer liquid in which the light quencher dimethylphthalate (DMP) is added to the scintillator solvent. The neutrino signature is a β-like spectrum due to current exchange.

Since the PC is an organic liquid scintillator, it is ^{12}C rich. Therefore, muons might create ^{11}C isotopes, which undergo β^+ decay with an average decay time of $\tau = 29.4$ minutes and a Q-value of 0.96 MeV. Due to positron annihilation, the visible spectrum is shifted to higher energies, covering a range between $\sim$0.8 and $\sim$2 MeV.

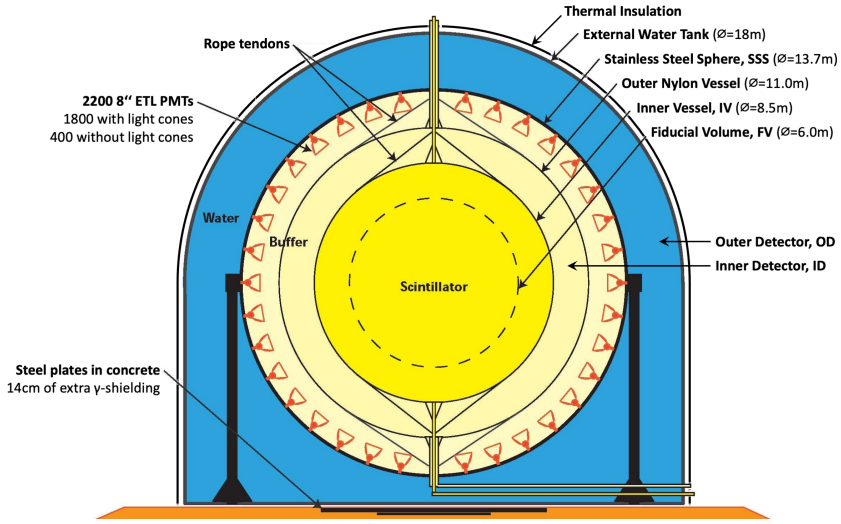


Figure 1. Sketch of the Borexino detector.

The dominant ^{11}C production is through muon spallation processes:

$$\mu + {}^{12}\mathrm{C} \rightarrow \mu + {}^{11}\mathrm{C} + n \tag{1}$$

$$n + p \rightarrow D + \gamma_{2.2\,\mathrm{MeV}} \tag{2}$$

$${}^{11}\mathrm{C} \rightarrow {}^{11}\mathrm{B} + e^{+} + \nu_e. \tag{3}$$

The Equation (2), the neutron capture, has an average time of 250 μs. That is not the only possible neutron interaction, but it is the most dominant. In Figure 2, a realistic Monte Carlo simulation with Geant4 of the ^{11}C decay energy spectrum (magenta line), compared to other identifiable backgrounds and the signals from neutrino current exchanges (red lines).

While we can simulate the spectrum, the full physics of the possible channels of its production and neutron capture is not very well understood. That makes a full simulation of the processes impossible. In the next sections, we will discuss the methodologies used by Borexino to treat this background: the Three-Fold Coincidence (TFC) and a new method to improve its performance named Burst Identification (BI).

THREE-FOLD COINCIDENCE (TFC)

The idea to treat this cosmogenic background is to split the data set into two subsets: one enriched of ^{11}C (Figure 3(a)) and one depleted (Figure 3(b)). These subsets have identical relative spectral components except the ^{11}C contents. As shown in Figure 3, a simultaneous fit of both subsets makes it possible to constrain all the amplitudes of the different spectra.

To achieve this goal, we select regions inside the detector where β-like events will be considered to belong to the enriched subsets. These vetoed regions are built using the knowledge of the ^{11}C formations: Equation (1) and Equation (2); the volume thus made will stay active for $5\tau \simeq 2.5$ hours to increase the possibility that the e^{+} from Equation (3) will likely fall inside it. That is the principle of the so-called *Three-Fold Coincidence*.

We create two different approaches to evaluate these space-time vetoed regions (for details, see [4]):

- **Hard Cut** (HC-TFC), depicted in Figure 4 as an illustrative example: the muon track (from Equation (1)) contributes with a cylindrical region (blue) of 0.7 m radius, neutron captures (from Equation (2)) and other related neutron events introduces spheres (shaded and green respectively) of 1.2 m radius.

- **Likelihood** (LH-TFC) estimates the Likelihood $\mathscr{L}$ of an event to have a coincidence as a TFC. Once a likelihood threshold is set, events with a higher $\mathscr{L}$ are tagged for the enriched subset. Probability density functions to

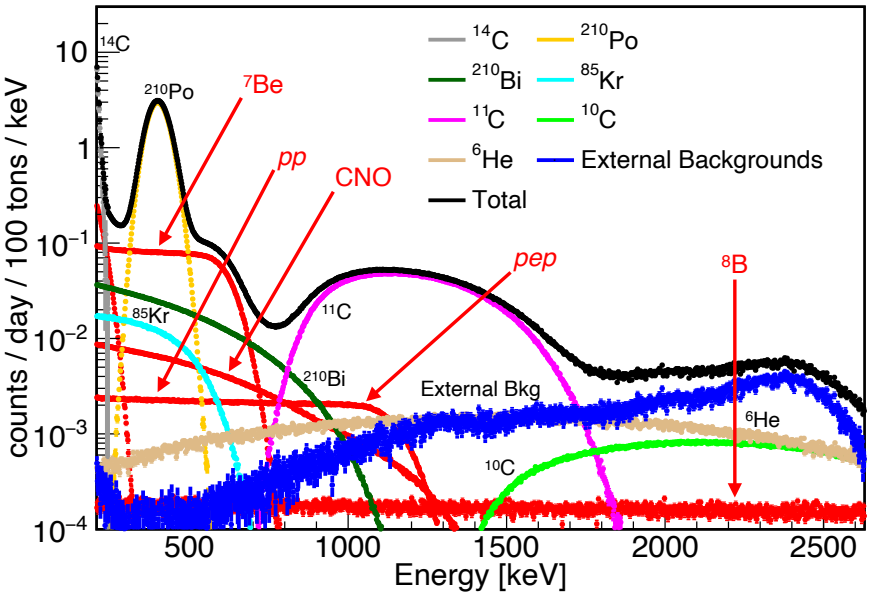


Figure 2. Expected spectrum in Borexino from Monte Carlo simulations. Electron recoil energy due to neutrino interactions (red lines) and background components (other colors).

estimate the likelihoods are data-driven, based on the event-to-muon time and space distributions and event-to-neutron vertices distributions.

When boards are saturated because or passing muons of in-between runs, we might miss some information to build the TFC regions. To be conservative, we veto the entire detector (i.e., all events belong to the enriched subsets) for the whole of the 5τ period; this is the Full Volume Veto (FVV).

Due to the dynamism of the TFC, we introduce two quantities to evaluate the performance of the method properly:

- **Exposure rate $\varepsilon_{\text{depleted}}$**: fraction of the total events that belong to the depleted subsets. Since we are estimating an exposure, what kind of event does not matter; therefore, we estimate it with a toy Monte Carlo, creating random fake events (~ 1 per second) and estimating how many are not vetoed.

- **Tagging efficiency e_{tagged}**: how many ^{11}C are identified (tagged), normalized by the exposure rate. Calling $^{11}\text{C}_{\text{total}}$ the events of the total data set and $^{11}\text{C}_{\text{untagged}}$ the events of the depleted subset,

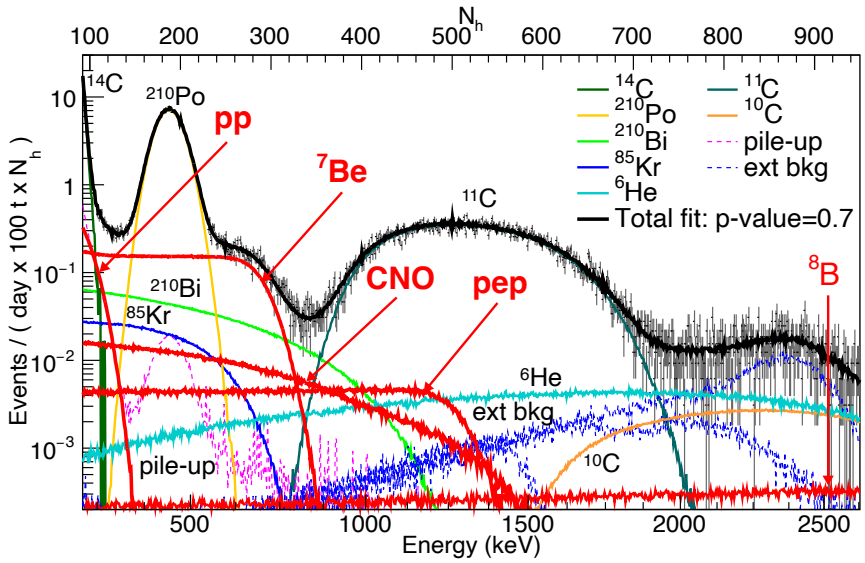
$$e_{\text{tagged}} = 1 - \frac{^{11}\text{C}_{\text{untagged}}}{^{11}\text{C}_{\text{total}}} \cdot \frac{1}{\varepsilon_{\text{depleted}}}. \tag{4}$$

If we restrict the event counting in an energy region without any ^{11}C, $e_{\text{tagged}} = 0$ since $^{11}\text{C}_{\text{untagged}}/^{11}\text{C}_{\text{total}} = \varepsilon_{\text{depleted}}$ by construction. We will measure the tagging performance if we restrict the energy region where almost pure ^{11}C is found (between 1300 and 1500 keV; see Figures 2 and 3). In the rest of the text, e_{tagged} refers to this latter region.

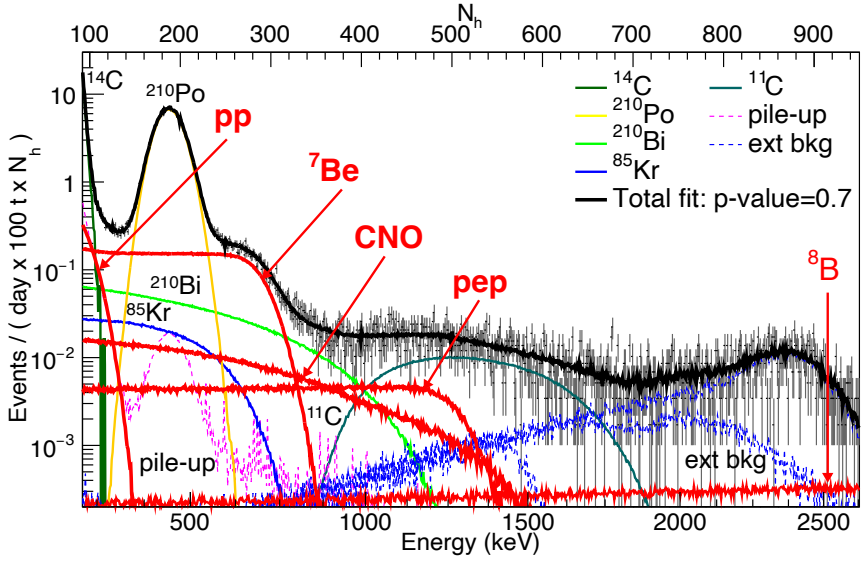
Both HC-TFC and LH-TFC have parameters to be optimized (e.g., radii in the HC-TFC and the likelihood threshold in the LH-TFC). The optimization is obtained by finding the best combination with high $\varepsilon_{\text{depleted}}$ and e_{tagged}. However, it is not evident which of the two must be preferred. Therefore, we simulate the spectrum with different $\varepsilon_{\text{depleted}}$ and e_{tagged} pairs, testing which combination had a better likelihood value from the simultaneous fit of the two subsets.

Naming the period from 14 December 2011 until 21 May 2016 as "Phase-II", and from 17 July 2016 until 2 January 2021 "Phase-III," the $\varepsilon_{\text{depleted}}$ and e_{tagged} of the two approaches are compared in Table I, showing similar results.

In recent Borexino works, such as [5, 6], the HC-TFC is used due to a slightly better stability of the multivariate fit in other background elements. The LH-TFC is left as a comparison for systematic uncertainties.



(a) ^{11}C-enriched.



(b) ^{11}C-depleted.

Figure 3. Example of multivariate simultaneous fit on data, presented in [3]. All lines are the elements (see legends) composing the total spectrum (black); in red, the neutrino current exchange signals from different sources of neutrinos.

BURST IDENTIFICATION (BI)

A muon can produce multiple ^{11}C by spallation as a cascade, i.e., a higher ^{11}C multiplicity is expected. We tried to exploit this *burst* production to eliminate FVV periods, which are quite demanding in terms of detector exposure, without using knowledge of muons and neutrons.

With a railing event procedure (i.e., every event repeats the process in succession), the algorithm starts opening a 4τ window from the current event to find ^{11}C candidates ($0.75 < E(\text{MeV}) < 1.87$). First, the time correlation is examined to see if the events are consistent in time with an exponential decay (black squares and blue open dots in Figure 5). Then, the combination of these ^{11}C candidates with the best spatial correlation is searched (black squares): they are used to build a correlation line (cyan dashed line) through linear regression. To remove energy dependence, the algorithm tags all events that are spatially correlated with the correlation line in the 4τ window (green dots). In

		Phase-II	Phase-III
Hard Cut (HC)	Tagging efficiency (e_{tagged})	90.2%	90.7%
	Exposure fraction ($\varepsilon_{\text{depleted}}$)	63.3%	63.6%
Likelihood (LH)	Tagging efficiency (e_{tagged})	89.8%	90.1%
	Exposure fraction ($\varepsilon_{\text{depleted}}$)	64.7%	65.6%

Table I. Tagging efficiency and exposure fraction for the two TFC approaches. Statistical uncertainty on all e_{tagged} is 0.5%, negligible on $\varepsilon_{\text{depleted}}$ values.

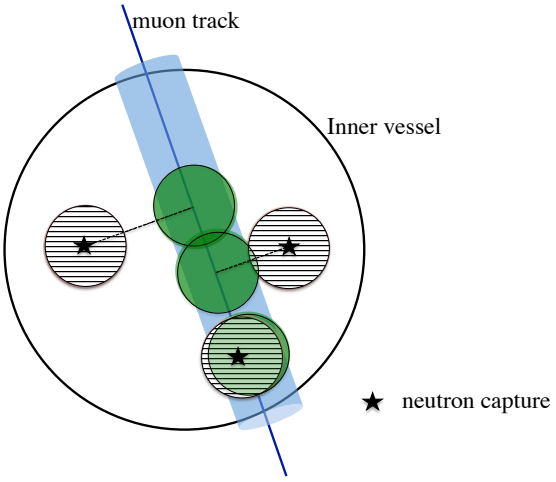


Figure 4. HC-TFC: upon the passage of a muon through the Inner Vessel, geometrical veto regions are shown. The picture is not to scale.

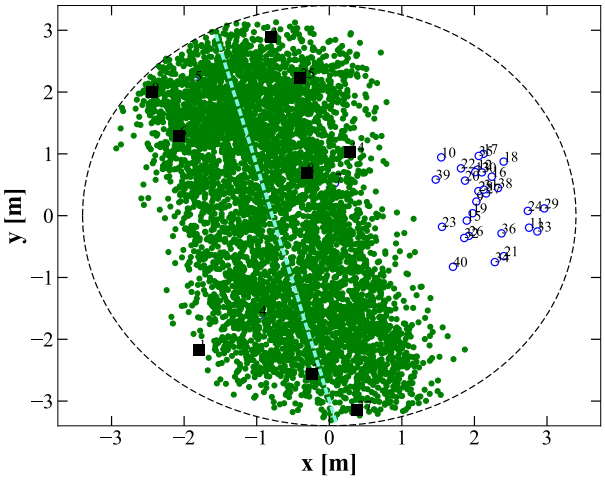


Figure 5. Projection in the *xy*-plane of a sample group of events considered for building a *burst*. The dashed circle shows a radius of 3.5 m. Numbers indicate a time sorting.

Figure 5, the open blue dots represent events not correlated in space but still in time. The are most likely a second coinciding ^{11}C *burst*, considered separately by the algorithm. See [4] for details.

In Figure 6, we show the daily rate of events in the ^{11}C energy range throughout a sample period of one year of Phase-III (black). The events identified as ^{11}C by the HC-TFC without FVV are in cyan. It misses a larger fraction of events with higher ^{11}C multiplicities. However, they are compensated by the BI (red). We note that the BI is not meant to be used on its own but rather via a logical OR operation with the TFC.

We tested the performance of the BI method with selected parameters to keep approximately the same e_{tagged} in Table I and test the minimum gain in terms of $\varepsilon_{\text{depleted}}$ as shown in Table II. We observe an increase of exposure of $\sim 3\%$. However, a simulation similar to the one used to optimize the two TFCs is ongoing. The best trade-off among e_{tagged} and $\varepsilon_{\text{depleted}}$ will define this method's best parameters and advantages.

CONCLUSIONS

Cosmogenic-induced isotopes might have a long decay time, which makes them hard to identify. However, they might be exploited as an advantage for the analysis, treating them instead of selecting them. We have shown Borexino's experience in the subject and how the ^{11}C background was treated. First, the dataset is split into two subsets: one ^{11}C-enriched and one depleted. Then, they are fit simultaneously, letting only the ^{11}C spectrum component differ. To select which event belongs to which subset, events are vetoed using space and time correlations to the isotope production, and byproduct decays using the Three-Fold Coincidence (TFC). Different approaches were devised based on the TFC, including using highly vetoed detector volumes (Hard Cut) or the probability of a single event to be a ^{11}C decay (Likelihood). Moreover, a complementary method has recently been designed to improve TFC performances further, the Burst Identification (BI). Still in the optimization phase, this approach uses the space-time coincidences of ^{11}C cascade production.

For future experiments, Borexino's experience is a pathfinder to solving similar issues related to long-lived cosmo-

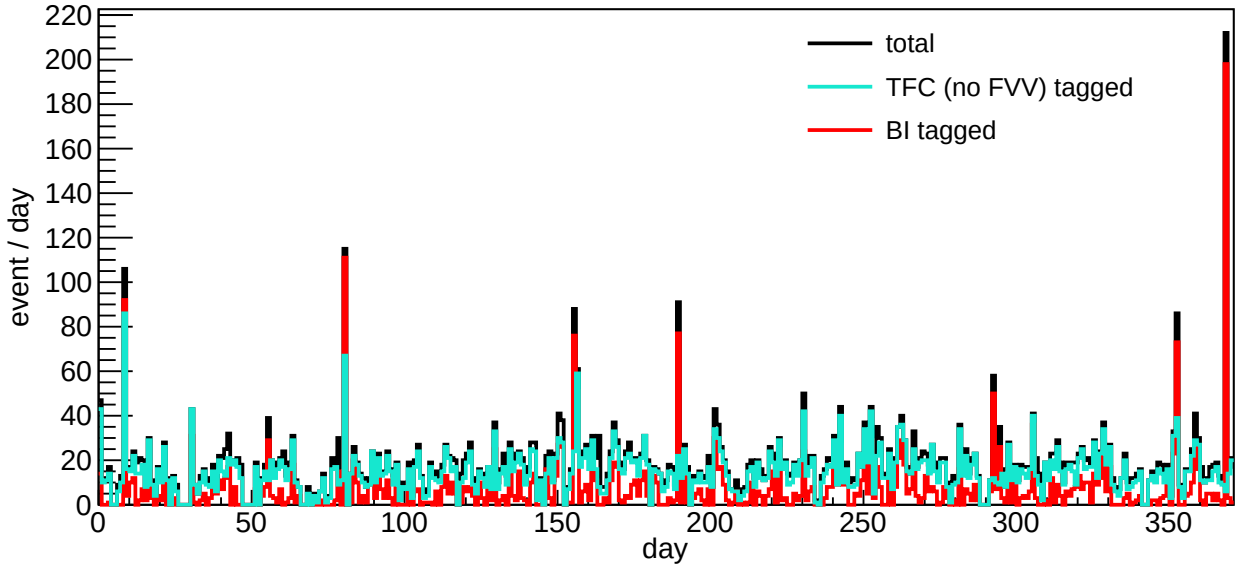


Figure 6. ^{11}C events per day in a year-long window of Phase-III: total in black, only TFC tagged in cyan, only BI tagged in red.

HC-TFC		Phase-II	Phase-III
no FVV	Tagging efficiency (e_{tagged})	79.5%	77.6%
	Exposure fraction ($\varepsilon_{\text{depleted}}$)	70.8%	69.7%
no FVV + BI	Tagging efficiency (e_{tagged})	89.3%	90.4%
	Exposure fraction ($\varepsilon_{\text{depleted}}$)	68.3%	66.7%

Table II. HC-TFC without Full Volume Veto (FVV) before and after the combination with the Burst Identification (BI) tag. Statistical uncertainty as per Table I.

genic backgrounds.

ACKNOWLEDGMENTS

The Borexino program is made possible by funding from Istituto Nazionale di Fisica Nucleare (INFN) (Italy), National Science Foundation (NSF) (USA), Deutsche Forschungsgemeinschaft (DFG) and Helmholtz-Gemeinschaft (HGF) (Germany), Russian Foundation for Basic Research RFBR (Grant 19-02-00097 A), RSF (Grant 21-12-00063) (Russia), and Narodowe Centrum Nauki (NCN) (Grant No. UMO 2017/26/M/ST2/00915) (Poland). This research was supported in part by PLGrid Infrastructure. We acknowledge the generous hospitality and support of the Laboratory Nazionali del Gran Sasso (Italy).

REFERENCES

1. G. Alimonti *et al.* (Borexino), "The Borexino detector at the Laboratori Nazionali del Gran Sasso," Nucl. Instrum. Meth. A **600**, 568–593 (2009), arXiv:0806.2400 [physics.ins-det].
2. G. Bellini *et al.* (Borexino), "Muon and Cosmogenic Neutron Detection in Borexino," JINST **6**, P05005 (2011), arXiv:1101.3101 [physics.ins-det].
3. M. Agostini *et al.* (Borexino), "First Simultaneous Precision Spectroscopy of *pp*, ^{7}Be, and *pep* Solar Neutrinos with Borexino Phase-II," Phys. Rev. D **100**, 082004 (2019), arXiv:1707.09279 [hep-ex].
4. M. Agostini *et al.* (Borexino), "Identification of the cosmogenic ^{11}C background in large volumes of liquid scintillators with Borexino," Eur. Phys. J. C **81**, 1075 (2021), arXiv:2106.10973 [physics.ins-det].
5. M. Agostini *et al.* (Borexino), "Comprehensive measurement of *pp*-chain solar neutrinos," Nature **562**, 505–510 (2018).
6. M. Agostini *et al.* (Borexino), "Experimental evidence of neutrinos produced in the CNO fusion cycle in the Sun," Nature **587**, 577–582 (2020), arXiv:2006.15115 [hep-ex].